

Angela Luna Hernandez

Estimação em pesquisas repetidas empregando o filtro GLS

Campinas
2012

Universidade Estadual de Campinas
Instituto de Matemática, Estatística e
Computação Científica - IMECC

Angela Luna Hernandez



Estimação em pesquisas repetidas empregando o filtro GLS

Dissertação de mestrado apresentada ao Instituto de Matemática, Estatística e Computação Científica da UNICAMP, para obtenção do Título de Mestra em Estatística.

Orientador: Prof. Dr. Luiz Koodi Hotta

Coorientador: Prof. Dr. Fernando Antônio da Silva Moura

Este exemplar corresponde à versão final da dissertação defendida pela aluna Angela Luna Hernandez e orientada pelo Prof. Dr. Luiz Koodi Hotta.


Luiz Koodi Hotta
Orientador
Fernando Antônio da Silva Moura
Coorientador

Campinas, 2012

FICHA CATALOGRÁFICA ELABORADA POR
ANA REGINA MACHADO - CRB8/5467
BIBLIOTECA DO INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E
COMPUTAÇÃO CIENTÍFICA - UNICAMP

L971e Luna Hernandez, Angela, 1980-
Estimação em pesquisas repetidas empregando o filtro GLS /
Angela Luna Hernandez. – Campinas, SP : [s.n.], 2012.

Orientador: Luiz Koodi Hotta.

Coorientador: Fernando Antônio da Silva Moura.

Dissertação (mestrado) – Universidade Estadual de Campinas,
Instituto de Matemática, Estatística e Computação Científica.

1. Kalman, Filtragem de. 2. Métodos de espaço de estados. 3.
Mínimos quadrados. I. Hotta, Luiz Koodi, 1952-. II. Moura, Fernando
Antônio da Silva. III. Universidade Estadual de Campinas. Instituto
de Matemática, Estatística e Computação Científica. IV. Título.

Informações para Biblioteca Digital

Título em inglês: Estimation on repeated surveys using the GLS filter

Palavras-chave em inglês:

Kalman filtering

State-space methods

Least squares

Área de concentração: Estatística

Titulação: Mestre em Estatística

Banca examinadora:

Luiz Koodi Hotta [Orientador]

Caio Lucidius Naberezny Azevedo

Denise Britz do Nascimento Silva

Data de defesa: 05-07-2012

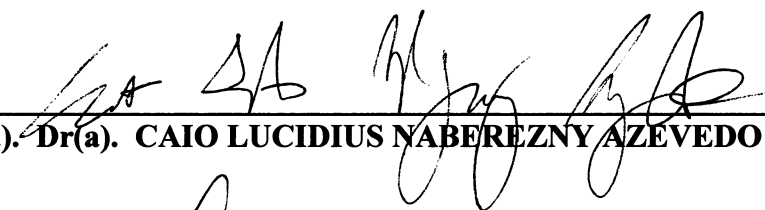
Programa de Pós-Graduação: Estatística

Dissertação de Mestrado defendida em 05 de julho de 2012 e aprovada


Pela Banca Examinadora composta pelos Profs. Drs.



Prof(a). Dr(a). LUIZ KOODI HOTTA



Prof(a). Dr(a). CAIO LUCIDIUS NABEREZNY AZEVEDO



Prof(a). Dr(a). DENISE BRITZ DO NASCIMENTO SILVA

*A minha família e amigos
porque com amor,
percorreram este caminho comigo.*

Agradecimentos

Ao longo destes dois anos, foram muitas as pessoas que participaram de uma ou de outra forma nesta maravilhosa experiência que foi para mim fazer o mestrado na UNICAMP. A todos eles, meu agradecimento.

Aos professores Luiz Hotta e Fernando Moura, obrigada por compartilhar seu conhecimento comigo, pela paciência que mostraram com minhas muitas dúvidas e pela dedicação e amor com que revisaram e corrigiram uma e outra vez, cada parte deste trabalho. Espero contar com a boa fortuna de continuar trabalhando com vocês.

A minha mãe, meu irmão, meu sobrinho, aos meus ‘amigos de lá’, agradeço por aceitarem meu desejo de começar de novo, porque dia a dia suportam minhas ausências e se esforçam por entender minhas paixões e razões.

A Juan Pablo, muito obrigada pela confiança e apoio que sempre me brindastes, por me dar razões para continuar nos momentos mais difíceis, pelo maravilhoso tempo que compartilhamos.

A Eduardo, Pedro, Rafa, obrigada por ter-me abençoado com seu afeto sem restrições. Obrigada pela sua generosidade, pelos muitos bate-papos e almoços, pela compreensão e sorrisos, por implicar comigo sempre que esqueci das coisas verdadeiramente importantes. Du: obrigada pelo café de cada manhã, pelas incansáveis correções a todos meus textos. Rafa: obrigada, entre muitas outras coisas, pelos aportes para melhorar minha prosódia. Pedro: obrigada pela entrega e amizade incondicional. Vocês tocaram minha vida e ficam sempre no meu coração.

À turma do IMECC: Michael, Carole, Larissa, Désirée, Roberto, Nívea, Guilherme,

obrigada por me acolher com tanto carinho, porque nunca me senti estrangeira entre vocês. Obrigada pela ajuda sempre que precisei, pelas pacientes explicações, pelos momentos de lazer, por sua amizade. Carole, obrigada pela força e pelo excelente conselho na escolha de orientador. Lari, obrigada por me escutar pacientemente e me tirar várias vezes do círculo de perguntas sem sentido. Guilherme, obrigada por ser tão generoso na hora de compartilhar seu conhecimento e sua inspiração. Tudo de bom para vocês!

A Andrés Arias, porque o projeto de vir para o Brasil dificilmente teria se materializado se não tivesse contado com sua amizade e apoio. Obrigada amigo. A Ricardo e Catalina, e a Dani, obrigada por me receber na sua casa sempre que precisei ir para o Rio de Janeiro. Agradeço também ao Laboratório de Epidemiologia e Fisiologia Matemáticas, por proporcionar os equipamentos necessários para desenvolver todo o trabalho de simulação.

Aos membros da banca de minha defesa: professores Caio e Denise, obrigada por ter dedicado o tempo para ler esta dissertação e pelos valiosos aportes que fizeram à versão final. À CAPES pelo suporte financeiro para o desenvolvimento deste trabalho.

*“Não vim de tão longe
para morrer diante dos muros de Lisboa.”
José Saramago
História do Cerco de Lisboa*

Resumo

A dissertação apresenta o processo de estimação em pesquisas repetidas sob o enfoque de séries temporais, empregando Modelos de Espaço de Estados e o filtro dos mínimos quadrados generalizados ou *filtro GLS*. Este filtro permite o tratamento de modelos com erros de observação autocorrelacionados de forma mais simples do que utilizando o filtro de Kalman e, além disso, possibilita a modelagem conjunta de várias subpopulações ou domínios sob restrições de *benchmark* obtidas a partir da mesma pesquisa. Isto não só permite manter a coerência entre as estimativas obtidas pelo método, e estimativas agregadas baseadas no planejamento da amostra, como ajuda na proteção contra possíveis erros de especificação dos modelos.

Considerando o caso de amostras com rotação, também é abordado o processo de estimação da estrutura de autocorrelação dos erros amostrais empregando o método dos *pseudo-erros*. Via simulação, é replicado todo o procedimento de estimação, comparando resultados obtidos empregando os filtros GLS e de Kalman. Adicionalmente, é ilustrada a aplicação do filtro sob restrições de benchmark empregando a série de taxa de desemprego da Pesquisa Mensal de Emprego do IBGE, no período de março de 2002 a fevereiro de 2012.

Abstract

This work presents the estimation process in repeated surveys using State Space Models and the generalized linear squares filter, *GLS filter*, under the time series approach. This filter deals with autocorrelated errors in the observation equation, in a simpler way than the well-known Kalman filter. Additionally, it allows for modeling jointly several domains under benchmark constraints obtained from the same survey. The benchmarking not only achieves coherence between the model-based estimates and the corresponding design-based aggregated estimates, but also provides protection against possible model failures.

For the scenario of samples with a rotation scheme, the estimation of the autocorrelation structure of observational errors, using the *pseudo-errors* method, is also addressed. Simulation are used to compare the GLS and Kalman filter estimators. Moreover, the application of GLS filter under benchmark restrictions is illustrated, using the unemployment rate time serie from the Brazilian monthly labor force survey, from March 2002 to February 2012.

Sumário

Lista de Tabelas	xxi
Lista de Figuras	xxiv
Introdução	1
1 Preliminares	7
1.1 Tipos de amostras para levantamentos repetidos	9
1.1.1 Amostras em painel	9
1.1.2 Amostras transversais	10
1.1.3 Amostras com rotação	10
1.2 Estimação em levantamentos repetidos sob o enfoque de séries temporais	15
1.3 Modelos de espaço de estados - MEE	18
1.3.1 Modelo estrutural básico (BSM)	19
1.4 O filtro de Kalman	22
1.4.1 Equações do filtro de Kalman	22

Sumário

1.4.2	Decomposição da verossimilhança através dos erros de previsão	24
1.4.3	Inicialização e aspectos computacionais	24
2	Modelagem do erro amostral	27
2.1	Sobreposição de unidades e autocorrelação	28
2.1.1	Amostras independentes	30
2.1.2	Amostras com sobreposição total de unidades	31
2.1.3	Um exemplo de amostra com sobreposição parcial de unidades .	32
2.1.4	Um exemplo de amostra com rotação	33
2.2	Estimação das autocovariâncias dos e_t . Método dos pseudo-erros	35
2.2.1	Estimativas e erros de painel	36
2.2.2	Covariância de lag k dos erros de painel	37
2.2.3	Pseudo-erros de painel e sua covariância	39
2.2.4	Um estimador para $\rho(k)$	41
3	Estimação em MEE com erros de observação autocorrelacionados	43
3.1	Erros de observação ARMA e o filtro de Kalman	44
3.1.1	BSM com erros de observação AR(1)	46
3.2	O filtro GLS	47
3.2.1	Um preditor empregando mínimos quadrados generalizados . . .	48

4	Exercícios de Simulação	57
4.1	Mecanismo de geração de dados	57
4.1.1	Geração das séries $\{\theta_t\}$	58
4.1.2	Geração das séries $\{e_t\}$	59
4.1.3	Geração das séries $\{y_t\}$	63
4.2	Procedimentos de estimação empregados	64
4.2.1	Estimação de $\rho(k)$	64
4.2.2	Estimação dos hiperparâmetros do modelo	64
4.2.3	Estimação do parâmetro de interesse	65
4.3	Resultados	67
4.3.1	Estimação de $\rho(k)$	67
4.3.2	Estimação dos hiperparâmetros do modelo	70
4.3.3	Estimação dos parâmetros de interesse	71
5	Estimação com o filtro GLS sob restrições de benchmark	77
5.1	As restrições de benchmark	78
5.2	Um MEE para todos os domínios	79
5.3	O filtro GLS sob restrições de benchmark	80
5.4	Aplicação	83
5.4.1	A pesquisa mensal de emprego - PME	84

Sumário

5.4.2	Ajuste simultâneo das séries de taxa de desemprego por RM . . .	88
6	Considerações finais	99
6.1	Trabalhos futuros	100
	Referências Bibliográficas	101
A	Representação de um modelo ARMA em espaço de estados	105
B	Proposta para a estimação das $C^j(k)$	107
C	Equações do filtro GLS	109
D	Resultados adicionais do exercício de simulação	111
D.1	Avaliação do estimador $\hat{\rho}(k)$	112
D.2	Ajuste das funções de autocorrelação dos erros com modelos AR.	118
D.3	Estimação das variâncias dos erros de evolução	119
D.4	Filtros de Kalman com erros ARMA e GLS	123
E	Resultados adicionais da aplicação	141

Lista de Tabelas

1.1	Exemplo de amostra com esquema de rotação 2-2(2).	12
1.2	Esquema de rotação 4-8(2) balanceado a duas vias.	15
2.1	Ilustração da notação introduzida num esquema 4-0(1).	39
4.1	Grupos de rotação com amostra sobreposta num esquema 4-8(2)	61
4.2	Matriz de correlações entre os erros de um grupo de rotação.	62
4.3	Estrutura de autocorrelação simulada para o cenário ad hoc.	63
4.4	Vício médio e EQM na estimação pelo método dos pseudo-erros.	70
4.5	Resumo das estimativas para $\{\theta_t\}$ a partir do segundo ano. Cenário 1. .	75
4.6	Resumo das estimativas para $\{\theta_t\}$ a partir do segundo ano. Cenário 2. .	76
5.1	Tamanho e composição da amostra, segundo Regiões Metropolitanas. .	86
5.2	Média dos CV publicados para a taxa de desemprego, segundo RM. . .	87
5.3	Ajuste de modelos AR(12) às séries de erros de observação.	89
5.4	Variâncias estimadas para os erros de evolução.	91

Lista de Tabelas

D.1	Estimação pelo método dos pseudo-erros. Cenário 1. $\phi = 0.4$	112
D.2	Estimação pelo método dos pseudo-erros. Cenário 1. $\phi = 0.6$	113
D.3	Estimação pelo método dos pseudo-erros. Cenário 1. $\phi = 0.8$	114
D.4	Estimação pelo método dos pseudo-erros. Cenário 2. $\phi = 0.4$	115
D.5	Estimação pelo método dos pseudo-erros. Cenário 2. $\phi = 0.6$	116
D.6	Estimação pelo método dos pseudo-erros. Cenário 2. $\phi = 0.8$	117
D.7	Estimação do modelo AR(1) no cenário 1.	118
D.8	Ajuste de um modelo AR(1) no cenário 2.	118
D.9	Ajuste de um modelo AR(12) no cenário 2.	119
D.10	Estimativas para $\sigma_L^2, \sigma_R^2, \sigma_S^2, \sigma_I^2$. Cenário 1. Erros AR(1)	119
D.11	Estimativas para $\sigma_L^2, \sigma_R^2, \sigma_S^2, \sigma_I^2$. Cenário 2. Erros AR(1) e AR(12). . .	121
D.12	Estimativas para $\{\theta_t\}$. Cenário 1 com $\phi = 0.4$ e $T = 60$	123
D.13	Estimativas para $\{\theta_t\}$. Cenário 1 com $\phi = 0.4$ e $T = 120$	124
D.14	Estimativas para $\{\theta_t\}$. Cenário 1 com $\phi = 0.4$ e $T = 240$	125
D.15	Estimativas para $\{\theta_t\}$. Cenário 1 com $\phi = 0.6$ e $T = 60$	126
D.16	Estimativas para $\{\theta_t\}$. Cenário 1 com $\phi = 0.6$ e $T = 120$	127
D.17	Estimativas para $\{\theta_t\}$. Cenário 1 com $\phi = 0.6$ e $T = 240$	128
D.18	Estimativas para $\{\theta_t\}$. Cenário 1 com $\phi = 0.8$ e $T = 60$	129
D.19	Estimativas para $\{\theta_t\}$. Cenário 1 com $\phi = 0.8$ e $T = 120$	130

D.20 Estimativas para $\{\theta_t\}$. Cenário 1 com $\phi = 0.8$ e $T = 240$	131
D.21 Estimativas para $\{\theta_t\}$. Cenário 2 com $\phi = 0.4$ e $T = 60$	132
D.22 Estimativas para $\{\theta_t\}$. Cenário 2 com $\phi = 0.4$ e $T = 120$	133
D.23 Estimativas para $\{\theta_t\}$. Cenário 2 com $\phi = 0.4$ e $T = 240$	134
D.24 Estimativas para $\{\theta_t\}$. Cenário 2 com $\phi = 0.6$ e $T = 60$	135
D.25 Estimativas para $\{\theta_t\}$. Cenário 2 com $\phi = 0.6$ e $T = 120$	136
D.26 Estimativas para $\{\theta_t\}$. Cenário 2 com $\phi = 0.6$ e $T = 240$	137
D.27 Estimativas para $\{\theta_t\}$. Cenário 2 com $\phi = 0.8$ e $T = 60$	138
D.28 Estimativas para $\{\theta_t\}$. Cenário 2 com $\phi = 0.8$ e $T = 120$	139
D.29 Estimativas para $\{\theta_t\}$. Cenário 2 com $\phi = 0.8$ e $T = 240$	140
E.1 Estimativas da autocorrelação dos e_t por RM.	141
E.2 Estimativas da autocorrelação parcial dos e_t por RM.	143
E.3 P-valores do teste de Jarque-Bera.	146
E.4 P-valores do teste de Ljung-Box.	147
E.5 Autocorrelações estimadas dos erros de previsão por RM.	147

Lista de Tabelas

Lista de Figuras

4.1	Taxa de desemprego (y_t) vs Variância estimada. Curva ajustada.	60
4.2	Estimativas pelo método dos pseudo erros. Cenário 1.	68
4.3	Estimativas pelo método dos pseudo erros. Cenário 1.	69
5.1	Autocorrelações estimadas para os erros amostrais por RM.	90
5.2	Séries ajustadas com e sem benchmark. Série Filtrada.	93
5.3	Séries ajustadas com e sem benchmark. Nível local (Tendência).	94
5.4	Séries ajustadas com e sem benchmark. Componente Sazonal.	95
5.5	Séries ajustadas com e sem benchmark. Série Desazonalizada.	96
5.6	Séries ajustadas com e sem benchmark. Variâncias.	97
5.7	Séries ajustadas com e sem benchmark. Conjunto 6 RM.	98
D.1	Ajuste das autocorrelações dos e_t com modelos AR. Cenário 2.	120
D.2	Valores estimados para σ_L^2 , σ_R^2 , σ_S^2 e σ_I^2 . Cenário 1, $\phi = 0.8$	122
E.1	Boxplots dos Pseudo-erros observados por RM.	142

Lista de Figuras

E.2	Séries ajustadas com e sem benchmark. Inclinação local.	144
E.3	Séries ajustadas com e sem benchmark. Componente Irregular.	145
E.4	Séries ajustadas com e sem benchmark. Adicionais Conjunto 6 RM. . .	146

Introdução

Em diferentes áreas do conhecimento existe a necessidade de monitorar ao longo do tempo a evolução de indicadores relacionados com alguma característica populacional. Com frequência, este problema é contornado na prática mediante a aplicação periódica de pesquisas por amostragem, contexto no qual a maior disponibilidade de informação torna relevante questionar qual a melhor maneira de combinar as informações dos distintos períodos, com a finalidade de obter estimadores mais eficientes para os parâmetros de interesse.

Na área de estimação em pesquisas periódicas, isto tem sido abordado através de dois enfoques: inicialmente, sob o enfoque clássico e, mais recentemente, sob o enfoque de séries temporais. O primeiro enfoque baseia-se na premissa que as observações de uma mesma característica, num mesmo indivíduo, ao longo do tempo, apresentam correlação geralmente positiva¹. Assim, no caso de esquemas de amostragem com sobreposição parcial ou total de unidades nas distintas repetições da pesquisa, a utilização de estimadores que aproveitem esta dependência pode auxiliar a estimação de parâmetros tais como o nível do indicador em cada momento, ou sua diferença entre dois períodos de tempo. Note que, neste enfoque, a inferência é feita com relação à aleatorização da amostra, considerando os parâmetros de cada momento como fixos e sem fazer nenhuma suposição sobre sua evolução.

No entanto, como observado por Scott et al. (1974), na análise é comum supor que o parâmetro de interesse evolui temporalmente através de alguma estrutura estocástica, o que permite aos pesquisadores estudar características como sua tendência e sazonalidade.

¹No entanto, vários tipos de características fogem desta generalização: uma variável como *gasto mensal em compra de automóveis*, por exemplo, não apresenta autocorrelação positiva a nível de indivíduo.

Introdução

dade, ou fazer previsões. Assim, empregando ideias provenientes da teoria de extração de sinais na presença de ruído, uma nova abordagem para este problema é proposta utilizando modelos de séries temporais. Neste caso, o parâmetro de interesse θ_t faz o papel de *signal* que está sendo observado com *ruído* através da estimativa amostral y_t , enquanto o erro amostral e_t faz o papel de erro de observação. Com isto,

$$y_t = \theta_t + e_t. \quad (1)$$

Sob uma perspectiva de inferência baseada em modelos, se supormos alguma estrutura estocástica para explicar a evolução dos θ_t , a equação (1) permite induzir uma estrutura de autocorrelação na série de estimativas y_t que pode ser empregada para obter melhores estimadores, sem precisar de sobreposição de unidades entre as amostras dos distintos períodos.

Os primeiros desenvolvimentos sob esta abordagem foram apresentados em Blight e Scott (1973), Scott et al. (1974) e Jones (1980). Porém, a idéia só se popularizou quando foi desenvolvida uma metodologia de implementação usando Modelos de Espaço de Estados (MEE) e o filtro de Kalman (Kalman, 1960). De forma simplificada, este método objetiva reescrever o problema de estimação do parâmetro de interesse, θ_t , usando uma dupla de equações do tipo:

$$\begin{aligned} y_t &= \mathbf{Z}_t \boldsymbol{\alpha}_t + e_t \\ \boldsymbol{\alpha}_t &= \mathbf{T}_t \boldsymbol{\alpha}_{t-1} + \mathbf{G}_t \boldsymbol{\eta}_t, \end{aligned}$$

com $\theta_t = \mathbf{Z}_t \boldsymbol{\alpha}_t$. A primeira equação é chamada *de observação* e a segunda *de evolução*; y_t é um estimador não viesado para θ_t ; e_t é seu erro amostral associado; $\boldsymbol{\alpha}_t$ é chamado *de vetor de estados*; $\boldsymbol{\eta}_t$ é o seu erro de evolução; e \mathbf{Z}_t e \mathbf{T}_t podem ser constantes ou não no tempo e conter ou não componentes desconhecidas.

Usualmente, assume-se que os erros e_t e $\boldsymbol{\eta}_t$ têm distribuição normal com média zero e variância constante, possivelmente desconhecida, são não autocorrelacionados ao longo do tempo e não correlacionados entre si². Sob estas hipóteses, quando \mathbf{Z}_t , \mathbf{T}_t e \mathbf{G}_t são conhecidas, o filtro de Kalman permite calcular recursivamente o estimador de $\boldsymbol{\alpha}_t$ com menor erro quadrático médio (EQM), que corresponde à esperança da sua

²Geralmente, y_t é um estimador direto baseado no desenho amostral, como o de Horvitz-Thompson (Särndal et al. (1992, pág. 42)), ou o de Hansen-Hurwitz (Särndal et al. (1992, pág. 51)), que emprega informação só do t -ésimo período, justificando a suposição de e_t e $\boldsymbol{\eta}_t$ não serem correlacionados.

distribuição condicionada nas estimativas y_t obtidas para todos os períodos disponíveis até o momento atual. Sem supor a normalidade³, este estimador segue sendo o de menor Erro Quadrático Médio (EQM) na família de estimadores lineares de θ_t . α_t e T_t não têm uma formulação única, mas é interessante escolhê-los de forma que o modelo resultante seja interpretável, como por exemplo, no caso da decomposição estrutural da série temporal em suas tendência, sazonalidade, ciclo, etc.

Pela sua utilidade, é possível encontrar diversas aplicações desta metodologia em pesquisas repetidas nos últimos 20 anos, especialmente no âmbito de indicadores de força de trabalho. Neste contexto, a primeira aplicação feita por Tiller (1992) sobre a pesquisa de emprego dos Estados Unidos foi seguida por aplicações para as pesquisas da Austrália (Pfeffermann et al., 1996), do Brasil (Silva e Cruz, 2002; Cruz et al., 2006), dos Países Baixos (Van den Brakel e Krieg, 2009) e mais recentemente do Reino Unido (Chow et al., 2010).

Contudo, já na primeira aplicação feita por Tiller (1992) é evidente a necessidade de produzir não só estimativas para o país, mas também, para sub-regiões ou outros domínios de interesse. Uma alternativa é considerar cada domínio individualmente, mas, em termos de precisão, pode ser mais vantajoso encará-lo como um problema de estimação de séries multivariadas, aproveitando, assim, informações de todos os domínios como é comum na abordagem de *estimação em áreas pequenas* (Pfeffermann, 2002). Mesmo que a proposta original usando o filtro de Kalman possa, em princípio, ser utilizada para este fim, nem sempre é essa a melhor solução na presença de esquemas complexos de sobreposição de amostra, como evidenciado por Pfeffermann e Tiller (2006).

Assim, motivados pelo problema de ajuste simultâneo das taxas de desemprego para os Estados e divisões censitárias na pesquisa do Bureau de Censos dos Estados Unidos, Pfeffermann e Tiller (2006) desenvolveram um novo filtro para estimar α_t , no caso de erros observacionais autocorrelacionados, baseado no método de mínimos quadrados generalizados pelo que é denominado *filtro GLS*. Embora, diferentemente do filtro de Kalman, este filtro não conduza ao estimador com menor EQM baseado nas estimativas individuais, não é preciso escrever a estrutura de evolução dos erros observacionais na

³A suposição de não autocorrelação na série de erros amostrais também pode ser relaxada, como é evidenciado no Capítulo 3, sempre que a estrutura de autocorrelação for razoavelmente explicada por um modelo com representação em espaço de estados. Os modelos ARMA, por exemplo, satisfazem esta condição (ver Apêndice A).

Introdução

forma em espaço de estados. O filtro pode ser calculado mesmo nos casos onde a autocorrelação dos erros só pode ser especificada pontualmente, o que facilita o tratamento de modelos com esquemas de sobreposição de amostra complexos.

Além disso, para o caso de estimação conjunta para vários domínios, tratando y_t como uma série multivariada e utilizando o filtro GLS, Pfeiffermann e Tiller (2006) apresenta uma metodologia simples para a inclusão de restrições de *benchmark*⁴ no processo de estimação, a qual pode ser empregada para garantir consistência entre as estimativas geradas pelo modelo para os domínios e as baseadas no planejamento, usualmente publicadas a níveis de agregação maiores. Tal metodologia não só resolve o problema de coerência na publicação de resultados, como também ajuda na proteção contra possíveis erros de especificação do modelo.

O objetivo deste trabalho é apresentar o processo de estimação em pesquisas repetidas mediante modelos de espaço de estados, fazendo uso dos filtros de Kalman e GLS. Com este intuito, esta dissertação está organizada da seguinte maneira: no primeiro Capítulo, é apresentada uma revisão de literatura sobre estimação em pesquisas repetidas com foco na inferência baseada em modelos, e são introduzidos os MEE e o filtro de Kalman. No segundo Capítulo, é abordado o tema da modelagem do erro amostral. Inicialmente, é apresentada uma ilustração de como a sobreposição de unidades nas amostras de pesquisas repetidas pode introduzir autocovariância nas correspondentes séries de erros amostrais. Posteriormente, é introduzido o método proposto por Pfeiffermann et al. (1998) para a estimação das autocorrelações da série de erros amostrais, no caso de amostras com rotação.

No Capítulo três, são expostos dois métodos para a estimação em MEE com erros observacionais autocorrelacionados: o filtro de Kalman com erros que seguem um processo autorregressivo e de médias móveis (ARMA, Box et al., 1994) e o filtro GLS. Tanto a metodologia de estimação das autocorrelações introduzida no Capítulo dois, quanto os estimadores apresentados no Capítulo três foram avaliados através de simulação, os resultados deste exercício são apresentados no Capítulo quatro. No Capítulo cinco, dar-se-á atenção à metodologia proposta por Pfeiffermann e Tiller (2006) para a estimação com o filtro GLS sob uma restrição de benchmark. Para efeitos de ilustração, nesse

⁴Isto é, restrições onde uma soma ponderada das séries que compõem y_t é igualada ao valor de uma outra série temporal.

último capítulo é apresentada uma aplicação deste método na estimação da taxa de desemprego do Brasil, utilizando informação das regiões metropolitanas participantes da Pesquisa Mensal de Emprego (PME), conduzida pelo Instituto Brasileiro de Geografia e Estatística (IBGE) no período de março de 2002 a fevereiro de 2012. Finalmente, no Capítulo seis, são feitas algumas considerações finais e são apresentadas propostas de trabalhos futuros.

Introdução

Capítulo 1

Preliminares

O objetivo de acompanhar a evolução de algumas características de interesse numa população dada é frequentemente abordado mediante a execução repetida de pesquisas por amostragem. Diferentemente das pesquisas que são feitas uma única vez, neste caso é preciso considerar a dimensão temporal para identificar que tipo de plano amostral se adapta melhor aos objetivos e à periodicidade do estudo, por exemplo: amostras de tipo *painel*, onde todas as unidades selecionadas participem em todas as repetições do estudo; amostras com sobreposição parcial de unidades entre períodos; ou amostras em que as unidades finais a serem pesquisadas são sempre diferentes, que denominaremos *transversais*¹.

Sob a hipótese que medições para um mesmo indivíduo em distintos períodos de tempo estão positivamente correlacionadas e, mais ainda, que esta correlação pode ser transmitida aos estimadores através de amostras com unidades sobrepostas, a perspectiva de inferência clássica, baseada no planejamento da amostra, demonstra que o método escolhido pode incidir positivamente na precisão dos estimadores usualmente empregados. Por exemplo, como indicado em Cochran (1977, Seção 12.10 et seq.), ou Kish (1965, Seção 12.4 et seq.), os métodos com algum tipo de sobreposição de amostra poderiam conduzir a uma estimação mais precisa das diferenças de médias ou totais, dado que para dois estimadores \hat{y}_t e \hat{y}_{t-k} , a variância de sua diferença decresce quando

¹Consideraremos que não é de interesse do estudo a evolução dos elementos da população individualmente, i.e., que qualquer das três metodologias mencionadas pode ser aplicável.

Preliminares

se incrementa a covariância entre eles,

$$V(\hat{y}_t - \hat{y}_{t-k}) = V(\hat{y}_t) + V(\hat{y}_{t-k}) - 2\text{Cov}(\hat{y}_t, \hat{y}_{t-k}).$$

Do mesmo modo, para estimar o valor do parâmetro em cada período, a maior disponibilidade de informação para os elementos nas amostras sobrepostas possibilita o emprego de estimadores potencialmente mais precisos². A partir deste enfoque, a discussão sobre como melhorar a estimação centra-se na identificação dos esquemas e frações ótimas de sobreposição de amostra para um estimador dado, às vezes supondo modelos para a evolução no tempo das observações comuns, enquanto nada é suposto para a evolução do parâmetro, considerado fixo em cada período de tempo.

Contudo, como observado por Scott et al. (1974), na etapa de análise de dados é comum supor que o parâmetro de interesse θ_t evolui como uma série temporal com alguma estrutura estocástica, o que permite aos pesquisadores estudar características como sua tendência e sazonalidade e fazer previsões. Daí, é possível trazer este problema para outra perspectiva de inferência, baseada no uso de séries temporais, onde a sequência de parâmetros $\{\theta_t\}$ é considerada como resultado de um processo estocástico. Caso o modelo seja adequado para explicar sua evolução, é possível obter estimadores e previsões com menor EQM.

Os primeiros desenvolvimentos neste sentido apareceram em Blight e Scott (1973), Scott et al. (1974) e Jones (1980), mas foi o aparecimento de uma proposta de implementação usando MEE e o Filtro de Kalman, (Kalman, 1960) que facilitou a popularização deste enfoque.

Na próxima seção, são abordados alguns tipos de amostras empregadas para levantamentos repetidos, dando ênfase nas denominadas *amostras com rotação*, a serem utilizadas no capítulo seguinte. Na Seção 1.2, são apresentados os aspectos mais relevantes da evolução histórica do enfoque de estimação usando séries temporais³, depois, na Seção 1.3 são introduzidos os MEE e, por fim, na Seção 1.4, é apresentado o filtro de Kalman.

²Por exemplo, os estimadores compostos apresentado em Wolter (1979).

³Para uma revisão mais exaustiva podem ser consultados Binder e Hidiroglou (1988) ou Silva e Cruz (2002, cap. 2).

1.1 Tipos de amostras para levantamentos repetidos

Nesta seção, são introduzidos três tipos de amostras amplamente utilizadas neste contexto: as de *painel*, as *transversais* e as *amostras com rotação*. Os dois primeiros tipos são brevemente tratados seguindo o trabalho de Duncan e Kalton (1987)⁴. O terceiro é desenvolvido com mais detalhe visando seu uso ao longo deste trabalho.

1.1.1 Amostras em painel

Neste método a mesma amostra é pesquisada em todas as repetições do estudo. Isto possibilita a estimação não só da evolução do parâmetro de interesse, mas das componentes individuais dessa evolução, também conhecidas como *fluxos*: por exemplo, a proporção de elementos na população que, estando num estado A em um período passa para o estado B no seguinte.

Já, para a estimação dos valores dos parâmetros em cada momento ou suas diferenças entre períodos de tempo, a total sobreposição de unidades nas amostras empregadas têm o intuito de melhorar a precisão dos estimadores, com referência ao caso de amostras sem sobreposição. Tal intuito baseia-se na premissa de que observações de um mesmo indivíduo estão positivamente correlacionadas e que, pela via da sobreposição de amostra, esta correlação poderia ser introduzida entre os estimadores dos distintos períodos. Este tema é estudado por autores clássicos como Cochran (1977, Seção 12.10 et seq.) ou Kish (1965, Seção 12.4 et seq.).

Contudo, nem sempre é desejável possuir uma amostra estática quando a pesquisa planeja ser repetida muitas vezes: a informação do painel pode ter sua validade comprometida, tanto por não poder incorporar as mudanças que vão acontecendo com a composição da população, o que é particularmente importante em populações muito dinâmicas, quanto porque a entrevista repetida das mesmas unidades pode conduzir a perdas importantes de amostra, como indicado por Duncan e Kalton (1987).

⁴O termo *transversais* substitui uma tradução mais literal do original *repeated surveys* empregado pelo autor, pois este tem um sentido muito mais geral em outros textos. Ver por exemplo Kish (1965, Seção 12.4).

Tipos de amostras para levantamentos repetidos

1.1.2 Amostras transversais

Este tipo de amostra tem como característica que, no seu planejamento, não se procura sobreposição de unidades finais⁵ entre as amostras dos distintos períodos do estudo. Isto é, quando empregada amostragem em um único estágio não existem unidades intencionalmente comuns entre amostras e, nos casos de amostras em vários estágios, embora seja possível a existência de unidades comuns nas primeiras etapas de seleção, as unidades finais sobre as quais é levantada a informação são, em princípio, diferentes.

Na perspectiva clássica de inferência baseada no planejamento, a não sobreposição intencional entre as amostras tem como objetivo evitar a correlação entre os estimadores de distintos períodos, ou ainda introduzir correlação negativa entre eles, o que poderia favorecer a estimação em casos em que o parâmetro tem forma de média móvel. Também, este tipo de amostra pode ser empregado para a medição de domínios ou subpopulações raras, quando agregada a informação de várias repetições da pesquisa, considerando que as características de interesse sejam estáveis no tempo. Note, no entanto, como nenhuma das unidades finais de amostragem participa na pesquisa mais de uma vez, este método não permite estimar fluxos.

1.1.3 Amostras com rotação

A ideia de empregar amostras com sobreposição parcial de unidades para a estimação em levantamentos repetidos, foi proposta inicialmente por autores como Yates (1949), Cochran (1977) ou Kish (1965) no conceito de *amostragem em ocasiões sucessivas*, e por Eckler (1955) e Rao e Graham (1964), entre outros, sob o termo *amostras com rotação*. Nos dois casos, procura-se aproveitar, ao menos parcialmente, as vantagens das pesquisas de painel, ao mesmo tempo que são reduzidos os riscos para a validade das informações levantadas nestes estudos, como mencionado em 2.1.2.

Em linhas gerais, uma amostra com rotação é uma amostra onde grupos de unidades se alternam entrando e saindo da pesquisa, ao longo do tempo, de acordo a um esquema preestabelecido denominado *esquema de rotação*, de forma a garantir frações de amostra

⁵Isto é, as unidades sobre as quais é levantada informação, e.g. domicílios, pessoas, etc.

sobrepostas entre períodos de tempo previamente fixados.

Para ilustrar este conceito, suponha que se deseja fazer um estudo com periodicidade mensal, para estabelecer como evolui a opinião dos moradores de um determinado bairro com relação a certas medidas da prefeitura. Empregando um mapa do bairro, quatro blocos são selecionados aleatoriamente, e cinco domicílios, também escolhidos aleatoriamente em cada bloco, serão pesquisados cada mês. Deseja-se garantir uma sobreposição de 50 % entre as subamostras de domicílios entre meses sucessivos. Além disso, com a finalidade de reduzir a possível não resposta, pretende-se que os moradores dos domicílios selecionados respondam o questionário dois meses consecutivos, parem por dois meses e voltem a respondê-lo por mais dois meses, para depois saírem definitivamente da pesquisa. Note que, como os quatro blocos originalmente selecionados devem estar presentes na amostra em todos os meses, mas um domicílio não pode responder o questionário mais do que dois meses consecutivos, quando uma subamostra de domicílios for visitada pela segunda ou pela quarta vez, ela deve ser substituída por outra selecionada do mesmo bloco para o seguinte mês. Para a seleção destas subamostras, pode-se empregar, por exemplo, amostragem sistemática sobre uma lista dos domicílios de cada bloco.

Várias configurações podem satisfazer o esquema de rotação mencionado, uma delas é apresentada na Tabela 1.1. Os números nas células indicam quantas vezes, incluída esta, têm sido pesquisados os domicílios na subamostra correspondente a esse bloco nesse período, e o símbolo ao redor do número indica qual subamostra esta sendo empregada nesse mês: se a primeira (nenhum símbolo), a segunda (círculo vazio), a terceira (círculo cheio), e assim por diante. No entanto, para que as primeiras subamostras dos blocos 3 e 4 sejam pesquisadas pela segunda vez em janeiro, note que é preciso começar o estudo com um mês de antecedência e a informação deles nessa ocasião não é utilizada.

Em geral, denotaremos por $r_1 - r_2(m)$ um esquema de rotação onde as unidades permanecem na amostra por r_1 períodos consecutivos e saem por r_2 , repetindo este procedimento m vezes; assim, o exemplo na Tabela 1.1 é um esquema 2-2(2). Na prática, a escolha destes valores r_1 , r_2 e m visa garantir uma determinada fração de amostra sobreposta entre períodos específicos, para os quais se desejam estimar diferenças nos níveis dos indicadores. Como exemplo, Gambino e do Nascimento Silva (2009) evidenciam que, para o contexto de pesquisas mensais de mercado de trabalho, são empregados

Tipos de amostras para levantamentos repetidos

Tabela 1.1: Exemplo de amostra com esquema de rotação 2-2(2).

Nas linhas: amostra selecionada para cada mês. Nas colunas: blocos selecionados. Texto nas células: número de visita em que se encontra o conjunto de domicílios selecionado nesse bloco nesse período. Símbolo ao redor do texto: para cada bloco, identifica os membros de uma mesma subamostra.

Período	Bloco			
	1	2	3	4
Janeiro	1	1	2	2
Fevereiro	2	2	①	①
Março	①	①	②	②
Abril	②	②	3	3
Maio	3	3	4	4
Junho	4	4	③	③
Julho	③	③	④	④
Agosto	④	④	● 1	● 1
Setembro	● 1	● 1	● 2	● 2
...				

esquemas 4-8(2), 6-0(1) e 8-0(1), entre outros, sendo que o primeiro deles está orientado a melhorar a estimação de diferenças mensais, trimestrais e anuais e os dois últimos a de diferenças mensais e trimestrais.

No exemplo anteriormente empregado, foi assumido que os blocos inicialmente selecionados estavam na amostra ao longo de toda a pesquisa. Na prática, de fato, acontece comumente que em planejamentos de amostra mais complexos, que envolvam vários estágios de seleção, a mesma amostra de unidades primárias (PSU) ou secundárias (SSU) seja mantida ao longo de todas as repetições da pesquisa, trocando em cada período somente as unidades correspondentes aos últimos estágios de seleção. Como indicado por Gambino e do Nascimento Silva (2009), isto permite aproveitar recursos e tempo associados com a seleção e preparação de material para a amostra nos primeiros estágios.

Grupos de rotação e painéis

Para implementar uma amostra com rotação em estudos de grande porte, as PSU ou SSU selecionadas são distribuídas em grupos, denominados *de rotação*, para alternar

suas entradas e saídas da pesquisa simultaneamente. Sob esta definição, na Tabela 1.1, os blocos 1 e 2 constituem um grupo de rotação e os blocos 3 e 4 outro. Além disso, cada subamostra de unidades finais selecionada no mesmo grupo de rotação é chamada de *painel*⁶.

Voltando novamente à Tabela 1.1, em cada grupo de rotação, as células com o mesmo símbolo (nenhum, círculo vazio, círculo cheio), identificam um painel. Assim, por exemplo, o primeiro painel do primeiro grupo de rotação (blocos 1-2) ingressa pela primeira vez na amostra em janeiro, sai em março para dar passo ao segundo painel, volta em maio e sai definitivamente em julho. Também, veja que a amostra do mês de abril, por exemplo, está composta por: i) o segundo painel do primeiro grupo de rotação, que está sendo pesquisado pela segunda vez, e ii) o primeiro painel do segundo grupo de rotação, que está sendo pesquisado pela terceira vez.

Vício de grupos de rotação e amostras balanceadas

Associado com o fato de entrevistar repetidamente os mesmos indivíduos, existe a possibilidade de que o processo de estimação seja afetado pelo denominado *vício de grupos de rotação*. Como indicado por Kalton (2009), este vício consiste em diferenças sistemáticas que podem ocorrer nas estimações do parâmetro de interesse obtidas a partir dos distintos painéis que compõem a amostra num momento dado, em função do número de visita na que eles se encontram. Por exemplo, veja que no esquema apresentado na Tabela 1.1, em caso de existir este tipo de vício, a comparação de estimativas obtidas nos meses de janeiro a março com as de maio a julho poderia ser comprometida, porque enquanto as primeiras são baseadas em painéis pesquisados pela primeira e segunda vez, as outras, são baseadas em painéis na terceira e quarta visitas.

Uma forma de contornar este problema é garantir que em todos os períodos a amostra da pesquisa tenha a mesma composição com relação ao número de visitas. Este conceito foi formalizado por Park et al. (2001) da seguinte forma:

Definição: Um planejamento amostral é $r_1 - r_2(m)$ se:

⁶Por esta razão também é possível encontrar a expressão *amostra de painéis rotativos* fazendo referência às amostras com rotação, como por exemplo em Silva e Cruz (2002).

Tipos de amostras para levantamentos repetidos

- i) tem esquema de rotação $r_1 - r_2(m)$ e*
- ii) existem mr_1 grupos de rotação e a amostra de cada período está composta por um painel de cada um deles.*

Definição: Um planejamento amostral $r_1 - r_2(m)$ é balanceado a duas vias⁷ (no grupo de rotação e no tempo na amostra) se:

- i) em cada período, um painel está sendo entrevistado pela primeira vez, outro pela segunda, e assim por diante, e*
- ii) num lapso de mr_1 períodos, cada grupo de rotação tem participado mr_1 vezes na amostra, um período pela primeira vez, um pela segunda e assim por diante.*

Também, em Park et al. (2001) é apresentado um algoritmo para a construção de esquemas deste tipo.

Esquema de rotação 4-8(2) balanceado a duas vias

Este esquema de rotação, apresentado na Tabela 1.2, é utilizado para estimar as diferenças mensais, trimestrais e anuais no nível dos indicadores, nas pesquisas de emprego do IBGE, e do Bureau de Census dos Estados Unidos. É empregado também nos exercícios de simulação apresentados nesta dissertação.

Como veremos na Tabela 1.2, de acordo com o balanceamento a duas vias, a amostra de cada período está composta por um painel de cada grupo de rotação. Para o primeiro período, por exemplo, o painel do primeiro grupo está sendo pesquisado pela oitava vez, o do segundo grupo pela sétima, e assim sucessivamente, enquanto o painel correspondente ao oitavo grupo de rotação participa na pesquisa pela primeira vez. Acompanhando a evolução ao longo do tempo deste último painel, pode-se observar que cada painel permanece na pesquisa por quatro períodos consecutivos (neste caso desde 1 até 4), sai por oito períodos, tempo durante o qual é substituído por outros painéis

⁷Em inglês, *two way balanced rotation design*.

Tabela 1.2: Esquema de rotação 4-8(2) balanceado a duas vias.

Nas linhas: composição da amostra de cada mês. Nas colunas: Grupos de rotação. Texto nas células: número de visita em que se encontra o painel nesse período. Símbolo ao redor do texto (nenhum, círculo vazio, círculo sombreado, círculo cheio): para cada grupo de rotação, identifica um painel.

Período	Grupo de rotação							
	1	2	3	4	5	6	7	8
1	8	7	6	5	④	③	②	①
2	①	8	7	6	5	④	③	②
3	②	①	8	7	6	5	④	③
4	③	②	①	8	7	6	5	④
5	④	③	②	①	8	7	6	5
6	⑤	④	③	②	①	8	7	6
7	⑥	⑤	④	③	②	①	8	7
8	⑦	⑥	⑤	④	③	②	①	8
9	⑧	⑦	⑥	⑤	④	③	②	①
10	①	⑧	⑦	⑥	⑤	④	③	②
11	②	①	⑧	⑦	⑥	⑤	④	③
12	③	②	①	⑧	⑦	⑥	⑤	④
13	④	③	②	①	⑧	⑦	⑥	⑤
14	⑤	④	③	②	①	⑧	⑦	⑥
15	⑥	⑤	④	③	②	①	⑧	⑦
16	⑦	⑥	⑤	④	③	②	①	⑧
⋮								

do mesmo grupo de rotação (de 5 até 8 por um painel e de 9 até 12 por outro), e logo volta por mais quatro períodos para sair definitivamente da amostra (o que ocorre no período 16 neste caso).

1.2 Estimação em levantamentos repetidos sob o enfoque de séries temporais

A primeira proposta de estimação em levantamentos repetidos sob o enfoque de séries temporais foi apresentada por Blight e Scott (1973), a partir de um modelo elaborado

Estimação em levantamentos repetidos sob o enfoque de séries temporais

por Patterson (1950) sob o enfoque de inferência baseada no planejamento da amostra. O modelo de Patterson explicava a evolução das observações dos indivíduos numa amostra com sobreposição de unidades; já a proposta de Blight e Scott acrescentou um modelo autorregressivo de primeira ordem para explicar a evolução do parâmetro de interesse, θ_t . O modelo proposto é:

$$\begin{aligned} y_{t,i} - \theta_t &= \begin{cases} \rho(y_{t-1,i} - \theta_{t-1}) + e_{t,i} & i \in M_t \\ e_{t,i} & i \in M'_t \end{cases} \\ \theta_t - \mu &= \lambda(\theta_{t-1} - \mu) + \eta_t, \end{aligned} \quad (1.1)$$

onde $y_{t,i}$ é a observação no momento t para o i -ésimo elemento da amostra, M_t é o conjunto de índices dos elementos comuns entre t e $t-1$ e M'_t seu complemento; λ , μ e ρ são constantes que podem ou não ser conhecidas e as sequências de erro, $\{\eta_t\}$ e $\{e_{t,i}\}$, são não autocorrelacionadas temporalmente e independentes entre elas, com média zero e variâncias constantes.

Supondo normalidade das sequências de erro, e sendo \bar{y}_{t,M_t} e \bar{y}_{t-1,M_t} as médias das observações com índices em M_t nos momentos t e $t-1$, respectivamente, e \bar{y}_{t,M'_t} a média em t da amostra não pareada, Blight e Scott salientam que o vetor de médias $\bar{y}_t = \{\bar{y}_{j,M_j}, \bar{y}_{j-1,M_j}, \bar{y}_{j,M'_j}; j = 2, 3, \dots, t\}$ é uma estatística suficiente para a sequência de parâmetros $\{\theta_t\}$, o que na prática, permite resumir apropriadamente as informações empregadas para a estimação. Assim, assumindo normalidade, os autores apresentam o estimador de menor EQM para θ_t sob o modelo (1.1), que adicionalmente, pode ser calculado recursivamente. Sem a hipótese de normalidade, ele ainda é o de menor EQM na família dos estimadores lineares de θ_t .

Como pode-se observar na equação (1.1), o modelo considerado por Blight e Scott só faz sentido nos casos em que existe sobreposição entre as amostras dos distintos períodos. O caso de amostras independentes foi abordado por Scott et al. (1974) utilizando a metodologia de extração de sinais na presença de ruído. Seja y_t um estimador não viesado para θ_t , função somente da informação correspondente a esse período e com erro de estimação denotado por e_t . É possível escrever:

$$y_t = \theta_t + e_t, \quad (1.2)$$

sendo e_t uma variável aleatória com média zero e variância σ_t^2 . Se a sequência $\{\theta_t\}$ for uma realização de um processo estocástico, através de (1.2) ela induz uma estrutura de autocorrelação entre os estimadores y_t que pode ser aproveitada na estimação, tanto no caso de amostras independentes quanto no de amostras sobrepostas. Nesta proposta, $\{\theta_t\}$ é assumido estacionário, ou que possa ser transformado em um processo estacionário usando diferenciação.

Segundo Scott et al. (1974), se as amostras foram independentes, o estimador linear de menor EQM para θ_t sob o modelo 1.2 tem a forma:

$$\hat{\theta}_t = (1 - \pi_t)y_t + \pi_t\hat{y}_t, \quad (1.3)$$

onde \hat{y}_t é a melhor previsão linear de y_t que pode ser feita usando as estimativas obtidas nos períodos anteriores, arranjadas no vetor $\mathbf{Y}_{t-1} = \{y_{t-1}, y_{t-2}, \dots, y_1\}$, e $\pi_t = \sigma_t^2 / V(y_t | \mathbf{Y}_{t-1})$ é a relação ruído-sinal do sistema.

Um marco unificado de trabalho para estes e outros modelos foi proposto por Jones (1980), arranjando convenientemente os y_t, θ_t e e_t nos períodos observados numa formulação matricial e aplicando o método de mínimos quadrados estocásticos para encontrar o estimador linear de menor EQM para θ_t . Embora esta abordagem permita estender a ideia de estimação em levantamentos repetidos usando séries temporais a um conjunto muito mais amplo de modelos para θ_t , como salienta o próprio autor, nem sempre é possível achar métodos recursivos de cálculo como os apresentados em Scott et al. (1974) ou Blight e Scott (1973), o que na prática, limitou a implementação desta metodologia.

Na busca de formulações que aproveitassem a informação prévia mantendo a simplicidade nos cálculos de estimativas, vários autores, entre eles o próprio Jones (1980), Tam (1987) e Binder e Hidioglou (1988), propuseram o uso de MEE e o filtro de Kalman, assumindo que $\{\theta_t\}$ é um processo markoviano de primeira ordem. Estas duas ferramentas serão apresentadas na próxima seção.

1.3 Modelos de espaço de estados - MEE

Os MEE têm sido aplicados em diversos campos, dentre eles na modelagem de séries temporais, como pode ser visto em Harvey (1989), Durbin e Koopman (2001) ou Morettin e Toloi (2006, cap. 13). Entre as maiores vantagens com relação a metodologias mais conhecidas como os modelos ARIMA (Box et al., 1994) estão: i) a possibilidade de se fazer uma análise estrutural das séries, modelando separadamente componentes como tendência, sazonalidade, ciclos, etc.; e ii) a facilidade de incluir variáveis explicativas no modelo, ou considerar dados multivariados, como extensões naturais dos casos mais simples.

Como destacado no livro de Durbin e Koopman (2001, Sec. 3.3), os MEE são suficientemente flexíveis para contemplar toda a classe dos modelos ARIMA e permitir modificações, ao longo do tempo, na estrutura assumida para a evolução da série temporal; fato particularmente útil quando os mecanismos responsáveis da evolução do fenômeno em estudo mudam durante o tempo considerado para análise. Estudos específicos sobre o uso destes modelos na análise de séries temporais podem ser encontrados nos livros de Harvey (1989) ou Durbin e Koopman (2001), ou, sob nome de *Modelos Lineares Dinâmicos*, no de West e Harrison (1997), onde é empregada uma abordagem bayesiana.

Suponha que se deseja fazer inferência sobre um processo multivariado denotado por $\{\alpha_t\}$, mas que é observado apenas indiretamente através da série temporal multivariada $\{y_t\}$, com erro de observação $\{e_t\}$, onde para cada t , y_t é um estimador não viesado do parâmetro de interesse, $Z_t\alpha_t$. Sob suposição que $\{\alpha_t\}$ é markoviano de primeira ordem, o modelo:

$$\begin{aligned} y_t &= Z_t\alpha_t + e_t & E(e_t) &= 0; & E(e_te_t') &= \Sigma_{tt} \\ \alpha_t &= T_t\alpha_{t-1} + G_t\eta_t & E(\eta_t) &= 0; & E(\eta_t\eta_t') &= Q_t, \end{aligned} \tag{1.4}$$

é sua representação em espaço de estados. A primeira linha de (1.4) é denominada *equação de observação* e a segunda *equação de evolução* ou *de transição*; α_t é conhecido como *vetor de estados* e as matrizes Z_t , Σ_{tt} , T_t , G_t e Q_t se denominam *matrizes do sistema*.

Geralmente, assume-se que os termos de erro \mathbf{e}_t e $\boldsymbol{\eta}_t$ satisfazem

$$E(\mathbf{e}_\tau \mathbf{e}_t') = \mathbf{0} \quad \text{e} \quad E(\boldsymbol{\eta}_\tau \boldsymbol{\eta}_t') = \mathbf{0} \quad \forall \tau \neq t \quad \text{e} \quad E(\mathbf{e}_\tau \boldsymbol{\eta}_t') = \mathbf{0} \quad \forall \tau, t.$$

Assume-se também que o vetor de estados inicial $\boldsymbol{\alpha}_0$ tem esperança \mathbf{a}_0 e variância \mathbf{P}_0 e é independente de \mathbf{e}_t e $\boldsymbol{\eta}_t$ para todo t . A não autocorrelação dos \mathbf{e}_t será flexibilizada posteriormente para modelar dados provenientes de levantamentos repetidos, nos casos em que exista sobreposição nas amostras dos distintos períodos.

Note que em (1.4), só \mathbf{y}_t é necessariamente observado, e que no caso de uma série temporal univariada, tanto y_t quanto e_t são escalares para cada t . Nesta formulação, as matrizes do sistema podem variar no tempo, mas são assumidas como não estocásticas e em particular, independentes dos \mathbf{y}_i para $i = 1, 2, \dots, t$.

Na literatura, existem formulações alternativas destes modelos, com e sem covariáveis, e podendo adotar diferentes hipóteses como, por exemplo: normalidade das sequências de erro, matrizes do sistema constantes no tempo⁸ ou considerando a possibilidade de correlação entre os \mathbf{e}_t ou entre estes e os $\boldsymbol{\eta}_t$. Da mesma forma, a matriz \mathbf{G}_t não é sempre explicitada. No entanto, sua inclusão é conveniente no caso em que alguns dos elementos de $\boldsymbol{\alpha}_t$ não possuem um erro associado na equação de evolução.

Note, no entanto, que a representação de um modelo na forma de espaço de estados não é única. Dois modelos particulares foram empregados nesta dissertação: a representação em espaço de estados de um modelo ARMA, univariado ou multivariado, que se encontra no Apêndice A; e o *Modelo Estrutural Básico* (BSM) como em Harvey (1989, p.47, p.172), que visa produzir uma decomposição estrutural da série $\theta_t = \mathbf{Z}_t \boldsymbol{\alpha}_t$ e é apresentado na Subseção 1.3.1.

1.3.1 Modelo estrutural básico (BSM)

Seja y_t uma série univariada mensal com sazonalidade anual e e_t o seu erro de observação associado, o BSM está definido pelas equações:

⁸O que é conhecido como um modelo invariante.

$$\begin{aligned}
 y_t &= L_t + S_t + e_t \\
 L_t &= L_{t-1} + R_{t-1} + \eta_{Lt} \\
 R_t &= R_{t-1} + \eta_{Rt} \\
 S_t &= - \sum_{j=1}^{11} S_{t-j} + \eta_{St},
 \end{aligned} \tag{1.5}$$

a componente L_t corresponde ao nível local, que evolui no tempo segundo a inclinação local, R_t ; S_t é a componente sazonal modelada através de variáveis indicadoras, assumida como estocástica se $V(\eta_{St}) > 0$ e determinística, caso contrário. Também pode-se incluir uma componente irregular no vetor de estados, denotada por I_t e modelada através de um ruído branco⁹.

Como alternativa para a decomposição da sazonalidade pode-se empregar uma formulação trigonométrica dada por:

$$\begin{aligned}
 S_t &= \sum_{j=1}^6 S_{j,t} \\
 S_{j,t} &= S_{j,t-1} \cos(\lambda_j) + S_{j,t-1}^* \sin \lambda_j + \omega_{j,t} \\
 S_{j,t}^* &= S_{j,t-1}^* \cos(\lambda_j) - S_{j,t-1} \sin \lambda_j + \omega_{j,t}^*,
 \end{aligned}$$

sendo $\lambda_j = \frac{\pi j}{6}$, e $\omega_{j,t}$ e $\omega_{j,t}^*$ ruídos brancos. A sazonalidade será determinística se $V(\omega_{j,t}) = V(\omega_{j,t}^*) = 0 \quad \forall j, t$, e estocástica, caso contrário.

O modelo em (1.5) com componente irregular, pode ser escrito na notação de (1.4) empregando as matrizes:

$$\begin{aligned}
 \mathbf{Z} &= \begin{bmatrix} 1 & 0 & 1 & 0 & \dots & 0 & 1 \end{bmatrix}, \\
 \boldsymbol{\alpha}'_t &= \begin{bmatrix} L_t & R_t & S_t & S_{t-1} & \dots & S_{t-10} & I_t \end{bmatrix},
 \end{aligned}$$

⁹No entanto, nem sempre é possível estimar as variâncias tanto dos e_t quanto dos I_t via máxima verossimilhança com o filtro de Kalman. Por exemplo, este modelo não é identificável se os dois componentes tiverem distribuições gaussianas. Uma forma de contornar este problema é estimar a variância dos e_t por outro método diferente ao filtro de Kalman e, posteriormente, assumir esta variância como conhecida para estimar por máxima verossimilhança a variância dos I_t empregando o filtro, como será apresentado na Subseção 3.1.

$$\mathbf{T} = \left[\begin{array}{cc|ccccc|c} 1 & 1 & & & & & & 0 \\ 0 & 1 & & & & & & 0 \\ \hline & & \mathbf{0}_{(2 \times 11)} & & & & & \\ \hline & & -1 & -1 & \dots & -1 & -1 & 0 \\ & & 1 & 0 & \dots & 0 & 0 & 0 \\ \mathbf{0}_{(11 \times 2)} & & 0 & 1 & \dots & 0 & 0 & 0 \\ & & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ & & 0 & 0 & \dots & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \end{array} \right], \quad \mathbf{G} = \left[\begin{array}{c|c} & 0 \\ & 0 \\ & 0 \\ \hline & \\ \mathbf{0}_{(10 \times 3)} & \mathbf{0}_{(10 \times 1)} \\ \hline 0 & 0 & 0 & 1 \end{array} \right],$$

$$\boldsymbol{\eta}'_t = \begin{bmatrix} \eta_{L_t} & \eta_{R_t} & \eta_{S_t} & I_t \end{bmatrix},$$

onde $\mathbf{0}_{m \times n}$ e \mathbf{I}_m correspondem à matriz zero de tamanho $(m \times n)$ e à identidade de tamanho m , respectivamente. Note que se as matrizes $\boldsymbol{\Sigma}_{tt}$ e \mathbf{Q}_t foram constantes, este modelo resulta invariante no tempo.

No caso de uma decomposição trigonométrica da sazonalidade, uma possível representação matricial esta dada por:

$$\begin{aligned} \boldsymbol{\alpha}'_t &= \begin{bmatrix} L_t & R_t & S_{1,t} & S_{1,t}^* & \dots & S_{6,t} & I_t \end{bmatrix}, \\ \mathbf{Z} &= \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & \dots & 1 & 1 \end{bmatrix}, \\ \mathbf{G} &= \mathbf{I}_{14}, \\ \boldsymbol{\eta}'_t &= \begin{bmatrix} \eta_{L_t} & \eta_{R_t} & \eta_{S_{1,t}} & \eta_{S_{1,t}^*} & \dots & \eta_{S_{6,t}} & I_t \end{bmatrix}, \end{aligned}$$

com \mathbf{T} definida anteriormente, exceto pelo seu bloco central que corresponde a:

O filtro de Kalman

$$\begin{bmatrix} \mathbf{C}_1 & & & & \\ & \mathbf{C}_2 & & & \\ & & \ddots & & \\ & & & \mathbf{C}_5 & \\ 0 & & & & -1 \end{bmatrix}, \quad \text{onde } \mathbf{C}_j = \begin{bmatrix} \cos \lambda_j & \sin \lambda_j \\ -\sin \lambda_j & \cos \lambda_j \end{bmatrix}.$$

Como já mencionado, a estimação do vetor de estados α_t em cada momento do tempo é feita empregando um algoritmo denominado Filtro de Kalman, que recebe o nome de *filtro* devido a seu surgimento na teoria de extração de sinais na presença de ruído. As equações que descrevem este processo e algumas das suas características são apresentados a seguir.

1.4 O filtro de Kalman

O filtro de Kalman (Kalman, 1960) é um algoritmo recursivo empregado para levar a cabo a estimação do vetor de estados num MEE. Quando as matrizes do sistema são conhecidas, ele permite calcular os melhores preditores lineares no sentido do EQM para α_t e y_t e suas matrizes de covariâncias, dado um conjunto de estimativas disponíveis até o momento $t - 1$. Também, quando disponível a estimativa para o momento t , y_t , pode-se atualizar a previsão anterior incluindo a nova informação e obter o estimador linear de menor EQM para α_t e sua matriz de covariâncias.

1.4.1 Equações do filtro de Kalman

Considere novamente o modelo (1.4). Denote por $\mathbf{Y}_{t-1} = \{y_{t-1}, y_{t-2}, \dots, y_1\}$ o conjunto de informações disponíveis até o momento $t - 1$ e suponha que o vetor de estados inicial α_0 tem média \mathbf{a}_0 e variância \mathbf{P}_0 conhecidas, independentes de \mathbf{e}_t e η_t para todo t . O filtro de Kalman é descrito pelas equações:

Previsão para α_t e y_t : os preditores lineares de menor EQM para α_t e y_t , baseados em Y_{t-1} estão dados por:

$$\hat{\alpha}_{t|t-1} = T_t \hat{\alpha}_{t-1}; \quad \hat{y}_t = Z_t \hat{\alpha}_{t|t-1},$$

com matrizes de covariância:

$$P_{t|t-1} = T_t P_{t-1} T_t' + G_t Q_t G_t', \quad e \\ F_t = Z_t P_{t|t-1} Z_t' + \Sigma_{tt},$$

respectivamente.

Atualização da previsão para α_t : Sendo $Y_t = \{y_t, Y_{t-1}\}$, o estimador linear de menor EQM para α_t baseado em Y_t e sua matriz de covariância, P_t são:

$$\hat{\alpha}_t = \hat{\alpha}_{t|t-1} + K_t \nu_t; \quad \text{com} \quad K_t = P_{t|t-1} Z_t' F_t^{-1} \quad e \quad \nu_t = y_t - Z_t \hat{\alpha}_{t|t-1}, \quad e \\ P_t = P_{t|t-1} - K_t Z_t P_{t|t-1}.$$

Aspectos relacionados com estes estimadores serão mencionados a seguir. As demonstrações destas equações podem ser encontradas em Harvey (1989, págs. 109-110), ou Durbin e Koopman (2001, págs. 65-67).

Como mencionado, sob a hipótese que as matrizes do sistema são conhecidas, o estimador $\hat{\alpha}_t$ resultante do filtro de Kalman é o de menor EQM na família dos estimadores lineares não viesados de α_t , baseados em Y_t . Além disso, sob a hipótese de normalidade das sequências de erro $\{e_t\}$ e $\{\eta_t\}$ e do vetor de estados inicial α_0 , este estimador é a esperança da distribuição a posteriori de α_t , sendo então o de menor EQM no sentido geral.

Por outro lado, veja que:

$$\begin{aligned} \hat{\alpha}_t &= \hat{\alpha}_{t|t-1} + K_t \nu_t; \\ &= \hat{\alpha}_{t|t-1} + K_t (y_t - Z_t \hat{\alpha}_{t|t-1}) \\ &= (I - K_t Z_t) \hat{\alpha}_{t|t-1} + K_t y_t. \end{aligned}$$

Assim, $\hat{\alpha}_t$ pode ser calculado recursivamente como uma média ponderada entre a pre-

O filtro de Kalman

visão para α_t feita com base na informação disponível até $t - 1$ e a nova informação y_t . A ponderação que recebe y_t , determinada pela matriz K_t , é conhecida como *ganho de Kalman* e depende da relação entre as variâncias dos erros, de observação e evolução, do sistema definido por (1.4).

1.4.2 Decomposição da verossimilhança através dos erros de previsão

Os *erros de previsão*, $\nu_t = y_t - Z_t \hat{\alpha}_{t|t-1}$, também conhecidos como *inovações*, representam a parte de y_t que não pode ser explicada a partir da informação existente no modelo até o momento $t - 1$. Além disso, note que sob hipótese de normalidade das sequências de erros e do vetor de estados inicial, $(y_t | Y_{t-1})$ tem distribuição normal com média $Z_t \hat{\alpha}_{t|t-1}$ e variância F_t ; com isto, os ν_t formam uma sequência de variáveis aleatórias independentes com média zero e variância F_t .

Assim, através das inovações ν_t , é possível decompor a verossimilhança do modelo de uma forma computacionalmente conveniente, com o objetivo de estimar seus componentes desconhecidos, tais como elementos das matrizes Q ou Σ ou ainda, no caso de modelos ARMA como apresentados no Apêndice A, elementos das matrizes T e G . Em Durbin e Koopman (2001, cap. 7) é apresentada uma revisão mais detalhada deste tema.

1.4.3 Inicialização e aspectos computacionais

Foi suposto em (1.4.1) que a média e a variância do vetor de estados inicial α_0 fossem conhecidas, mas na prática nem sempre contamos com este tipo de informação. Duas abordagens simples podem ser utilizadas para contornar este problema:

- Fixar P_0 como cI , um múltiplo da matriz identidade, com c grande para refletir a incerteza da informação inicial. Note que a_0 pode ser fixado em 0 pois não tem importância prática para c suficientemente grande.
- Assumir α_0 como fixa, mas desconhecida, e estimá-la via máxima verossimilhança a partir Y_t .

Nos modelos mais simples, como salientado por Durbin e Koopman (2001, p.118), a sequência de estimadores $\hat{\alpha}_t$ obtidos pelos dois métodos não difere muito para séries com um número grande de observações. Um estudo completo sobre estes e outros métodos de inicialização pode ser encontrado no Capítulo 5 de Durbin e Koopman (2001).

Por outro lado, nem sempre a formulação apresentada é a forma mais eficiente de implementar o filtro de Kalman computacionalmente. Segundo Harvey (1989), entre os algoritmos alternativos para sua implementação se encontram: i) o *information filter* ou *filtro de informação*, que estima a matriz de informação \mathbf{P}_t^{-1} e $\hat{\mathbf{b}}_t = \mathbf{P}_t^{-1}\hat{\alpha}_t$ em lugar de \mathbf{P}_t e $\hat{\alpha}_t$; e ii) o *filtro da raiz quadrada*, que decompõe a matriz \mathbf{P}_t como $\mathbf{P}_t = \mathbf{P}_t^* \mathbf{P}_t^{*'}.$ A primeira formulação é conveniente nos casos em que não se conta com informação inicial e vai ser empregado o método com $\mathbf{P}_0 = c\mathbf{I}$, enquanto o segundo permite contornar situações onde \mathbf{P}_t pode ser não definida positiva, o que pode acontecer por erros de arredondamento quando a escala de \mathbf{Q}_t é muito pequena em relação à de Σ_{tt} .

O filtro de Kalman

Capítulo 2

Modelagem do erro amostral

Como mencionado anteriormente, distintos tipos de estudo são empregados para estimação em pesquisas repetidas. Em alguns casos, as amostras são planejadas para garantir um certo nível de sobreposição de unidades entre períodos específicos, como nos casos de estudos de painel ou as amostras com rotação. Em outros casos, principalmente quando se deseja estimar médias que envolvam vários períodos de tempo, podem ser de interesse amostras independentes, ou ainda, mutuamente exclusivas, com o objetivo de acumulá-las para detectar subpopulações de baixa frequência.

Nos casos em que as amostras das diferentes repetições da pesquisa não são independentes, a série de estimadores de planejamento¹ usualmente empregados para o nível do indicador, mesmo que baseados só na informação da amostra de cada período, apresentam autocorrelação, como será ilustrado na Seção 2.1. Outros estimadores que aproveitam as informações de vários períodos, por exemplo, os estimadores compostos, têm causas adicionais para serem autocorrelacionados.

Retomando a notação do capítulo anterior, suponha que existe uma série univariada de estimativas $\{y_t\}$ tal que para $t = 1, 2, \dots$

$$y_t = \theta_t + e_t,$$

¹Entende-se por *estimador de planejamento*, um estimador no marco da inferência clássica, baseada no planejamento da amostra. Estes estimadores também são conhecidos como *estimadores diretos* do desenho amostral.

Sobreposição de unidades e autocorrelação

onde θ_t é o parâmetro de interesse no momento t e e_t é o erro amostral associado a y_t , com esperança igual a zero. No caso de existir uma estrutura de autocorrelação na série de estimadores, esta estrutura é refletida também pelos erros amostrais, pois, como apresentado em Silva e Cruz (2002),

$$\begin{aligned}\gamma_{y|\theta}(k) &= \text{Cov}(y_t, y_{t-k} | \theta_t, \theta_{t-k}) \\ &= \text{Cov}(\theta_t + e_t, \theta_{t-k} + e_{t-k} | \theta_t, \theta_{t-k}) \\ &= \text{Cov}(e_t, e_{t-k} | \theta_t, \theta_{t-k}) = \gamma_{e|\theta}(k),\end{aligned}$$

supondo que o processo em questão seja estacionário de segunda ordem.

No entanto, a proposta de estimação baseada no filtro de Kalman, como introduzida na Seção 1.4, pressupõe que os $\{e_t\}$ são um ruído branco, o que geralmente não é verdadeiro no caso de pesquisas com amostras não independentes. Existem na literatura propostas para adaptar este filtro quando os erros da equação de observação são autocorrelacionados, tema que será abordado no Capítulo 3. Porém, estes métodos precisam que a estrutura de autocorrelação dos $\{e_t\}$ seja identificada, funcionalmente, mediante um modelo, ou pelo menos pontualmente. Com esta finalidade será apresentada na Seção 2.2 a metodologia proposta por Pfeiffermann et al. (1998) para amostras com rotação. Posteriormente, no Capítulo 4 este método será avaliado mediante simulação.

2.1 Sobreposição de unidades e autocorrelação

Com o objetivo de esclarecer porque, sob a abordagem de inferência baseada no planejamento de amostra², a sobreposição de unidades entre as amostras numa pesquisa repetida introduz covariância entre os estimadores dos distintos períodos, vamos ilustrar a estimação do total de uma variável de interesse nos casos de amostras independentes e com sobreposição de unidades. Empregaremos a notação de Särndal et al. (1992, Capítulo 1, seções 2.3 a 2.6 e 2.8), como resumido a seguir³.

²Isto é, sob a abordagem clássica de inferência em amostragem de populações finitas.

³São mudados os símbolos t e \hat{t} empregados em Särndal et al. (1992) para o parâmetro de interesse e seu estimador com a finalidade de evitar confusões com a notação utilizada para o período da pesquisa.

Seja U uma população finita de tamanho N , para a qual se quer estimar o total de uma variável no momento t , utilizando uma amostra aleatória probabilística. Denotaremos por $y_{t,i}$ a observação do i -ésimo indivíduo nesse momento e por $\theta_t = \sum_{i=1}^N y_{t,i}$ o parâmetro de interesse. Formalmente, o processo de seleção de amostra pode ser modelado pela tríade $(U, \mathcal{S}, p_t(\cdot))$, onde U é a população, \mathcal{S} é o conjunto das 2^N possíveis amostras de U e $p_t(\cdot)$, denominado *plano amostral*, é a função que associa a cada amostra $s \in \mathcal{S}$ uma probabilidade de ocorrer.

Neste enfoque de inferência, é suposto que todas as amostras de \mathcal{S} têm uma probabilidade conhecida de serem selecionadas e a única componente aleatória está dada por qual delas será escolhida. Como estimador empregaremos $y_t = \sum_{i \in s} w_{t,i} y_{t,i}$, com $w_{t,i}$ uma variável de ponderação. Introduzindo, para $i = 1, \dots, N$ a variável aleatória:

$$I_{t,i} = \begin{cases} 1 & \text{quando o } i\text{-ésimo indivíduo pertence à amostra selecionada no } t\text{-ésimo período,} \\ 0 & \text{em caso contrário,} \end{cases}$$

é possível reescrever convenientemente $y_t = \sum_{i \in U} I_{t,i} w_{t,i} y_{t,i}$ e basear a inferência na distribuição das $I_{t,i}$ induzida por $p_t(\cdot)$. A variável $I_{t,i}$ é denominada, *indicadora do i -ésimo elemento na amostra s_t* .

A seguir, é ilustrado o impacto na covariância dos estimadores, produzido pela seleção de amostras independentes ou não, nas distintas ocasiões da pesquisa. Para isto, serão considerados os seguintes cenários: i) amostras independentes, ii) amostras com sobreposição total de unidades, iii) amostras com sobreposição parcial de unidades, e iv) amostras com rotação.

Por simplicidade, suponha que a população U seja composta pelas mesmas unidades durante dois períodos de tempo, nos quais é pesquisada utilizando amostragem aleatória simples sem reposição e o π -estimador ou estimador de Horvitz-Thompson (Särndal et al., 1992, pág. 42). Assim, denote por $\theta_1 = \sum_{i \in U} y_{1,i}$ o parâmetro de interesse no primeiro período, $I_{1,i}$ a indicadora do indivíduo i na amostra desse período, s_1 e

$$y_{1\pi} = \sum_{i \in s_1} \frac{y_{1,i}}{E(I_{1,i})},$$

Sobreposição de unidades e autocorrelação

o π -estimador correspondente. Veja que sob este plano amostral, no caso em que n unidades são selecionadas,

$$E(I_{1,i}) = P(I_{1,i} = 1) = \sum_{\substack{s_1 \in \mathcal{S} \\ i \in s_1}} p_1(s_1) = \frac{n}{N},$$

com o que o π -estimador é

$$y_{1\pi} = \frac{N}{n} \sum_{i \in s_1} y_{1,i}.$$

Veja também que

$$E(I_{1,i}I_{1,j}) = \sum_{\substack{s_1 \in \mathcal{S} \\ i,j \in s_1}} p_1(s_1) = \begin{cases} \frac{n}{N} \frac{n-1}{N-1} & \text{se } i \neq j \\ \frac{n}{N} & \text{se } i = j, \end{cases}$$

o que será empregado para calcular a variância do estimador. Definições análogas são utilizadas para o segundo período.

2.1.1 Amostras independentes

No caso em que as duas amostras são selecionadas independentemente, usando em cada caso amostragem aleatória simples sem reposição, de tamanho n , a covariância entre os π -estimadores dos dois momentos está dada por:

$$\text{Cov}(y_{1\pi}, y_{2\pi}) = \text{Cov} \left[\frac{N}{n} \sum_{i \in s_1} y_{1,i}, \frac{N}{n} \sum_{j \in s_2} y_{2,j} \right] = \frac{N^2}{n^2} \sum_{i \in U} \sum_{j \in U} y_{1,i} y_{2,j} \text{Cov}(I_{1,i}, I_{2,j}),$$

onde $I_{1,i}$ e $I_{2,j}$ são as indicatoras correspondentes aos indivíduos i e j nas amostras s_1 e s_2 , respectivamente. Como as amostras são independentes, as indicatoras $I_{1,i}$ e $I_{2,j}$ têm covariância igual a zero para todo i, j , logo

$$\text{Cov}(y_{1\pi}, y_{2\pi}) = 0.$$

Note que isto continua sendo verdade, mesmo que por acaso exista alguma sobreposição de unidades. Sob este plano amostral isto acontece com probabilidade $1 - \left[\binom{N-n}{n} / \binom{N}{n}\right]$, sendo pouco possível para N muito maior do que n .

2.1.2 Amostras com sobreposição total de unidades

Suponha agora que para o segundo período se emprega a mesma amostra do primeiro período, isto é $s_2 = s_1$; assim

$$\text{Cov}(y_{1\pi}, y_{2\pi}) = \text{Cov} \left[\frac{N}{n} \sum_{i \in s_1} y_{1,i}, \frac{N}{n} \sum_{j \in s_1} y_{2,j} \right] = \frac{N^2}{n^2} \sum_{i \in U} \sum_{j \in U} y_{1,i} y_{2,j} \text{Cov}(I_{1,i}, I_{1,j}),$$

mas

$$\begin{aligned} \text{Cov}(I_{1,i}, I_{1,j}) &= E(I_{1,i} I_{1,j}) - E(I_{1,i})E(I_{1,j}) \\ &= \begin{cases} -\frac{n}{N^2} \left(\frac{N-n}{N-1} \right) & \text{se } i \neq j \\ \frac{n}{N^2} (N-n) & \text{se } i = j, \end{cases} \end{aligned}$$

logo,

$$\begin{aligned} \text{Cov}(y_{1\pi}, y_{2\pi}) &= \frac{N-n}{n} \frac{1}{N-1} \left[(N-1) \sum_{i \in U} y_{1,i} y_{2,j} - \sum_{i \in U} y_{1,i} \sum_{\substack{j \in U \\ j \neq i}} y_{2,j} \right] \\ &= \frac{N-n}{n} \frac{1}{N-1} \left[(N-1) \sum_{i \in U} y_{1,i} y_{2,j} - \sum_{i \in U} y_{1,i} (\theta_2 - y_{2,i}) \right] \quad (2.1) \\ &= \frac{N}{n} (N-n) \left[\frac{1}{N-1} \sum_{i \in U} (y_{1,i} - \bar{y}_1)(y_{2,i} - \bar{y}_2) \right] \\ &= \frac{N}{n} (N-n) S_{y_1 y_2 U}, \end{aligned}$$

onde \bar{y}_1 e \bar{y}_2 são, respectivamente, as médias das observações no primeiro e segundo períodos e $S_{y_1 y_2 U}$ é sua covariância observada. Assim, no caso de existir uma associação

Sobreposição de unidades e autocorrelação

positiva entre os valores da variável de interesse nos dois períodos de tempo considerados, ela vai ser refletida pelos estimadores, o que pode ser empregado, por exemplo, para melhorar a estimação de $\theta_2 - \theta_1$.

2.1.3 Um exemplo de amostra com sobreposição parcial de unidades

Para exemplificar o caso de amostras com sobreposição parcial de unidades, suponha que s_2 é uma amostra aleatória simples sem reposição, de tamanho $n\mu$, selecionada de s_1 , com $0 < \mu < 1$, e que a estimação do segundo período será feita com o estimador

$$y_{2\pi} = \frac{N}{n\mu} \sum_{i \in s_2} y_{2,i}.$$

A covariância dos estimadores $y_{1\pi}$ e $y_{2\pi}$ esta dada por:

$$\text{Cov}(y_{1\pi}, y_{2\pi}) = \text{Cov} \left[\frac{N}{n} \sum_{i \in s_1} y_{1,i}, \frac{N}{n\mu} \sum_{j \in s_2} y_{2,j} \right] = \frac{N^2}{n^2\mu} \sum_{i \in U} \sum_{j \in U} y_{1,i} y_{2,j} \text{Cov}(I_{1,i}, I_{2,j}),$$

onde $I_{1,i}$ e $I_{2,j}$ são as indicadoras relativas aos indivíduos i e j nas amostras s_1 e s_2 , respectivamente.

Veja que

$$\begin{aligned} \text{Cov}(I_{1,i}, I_{2,j}) &= E(I_{1,i}I_{2,j}) - E(I_{1,i})E(I_{2,j}) \\ &= P(I_{1,i}I_{2,j} = 1) - P(I_{1,i} = 1)P(I_{2,j} = 1) \\ &= \begin{cases} P(I_{2,j} = 1|I_{1,j} = 1; I_{1,i} = 1)P(I_{1,j} = 1; I_{1,i} = 1) - P(I_{1,i} = 1)P(I_{2,j} = 1) & \text{se } i \neq j \\ P(I_{2,i} = 1|I_{1,i} = 1)P(I_{1,i} = 1) - P(I_{1,i} = 1)P(I_{2,i} = 1) & \text{se } i = j \end{cases} \\ &= \begin{cases} -\frac{n\mu}{N^2} \frac{(N-n)}{N-1} & \text{se } i \neq j \\ \frac{n\mu}{N^2} (N-n) & \text{se } i = j, \end{cases} \end{aligned}$$

assim,

$$\begin{aligned} \text{Cov}(y_{1\pi}, y_{2\pi}) &= \frac{N^2}{n^2\mu} \sum_{i \in U} \sum_{j \in U} y_{1,i} y_{2,j} \text{Cov}(I_{1,i}, I_{2,j}), \\ &= \frac{\mu(N-n)}{n} \frac{1}{N-1} \left[(N-1) \sum_{i \in U} y_{1,i} y_{2,i} - \sum_{i \in U} y_{1,i} \sum_{\substack{j \in U \\ j \neq i}} y_{2,j} \right]. \end{aligned}$$

Procedendo como na equação (2.1), pode-se ver que

$$\text{Cov}(y_{1\pi}, y_{2\pi}) = \mu \frac{N}{n} (N-n) S_{y_1 y_{2U}}.$$

Então, neste caso específico, a covariância existente entre os estimadores dos dois períodos corresponde a μ vezes a que poderia ser obtida no caso de sobreposição total, com $0 \leq \mu \leq 1$.

2.1.4 Um exemplo de amostra com rotação

Para selecionar amostras com sobreposição parcial de unidades entre dois períodos, mantendo o tamanho de amostra n constante ao longo do tempo, uma alternativa é fazer $s_2 = s_{2,1} \cup s_{2,2}$, onde $s_{2,1}$ é uma amostra de $n\mu$ unidades selecionada de s_1 e $s_{2,2}$ é uma amostra de $n(1-\mu)$ unidades selecionada de $s_1^c = U - s_1$, com $0 < \mu < 1$. No entanto, a estrutura de probabilidades condicionais para as variáveis $I_{t,i}$ na qual se baseia a inferência neste esquema, ilustrada na seção anterior, pode resultar complexa no caso de uma pesquisa com muitos períodos.

Uma outra possibilidade nesses casos é dividir a população em grupos disjuntos e selecionar independentemente uma amostra de cada um deles no primeiro período. Para as amostras dos períodos subsequentes, alguns dos grupos podem ser mantidos em painel, como na Subseção 2.1.2, enquanto nos outros grupos são selecionadas amostras independentes. Este procedimento, descrito anteriormente na Subseção 1.1.3, permite induzir esquemas complexos de sobreposição, sem dificultar o cálculo das esperanças

Sobreposição de unidades e autocorrelação

e covariâncias das indicadoras $I_{t,i}$ dos distintos períodos, usualmente empregadas no procedimento de estimação.

Para exemplificar esta situação suponha que a população U é dividida em 2 grupos disjuntos de igual tamanho, U_1 e U_2 ; de cada um deles é selecionada independentemente uma amostra aleatória simples de tamanho $n/2$ para o primeiro período, sendo então $s_1 = s_{11} \cup s_{12}$, com s_{11} a amostra do primeiro período para o primeiro grupo e s_{12} a correspondente a este mesmo período no segundo grupo. Para o seguinte período, a amostra do primeiro grupo é mantida em painel enquanto que a do segundo grupo é trocada por uma nova, s_{22} , selecionada de U_2 independentemente de s_{12} . Se usarmos como estimador para θ_1 ,

$$y_{1\pi} = \frac{N}{n} \sum_{i \in s_{11}} y_{1,i} + \frac{N}{n} \sum_{j \in s_{12}} y_{1,j},$$

e para para θ_2 ,

$$y_{2\pi} = \frac{N}{n} \sum_{i \in s_{11}} y_{2,i} + \frac{N}{n} \sum_{j \in s_{22}} y_{2,j},$$

temos

$$\text{Cov}(y_{1\pi}, y_{2\pi}) = \text{Cov} \left[\frac{N}{n} \sum_{i \in s_{11}} y_{1,i}, \frac{N}{n} \sum_{j \in s_{11}} y_{2,j} \right],$$

pois as outras covariâncias são zero. Como s_{11} é um painel, podemos usar o resultado da Subseção 2.1.2 para obter:

$$\begin{aligned} \text{Cov}(y_{1\pi}, y_{2\pi}) &= \frac{N(N-n)}{n} \frac{1}{2} \left[\frac{1}{(N/2) - 1} \sum_{i \in U_1} (y_{1,i} - \bar{y}_{1U_1})(y_{2,i} - \bar{y}_{2U_1}) \right] \\ &= \frac{N(N-n)}{n} \frac{1}{2} S_{y_1 y_2 U_1}, \end{aligned}$$

onde \bar{y}_{1U_1} e \bar{y}_{2U_1} são as médias das $y_{1,i}$ e $y_{2,i}$ para $i \in U_1$. Assim, no caso de existir uma associação positiva entre estes valores nos indivíduos em U_1 , este esquema permite melhorar a precisão da estimação de $\theta_2 - \theta_1$ em relação ao caso de amostras independentes, e ao mesmo tempo, renovar a amostra periodicamente com a finalidade de evitar os vícios associados aos esquemas de painel referidos na Subseção 2.1.2.

2.2 Estimação das autocovariâncias dos e_t . Método dos pseudo-erros

Como mencionado anteriormente, quando numa pesquisa repetida as amostras dos distintos períodos não são independentes, os erros amostrais $e_t = y_t - \theta_t$ associados aos estimadores usualmente empregados apresentam autocorrelação. Nestes casos, para a implementação de uma metodologia de estimação usando séries temporais como a introduzida na Seção 1.4, é preciso identificar a forma desta autocorrelação, o que não é imediato, dado que os e_t constituem uma série não observável.

Para amostras com rotação num esquema $r_1 - r_2(m)$, segundo definidas na Subseção 1.1.3, Pfeiffermann et al. (1996, 1998) propuseram uma metodologia de estimação pontual da função de autocorrelação dos erros amostrais, $\rho(k)$, baseada nas covariâncias observadas entre os *pseudo-erros*, definidos como os desvios de estimativas calculadas a partir de cada painel com relação à estimativa total. Em linhas gerais, esta proposta está baseada em três hipóteses:

1. Para cada t , o erro amostral e_t pode ser aproximado razoavelmente a partir da média dos erros amostrais dos painéis, $e_t^{(j)} = (y_t^{(j)} - \theta_t)$, onde $y_t^{(j)}$ é uma estimativa direta do desenho amostral, calculada a partir da informação do painel que está participando na pesquisa pela j -ésima vez em t . Isto permite decompor a autocovariância de lag k da série $\{e_t\}$ como uma média das covariâncias desse mesmo lag entre os erros dos painéis envolvidos em cada momento.
2. Erros de painéis selecionados de grupos de rotação diferentes são não correlacionados.
3. A covariância entre dois erros de painel de um mesmo grupo de rotação depende só da quantidade de períodos entre eles e do número de períodos que eles participaram na pesquisa até aquele momento.

Sob estas hipóteses e dado que os pseudo-erros expressam a relação entre os erros de painel e o erro amostral de cada período, como será evidenciado na Subseção 2.2.3, as covariâncias observadas entre os pseudo-erros são empregadas para construir um

estimador de $\rho(k)$. Nesta seção, apresentaremos em detalhe esta metodologia e posteriormente, no Capítulo 5, a avaliaremos via simulação. A estrutura desta seção é a seguinte: em 2.2.1 são definidas as estimativas e erros de painel e é estabelecida a notação requerida nas seções subsequentes; em 2.2.2 são enunciadas as hipóteses sobre a estrutura da covariância dos erros de painel requeridas para o desenvolvimento do estimador; em 2.2.3 são definidos os pseudo-erros de painel e sua covariância de lag k e, por fim, em 2.2.4, é apresentado o estimador pontual proposto por Pfeiffermann et al. (1998) para a função de autocorrelação dos erros amostrais.

2.2.1 Estimativas e erros de painel

Considere uma pesquisa repetida periodicamente com a finalidade de estimar o parâmetro θ_t , para $t = 1, 2, \dots, T$, empregando uma amostra com g grupos de rotação. Mesmo que não explicitado, em Pfeiffermann et al. (1998) é assumido que a amostra de cada período está composta por g painéis, um de cada grupo de rotação, sendo que um deles está sendo pesquisado pela primeira vez, outro pela segunda vez, e assim por diante. Em particular, amostras com rotação balanceadas a duas vias segundo a definição dada na Subseção 1.1.3 satisfazem este requerimento.

Suponha que para essa pesquisa se dispõe de uma série de estimativas de planejamento, denotada por $\{y_t\}$, onde para cada t , y_t é não viesada para θ_t e envolve só informação desse período. Além dessas estimativas, suponha também que em cada período t é possível dispor de g estimativas para θ_t , cada uma calculada a partir da informação de um dos g painéis, sendo que a estimativa obtida com a amostra completa, y_t , corresponde à média das estimativas dos painéis, ao menos para uma aproximação razoável.

Para cada período t , tanto as estimativas dos painéis presentes na amostra em t quanto as dos períodos prévios, são denotadas fazendo-se referência ao número de períodos que leva, na amostra desse momento, o painel que está representando seu grupo de rotação. Assim, por exemplo, para um dado t , as estimativas desse momento são denotadas $y_t^{(j)}$, sendo que $y_t^{(1)}$ é estimativa correspondente ao painel pesquisado pela primeira vez nesse momento, $y_t^{(2)}$ a estimativa daquele pesquisado pela segunda vez e assim por diante. As estimativas dos períodos prévios, são denotadas por $y_{t-k}^{j,t}$, onde o subíndice $t-k$ indica o período, para $k = 1, 2, \dots, t-1$, e o superíndice j, t sinaliza que essa estimativa

corresponde ao mesmo grupo de rotação, cujo painel estava sendo pesquisado pela j -ésima vez em t . Dependendo do esquema de rotação escolhido e do valor de k , $y_t^{(j)}$ e $y_{t-k}^{j,t}$ podem corresponder ao mesmo painel ou não, como ilustrado a seguir.

Considere um esquema onde as unidades permanecem durante quatro meses na amostra e depois saem definitivamente. Para o grupo de rotação pesquisado pela, digamos, terceira vez em t , as estimativas dos períodos $t, t-1$ e $t-2$ são denotadas com relação a t como sendo $y_t^{(3)}$, $y_{t-1}^{3,t}$ e $y_{t-2}^{3,t}$, respectivamente. Veja que, por exemplo, $y_t^{(3)}$ e $y_{t-2}^{4,t}$ são estimativas que correspondem a distintos grupos de rotação, pois um painel não pode ser pesquisado simultaneamente pela terceira e quarta vez num mesmo período. Também, note que conforme o esquema de rotação proposto, $y_t^{(3)}$ e $y_{t-k}^{3,t}$ são obtidas do mesmo painel para $k = 0, 1$ ou 2 , e de painéis diferentes selecionados do mesmo grupo de rotação para $k > 2$.

Associados com as estimativas dos painéis, pode-se definir os *erros de painel* como $e_t^{(j)} = (y_t^{(j)} - \theta_t)$, para o momento t , e por $e_{t-k}^{j,t} = (y_{t-k}^{j,t} - \theta_{t-k})$, para os períodos anteriores a t . Sob esta definição, supondo que y_t pode ser aproximada pela média das $y_t^{(j)}$ para $j = 1, 2, \dots, g$, tem-se que

$$e_t = \frac{1}{g} \sum_{j=1}^g e_t^{(j)} \quad \text{e} \quad e_{t-k} = \frac{1}{g} \sum_{m=1}^g e_{t-k}^{m,t}.$$

Será assumido também que os $\{e_t\}$ provém de um processo estacionário de segunda ordem com função de autocovariância $\gamma(k)$.

2.2.2 Covariância de lag k dos erros de painel

Estabeleceremos agora algumas hipóteses sobre a estrutura da covariância de lag k entre os erros de painel. Suponha que:

1. Erros de painel que provém de grupos de rotação diferentes são não autocorrelacionados, isto é, $\text{Cov} \left(e_t^{(j)}, e_{t-k}^{m,t} \right) = 0$, $k \geq 0$, sempre que $m \neq j$, e
2. A covariância de lag k entre dois erros de painel do mesmo grupo de rotação, depende só de k e do número de vezes que eles tenham sido pesquisados, mais

Estimação das autocovariâncias dos e_t . Método dos pseudo-erros

não de t .

Sob estas hipóteses, observa-se que

$$\begin{aligned}
 \gamma(k) &= \text{Cov}(e_t, e_{t-k}) = \frac{1}{g^2} \text{Cov} \left(\sum_{j=1}^g e_t^{(j)}, \sum_{m=1}^g e_{t-k}^{m,t} \right) \\
 &= \frac{1}{g^2} \sum_{j=1}^g \text{Cov} \left(e_t^{(j)}, e_{t-k}^{j,t} \right) \\
 &= \frac{1}{g^2} \sum_{j=1}^g \gamma_k^{(j)},
 \end{aligned} \tag{2.2}$$

onde $\gamma_k^{(j)}$ denota a covariância entre dois pares de erros de um mesmo grupo de rotação, distanciados k períodos, o mais recente sendo pesquisado pela j -ésima vez. Na prática, a validade da primeira hipótese, que permite passar da primeira para a segunda linha de (2.2), decorre de que usualmente as amostras de grupos de rotação diferentes são independentes. A segunda hipótese, que garante a existência de um valor $\gamma_k^{(j)}$ não em função de t , tem como objetivo permitir que as covariâncias entre pares de erros variem em magnitude dependendo de: i) o tempo das unidades na amostra, o que pode estar associado com perdas progressivas de amostra ou vícios de rotação, por exemplo, e ii) se estes erros pertencem ao mesmo painel ou a painéis diferentes selecionados do mesmo grupo de rotação. Este último, como mencionado anteriormente, pode ser identificado a partir dos valores j e k .

Para melhor ilustrar esta situação, considere o exemplo da seção anterior, uma amostra com quatro grupos de rotação onde as unidades permanecem um igual número de meses na amostra e depois saem definitivamente, seguindo a estrutura de rotação dada pela Tabela 2.1(a). Para cada período, os números nas células indicam quantas vezes (incluindo essa) o painel tem participado da pesquisa. Para cada grupo de rotação, os símbolos ao redor dos números (nenhum, círculo vazio) identificam a um mesmo painel. Assim, para o grupo de rotação 3, por exemplo, um painel participa da pesquisa nos períodos 1 e 2, e outro nos 3, 4 e 5.

Para $j = 2$ e $k = 1$, por exemplo, o termo $\gamma_1^{(2)} = \text{Cov}(e_t^{(2)}, e_{t-1}^{2,t})$ representa a covariância entre os pares de erros de um mesmo grupo de rotação, com um período de distância,

Tabela 2.1: Esquema de rotação 4-0(1). Ilustração da notação introduzida para a definição de $\gamma_1^{(2)}$.

(a) Estrutura de rotação dos painéis. Por linhas: os números nas células indicam a quantidade de vezes que cada painel tem estado na amostra, incluindo esse período. Por colunas: um mesmo símbolo ao redor do número (nenhum, círculo vazio) identifica um mesmo painel.

(b) Por colunas: Esquematização dos pares de erros cuja covariância está representada por $\gamma_1^{(2)}$.

Período	Grupo de rotação			
t	1	2	3	4
1	①	4	3	2
2	②	①	4	3
3	③	②	①	4
4	④	③	②	①
5	1	④	③	②
...				

(a)

Período	Grupo de rotação			
t	1	2	3	4
1	$e_1^{2,2}$			
2	$e_2^{(2)}$	$e_2^{2,3}$		
3		$e_3^{(2)}$	$e_3^{2,4}$	
4			$e_4^{(2)}$	$e_4^{2,5}$
5				$e_5^{(2)}$
...				

(b)

indicados por colunas na Tabela 2.1(b). Note assim que $\gamma_k^{(j)}$ não é simétrica e não pode ser considerada uma função de autocovariância no sentido usual. Também, para $k > 0$, $\gamma_k^{(1)}$ mede a covariância entre erros de dois painéis diferentes para todo k , enquanto que $\gamma_k^{(3)}$ mede a covariância com lag k entre erros do mesmo painel para $k = 1, 2$; e entre erros de painéis diferentes nos outros casos, isto, independentemente do grupo de rotação ao qual esse par de erros pertençam.

Embora possa-se pensar que $\gamma_k^{(j)}$ fosse igual a zero para combinações j, k em que os erros considerados pertençam a painéis diferentes, como indicado em Pfeiffermann et al. (1998), em muitos casos a substituição de painéis ao interior de um grupo de rotação dá-se em áreas geográficas contíguas, o que cria uma correlação entre as estimativas que é preciso levar em conta.

2.2.3 Pseudo-erros de painel e sua covariância

A equação (2.2) relaciona a função de covariância dos erros amostrais com a covariância entre os erros amostrais dos painéis. No entanto, estes erros também são não observá-

Estimação das autocovariâncias dos e_t . Método dos pseudo-erros

veis pois dependem de θ que é desconhecido. Para contornar este problema, definem-se os *pseudo-erros de painel* como sendo $\check{e}_t^{(j)} = (y_t^{(j)} - y_t)$, ou seja, os desvios das estimativas dos painéis com relação à estimativa obtida da amostra total, para $t = 1, 2, \dots, T$. Estes erros são observáveis e veja que

$$\check{e}_t^{(j)} = (y_t^{(j)} - y_t) = (y_t^{(j)} - \theta_t) - (y_t - \theta_t) = (e_t^{(j)} - e_t). \quad (2.3)$$

Da mesma forma, $\check{e}_{t-k}^{j,t} = (e_{t-k}^{j,t} - e_{t-k})$. Denotaremos por $C_k^{(j)}$ o termo de covariância análogo a $\gamma_k^{(j)}$, entre os pseudo-erros de painel, e empregaremos a relação entre estes dois termos para construirmos um estimador da função de autocorrelação dos erros amostrais, $\rho(k)$, baseado nos pseudo-erros. Temos:

$$\begin{aligned} C_k^{(j)} &= \text{Cov}(\check{e}_t^{(j)}, \check{e}_{t-k}^{j,t}) = \text{Cov}(e_t^{(j)} - e_t, e_{t-k}^{j,t} - e_{t-k}) \\ &= \text{Cov}(e_t^{(j)}, e_{t-k}^{j,t}) - \text{Cov}(e_t^{(j)}, e_{t-k}) - \text{Cov}(e_t, e_{t-k}^{j,t}) + \text{Cov}(e_t, e_{t-k}) \\ &= \gamma_k^{(j)} - \frac{1}{g} \text{Cov}\left(e_t^{(j)}, \sum_{m=1}^g e_{t-k}^{(m)}\right) - \frac{1}{g} \text{Cov}\left(\sum_{m=1}^g e_t^{(m)}, e_{t-k}^{j,t}\right) + \frac{1}{g^2} \sum_{m=1}^g \gamma_k^{(m)}, \end{aligned}$$

onde o último termo da soma decorre de (2.2). Como os erros de painéis de grupos de rotação diferentes foram assumidos como independentes, i.e., $\text{Cov}(e_t^{(j)}, e_{t-k}^{m,t}) = 0$, para $m \neq j, k \geq 0$,

$$\begin{aligned} C_k^{(j)} &= \gamma_k^{(j)} - \frac{2}{g} \gamma_k^{(j)} + \frac{1}{g^2} \sum_{m=1}^g \gamma_k^{(m)} \\ &= \left[1 - \frac{2}{g} + \frac{1}{g^2}\right] \gamma_k^{(j)} + \frac{1}{g^2} \sum_{\substack{m=1 \\ m \neq j}}^g \gamma_k^{(m)} \\ &= \left[\frac{g-1}{g}\right]^2 \gamma_k^{(j)} + \frac{1}{g^2} \sum_{\substack{m=1 \\ m \neq j}}^g \gamma_k^{(m)}. \end{aligned}$$

Finalmente, relacionaremos $C_k^{(j)}$ com $\rho(k) = \frac{\gamma(k)}{\gamma(0)}$, a função de autocorrelação dos verdadeiros erros amostrais e_t . Veja que

$$\begin{aligned}
 \sum_{j=1}^g C_k^{(j)} &= \left[\frac{g-1}{g} \right]^2 \sum_{j=1}^g \gamma_k^{(j)} + \frac{1}{g^2} \sum_{j=1}^g \sum_{\substack{m=1 \\ m \neq j}}^g \gamma_k^{(m)} \\
 &= \left[\frac{g-1}{g} \right]^2 \sum_{j=1}^g \gamma_k^{(j)} + \left[\frac{g-1}{g^2} \right] \sum_{j=1}^g \gamma_k^{(j)} \\
 &= \left[\frac{g(g-1)}{g^2} \right] \sum_{j=1}^g \gamma_k^{(j)} \\
 &= g(g-1)\gamma(k),
 \end{aligned} \tag{2.4}$$

sendo que a última linha vem de (2.2). Assim, tem-se que

$$\rho(k) = \frac{\sum_{j=1}^g C_k^{(j)}}{\sum_{j=1}^g C_0^{(j)}},$$

com $C_k^{(j)}$ a covariância entre dois pseudo-erros de um mesmo grupo de rotação, com k períodos de distância, sendo que o mais recente está sendo pesquisado pela j -ésima vez.

2.2.4 Um estimador para $\rho(k)$

Segundo Pfeiffermann et al. (1998), $\rho(k)$ pode ser estimada consistentemente numa pesquisa com T períodos utilizando o estimador:

$$\hat{\rho}(k) = \frac{\sum_{j=1}^g \hat{C}_k^{(j)}}{\sum_{j=1}^g \hat{C}_0^{(j)}}, \tag{2.5}$$

onde

$$\hat{C}_k^{(j)} = \frac{1}{T} \sum_{t=k+1}^T (\check{e}_t^{(j)} - \check{e}^{(j)}) (\check{e}_{t-k}^{j,t} - \check{e}^{j,k}), \tag{2.6}$$

Estimação das autocovariâncias dos e_t . Método dos pseudo-erros

com

$$\check{e}^{(j)} = \frac{1}{T} \sum_{t=1}^T \check{e}_t^{(j)} \quad \text{e} \quad \check{e}^{j,k} = \frac{1}{T} \sum_{t=1}^T \check{e}_{t-k}^{j,t}.$$

As considerações feitas anteriormente para as $\gamma_k^{(j)}$, no sentido de não corresponder a funções de autocorrelação propriamente ditas, são válidas também para as $C_k^{(j)}$. Assim, sua estimação não pode ser feita usando diretamente as séries de pseudo-erros de cada grupo de rotação e uma ferramenta como a função *acf* do R. No Apêndice B é descrita uma proposta para implementar este cálculo e é apresentada uma sintaxe desenvolvida no R, para o caso de um esquema 4-8(2) balanceado a duas vias, como o empregado nos exercícios de simulação no Capítulo 4 e na aplicação do Capítulo 5.

Estimar as covariâncias $C_k^{(j)}$ pode ser útil na prática para verificar o comportamento do esquema de rotação. No entanto, nem sempre é esse o objetivo principal da aplicação desta metodologia. Para os casos em que o interesse é só a estimação da função $\rho(k)$, uma alternativa que simplifica o cálculo deriva do fato que

$$\begin{aligned} \sum_{j=1}^g \hat{C}_k^{(j)} &= \sum_{j=1}^g \frac{1}{T} \sum_{t=k+1}^T (\check{e}_t^{(j)} - \check{e}^{(j)})(\check{e}_{t-k}^{j,t} - \check{e}^{j,k}) \\ &= \sum_{r=1}^g \frac{1}{T} \sum_{t=k+1}^T (e_{r,t}^*)(e_{r,t-k}^*) \\ &= \sum_{r=1}^g \hat{C}_r^*(k); \end{aligned} \tag{2.7}$$

onde $\hat{C}_r^*(k)$ é a estimativa da função de autocovariância da série de pseudo-erros, propriamente centrados, do r -ésimo grupo de rotação. Note que mesmo que as autocorrelações sejam calculadas por grupo de rotação, o pseudo-erro de um painel que está sendo pesquisado pela j -ésima vez é centrado com relação à média de todos os pseudo-erros sendo pesquisados esse número de vezes, e não com relação à de seu respectivo grupo de rotação.

Capítulo 3

Estimação em MEE com erros de observação autocorrelacionados

Como mencionado no Capítulo 1, as primeiras propostas de estimação para pesquisas repetidas baseadas em métodos de séries temporais apareceram nos trabalhos de Blight e Scott (1973), Scott et al. (1974) e Jones (1980). Porém, este enfoque só ganhou popularidade após Tam (1987) e Binder e Hidiroglou (1988) propuserem sua implementação empregando MEE e o filtro de Kalman, (Kalman, 1960).

Em linhas gerais, esta proposta supõe que o parâmetro de interesse evolui no tempo como um processo estocástico, possivelmente multivariado, denotado por $\{\boldsymbol{\theta}_t\}$. Para estimá-lo, é conduzida uma pesquisa repetida em $t = 1, 2, \dots$ obtendo assim uma série de estimativas $\{\mathbf{y}_t\}$, sendo cada uma delas não viesada para o $\boldsymbol{\theta}_t$ correspondente. Se \mathbf{e}_t denota o erro de observação do t -ésimo período, para cada t se tem então: $\mathbf{y}_t = \boldsymbol{\theta}_t + \mathbf{e}_t$.

É suposto que a evolução de $\{\boldsymbol{\theta}_t\}$ pode ser modelada empregando um MEE como o apresentado na Seção 1.3, i.e., que existam uma matriz \mathbf{Z}_t não estocástica e conhecida e um vetor de estados $\boldsymbol{\alpha}_t$, tal que para cada t , $\boldsymbol{\theta}_t = \mathbf{Z}_t \boldsymbol{\alpha}_t$. Assim, a relação $\mathbf{y}_t = \boldsymbol{\theta}_t + \mathbf{e}_t$ pode ser reescrita como no modelo (1.4):

$$\begin{aligned} \mathbf{y}_t &= \mathbf{Z}_t \boldsymbol{\alpha}_t + \mathbf{e}_t & E(\mathbf{e}_t) &= \mathbf{0}; & E(\mathbf{e}_t \mathbf{e}_t') &= \boldsymbol{\Sigma}_{tt} \\ \boldsymbol{\alpha}_t &= \mathbf{T}_t \boldsymbol{\alpha}_{t-1} + \mathbf{G}_t \boldsymbol{\eta}_t & E(\boldsymbol{\eta}_t) &= \mathbf{0}; & E(\boldsymbol{\eta}_t \boldsymbol{\eta}_t') &= \mathbf{Q}_t. \end{aligned}$$

Erros de observação ARMA e o filtro de Kalman

Como indicado na Seção 1.4, se todos os componentes das matrizes do sistema foram conhecidos e supormos que os termos de erro \mathbf{e}_t e $\boldsymbol{\eta}_t$ são não autocorrelacionados e não correlacionados entre si, esta formulação e o algoritmo do filtro de Kalman dado na Subseção 1.4.1 podem ser empregados para obter, em cada período, o estimador linear de $\boldsymbol{\theta}_t$ de mínimo EQM, assim como a correspondente matriz de variâncias e covariâncias dos erros de estimação (MVEE).

No entanto, no contexto de pesquisas repetidas no tempo, é comum que as amostras dos distintos períodos apresentem algum tipo de dependência. Como ilustrado no Capítulo 2, isto induz autocorrelação na série de erros amostrais $\{\mathbf{e}_t\}$ a qual, se não for apropriadamente considerada no modelo¹, pode conduzir a estimativas viciadas pela existência de tendências espúrias associadas a esta autocorrelação (Pfeffermann et al., 1998). Duas das alternativas propostas para contornar este problema serão estudadas nesta dissertação: o uso do filtro de Kalman quando os erros são razoavelmente explicados por um modelo ARMA, e o filtro GLS proposto em Pfeffermann e Tiller (2006). Soluções alternativas podem ser encontradas em Wang et al. (2010), entre outros.

3.1 Erros de observação ARMA e o filtro de Kalman

Quando os erros da equação de observação apresentam uma estrutura de autocorrelação que pode ser ajustada empregando um modelo ARMA (Box et al., 1994), é possível modificar o modelo (1.4) para continuar utilizando o filtro de Kalman na estimação. Esta abordagem, que será apresentada a seguir, é encontrada em diversas aplicações, entre elas: Silva e Cruz (2002), Pfeffermann (2011), Van den Brakel e Krieg (2009) e Chow et al. (2010).

Suponha que \mathbf{y}_t satisfaz o MEE dado por:

$$\begin{aligned}\mathbf{y}_t &= \mathbf{Z}_t^{(y)} \boldsymbol{\alpha}_t^{(y)} + \mathbf{e}_t \\ \boldsymbol{\alpha}_t^{(y)} &= \mathbf{T}_t^{(y)} \boldsymbol{\alpha}_{t-1}^{(y)} + \mathbf{G}_t^{(y)} \boldsymbol{\eta}_t^{(y)},\end{aligned}\tag{3.1}$$

com $E(\boldsymbol{\eta}_t^{(y)}) = \mathbf{0}$; $V(\boldsymbol{\eta}_t^{(y)}) = \mathbf{Q}_t^{(y)}$. Suponha que o erro da equação de observação, \mathbf{e}_t ,

¹Por exemplo empregando o Filtro de Kalman na formulação da Seção 1.4.

segue um modelo ARMA univariado ou multivariado, com representação em espaço de estados como no Apêndice A, dado por:

$$\begin{aligned} \mathbf{e}_t &= \mathbf{Z}_t^{(e)} \boldsymbol{\alpha}_t^{(e)} \\ \boldsymbol{\alpha}_t^{(e)} &= \mathbf{T}_t^{(e)} \boldsymbol{\alpha}_{t-1}^{(e)} + \mathbf{G}_t^{(e)} \boldsymbol{\eta}_t^{(e)}, \end{aligned} \quad (3.2)$$

sendo que $E(\boldsymbol{\eta}_t^{(e)}) = \mathbf{0}$ e $V(\boldsymbol{\eta}_t^{(e)}) = \mathbf{Q}_t^{(e)}$. Após alguma álgebra é possível verificar que as equações (3.1) e (3.2) podem ser consideradas conjuntamente no modelo:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{Z}_t^{(*)} \boldsymbol{\alpha}_t^{(*)} \\ \boldsymbol{\alpha}_t^{(*)} &= \mathbf{T}_t^{(*)} \boldsymbol{\alpha}_{t-1}^{(*)} + \mathbf{G}_t^{(*)} \boldsymbol{\eta}_t^{(*)}, \end{aligned} \quad (3.3)$$

com

$$\boldsymbol{\alpha}_t^{(*)'} = \begin{bmatrix} \boldsymbol{\alpha}_t^{(y)'} & | & \boldsymbol{\alpha}_t^{(e)'} \end{bmatrix}, \quad \mathbf{Z}_t^{(*)} = \begin{bmatrix} \mathbf{Z}_t^{(y)} & | & \mathbf{Z}_t^{(e)} \end{bmatrix}, \quad \boldsymbol{\eta}_t^{(*)'} = \begin{bmatrix} \boldsymbol{\eta}_t^{(y)'} & | & \boldsymbol{\eta}_t^{(e)'} \end{bmatrix},$$

$\mathbf{T}_t^{(*)} = \mathbf{T}_t^{(y)} \otimes \mathbf{T}_t^{(e)}$, $\mathbf{G}_t^{(*)} = \mathbf{G}_t^{(y)} \otimes \mathbf{G}_t^{(e)}$, $E(\boldsymbol{\eta}_t^{(*)}) = \mathbf{0}$ e $V(\boldsymbol{\eta}_t^{(*)}) = \mathbf{Q}_t^{(*)} = \mathbf{Q}_t^{(y)} \otimes \mathbf{Q}_t^{(e)}$, e com $\boldsymbol{\theta}_t = [\mathbf{Z}_t^{(y)} \mid \mathbf{0}] \boldsymbol{\alpha}_t^{(*)}$. A notação \otimes é empregada no seguinte sentido: $\mathbf{T}_t^{(y)} \otimes \mathbf{T}_t^{(e)}$ denota uma matriz diagonal por blocos com $\mathbf{T}_t^{(y)}$ no primeiro bloco e $\mathbf{T}_t^{(e)}$ no segundo. Note que neste novo modelo, que não contém erro na equação de observação, pode ser estimado utilizando o filtro de Kalman apresentado na Seção 1.4, sempre que prefixarmos a variância do erro de observação em zero.

Os coeficientes do modelo ARMA para o vetor de erros amostrais podem ser estimados da mesma forma que os componentes desconhecidos das matrizes de \mathbf{Q}_t ou $\boldsymbol{\Sigma}_{tt}$, isto é, usando a decomposição da verossimilhança através dos erros de previsão². Porém, como indicado em Silva e Cruz (2002, p.59, p.75), no contexto de pesquisas repetidas é comum estimá-los separadamente, obtendo uma estimativa pontual da função de autocovariância da série $\{\mathbf{e}_t\}$ através de um método como o apresentado na Seção 2.2 e posteriormente, empregando estas estimativas para estimar a função de autocorrelação parcial usando as equações de Yule-Walker. Isto permite evitar possíveis problemas de identificabilidade no modelo. Após este procedimento, os coeficientes estimados para o modelo ARMA são tratados como conhecidos na aplicação do filtro.

A seguir, apresentaremos a forma do modelo (3.3) no caso univariado, quando a série

²Ver Subseção 1.4.2.

Erros de observação ARMA e o filtro de Kalman

$\{\theta_t\}$ é decomposta empregando um modelo estrutural básico e os erros de observação seguem um modelo AR(1). Este modelo será empregado posteriormente nos exercícios de simulação do Capítulo 4.

3.1.1 BSM com erros de observação AR(1)

Uma formulação na forma de espaço de estados dada em (3.3), para um BSM acrescentando um termo irregular denotado por I_t ao vetor de estados, e onde os erros da equação de observação, e_t , são ajustados com o modelo AR(1):

$$e_t = \phi e_{t-1} + \xi_t,$$

com $E(\xi_t) = 0$, $V(\xi_t) = \sigma_\xi^2$ e $\text{Cov}(\xi_t, \xi_{t-k}) = 0$ para todo $k \neq 0$, é dada por:

$$\mathbf{Z}_t^{(*)} = \begin{bmatrix} 1 & 0 & 1 & 0 & \dots & 0 & 1 & 1 \end{bmatrix},$$

$$\boldsymbol{\alpha}_t^{(*)'} = \begin{bmatrix} L_t & R_t & S_t & S_{t-1} & \dots & S_{t-10} & I_t & e_t \end{bmatrix},$$

$$\boldsymbol{\eta}_t^{(*)'} = \begin{bmatrix} \eta_{L_t} & \eta_{R_t} & \eta_{S_t} & I_t & \xi_t \end{bmatrix},$$

$$\mathbf{T}_t^{(*)} = \left[\begin{array}{cc|cccc|cc} 1 & 1 & & & & & & \\ 0 & 1 & & & & & & \\ \hline & & \mathbf{0}_{(2 \times 11)} & & & & & \mathbf{0}_{(2 \times 2)} \\ \hline & & -1 & -1 & \dots & -1 & -1 & \\ & & 1 & 0 & \dots & 0 & 0 & \\ \mathbf{0}_{(11 \times 2)} & & 0 & 1 & \dots & 0 & 0 & \mathbf{0}_{(11 \times 2)} \\ & & \vdots & \vdots & \ddots & \vdots & \vdots & \\ & & 0 & 0 & \dots & 1 & 0 & \\ \hline \mathbf{0}_{(2 \times 2)} & & \mathbf{0}_{(2 \times 11)} & & & & & \begin{matrix} 0 & 0 \\ 0 & \phi \end{matrix} \end{array} \right], \quad \mathbf{G}_t^{(*)} = \left[\begin{array}{cc|cc} \mathbf{I}_{(3)} & & \mathbf{0}_{(3 \times 2)} & \\ \hline & & & \\ \mathbf{0}_{(10 \times 3)} & & \mathbf{0}_{(10 \times 2)} & \\ \hline \mathbf{0}_{(2 \times 3)} & & \mathbf{I}_{(2)} & \end{array} \right],$$

com $\mathbf{Q}_t^{(*)} = \text{diag}(\sigma_{L_t}^2, \sigma_{R_t}^2, \sigma_{S_t}^2, \sigma_{I_t}^2, \sigma_{\xi_t}^2)$.

3.2 O filtro GLS

O filtro GLS foi proposto em Pfeiffermann e Tiller (2006) como resposta ao problema de estimação por domínios geográficos na *Current Population Survey*, a pesquisa mensal de emprego dos Estados Unidos. Segundo os autores, a estimação da taxa de desemprego nesta pesquisa era feita rotineiramente empregando a metodologia apresentada na Seção 3.1. No entanto, com a finalidade de se proteger contra possíveis erros de especificação do modelo assumido, existia interesse em ajustar as estimativas obtidas por este método, utilizando alguma média ponderada das estimativas diretas do desenho amostral, que pudesse ser considerada suficientemente confiável.

Para contornar este problema, dado um conjunto de domínios, os autores propõem que em cada período seja igualada a média das estimativas de modelo obtidas para os domínios à média das estimativas de planejamento correspondentes. Isto, na prática, sugere ajustar simultaneamente os domínios considerados, como numa série multivariada, e incluir uma restrição linear de benchmark para igualar as médias acima mencionadas. O ajuste simultâneo de vários domínios pode, em princípio, ser feito empregando a metodologia apresentada na Seção 3.1. No entanto, esta não parece ser a melhor solução neste caso, dado que são requeridas estimações para um grande número de domínios e o modelo empregado para a estimação é complexo³. Da mesma forma, esta metodologia não suporta a inclusão de uma restrição com as características apontadas acima.

Assim, em Pfeiffermann e Tiller (2006) é proposto um filtro, que é denominado *GLS*, por estar baseado no método dos mínimos quadrados generalizados, como uma alternativa de estimação que emprega diretamente as autocovariâncias estimadas da série de erros amostrais, tornando desnecessário acrescentar componentes ao vetor de estados (Ver Seção 3.1). Os casos onde a estrutura de autocorrelação dos erros observacionais é complexa, podem ser melhor tratados empregando este filtro. Adicionalmente, é fácil a inclusão de uma restrição como a mencionada acima e, no caso de erros independentes, o novo método coincide com o filtro de Kalman. Como desvantagem, o estimador obtido através deste filtro não é o de menor EQM na família de estimadores lineares

³Segundo os autores, era utilizado um BSM com erros de observação AR(15) que deriva num vetor de estados com 29 componentes, 15 das quais correspondem ao erro amostral. Por exemplo, obter estimativas para as 9 divisões censitárias do país com este método, considerando que o modelo de cada divisão fosse independente dos modelos das outras, requeria filtrar um MEE com um vetor de estados com 261 componentes.

O filtro GLS

não viesados baseados nas observações y_t individuais. Porém, ele é o de menor EQM quando se consideram somente os estimadores baseados em $\hat{\alpha}_{t-1}$, a estimativa do vetor de estados no período imediatamente anterior, e a nova observação y_t .

Nesta seção, será apresentado o filtro GLS para o caso de uma série temporal y_t multivariada, expressando as autocovariâncias da série de erros amostrais através de uma sequência de matrizes, como em Pfeiffermann e Tiller (2006). Uma comparação via simulação dos dois tipos de filtros apresentados, empregando dois cenários de séries univariadas com distintos esquemas de autocorrelação, assim como um resumo das fórmulas do filtro GLS são apresentados no Capítulo 4 e no Apêndice C, respectivamente. Por fim, no Capítulo 5, é apresentada a proposta para fazer benchmark empregando o filtro GLS.

3.2.1 Um preditor empregando mínimos quadrados generalizados

Seja o MEE dado em (1.4), com vetor de estados inicial $\hat{\alpha}_0$, independente das observações y_t para todo t , e com matriz de covariâncias P_0 . Sem perda de generalidade, vamos supor que as matrizes T_t , G_t e Q_t são fixas ao longo do tempo e relaxar a hipótese de não autocorrelação temporal da série de erros de observação, denotando por $\Sigma_{\tau t}$ a matriz de covariâncias $\text{Cov}(e_\tau, e'_t)$. Nestas condições, o filtro GLS propõe estimar α_t resolvendo por mínimos quadrados generalizados o sistema

$$\begin{pmatrix} \hat{\alpha}_{t|t-1} \\ y_t \end{pmatrix} = \begin{pmatrix} I \\ Z_t \end{pmatrix} \alpha_t + \begin{pmatrix} u_{t|t-1} \\ e_t \end{pmatrix},$$

onde $\hat{\alpha}_{t|t-1} = T_t \hat{\alpha}_{t-1}$ é uma previsão para α_t , $\hat{\alpha}_{t-1}$ é um estimador não viesado de α_{t-1} e $u_{t|t-1} = \hat{\alpha}_{t|t-1} - \alpha_t$ é o erro de previsão associado.

Para isto, observe que $E[(u_{t|t-1} \ e_t)'] = 0$, pois

$$E(u_{t|t-1}) = E[T_t \hat{\alpha}_{t-1} - (T_t \alpha_{t-1} + G_t \eta_t)] = 0,$$

dado que $\hat{\alpha}_{t|t-1}$ é não viesado para α_{t-1} e $E(e_t) = 0$. Assim, se denotarmos por V_t

a matriz de covariâncias de $(\mathbf{u}_{t|t-1} \ \mathbf{e}_t)'$, temos que

$$\hat{\boldsymbol{\alpha}}_t = \left[\begin{pmatrix} \mathbf{I} & \mathbf{Z}'_t \end{pmatrix} \mathbf{V}_t^{-1} \begin{pmatrix} \mathbf{I} \\ \mathbf{Z}_t \end{pmatrix} \right]^{-1} \begin{pmatrix} \mathbf{I} & \mathbf{Z}'_t \end{pmatrix} \mathbf{V}_t^{-1} \begin{pmatrix} \hat{\boldsymbol{\alpha}}_{t|t-1} \\ \mathbf{y}_t \end{pmatrix}, \quad (3.4)$$

é um estimador linear não viesado para $\boldsymbol{\alpha}_t$, com matriz de covariâncias

$$\mathbf{P}_t = \left[\begin{pmatrix} \mathbf{I} & \mathbf{Z}'_t \end{pmatrix} \mathbf{V}_t^{-1} \begin{pmatrix} \mathbf{I} \\ \mathbf{Z}_t \end{pmatrix} \right]^{-1}. \quad (3.5)$$

De fato, $\hat{\boldsymbol{\alpha}}_t$ é o estimador linear não viesado de mínimo EQM na família de estimadores baseados nas informações $\hat{\boldsymbol{\alpha}}_{t-1}$ e \mathbf{y}_t , como demonstrado no Apêndice A de Pfeiffermann e Tiller (2006). Além disso, \mathbf{V}_τ não é função de \mathbf{y}_t e portanto nem de \mathbf{e}_t , para todo τ, t , como mostrado no final desta seção. Com isto, \mathbf{P}_τ também não é função de \mathbf{e}_t para todo τ, t . Para o cálculo de estimações só resta então determinar \mathbf{V}_t para cada t . Note que

$$V(\mathbf{u}_{t|t-1}) = V(\hat{\boldsymbol{\alpha}}_{t|t-1}) = \mathbf{T}\mathbf{P}_{t-1}\mathbf{T}' + \mathbf{G}\mathbf{Q}\mathbf{G}' = \mathbf{P}_{t|t-1},$$

pois $\text{Cov}(\boldsymbol{\alpha}_\tau, \mathbf{e}_t) = \mathbf{0}$ para todo τ, t pela independência de $\boldsymbol{\eta}_\tau$ e \mathbf{e}_t ; veja também que $V(\mathbf{e}_t) = \boldsymbol{\Sigma}_{tt}$. Assim, denotando por \mathbf{C}_t a covariância entre $\mathbf{u}_{t|t-1}$ e \mathbf{e}_t , esta matriz tem a forma geral

$$\mathbf{V}_t = \begin{pmatrix} \mathbf{P}_{t|t-1} & \mathbf{C}_t \\ \mathbf{C}'_t & \boldsymbol{\Sigma}_{tt} \end{pmatrix}.$$

Note que pela característica sequencial do método, para cada t , $\hat{\boldsymbol{\alpha}}_t$ combina as informações de $\hat{\boldsymbol{\alpha}}_{t|t-1}$ e \mathbf{y}_t ; por sua vez, $\hat{\boldsymbol{\alpha}}_{t|t-1}$, baseada em $\hat{\boldsymbol{\alpha}}_{t-1}$, combina as de $\hat{\boldsymbol{\alpha}}_{t-1|t-2}$ e \mathbf{y}_{t-1} , e assim sucessivamente. Portanto, se os \mathbf{e}_t são autocorrelacionados, também \mathbf{y}_{t-k} e \mathbf{e}_t são correlacionados para $k > 0$; daí, \mathbf{e}_t é também correlacionado com $\hat{\boldsymbol{\alpha}}_{t|t-1}$ e $\mathbf{u}_{t|t-1}$ para cada t . Em resumo, no caso de erros autocorrelacionados, as matrizes \mathbf{C}_t são não nulas e sua forma depende das matrizes $\boldsymbol{\Sigma}_{\tau t}$.

Pela independência entre $\hat{\boldsymbol{\alpha}}_0$ e \mathbf{e}_1 , a estimação de $\boldsymbol{\alpha}_1$ pode ser feita empregando as equações do filtro de Kalman como apresentadas na Subseção 1.4.1; já para $t > 1$, é necessário estabelecer a forma de \mathbf{C}_t . A seguir, exemplificaremos isto considerando os

O filtro GLS

casos $t = 2$ e 3 e apresentaremos finalmente, a forma geral desta matriz.

Para $t=2$:

Veja que $\mathbf{C}_2 = \text{Cov}(\mathbf{u}_{2|1}, \mathbf{e}_2) = \text{Cov}(\hat{\boldsymbol{\alpha}}_{2|1} - \boldsymbol{\alpha}_2, \mathbf{e}_2) = \mathbf{T}\text{Cov}(\hat{\boldsymbol{\alpha}}_1, \mathbf{e}_2)$, pela independência de $\boldsymbol{\eta}_\tau$ e \mathbf{e}_t , para todo τ, t . Como $\hat{\boldsymbol{\alpha}}_1$ provém do filtro de Kalman, temos

$$\text{Cov}(\mathbf{u}_{2|1}, \mathbf{e}_2) = \mathbf{T}(\mathbf{I} - \mathbf{K}_1\mathbf{Z}_1)\mathbf{T}\text{Cov}(\hat{\boldsymbol{\alpha}}_0, \mathbf{e}_2) + \mathbf{T}\mathbf{K}_1\text{Cov}(\mathbf{y}_1, \mathbf{e}_2),$$

com \mathbf{K}_1 não em função de \mathbf{e}_2 e definida como em 1.4.1. O primeiro termo à direita é zero pois $\hat{\boldsymbol{\alpha}}_0$ é independente de \mathbf{y}_t para todo t . Além disso,

$$\text{Cov}(\mathbf{y}_{t-k}, \mathbf{e}_t) = \text{Cov}[(\mathbf{Z}_{t-k}\boldsymbol{\alpha}_{t-k} + \mathbf{e}_{t-k}), \mathbf{e}_t] = \text{Cov}(\mathbf{e}_{t-k}, \mathbf{e}_t) = \boldsymbol{\Sigma}_{t-k,t},$$

então,

$$\mathbf{C}_2 = \text{Cov}(\mathbf{u}_{2|1}, \mathbf{e}_2) = \mathbf{T}\mathbf{K}_1\boldsymbol{\Sigma}_{12}.$$

Note que tanto \mathbf{C}_2 quanto \mathbf{P}_1 não estão em função das observações \mathbf{y}_t ; com isto, \mathbf{V}_2 também não.

Para $t=3$:

Veja que $\mathbf{C}_3 = \text{Cov}(\mathbf{u}_{3|2}, \mathbf{e}_3) = \mathbf{T}\text{Cov}(\hat{\boldsymbol{\alpha}}_2, \mathbf{e}_3)$. Substituindo por (3.4),

$$\begin{aligned} \text{Cov}(\mathbf{u}_{3|2}, \mathbf{e}_3) &= \mathbf{T}\text{Cov}\left[\mathbf{P}_2 \begin{pmatrix} \mathbf{I} & \mathbf{Z}'_2 \end{pmatrix} \mathbf{V}_2^{-1} \begin{pmatrix} \hat{\boldsymbol{\alpha}}_{2|1} \\ \mathbf{y}_2 \end{pmatrix}, \mathbf{e}_3\right] \\ &= \mathbf{T}\text{Cov}\left[\mathbf{P}_2\mathbf{B}_2 \begin{pmatrix} \hat{\boldsymbol{\alpha}}_{2|1} \\ \mathbf{y}_2 \end{pmatrix}, \mathbf{e}_3\right], \end{aligned}$$

onde $\mathbf{B}_2 = (\mathbf{I} \ \mathbf{Z}'_2) \mathbf{V}_2^{-1}$. Veja que se \mathbf{y}_t for uma série p -variada e o vetor de estados tiver q componentes, \mathbf{Z}_t será uma matriz $p \times q$ e $(\mathbf{I} \ \mathbf{Z}'_t)$ terá dimensão $q \times (p+q)$. Da mesma forma, como $\mathbf{u}_{t|t-1}$ é um vetor de $p+q$ componentes, sua matriz de covariâncias \mathbf{V}_t é quadrada com esta mesma dimensão. Com isto, \mathbf{B}_2 é uma matriz $q \times (p+q)$ e pode ser particionada como $\mathbf{B}_2 = [\mathbf{B}_{21(q \times q)} | \mathbf{B}_{22(q \times p)}]$.

Temos então

$$\mathbf{B}_2 \begin{pmatrix} \hat{\boldsymbol{\alpha}}_{2|1} \\ \mathbf{y}_2 \end{pmatrix} = \mathbf{B}_{21}\mathbf{T}\hat{\boldsymbol{\alpha}}_1 + \mathbf{B}_{22}\mathbf{y}_2;$$

como \mathbf{V}_2 está em função de \mathbf{e}_3 , tanto \mathbf{P}_2 quanto \mathbf{B}_2 também não estão, assim

$$\begin{aligned}\text{Cov}(\mathbf{u}_{3|2}, \mathbf{e}_3) &= \mathbf{TP}_2 [\mathbf{B}_{21}\mathbf{TCov}(\hat{\boldsymbol{\alpha}}_1, \mathbf{e}_3) + \mathbf{B}_{22}\text{Cov}(\mathbf{y}_2, \mathbf{e}_3)] \\ &= \mathbf{A}_2\mathbf{TCov}(\hat{\boldsymbol{\alpha}}_1, \mathbf{e}_3) + \tilde{\mathbf{A}}_2\text{Cov}(\mathbf{y}_2, \mathbf{e}_3),\end{aligned}$$

onde $\mathbf{A}_2 = \mathbf{TP}_2\mathbf{B}_{21}$ e $\tilde{\mathbf{A}}_2 = \mathbf{TP}_2\mathbf{B}_{22}$. No entanto, como

$$\text{Cov}(\hat{\boldsymbol{\alpha}}_1, \mathbf{e}_3) = (\mathbf{I} - \mathbf{K}_1\mathbf{Z}_1)\mathbf{TCov}(\hat{\boldsymbol{\alpha}}_0, \mathbf{e}_3) + \mathbf{K}_1\text{Cov}(\mathbf{y}_1, \mathbf{e}_3) = \mathbf{K}_1\boldsymbol{\Sigma}_{13},$$

e $\text{Cov}(\mathbf{y}_2, \mathbf{e}_3) = \boldsymbol{\Sigma}_{23}$, denotando por $\tilde{\mathbf{A}}_1 = \mathbf{TK}_1$, obtemos finalmente,

$$\mathbf{C}_3 = \mathbf{A}_2\tilde{\mathbf{A}}_1\boldsymbol{\Sigma}_{13} + \tilde{\mathbf{A}}_2\boldsymbol{\Sigma}_{23}.$$

Para $t > 3$:

No caso geral, definem-se a matriz $\mathbf{B}_t = \begin{pmatrix} \mathbf{I} & \mathbf{Z}'_t \end{pmatrix} \mathbf{V}_t^{-1}$ e sua partição $[\mathbf{B}_{t1(q \times q)} | \mathbf{B}_{t2(q \times p)}]$. Com isto, é possível definir as sequências de matrizes:

$$\begin{cases} \mathbf{A}_j = \mathbf{TP}_j\mathbf{B}_{j1} \\ \tilde{\mathbf{A}}_j = \mathbf{TP}_j\mathbf{B}_{j2} \end{cases} \quad \text{para } j = 2, \dots, t-1,$$

e $\tilde{\mathbf{A}}_1 = \mathbf{TK}_1$, para $j = 1$. Assim,

$$\begin{aligned}\mathbf{C}_t &= \text{Cov}(\mathbf{u}_{t|t-1}, \mathbf{e}_t) \\ &= \mathbf{A}_{t-1}\mathbf{A}_{t-2} \dots \tilde{\mathbf{A}}_1\boldsymbol{\Sigma}_{1t} + \mathbf{A}_{t-1}\mathbf{A}_{t-2} \dots \tilde{\mathbf{A}}_2\boldsymbol{\Sigma}_{2t} + \dots + \mathbf{A}_{t-1}\tilde{\mathbf{A}}_{t-2}\boldsymbol{\Sigma}_{(t-2)t} + \tilde{\mathbf{A}}_{t-1}\boldsymbol{\Sigma}_{(t-1)t}.\end{aligned}$$

Uma expressão alternativa para $\hat{\boldsymbol{\alpha}}_t$

Nesta seção, apresentaremos uma expressão alternativa para o cálculo da estimativa $\hat{\boldsymbol{\alpha}}_t$. Esta expressão é mais similar à do estimador do filtro de Kalman e evidencia que, quando os erros de observação são não autocorrelacionados, os dois estimadores coincidem. Também serão apresentadas novas expressões para as matrizes \mathbf{A}_t e $\tilde{\mathbf{A}}_t$ envolvidas no cálculo de \mathbf{C}_t , com a finalidade de empregá-las no capítulo seguinte, onde, devido à restrição imposta no modelo, a formulação aqui apresentada não pode ser utilizada.

O filtro GLS

Inicialmente, considere novamente a matriz

$$\mathbf{V}_t = \begin{pmatrix} \mathbf{P}_{t|t-1} & \mathbf{C}_t \\ \mathbf{C}_t' & \Sigma_{tt} \end{pmatrix},$$

vamos achar uma forma para $\begin{pmatrix} \mathbf{I} & \mathbf{Z}_t' \end{pmatrix} \mathbf{V}_t^{-1} \begin{pmatrix} \mathbf{I} \\ \mathbf{Z}_t \end{pmatrix}$. Lembre que para uma matriz invertível \mathbf{A} , particionada da forma

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix},$$

com \mathbf{A}_{11} não singular, a inversa de \mathbf{A} é dada pela expressão (Seber, 2008, pág. 293),

$$\mathbf{A}^{-1} = \begin{pmatrix} \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1} \mathbf{A}_{12} \mathbf{H} \mathbf{A}_{21} \mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1} \mathbf{A}_{12} \mathbf{H} \\ -\mathbf{H} \mathbf{A}_{21} \mathbf{A}_{11}^{-1} & \mathbf{H} \end{pmatrix},$$

com $\mathbf{H} = [\mathbf{A}_{22} - \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{A}_{12}]^{-1}$. Portanto,

$$\mathbf{V}_t^{-1} = \begin{pmatrix} \mathbf{P}_{t|t-1}^{-1} + \mathbf{P}_{t|t-1}^{-1} \mathbf{C}_t \mathbf{H}_t \mathbf{C}_t' \mathbf{P}_{t|t-1}^{-1} & -\mathbf{P}_{t|t-1}^{-1} \mathbf{C}_t \mathbf{H}_t \\ -\mathbf{H}_t \mathbf{C}_t \mathbf{P}_{t|t-1}^{-1} & \mathbf{H}_t \end{pmatrix},$$

com

$$\mathbf{H}_t = [\Sigma_{tt} - \mathbf{C}_t' \mathbf{P}_{t|t-1}^{-1} \mathbf{C}_t]^{-1}, \quad (3.6)$$

uma matriz simétrica, pois tanto $\mathbf{P}_{t|t-1}^{-1}$ quanto Σ_{tt} são simétricas. Assim,

$$\begin{pmatrix} \mathbf{I} & \mathbf{Z}_t' \end{pmatrix} \mathbf{V}_t^{-1} = \begin{pmatrix} \mathbf{P}_{t|t-1}^{-1} - (\mathbf{Z}_t' - \mathbf{P}_{t|t-1}^{-1} \mathbf{C}_t) \mathbf{H}_t \mathbf{C}_t' \mathbf{P}_{t|t-1}^{-1} & (\mathbf{Z}_t' - \mathbf{P}_{t|t-1}^{-1} \mathbf{C}_t) \mathbf{H}_t \end{pmatrix}, \quad (3.7)$$

e finalmente,

$$\mathbf{P}_t = \left[\begin{pmatrix} \mathbf{I} & \mathbf{Z}_t' \end{pmatrix} \mathbf{V}_t^{-1} \begin{pmatrix} \mathbf{I} \\ \mathbf{Z}_t \end{pmatrix} \right]^{-1} = [\mathbf{P}_{t|t-1}^{-1} + (\mathbf{Z}_t' - \mathbf{P}_{t|t-1}^{-1} \mathbf{C}_t) \mathbf{H}_t (\mathbf{Z}_t - \mathbf{C}_t' \mathbf{P}_{t|t-1}^{-1})]^{-1}. \quad (3.8)$$

Agora, empregaremos (3.8) para expressar $\hat{\boldsymbol{\alpha}}_t$ de uma forma mais similar à do estimador

do filtro de Kalman. Veja que

$$\begin{aligned} \begin{pmatrix} \mathbf{I} & \mathbf{Z}'_t \end{pmatrix} \mathbf{V}_t^{-1} \begin{pmatrix} \mathbf{T}\hat{\boldsymbol{\alpha}}_{t-1} \\ \mathbf{y}_t \end{pmatrix} &= (\mathbf{P}_{t|t-1}^{-1} + \mathbf{P}_{t|t-1}^{-1} \mathbf{C}_t \mathbf{H}_t \mathbf{C}_t' \mathbf{P}_{t|t-1}^{-1} - \mathbf{Z}'_t \mathbf{H}_t \mathbf{C}_t' \mathbf{P}_{t|t-1}^{-1}) \mathbf{T}\hat{\boldsymbol{\alpha}}_{t-1} \\ &\quad + (\mathbf{Z}'_t \mathbf{H}_t - \mathbf{P}_{t|t-1}^{-1} \mathbf{C}_t \mathbf{H}_t) \mathbf{y}_t, \quad (3.9) \end{aligned}$$

somando e substraindo $\mathbf{Z}'_t \mathbf{H}_t \mathbf{Z}_t - \mathbf{P}_{t|t-1}^{-1} \mathbf{C}_t \mathbf{H}_t \mathbf{Z}_t$ ao primeiro termo da soma, e empregando (3.8) tem-se

$$\begin{aligned} \begin{pmatrix} \mathbf{I} & \mathbf{Z}'_t \end{pmatrix} \mathbf{V}_t^{-1} \begin{pmatrix} \mathbf{T}\hat{\boldsymbol{\alpha}}_{t-1} \\ \mathbf{y}_t \end{pmatrix} &= (\mathbf{P}_t^{-1} - \mathbf{Z}'_t \mathbf{H}_t \mathbf{Z}_t + \mathbf{P}_{t|t-1}^{-1} \mathbf{C}_t \mathbf{H}_t \mathbf{Z}_t) \mathbf{T}\hat{\boldsymbol{\alpha}}_{t-1} \\ &\quad + (\mathbf{Z}'_t \mathbf{H}_t - \mathbf{P}_{t|t-1}^{-1} \mathbf{C}_t \mathbf{H}_t) \mathbf{y}_t, \end{aligned}$$

e com isto, obtemos inicialmente de (3.4),

$$\hat{\boldsymbol{\alpha}}_t = \mathbf{T}\hat{\boldsymbol{\alpha}}_{t-1} + \mathbf{P}_t (\mathbf{Z}'_t - \mathbf{P}_{t|t-1}^{-1} \mathbf{C}_t) \mathbf{H}_t (\mathbf{y}_t - \mathbf{Z}_t \mathbf{T}\hat{\boldsymbol{\alpha}}_{t-1}). \quad (3.10)$$

Agora, expressaremos \mathbf{P}_t de uma forma alternativa. Veja que para matrizes \mathbf{R} , \mathbf{S} , \mathbf{T} e \mathbf{U} , tais que $\mathbf{R} + \mathbf{STU}$ seja invertível, sua inversa está dada por

$$[\mathbf{R} + \mathbf{STU}]^{-1} = \mathbf{R}^{-1} - \mathbf{R}^{-1} \mathbf{S} (\mathbf{T}^{-1} + \mathbf{UR}^{-1} \mathbf{S})^{-1} \mathbf{UR}^{-1},$$

sempre que as inversas envolvidas existam (Seber, 2008, pág. 309). Assim, da equação (3.8),

$$\begin{aligned} \mathbf{P}_t &= \left[\mathbf{P}_{t|t-1}^{-1} + (\mathbf{Z}'_t - \mathbf{P}_{t|t-1}^{-1} \mathbf{C}_t) \mathbf{H}_t (\mathbf{Z}_t - \mathbf{C}_t' \mathbf{P}_{t|t-1}^{-1}) \right]^{-1} \\ &= \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} (\mathbf{Z}'_t - \mathbf{P}_{t|t-1}^{-1} \mathbf{C}_t) \left[\mathbf{H}_t^{-1} + (\mathbf{Z}_t - \mathbf{C}_t' \mathbf{P}_{t|t-1}^{-1}) \mathbf{P}_{t|t-1}^{-1} (\mathbf{Z}'_t - \mathbf{P}_{t|t-1}^{-1} \mathbf{C}_t) \right]^{-1} \\ &\quad \times (\mathbf{Z}_t - \mathbf{C}_t' \mathbf{P}_{t|t-1}^{-1}) \mathbf{P}_{t|t-1}^{-1}, \end{aligned}$$

e substituindo \mathbf{H}_t por sua definição em (3.6) temos

$$\mathbf{P}_t = \mathbf{P}_{t|t-1} - (\mathbf{P}_{t|t-1} \mathbf{Z}'_t - \mathbf{C}_t) \mathbf{F}_t^{-1} (\mathbf{Z}_t \mathbf{P}_{t|t-1} - \mathbf{C}_t'), \quad (3.11)$$

com $\mathbf{F}_t = [\mathbf{Z}_t \mathbf{P}_{t|t-1} \mathbf{Z}'_t - \mathbf{C}_t' \mathbf{Z}'_t - \mathbf{Z}_t \mathbf{C}_t + \boldsymbol{\Sigma}_{tt}]$. De (3.10), substituindo \mathbf{P}_t pelo valor

O filtro GLS

em (3.11) e arranjando os termos, obtemos

$$\hat{\alpha}_t = \mathbf{T}\hat{\alpha}_{t-1} + \left[(\mathbf{P}_{t|t-1}\mathbf{Z}'_t - \mathbf{C}_t) - (\mathbf{P}_{t|t-1}\mathbf{Z}'_t - \mathbf{C}_t)\mathbf{F}_t^{-1}(\mathbf{Z}_t\mathbf{P}_{t|t-1} - \mathbf{C}'_t)(\mathbf{Z}'_t - \mathbf{P}_{t|t-1}^{-1}\mathbf{C}_t) \right] \\ \times \mathbf{H}_t(\mathbf{y}_t - \mathbf{Z}_t\mathbf{T}\hat{\alpha}_{t-1}).$$

Ou equivalentemente,

$$\hat{\alpha}_t = \mathbf{T}\hat{\alpha}_{t-1} + (\mathbf{P}_{t|t-1}\mathbf{Z}'_t - \mathbf{C}_t) \left[\mathbf{I} - \mathbf{F}_t^{-1}(\mathbf{Z}_t\mathbf{P}_{t|t-1} - \mathbf{C}'_t)(\mathbf{Z}'_t - \mathbf{P}_{t|t-1}^{-1}\mathbf{C}_t) \right] \\ \times \mathbf{H}_t(\mathbf{y}_t - \mathbf{Z}_t\mathbf{T}\hat{\alpha}_{t-1}), \quad (3.12)$$

mas veja que

$$\mathbf{F}_t^{-1}(\mathbf{Z}_t\mathbf{P}_{t|t-1} - \mathbf{C}'_t)(\mathbf{Z}'_t - \mathbf{P}_{t|t-1}^{-1}\mathbf{C}_t) = \mathbf{F}_t^{-1}(\mathbf{Z}_t\mathbf{P}_{t|t-1}\mathbf{Z}'_t - \mathbf{Z}_t\mathbf{C}_t - \mathbf{C}'_t\mathbf{Z}'_t + \mathbf{C}'_t\mathbf{P}_{t|t-1}^{-1}\mathbf{C}_t),$$

somando e subtraindo Σ_{tt}

$$\mathbf{F}_t^{-1}(\mathbf{Z}_t\mathbf{P}_{t|t-1} - \mathbf{C}'_t)(\mathbf{Z}'_t - \mathbf{P}_{t|t-1}^{-1}\mathbf{C}_t) = \mathbf{F}_t^{-1}(\mathbf{F}_t + \mathbf{C}_t\mathbf{P}_{t|t-1}^{-1}\mathbf{C}_t - \Sigma_{tt}).$$

Como $\mathbf{H}_t = (\Sigma_{tt} - \mathbf{C}'_t\mathbf{P}_{t|t-1}^{-1}\mathbf{C}_t)^{-1}$,

$$\mathbf{F}_t^{-1}(\mathbf{Z}_t\mathbf{P}_{t|t-1} - \mathbf{C}'_t)(\mathbf{Z}'_t - \mathbf{P}_{t|t-1}^{-1}\mathbf{C}_t) = \mathbf{I} - \mathbf{F}_t^{-1}\mathbf{H}_t^{-1}. \quad (3.13)$$

Substituindo em 3.12, obtemos finalmente,

$$\hat{\alpha}_t = \mathbf{T}\hat{\alpha}_{t-1} + (\mathbf{P}_{t|t-1}\mathbf{Z}'_t - \mathbf{C}_t)\mathbf{F}_t^{-1}(\mathbf{y}_t - \mathbf{Z}_t\mathbf{T}\hat{\alpha}_{t-1}). \quad (3.14)$$

De forma análoga, vamos desenvolver expressões alternativas para as matrizes $\mathbf{A}_j = \mathbf{TP}_j\mathbf{B}_{j1}$ e $\tilde{\mathbf{A}}_j = \mathbf{TP}_j\mathbf{B}_{j2}$ envolvidas no cálculo de \mathbf{C}_t . Da equação (3.7) veja que para cada t ,

$$\mathbf{B}_{t1} = \mathbf{P}_{t|t-1}^{-1} - (\mathbf{Z}'_t - \mathbf{P}_{t|t-1}^{-1}\mathbf{C}_t)\mathbf{H}_t\mathbf{C}'_t\mathbf{P}_{t|t-1}^{-1}$$

e

$$\mathbf{B}_{t2} = (\mathbf{Z}'_t - \mathbf{P}_{t|t-1}^{-1}\mathbf{C}_t)\mathbf{H}_t,$$

combinando isto com a expressão para \mathbf{P}_t dada na equação (3.11), obtemos

$$\begin{aligned}\mathbf{A}_t &= \mathbf{T} \left[\mathbf{P}_{t|t-1} - (\mathbf{P}_{t|t-1} \mathbf{Z}'_t - \mathbf{C}_t) \mathbf{F}_t^{-1} (\mathbf{Z}_t \mathbf{P}_{t|t-1}^{-1} \mathbf{C}'_t) (\mathbf{Z}_t \mathbf{P}_{t|t-1}^{-1} - \mathbf{C}'_t) \right] \\ &\quad \times \left[\mathbf{P}_{t|t-1}^{-1} - (\mathbf{Z}'_t - \mathbf{P}_{t|t-1}^{-1} \mathbf{C}_t) \mathbf{H}_t \mathbf{C}'_t \mathbf{P}_{t|t-1}^{-1} \right] \\ &= \mathbf{T} \left[\mathbf{I} - (\mathbf{P}_{t|t-1} \mathbf{Z}'_t - \mathbf{C}_t) \left[\mathbf{I} - \mathbf{F}_t^{-1} (\mathbf{Z}_t \mathbf{P}_{t|t-1}^{-1} - \mathbf{C}'_t) (\mathbf{Z}'_t - \mathbf{P}_{t|t-1}^{-1} \mathbf{C}_t) \right] \right] \mathbf{H}_t \mathbf{C}'_t \mathbf{P}_{t|t-1}^{-1} \\ &\quad + \mathbf{F}_t^{-1} (\mathbf{Z}_t - \mathbf{C}'_t \mathbf{P}_{t|t-1}^{-1})\end{aligned}$$

Empregando 3.13 temos finalmente

$$\mathbf{A}_t = \mathbf{T} \mathbf{M}_t, \quad (3.15)$$

com $\mathbf{M}_t = [\mathbf{I} - (\mathbf{P}_{t|t-1} \mathbf{Z}'_t - \mathbf{C}_t) \mathbf{F}_t^{-1} \mathbf{Z}_t]$.

Para $\tilde{\mathbf{A}}_t$ temos

$$\begin{aligned}\tilde{\mathbf{A}}_t &= \mathbf{T} \left[\mathbf{P}_{t|t-1} - (\mathbf{P}_{t|t-1} \mathbf{Z}'_t - \mathbf{C}_t) \mathbf{F}_t^{-1} (\mathbf{Z}_t \mathbf{P}_{t|t-1}^{-1} \mathbf{C}'_t) (\mathbf{Z}_t \mathbf{P}_{t|t-1}^{-1} - \mathbf{C}'_t) \right] \left[(\mathbf{Z}'_t - \mathbf{P}_{t|t-1}^{-1} \mathbf{C}_t) \mathbf{H}_t \right] \\ &= \mathbf{T} (\mathbf{P}_{t|t-1} \mathbf{Z}'_t - \mathbf{C}_t) \left[\mathbf{I} - \mathbf{F}_t^{-1} (\mathbf{Z}_t \mathbf{P}_{t|t-1}^{-1} - \mathbf{C}'_t) (\mathbf{Z}'_t - \mathbf{P}_{t|t-1}^{-1} \mathbf{C}_t) \right] \mathbf{H}_t,\end{aligned}$$

e empregando 3.13 novamente, obtemos

$$\tilde{\mathbf{A}}_t = \mathbf{T} \mathbf{K}_t, \quad (3.16)$$

com $\mathbf{K}_t = (\mathbf{P}_{t|t-1} \mathbf{Z}'_t - \mathbf{C}_t) \mathbf{F}_t^{-1}$.

O filtro GLS

Capítulo 4

Exercícios de Simulação

Com o objetivo de avaliar o desempenho dos estimadores introduzidos nos Capítulos 2 e 3 desta dissertação, foi conduzido um estudo de simulação que procura replicar o processo de estimação numa série univariada mensal de taxa de desemprego, satisfazendo um BSM em uma pesquisa com oito painéis rotativos, que segue um esquema 4-8(2) como o empregado pela PME do IBGE.

O capítulo está organizado da seguinte forma: inicialmente, na Seção 4.1, são explicados os cenários considerados e os mecanismos empregados para a geração dos dados. Depois, na Seção 4.2, são apresentados os procedimentos levados a cabo para a estimação, tanto da estrutura de autocorrelação dos erros amostrais e das variâncias desconhecidas do modelo quanto a dos parâmetros de interesse. Por fim, em 4.3 são apresentados os principais resultados desta simulação. Para cada cenário, foram feitas 2000 replicações.

4.1 Mecanismo de geração de dados

Como mencionado, é de interesse avaliar os métodos apresentados, em um cenário que simule uma taxa de desemprego estimada através de uma pesquisa mensal. Com este intuito, foram analisadas as estimativas obtidas pela PME na área metropolitana de Recife, entre março de 2002 e fevereiro de 2011, para estabelecer os hiperparâmetros

Mecanismo de geração de dados

dos modelos empregados para a geração de dados. Inicialmente, apresentaremos o mecanismo de geração das séries de parâmetros de interesse, $\{\theta_t\}$, depois, o das séries de erros amostrais, $\{e_t\}$, considerando a estrutura de painéis do planejamento amostral através de dois cenários de autocorrelação. Finalmente, apresentaremos o procedimento empregado para a geração das séries de estimativas $\{y_t\}$. Em todos os casos, foram geradas séries com $T = 60, 120$ e 240 observações, isto é, 5, 10 e 20 anos de informação, visando considerar o impacto que isto pode ter sobre os métodos em estudo.

4.1.1 Geração das séries $\{\theta_t\}$

Para cada replicação foi gerada uma série de parâmetros de interesse empregando uma modificação do BSM em (1.5), trocando a equação correspondente ao nível local por:

$$L_t = \phi_L L_{t-1} + \delta + R_{t-1} + \eta_{Lt}.$$

Esta nova equação tem a finalidade de manter a série gerada entre 3% e 20%, valores considerados plausíveis para os propósitos do exercício, ao mesmo tempo que se modifica o modelo o menos possível. Também foi incluída no vetor de estados uma componente irregular I_t que é um ruído branco com variância σ_I^2 . Os erros do modelo, η_{Lt} , η_{Rt} , e η_{St} foram gerados independentemente entre eles e ao longo do tempo, sob distribuição normal com média zero e variâncias

$$\sigma_L^2 = 3.17 \times 10^{-01}, \quad \sigma_R^2 = 5.13 \times 10^{-05} \quad \text{e} \quad \sigma_S^2 = 6.98 \times 10^{-08},$$

para I_t foi assumido $\sigma_I^2 = 1.29 \times 10^{-04}$. As séries L_t , R_t e I_t foram inicializadas usando:

$$L_0 = 12.4378, \quad R_0 = 0.0034, \quad I_0 = 0,$$

enquanto para S_t foi empregado:

$$\begin{aligned} S_{t-11} &= -0.6301, & S_{t-10} &= -0.0756, & S_{t-9} &= 0.4151, & S_{t-8} &= 0.8414, \\ S_{t-7} &= 0.4518, & S_{t-6} &= -0.2973, & S_{t-5} &= 0.2638, & S_{t-4} &= 0.4649, \\ S_{t-3} &= 0.3291, & S_{t-2} &= 0.1225, & S_{t-1} &= -0.3279. \end{aligned}$$

Tendo fixado $\phi_L = 0.97$ e $\delta = (L_0 * (1 - \phi_L))$ foram descartadas, em média, 5% das

séries por conter um ou mais valores fora dos limites estabelecidos, caso no qual foi repetido o procedimento com uma nova semente até gerar uma série válida. Note que a série L_t deixa de ser não estacionária e têm média L_0 . Note também que os dados assim gerados representam uma taxa de desemprego expressa em porcentagem.

4.1.2 Geração das séries $\{e_t\}$

Considerando que o nosso parâmetro de interesse é uma taxa, e que $V(e_t) = V(y_t|\theta_t)$, é de se esperar que os erros amostrais sejam heteroscedásticos, com variância proporcional a $\theta_t(100 - \theta_t)$. Isto é corroborado empiricamente pela Figura 4.1 na qual são apresentadas as estimativas de planejamento obtidas para a taxa de desemprego y_t na Área de Recife, no período mencionado, junto com suas respectivas estimativas de variância calculadas a partir dos coeficientes de variação, ambas cifras publicadas na página web do IBGE¹. Para estes dados, as variâncias estimadas podem ser razoavelmente ajustadas empregando um modelo da forma $\psi y_t(100 - y_t)$, com $\psi = 3.024 \times 10^{-04}$. Assim, para cada replicação, uma vez gerada a série de parâmetros θ_t , foi fixado que para cada período t :

$$V(e_t) = \sigma_{e_t}^2 = 3.024 \times 10^{-04}(\theta_t)(100 - \theta_t). \quad (4.1)$$

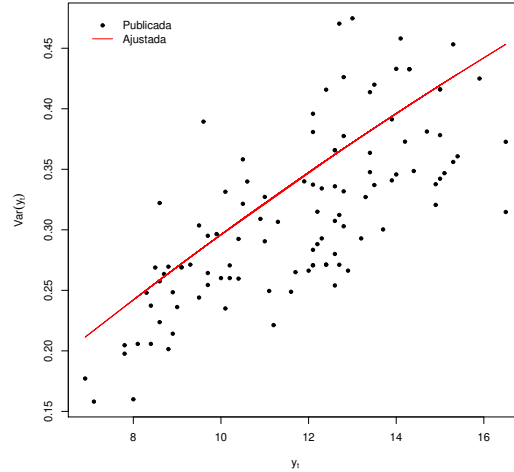
Por outro lado, como mencionado anteriormente, espera-se que o tipo de esquema de rotação assumido introduza autocorrelação entre os erros amostrais ao longo do tempo. Para simular isto, em cada replicação, ao invés de gerar diretamente as séries de erros amostrais, foram geradas oito séries de erros $\{e_t^{(j)}\}$, com $j = 1, 2, \dots, 8$, uma para cada grupo de rotação, sendo que a série $\{e_t\}$ corresponde à média destas oito séries. Para os $e_t^{(j)}$ foram considerados dois cenários de autocorrelação: um que supõe um modelo autorregressivo de primeira ordem e outro baseado no esquema de rotação da amostra. Estes dois cenários serão apresentados a seguir.

¹Consultado em:

http://www.ibge.gov.br/home/estatistica/indicadores/trabalhoerendimento/pme_nova/defaulttab_hist.shtm

Mecanismo de geração de dados

Figura 4.1: Estimativas de taxa de desemprego (y_t) vs Variância calculada a partir do coeficiente de variação publicado. Curva ajustada $3.024 \times 10^{-04}(y_t)(100 - y_t)$. Dados PME-IBGE, Área Metropolitana de Recife, março 2002 - fevereiro 2011.



Cenário 1. Modelo AR(1).

Para um valor de ϕ dado, a série do i -ésimo grupo de rotação satisfaz o modelo

$$e_t^{(i)} = \phi e_{t-1}^{(i)} + a_t^{(i)},$$

com $a_t^{(i)}$ distribuída normalmente com média zero e variância $\sigma_{a_t}^2$ para $i = 1, 2, \dots, 8$ e $t = 1, 2, \dots, T$. Com a finalidade que $V(e_t) = \sigma_{e_t}^2$, a variância dos $e_t^{(i)}$ foi fixada em $8\sigma_{e_t}^2$ e consequentemente, segundo o modelo AR(1), $\sigma_{a_t}^2 = 8(1 - \phi^2)\sigma_{e_t}^2$. Foram utilizados valores de $\phi = 0.4, 0.6$ e 0.8 . Note que neste cenário, os $e_t = \theta_t - y_t$ resultam heteroscedásticos com um modelo autorregressivo de primeira ordem.

Cenário 2. Modelo ad hoc.

A partir da Tabela 1.2 apresentada na Subseção 1.1.3, pode-se observar que sob o esquema de rotação 4-8(2) as amostras dos distintos períodos apresentam a seguinte estrutura de sobreposição:

Tabela 4.1: Quantidade de grupos de rotação com amostra sobreposta, segundo defasagem, num esquema de rotação 4-8(2).

Defasagem	1	2	3	4-8	9	10	11	12	13	14	15	≥ 16
Grupos	6	4	2	0	1	2	3	4	3	2	1	0

Assim, se para cada lag, as observações dos grupos de rotação que se encontram sobrepostos apresentam covariância positiva, é de se esperar que a função de autocorrelação dos erros amostrais apresente maiores valores nos casos de maior sobreposição. Note também que, mesmo que a quantidade de grupos sobrepostos seja igual nos lags 2 e 12, ou nos 3, 10 e 14, são esperadas correlações maiores para as amostras mais próximas no tempo considerando que a amostra tem painéis de domicílios, não de pessoas, e a mobilidade destas pode reduzir a sobreposição efetiva de indivíduos nos domicílios pesquisados, particularmente, após o recesso de 8 meses.

O segundo cenário tenta simular esta estrutura da seguinte forma: para um valor de ϕ prefixado, a correlação entre os erros de um grupo de rotação cuja amostra esteja sobreposta tem os valores de ϕ para o lag 1, ϕ^2 para o lag 2, ϕ^3 para os lag 3 e 9 a 15, entretanto, para amostras não sobrepostas é assumida uma correlação de ϕ^6 independentemente do lag. Esta última correlação procura modelar um fato mencionado em Pfeiffermann et al. (1998): que os distintos painéis de um mesmo grupo de rotação são usualmente selecionados de áreas próximas geograficamente, o que pode gerar alguma correlação entre as estimativas, mesmo sem sobreposição de amostra. Os erros de grupos de rotação diferentes foram considerados independentes. Analogamente com o primeiro cenário, foram empregados valores de $\phi = 0.4, 0.6$ e 0.8 . Um exemplo da estrutura de autocorrelação entre erros de um mesmo grupo de rotação é apresentada na Tabela 4.2. A autocorrelação introduzida por eles nos erros de observação é encontrada na Tabela 4.3. Note que diferentemente do primeiro cenário, neste caso os erros não correspondem a um modelo ARMA específico.

Tabela 4.2: Matriz de correlações entre os erros de um grupo de rotação. Simulação, cenário ad hoc. $\phi = 0.6$.
É presuposto que o painel que participa da pesquisa em t está sendo pesquisado pela oitava vez.

Painel	C			C	C	C	D	D	D	D	B	B	B	B	C	C	C	C	A	A	A	A
	Visita	8	7	6	5	4	3	2	1	8	7	6	5	4	3	2	1	8	7	6	5	
		Período	t	t-1	t-2	t-3	t-4	t-5	t-6	t-7	t-8	t-9	t-10	t-11	t-12	t-13	t-14	t-15	t-16	t-17	t-18	t-19
C	8	t	1	0.6	0.36	0.216	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.216	0.216	0.216	0.216	0.047	0.047	0.047	0.047	
C	7	t-1	0.6	1	0.6	0.36	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.216	0.216	0.216	0.216	0.047	0.047	0.047	0.047	
C	6	t-2	0.36	0.6	1	0.6	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.216	0.216	0.216	0.216	0.047	0.047	0.047	0.047	
C	5	t-3	0.216	0.36	0.6	1	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.216	0.216	0.216	0.216	0.047	0.047	0.047	0.047	
D	4	t-4	0.047	0.047	0.047	0.047	1	0.6	0.36	0.216	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	
D	3	t-5	0.047	0.047	0.047	0.047	0.6	1	0.6	0.36	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	
D	2	t-6	0.047	0.047	0.047	0.047	0.36	0.6	1	0.6	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	
D	1	t-7	0.047	0.047	0.047	0.047	0.216	0.36	0.6	1	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	
B	8	t-8	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	1	0.6	0.36	0.216	0.047	0.047	0.047	0.047	0.047	0.047	0.047	
B	7	t-9	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.6	1	0.6	0.36	0.047	0.047	0.047	0.047	0.047	0.047	0.047	
B	6	t-10	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.36	0.6	1	0.6	0.047	0.047	0.047	0.047	0.047	0.047	0.047	
B	5	t-11	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.216	0.36	0.6	1	0.047	0.047	0.047	0.047	0.047	0.047	0.047	
C	4	t-12	0.216	0.216	0.216	0.216	0.047	0.047	0.047	0.047	0.047	0.047	0.047	1	0.6	0.36	0.216	0.047	0.047	0.047	0.047	
C	3	t-13	0.216	0.216	0.216	0.216	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.6	1	0.6	0.36	0.047	0.047	0.047	0.047
C	2	t-14	0.216	0.216	0.216	0.216	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.36	0.6	1	0.6	0.047	0.047	0.047	0.047
C	1	t-15	0.216	0.216	0.216	0.216	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.216	0.36	0.6	1	0.047	0.047	0.047	0.047
A	8	t-16	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	1	0.6	0.36	0.216
A	7	t-17	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.6	1	0.6	0.36
A	6	t-18	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.36	0.6	1	0.6
A	5	t-19	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.216	0.36	0.6	1

Tabela 4.3: Estrutura de autocorrelação entre e_t e e_{t-k} simulada para o cenário ad hoc.

k	ϕ		
	0.8	0.6	0.4
1	0.66554	0.46166	0.30102
2	0.45107	0.20333	0.08205
3	0.32461	0.08899	0.01907
4 - 8	0.26214	0.04666	0.00410
9	0.29338	0.06782	0.01158
10	0.32461	0.08899	0.01907
11	0.35584	0.11016	0.02656
12	0.38707	0.13133	0.03405
13	0.35584	0.11016	0.02656
14	0.32461	0.08899	0.01907
15	0.29338	0.06782	0.01158
≥ 16	0.26214	0.04666	0.00410

Fixado o esquema de autocorrelação, para cada replicação foi calculada a matriz de covariâncias correspondente, dada a série $\{\theta_t\}$. A seguir, foram geradas as oito séries de erros por grupo de rotação empregando uma distribuição normal multivariada, com vetor de médias zero e a matriz de covariâncias descrita. Para que as ditas séries de erros correspondam ao esquema de rotação assumido e, então, uma delas esteja na oitava visita, outra na sétima e assim por diante, foram geradas séries de $T + 8$ observações. Para conformar a série de erros do i -ésimo grupo de rotação, as primeiras i observações foram desconsideradas.

4.1.3 Geração das séries $\{y_t\}$

Conforme o apresentado no Capítulo 2, para levar a cabo a estimação da estrutura de autocorrelação dos erros amostrais é preciso ter estimativas individuais para cada grupo de rotação, denotadas $y_t^{(j)}$. Assim, empregando os erros por grupo de rotação, $e_t^{(j)}$, estas estimativas foram geradas como

$$y_t^{(j)} = \theta_t + e_t^{(j)} \quad \text{para } j = 1, 2, \dots, 8,$$

Procedimentos de estimação empregados

e, finalmente, a estimativa global para o período t corresponde a

$$y_t = \frac{1}{8} \sum_{j=1}^8 y_t^{(j)}.$$

4.2 Procedimentos de estimação empregados

4.2.1 Estimação de $\rho(k)$

Para cada uma das 2000 replicações, foi estimada a função de autocorrelação dos erros de observação, empregando o método dos pseudo-erros proposto por Pfeiffermann et al. (1998) e apresentado no Capítulo 2. Como uma das hipóteses do método é que os erros de observação são estacionários, o que não é verdade neste caso dado que sua variância depende de θ_t , as séries de pseudo-erros foram padronizadas, dividindo-as por $\hat{\sigma}_{e_t}^2 = 3.024 \times 10^{-04}(y_t)(100 - y_t)$, previamente à estimação. Resultados preliminares desta simulação mostraram uma subestimação das autocorrelações a partir do lag 5, mais notória no cenário 2 e que se acentua quando o tamanho de amostra é menor. Isto sugeriu incluir como parte da análise uma modificação deste estimador, que denotaremos $\hat{\rho}^*(k)$, empregando $T - k$ como denominador da autocovariância de lag k .

4.2.2 Estimação dos hiperparâmetros do modelo

Tendo estimado a função de autocorrelação dos erros amostrais, o passo seguinte para aplicar os métodos apresentados no Capítulo 3 é estimar as variâncias dos erros de evolução σ_L^2 , σ_R^2 , σ_S^2 e σ_I^2 e, para avaliar o filtro de Kalman com erros ARMA, os parâmetros deste último modelo.

Assim, com base na estimativa da função de autocorrelação obtida na Subseção 4.2.1, foram ajustados modelos AR(1) e AR(12) às séries de erros amostrais de dados gerados com o primeiro e o segundo cenários, respectivamente². Posteriormente, estes

²O modelo AR(12) foi escolhido com a finalidade de considerar as autocovariâncias resultantes do esquema de rotação, para os lag 9 e subsequentes.

modelos foram empregados para estimar por máxima verossimilhança as variâncias dos erros de evolução, fazendo uso do filtro de Kalman. O detalhe destes procedimentos é apresentado a seguir.

Ajuste de modelos AR

O ajuste de modelos AR foi feito via minimização numérica da soma dos quadrados das diferenças para cada lag, entre a função de autocorrelação estimada na seção anterior e a correlação teórica esperada sob o modelo AR. Para os dois cenários, foram empregadas as autocorrelações dos lag desde 1 até 15 e a diferença entre as médias dos lags 16 a 20.

Estimativas para as variâncias dos erros de evolução

Empregando o modelo AR ajustado na Subseção 4.2.2, para cada replicação foram estimadas as variâncias dos erros de evolução, via máxima verossimilhança empregando o filtro de Kalman por meio do pacote *sspir* e a função *optim* do software R. Em todos os casos o filtro foi inicializado em forma difusa.

4.2.3 Estimação do parâmetro de interesse

Com o objetivo de enriquecer a análise, para cada cenário de autocorrelação dos erros de observação, foram formulados modelos com distintos níveis de complexidade, desde a estimativa de planejamento y_t que é a entrada dos filtros até modelos que sirvam de ponto de referência teórico, no sentido de descrever os dados da forma teoricamente mais correta possível, empregando os hiperparâmetros verdadeiros com os que estes foram gerados. Para o primeiro cenário foram considerados 7 estimadores e para o segundo 8 estimadores, a saber:

Procedimentos de estimação empregados

Cenário 1

- (1) Estimativa de desenho
- (2) Filtro de Kalman com e_t independentes
- (3) Filtro de Kalman com e_t AR(1). Hiperparâmetros estimados
- (4) Filtro de Kalman com e_t AR(1). Hiperparâmetros reais (Ponto de referência)
- (5) Filtro GLS com estrutura de autocorrelação dos erros $\hat{\rho}(k)$
- (6) Filtro GLS com estrutura de autocorrelação dos erros $\hat{\rho}(k)$ ajustada por um AR(1)
- (7) Filtro GLS com hiperparâmetros reais (Ponto de referência)

Cenário 2

- (1) Estimativa de desenho
- (2) Filtro de Kalman com e_t independentes
- (3) Filtro de Kalman com e_t AR(1). Hiperparâmetros estimados
- (4) Filtro de Kalman com e_t AR(12). Hiperparâmetros estimados
- (5) Filtro de Kalman com e_t AR(12) estimado e variâncias dos erros de evolução reais (Ponto de referência)
- (6) Filtro GLS com estrutura de autocorrelação dos erros $\hat{\rho}(k)$
- (7) Filtro GLS com estrutura de autocorrelação dos erros $\hat{\rho}(k)$ ajustada por um AR(12)
- (8) Filtro GLS com hiperparâmetros reais (Ponto de referência)

Para cada iteração, foram empregados os mesmos valores iniciais para todos os filtros, gerados a partir de uma observação da distribuição normal multivariada com vetor de médias igual aos valores reais indicados na Subseção 4.1.1 e matriz de covariâncias diagonal, contendo as verdadeiras variâncias dos erros de evolução correspondentes.

Como matriz de covariâncias P_0 deste vetor de estados inicial foi empregada uma matriz identidade.

A seguir, são apresentados os principais resultados da simulação. Resultados adicionais são apresentados no Apêndice D.

4.3 Resultados

4.3.1 Estimação de $\rho(k)$

A estimação de $\rho(k)$ foi avaliada empregando a média e EQM das estimativas obtidas para as distintas replicações em cada cenário, assim como o vício médio correspondente. Na tabela 4.4 são apresentados o vício médio e o EQM e nas figuras 4.2 e 4.3 o parâmetro de interesse e as médias das estimativas obtidas.

Em linhas gerais, pode-se concluir que nos dois cenários considerados, tanto o estimador $\hat{\rho}(k)$ quanto o $\hat{\rho}^*(k)$ estiveram razoavelmente perto dos parâmetros de interesse em todos os lags, sem mostrar evidência de vício e com um EQM pequeno, como pode-se observar na Tabela 4.4 e nas Figuras 4.2 e 4.3. Porém, como mencionado anteriormente, o estimador $\hat{\rho}(k)$ parece subestimar levemente as autocorrelações dos últimos lags (Ver Figura 4.3 (d), (g) e (h)) o que é corrigido pelo estimador $\hat{\rho}^*(k)$.

Note também que ambos os estimadores mostram indicações de consistência no sentido que seus EQM diminuam conforme o tamanho da série foi incrementado e o vício se manteve praticamente em zero. Como pode-se verificar nas figuras, para o primeiro cenário, os EQM resultam inversamente proporcionais a \sqrt{n} enquanto para o segundo cenário, a velocidade da diminuição é menor.

Resultados

Figura 4.2: Estimativas da autocorrelação dos erros pelo método dos pseudo-erros. Cenário 1. 2000 replicações. Nas linhas: $\phi = 0.4, 0.6$ e 0.8 . Nas colunas: $T = 60, 120$ e 240 . Linhas: preta, $\rho(k)$; azul contínua, $\hat{\rho}(k)$; vermelha tracejada, $\hat{\rho}^*(k)$; azul pontilhada, percentis 2.5 e 97.5 da distribuição simulada da estimativa $\hat{\rho}(k)$.

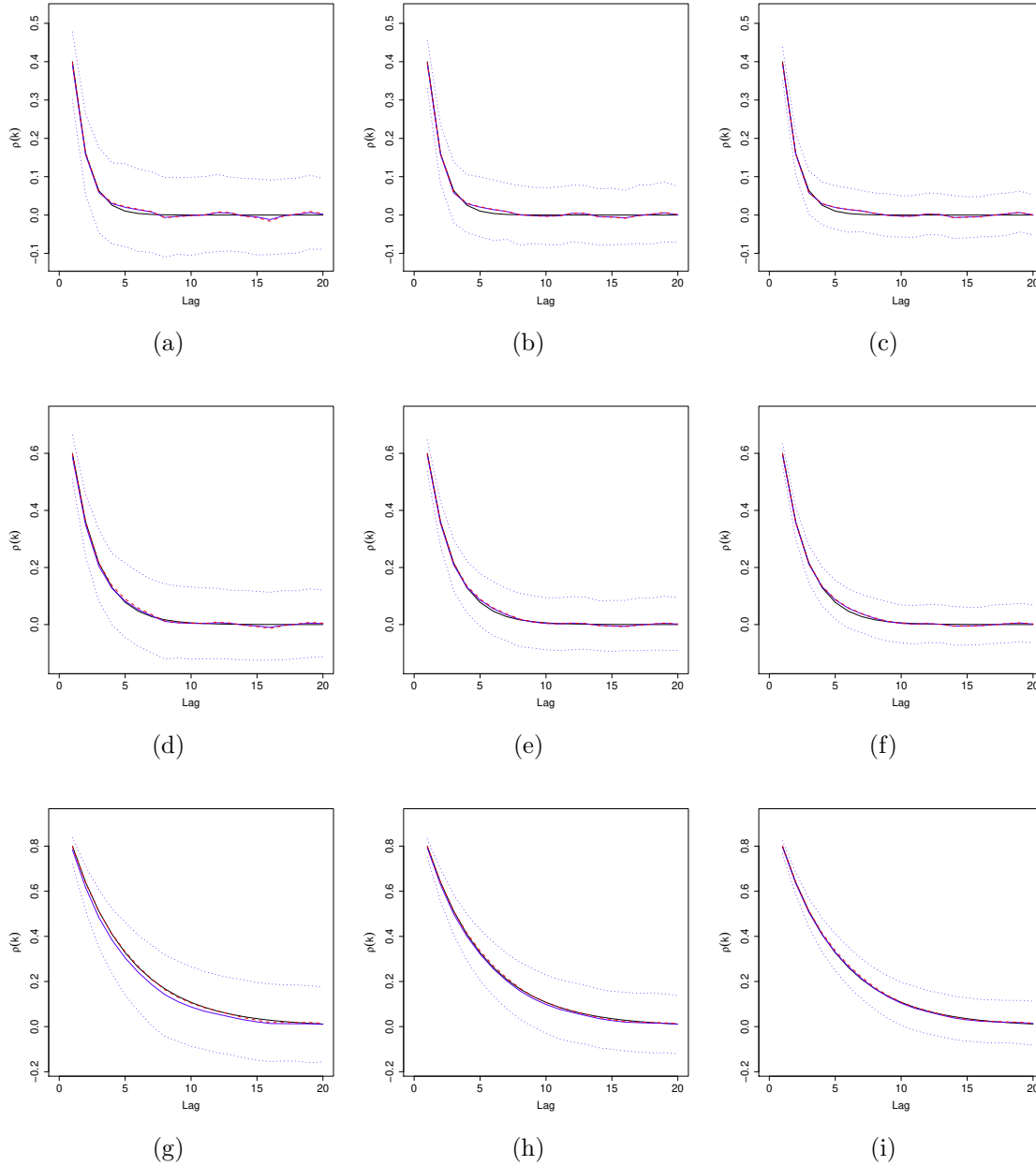
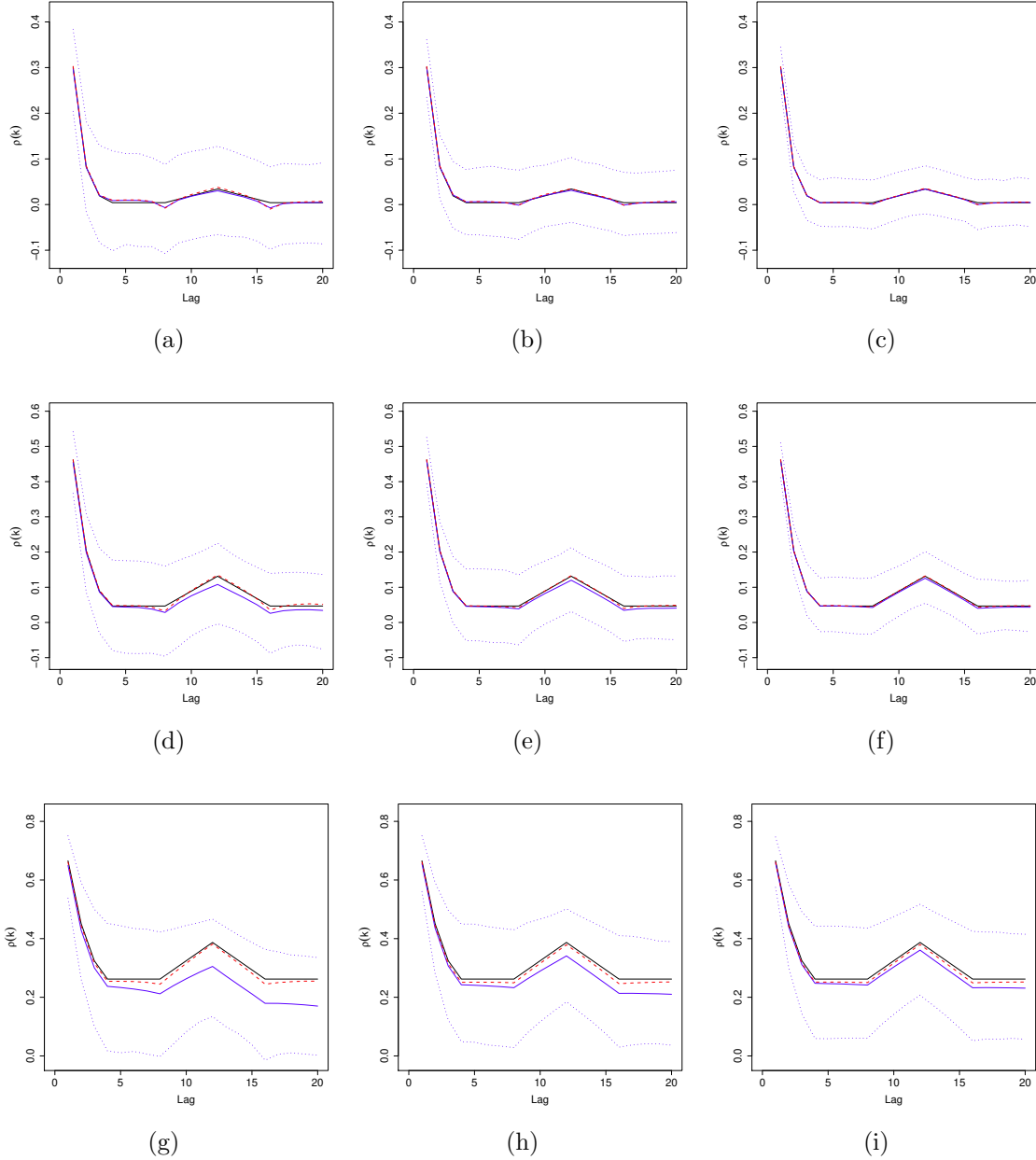


Figura 4.3: Estimativas da autocorrelação dos erros pelo método dos pseudo-erros. Cenário 2. 2000 replicações. Nas linhas: $\phi = 0.4, 0.6$ e 0.8 . Nas colunas: $T = 60, 120$ e 240 . Linhas: preta, $\rho(k)$; azul contínua, $\hat{\rho}(k)$; vermelha tracejada, $\hat{\rho}^*(k)$; azul pontilhada, percentis 2.5 e 97.5 da distribuição simulada da estimativa $\hat{\rho}(k)$.



Resultados

Tabela 4.4: Vício médio e EQM médio na estimação das autocorrelações da série de erros pelo método dos pseudo-erros. 2000 replicações.

Cenário	ϕ	Lags	Estimador	$T = 60$		$T = 120$		$T = 240$	
				Vício	EQM	Vício	EQM	Vício	EQM
1	0.4	1-5	$\hat{\rho}(k)$	-0.0004	0.0532	0.0002	0.0383	0.0002	0.0276
			$\hat{\rho}^*(k)$	0.0034	0.0560	0.0021	0.0392	0.0011	0.0279
	0.6	1-10	$\hat{\rho}(k)$	-0.0035	0.0624	0.0004	0.0453	0.0024	0.0328
			$\hat{\rho}^*(k)$	0.0028	0.0690	0.0036	0.0476	0.0040	0.0337
	0.8	1-15	$\hat{\rho}(k)$	-0.0200	0.0824	-0.0080	0.0592	-0.0022	0.0425
			$\hat{\rho}^*(k)$	-0.0011	0.0937	0.0015	0.0633	0.0026	0.0443
2	0.4	1-15	$\hat{\rho}(k)$	-0.0011	0.0639	-0.0004	0.0521	-0.0003	0.0430
			$\hat{\rho}^*(k)$	0.0018	0.0722	0.0011	0.0550	0.0004	0.0440
	0.6	1-15	$\hat{\rho}(k)$	-0.0116	0.0821	-0.0058	0.0706	-0.0025	0.0623
			$\hat{\rho}^*(k)$	0.0001	0.0920	0.0000	0.0744	0.0004	0.0639
	0.8	1-15	$\hat{\rho}(k)$	-0.0504	0.1255	-0.0303	0.1146	-0.0195	0.1087
			$\hat{\rho}^*(k)$	-0.0089	0.1313	-0.0097	0.1189	-0.0091	0.1113

4.3.2 Estimação dos hiperparâmetros do modelo

Ajuste de modelos AR à estrutura de autocorrelação dos erros de observação

Os resultados deste processo são apresentados na Seção D.2 nos apêndices. Para o cenário 1, segundo a Tabela D.7, não existem evidências de vício e os desvios padrão das estimativas são razoáveis, mesmo para os menores tamanhos de amostra. Para o cenário 2, os modelos ARMA são empregados como aproximação, com dois objetivos: i) conseguir estimativas das variâncias dos erros de evolução empregando o filtro de Kalman com erros ARMA e ii) comparar o desempenho deste filtro com o filtro GLS. Assim, neste cenário foram ajustados modelos AR(1) e AR(12), cujos resultados são apresentados nas Tabelas D.8 e D.9, respectivamente. Com relação ao ajuste empregando um AR(1), as estimativas parecem acumular-se ao redor de um mesmo ponto e seus EQM diminuem quando aumenta o tamanho da série. No entanto, como era de se esperar, salvo para $\phi = 0.4$, o ajuste da função de autocorrelação dos erros é inadequada, como pode ser observado na Figura D.1.

No ajuste empregando um AR(12), se considerarmos o desvio padrão das estimativas da simulação e uma aproximação normal, para $\phi = 0.4$, só a primeira componente autorregressiva resulta estatisticamente diferente de zero a 5% de significância. Para $\phi = 0.6$ nas séries com 240 observações, também a componente de ordem 12 resulta significativa e, no caso de $\phi = 0.8$, são estatisticamente significativas a 5% a primeira

componente, para $T = 60, 120$ e 240 , e a de ordem 12 nos dois últimos tamanhos de série. Nos modelos com todas as componentes, embora algumas delas não sejam estatisticamente significativas, os modelos AR(12) conseguem reproduzir razoavelmente a estrutura de autocorrelação escolhida para a simulação dos erros de observação.

Estimação das variâncias dos erros de evolução

Os resultados desta seção são apresentados na Seção D.3 do apêndice correspondente. Pode-se ver que nos ajustes com modelos AR(1) tanto para o primeiro quanto para o segundo cenários, as estimativas de σ_L^2 se acumulam perto do valor real, embora apresentem bastante variabilidade, mas isto não necessariamente acontece para as outras variâncias de magnitude muito menor. Além disso, note que a variabilidade das estimativas, expressa em termos de seu desvio padrão, nem sempre diminui com o aumento do tamanho da série. Em particular, para as componentes σ_L^2 e σ_R^2 nas séries de tamanho 240, parece aumentar. Analisando isto mais detalhadamente, foi observado que quando a série é mais longa, possivelmente pelo fato dela estar restrita ao intervalo (3% – 20%), o papel das componentes de nível e inclinação local, L_t e R_t , na variância total se confunde em alguns dos casos, produzindo uma maior variabilidade nas estimativas $\hat{\sigma}_L^2$ e $\hat{\sigma}_R^2$ obtidas. Isto é ilustrado na Figura D.2 empregando as estimativas obtidas para o cenário 1 com $\phi = 0.8$.

No ajuste com modelo AR(12), para dados gerados sob o cenário 2, as estimativas da variância do nível local estão ainda mais longe do valor real do que na aproximação com um AR(1). As outras componentes de variância aparecem melhor estimadas com este modelo.

4.3.3 Estimação dos parâmetros de interesse

Na revisão dos distintos métodos de estimação foram avaliadas: i) características da estimação da última observação da série, tais como seu EQM, seu vício médio e a probabilidade de cobertura de um intervalo com o 95% de confiança³ e, ii) características

³O intervalo é calculado através de uma aproximação normal empregando a variância do estimador obtida a partir da matriz de covariâncias P_t , no caso dos estimadores baseados em modelos, e a

Resultados

gerais da estimação da série tais como o EQM médio (sobre todos os períodos), o vício médio e a percentagem de vezes que é rejeitado o teste de Durbin-Watson para não autocorrelação dos erros de previsão, ao 5% de significância. Nessa avaliação, o primeiro ano foi desconsiderado destes cálculos por carregar ruído proveniente da inicialização dos filtros. Os resultados relacionados com a última observação são apresentados a seguir, nas Tabelas 4.5 e 4.6, para os cenários 1 e 2, respectivamente. Os resultados concernentes ao total da série podem ser encontrados na Seção D.4 do Apêndice D. Alguns comentários a seguir:

- Em alguns dos cenários os estimadores em estudo apresentam um vício médio significativamente distinto de zero a 5% de confiança, para a última observação da série. No entanto, este vício não é de grande magnitude, considerando que o objetivo é estimar uma taxa entre 3 e 20 pontos percentuais: no pior caso do cenário 1 o vício é da ordem de 0.04 pontos percentuais, e de 0.21 no pior caso do cenário 2. Como apresentado na Tabela 4.5, no cenário 1 o vício aparece nas séries de 240 observações tanto para os estimadores do filtro de Kalman quanto para os do filtro GLS. Para o cenário 2, segundo a Tabela 4.6, aparece principalmente para os estimadores derivados do Filtro de Kalman.
- Para os valores de $\phi = 0.4$ e 0.6 , as estimativas obtidas pelos métodos apresentados resultam mais precisas do que as obtidas pelo estimador de desenho, nos dois cenários, quando considerada a série total e quando analisada só a última observação. Estas diferenças são estatisticamente significativas a 5%.
- Para o caso de $\phi = 0.8$, alguns dos estimadores apresentam EQM próximos ou ainda maiores que o EQM do estimador de desenho. Isto acontece com o estimador obtido do filtro de Kalman com erros independentes, para o cenário 1 nos três tamanhos de série considerados e para o cenário 2 particularmente nas séries de 60 observações. Da mesma forma, no cenário 1 nas séries com 240 observações, encontra-se este comportamento: i) no estimador do filtro de Kalman com erros AR(1) e hiperparâmetros estimados e ii) para os dois estimadores do filtro GLS com hiperparâmetros estimados. Também, para o cenário 2, isto é observado para: i) o estimador do filtro de Kalman com erros AR(12) e hiperparâmetros

variância estimada calculável a partir dos coeficientes de variação publicados, no caso do estimador de desenho.

estimados nos três tamanhos de série considerados, e ii) para os dois estimadores do filtro GLS com hiperparâmetros estimados, nas séries com 240 observações. Considerando que estes comportamentos não são observados nos modelos com hiperparâmetros conhecidos, pode-se pensar que isto deriva-se da imprecisa estimação das variâncias dos erros de evolução sob o modelo AR(12).

- Para os dois cenários, quando comparado o estimador de desenho com os estimadores de referência, (5) e (7) no cenário 1 e (5) e (8) no cenário 2, que empregam os valores verdadeiros dos hiperparâmetros, observa-se melhoras consideráveis na precisão, especialmente para os menores valores de ϕ , que é quando os e_t estão menos autocorrelacionados.
- Exceto pelo estimador obtido do filtro de Kalman supondo erros independentes, as probabilidades de cobertura dos intervalos de 95% de confiança calculados a partir dos estimadores considerados são satisfatórias. Quando presupostos erros independentes, as baixas probabilidades de cobertura são consequência de uma subestimação da verdadeira variância do estimador.
- Tanto no cenário 1 quanto no cenário 2, comparando os EQM dos estimadores obtidos com o filtro GLS com hiperparâmetros estimados, isto é, os estimadores (6) e (7), não se observa uma perda significativa de precisão ao ajustar a estrutura de autocorrelação dos erros de observação com um modelo AR(12), que despreza as correlações para os lags superiores.
- Quando comparados os estimadores equivalentes obtidos pelos filtros de Kalman e GLS, isto é, os estimadores (4) e (6) e os (5) e (7), no cenário 1 para todos os valores de ϕ considerados; e os (4) e (5) e os (6) e (8), no cenário 2 para $\phi = 0.4$ e 0.6 ; pode-se observar uma menor precisão nos estimadores obtidos através do filtro GLS. Isto é derivado do fato que o estimador obtido pelo filtro de Kalman é o de menor EQM entre todos os estimadores lineares calculáveis a partir dos y_i para $i = 1, 2, \dots, t$, enquanto o GLS é o de menor EQM na subclasse de estimadores lineares calculados a partir da estimativa em $t-1$ e o y_t . No entanto, a diferença na precisão entre eles é menos importante quando os erros têm uma baixa correlação do que quando eles são altamente correlacionados. Este resultado já era esperado dado que o estimador do filtro GLS coincide com o do filtro de Kalman no caso de erros independentes, como mencionado em Pfeiffermann e Tiller (2006).

Resultados

- Comparados estes mesmos estimadores no cenário 2 para $\phi = 0.8$, a diferença em precisão entre os estimadores com hiperparâmetros conhecidos nos dois métodos é menor e o filtro GLS com hiperparâmetros estimados aparece com um EQM levemente menor que o correspondente com o filtro de Kalman. Estas diferenças apresentam significância estatística a 5%.
- Observando os resultados relacionados com o teste de Durbin-Watson inclusos nas Tabelas D.12 a D.29 no Apêndice D, é possível observar que só o filtro de Kalman com hiperparâmetros conhecidos consegue filtrar *completamente* a autocorrelação presente nos erros de observação. Nos outros casos, e particularmente para o filtro GLS, ainda persiste autocorrelação nos erros de previsão do modelo.
- Nos resultados desta simulação é observado que os ganhos em precisão obtidos pelos estimadores baseados em modelos, quando comparados com o estimador de desenho, são maiores para os menores valores de ϕ , nos dois cenários, o que sugeriria que é mais conveniente que os erros de observação tenham baixa autocorrelação. No entanto, esta é uma falsa conclusão pois na simulação de dados, a variância dos estimadores de desenho foi fixada no mesmo valor para os três valores de ϕ considerados e, na prática, sob a perspectiva de inferência baseada no desenho, o usual é que a variância dos estimadores de desenho seja menor para amostras com alta sobreposição, i.e. com erros de observação mais autocorrelacionados.

Tabela 4.5: Resumo das estimativas para $\{\theta_t\}$ a partir do segundo ano. Cenário 1. Modelos considerados: (1) Estimativa de desenho; Filtro de Kalman (2) com e_t independentes, (3) com e_t $AR(1)$, (4) com e_t $AR(1)$ com hiperparâmetros conhecidos; Filtro GLS com estrutura de autocorrelação (5) estimativas iniciais, (6) estimativas iniciais ajustadas $AR(1)$, (7) hiperparâmetros conhecidos.

ϕ	T	Estimador		(1)	(2)	(3)	(4)	(5)	(6)	(7)
0.4	60	EQM ult.	Média	0.3373	0.2717	0.2689	0.2608	0.2718	0.2715	0.2676
			Erro Padr.	0.0109	0.0090	0.0088	0.0086	0.0090	0.0090	0.0088
		Vício médio ult.	Média	0.0171	0.0063	-0.0033	0.0059	0.0101	0.0101	0.0167
			Erro Padr.	0.0130	0.0117	0.0116	0.0114	0.0117	0.0117	0.0116
		Prob. Cober. ult.	IC(95%)	94.90	91.40	94.20	95.40	94.60	94.65	95.30
	120	EQM ult.	Média	0.3194	0.2478	0.2432	0.2375	0.2495	0.2495	0.2478
			Erro Padr.	0.0106	0.0084	0.0081	0.0079	0.0084	0.0084	0.0083
		Vício médio ult.	Média	0.0085	-0.0118	-0.0189	-0.0089	-0.0080	-0.0081	-0.0019
			Erro Padr.	0.0126	0.0111	0.0110	0.0109	0.0112	0.0112	0.0111
		Prob. Cober. ult.	IC(95%)	94.20	91.55	94.45	95.15	94.50	94.55	95.15
	240	EQM ult.	Média	0.3162	0.2420	0.2436	0.2368	0.2484	0.2483	0.2440
			Erro Padr.	0.0106	0.0079	0.0087	0.0078	0.0088	0.0088	0.0080
		Vício médio ult.	Média	-0.0190	-0.0355	-0.0398	-0.0306	-0.0330	-0.0331	-0.0273
			Erro Padr.	0.0126	0.0110	0.0110	0.0109	0.0111	0.0111	0.0110
		Prob. Cober. ult.	IC(95%)	94.00	91.35	94.15	95.00	94.45	94.45	95.40
0.6	60	EQM ult.	Média	0.3399	0.3108	0.2924	0.2855	0.3005	0.3005	0.2983
			Erro Padr.	0.0109	0.0101	0.0094	0.0093	0.0098	0.0098	0.0096
		Vício médio ult.	Média	0.0233	0.0145	-0.0025	0.0062	0.0183	0.0182	0.0223
			Erro Padr.	0.0130	0.0125	0.0121	0.0119	0.0123	0.0123	0.0122
		Prob. Cober. ult.	IC(95%)	95.05	88.65	94.10	94.80	94.45	94.80	95.30
	120	EQM ult.	Média	0.3164	0.2850	0.2661	0.2587	0.2792	0.2795	0.2770
			Erro Padr.	0.0107	0.0097	0.0089	0.0086	0.0095	0.0095	0.0094
		Vício médio ult.	Média	0.0119	-0.0085	-0.0226	-0.0109	-0.0026	-0.0029	0.0018
			Erro Padr.	0.0126	0.0119	0.0115	0.0114	0.0118	0.0118	0.0118
		Prob. Cober. ult.	IC(95%)	94.45	90.00	94.55	94.95	94.30	94.45	95.15
	240	EQM ult.	Média	0.3137	0.2779	0.2746	0.2603	0.2828	0.2828	0.2744
			Erro Padr.	0.0104	0.0089	0.0097	0.0084	0.0097	0.0097	0.0089
		Vício médio ult.	Média	-0.0196	-0.0338	-0.0497	-0.0353	-0.0332	-0.0333	-0.0277
			Erro Padr.	0.0125	0.0118	0.0117	0.0114	0.0119	0.0119	0.0117
		Prob. Cober. ult.	IC(95%)	94.40	88.35	94.35	95.00	94.15	94.15	94.75
0.8	60	EQM ult.	Média	0.3443	0.3723	0.3230	0.3163	0.3331	0.3331	0.3310
			Erro Padr.	0.0110	0.0118	0.0103	0.0100	0.0106	0.0106	0.0105
		Vício médio ult.	Média	0.0273	0.0207	-0.0195	-0.0087	0.0206	0.0207	0.0227
			Erro Padr.	0.0131	0.0136	0.0127	0.0126	0.0129	0.0129	0.0129
		Prob. Cober. ult.	IC(95%)	94.10	84.45	93.30	94.25	93.70	93.70	93.80
	120	EQM ult.	Média	0.3160	0.3381	0.2915	0.2866	0.3054	0.3056	0.3044
			Erro Padr.	0.0106	0.0112	0.0096	0.0095	0.0102	0.0102	0.0102
		Vício médio ult.	Média	0.0127	-0.0059	-0.0449	-0.0301	-6e-04	-8e-04	0.0023
			Erro Padr.	0.0126	0.0130	0.0120	0.0120	0.0124	0.0124	0.0123
		Prob. Cober. ult.	IC(95%)	94.70	85.20	94.45	94.60	94.70	94.80	94.85
	240	EQM ult.	Média	0.3157	0.3419	0.3531	0.2913	0.3440	0.3443	0.3067
			Erro Padr.	0.0104	0.0113	0.0165	0.0093	0.0136	0.0136	0.0100
		Vício médio ult.	Média	-0.0128	-0.0242	-0.0825	-0.0499	-0.0345	-0.0344	-0.0216
			Erro Padr.	0.0126	0.0131	0.0132	0.0120	0.0131	0.0131	0.0124
		Prob. Cober. ult.	IC(95%)	94.45	84.70	92.00	94.65	92.75	92.70	94.70

Resultados

Tabela 4.6: Resumo das estimativas para $\{\theta_t\}$ a partir do segundo ano. Cenário 2.
Modelos considerados: (1) Estimativa de desenho; Filtro de Kalman (2) com e_t independentes, (3) com e_t AR(1), (4) com e_t AR(12), (5) com e_t AR(12) com hiperparâmetros conhecidos; Filtro GLS com estrutura de autocorrelação (6) estimativas iniciais (7), estimativas iniciais ajustadas AR(12), (8) hiperparâmetros conhecidos.

ϕ	T	Estimador		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0.4	60	EQM ult.	Média	0.3229	0.2709	0.2665	0.2706	0.2555	0.2712	0.2709	0.2597
			Erro Padr.	0.0101	0.0087	0.0086	0.0086	0.0081	0.0087	0.0086	0.0082
		Vicio médio ult.	Média	-0.0143	-0.0240	-0.0324	-0.0350	-0.0258	-0.0212	-0.0214	-0.0157
			Erro Padr.	0.0127	0.0116	0.0115	0.0116	0.0113	0.0116	0.0116	0.0114
		Prob. Cober. ult.	IC(95%)	94.50	91.25	93.40	93.70	95.20	92.90	93.10	94.15
	120	EQM ult.	Média	0.3226	0.2380	0.2348	0.2350	0.2318	0.2387	0.2386	0.2385
			Erro Padr.	0.0107	0.0077	0.0077	0.0077	0.0075	0.0078	0.0078	0.0077
		Vicio médio ult.	Média	-0.0057	-0.0249	-0.0307	-0.0349	-0.0238	-0.0238	-0.0239	-0.0168
			Erro Padr.	0.0127	0.0109	0.0108	0.0108	0.0108	0.0109	0.0109	0.0109
		Prob. Cober. ult.	IC(95%)	94.50	92.20	94.70	95.35	95.95	94.20	94.25	95.05
	240	EQM ult.	Média	0.3072	0.2319	0.2291	0.2307	0.2280	0.2321	0.2321	0.2318
			Erro Padr.	0.0104	0.0077	0.0076	0.0077	0.0076	0.0076	0.0076	0.0076
		Vicio médio ult.	Média	-0.0131	-0.0272	-0.0307	-0.0329	-0.0229	-0.0245	-0.0245	-0.0185
			Erro Padr.	0.0124	0.0108	0.0107	0.0107	0.0107	0.0108	0.0108	0.0108
		Prob. Cober. ult.	IC(95%)	94.65	92.75	94.80	95.05	95.80	94.65	94.70	95.15
0.6	60	EQM ult.	Média	0.3156	0.2780	0.2699	0.2758	0.2567	0.2772	0.2771	0.2680
			Erro Padr.	0.0103	0.0091	0.0087	0.0089	0.0084	0.0090	0.0090	0.0088
		Vicio médio ult.	Média	-0.0041	-0.0187	-0.0315	-0.0523	-0.0298	-0.0185	-0.0186	-0.0081
			Erro Padr.	0.0126	0.0118	0.0116	0.0117	0.0113	0.0118	0.0118	0.0116
		Prob. Cober. ult.	IC(95%)	95.10	91.00	94.35	94.75	96.40	93.80	93.85	95.25
	120	EQM ult.	Média	0.3113	0.2591	0.2491	0.2535	0.2485	0.2581	0.2581	0.2581
			Erro Padr.	0.0106	0.0085	0.0082	0.0082	0.0082	0.0085	0.0086	0.0086
		Vicio médio ult.	Média	0.0178	-0.0117	-0.0163	-0.0397	-0.0193	-0.0077	-0.0077	1e-04
			Erro Padr.	0.0125	0.0114	0.0112	0.0112	0.0111	0.0114	0.0114	0.0114
		Prob. Cober. ult.	IC(95%)	95.15	92.10	95.25	96.05	96.70	95.15	95.25	95.50
	240	EQM ult.	Média	0.3021	0.2552	0.2474	0.2572	0.2452	0.2618	0.2620	0.2539
			Erro Padr.	0.0096	0.0083	0.0082	0.0086	0.0080	0.0087	0.0087	0.0082
		Vicio médio ult.	Média	-0.0025	-0.0246	-0.0319	-0.0539	-0.0306	-0.0215	-0.0216	-0.0150
			Erro Padr.	0.0123	0.0113	0.0111	0.0113	0.0111	0.0114	0.0114	0.0113
		Prob. Cober. ult.	IC(95%)	95.95	91.30	94.85	95.25	96.50	94.65	94.65	95.55
0.8	60	EQM ult.	Média	0.2976	0.3091	0.2868	0.3171	0.2756	0.2871	0.2871	0.2793
			Erro Padr.	0.0095	0.0094	0.0091	0.0101	0.0087	0.0088	0.0088	0.0087
		Vicio médio ult.	Média	-0.0013	-0.0088	-0.0549	-0.1502	-0.0683	-0.0154	-0.0157	-0.0024
			Erro Padr.	0.0122	0.0124	0.0119	0.0121	0.0116	0.0120	0.0120	0.0118
		Prob. Cober. ult.	IC(95%)	95.65	88.39	95.20	94.64	97.60	95.15	95.15	95.95
	120	EQM ult.	Média	0.3068	0.2969	0.2893	0.3317	0.2860	0.2836	0.2828	0.2816
			Erro Padr.	0.0095	0.0092	0.0089	0.0105	0.0088	0.0087	0.0087	0.0087
		Vicio médio ult.	Média	-0.0022	-0.0270	-0.0695	-0.1992	-0.1055	-0.0271	-0.0261	-0.0153
			Erro Padr.	0.0124	0.0122	0.0119	0.0121	0.0117	0.0119	0.0119	0.0119
		Prob. Cober. ult.	IC(95%)	95.06	88.31	94.30	94.56	96.67	94.46	94.35	95.01
	240	EQM ult.	Média	0.2861	0.2779	0.2837	0.3394	0.2633	0.2834	0.2805	0.2643
			Erro Padr.	0.0095	0.0092	0.0105	0.0119	0.0086	0.0095	0.0093	0.0087
		Vicio médio ult.	Média	0.0042	-0.0102	-0.0659	-0.2146	-0.0999	-0.0101	-0.0109	-0.0039
			Erro Padr.	0.0120	0.0118	0.0118	0.0121	0.0113	0.0119	0.0118	0.0115
		Prob. Cober. ult.	IC(95%)	95.95	88.48	94.69	93.00	96.74	93.63	93.79	95.79

Capítulo 5

Estimação com o filtro GLS sob restrições de benchmark

No Capítulo 3 foi apresentado o filtro GLS, introduzido em Pfeiffermann e Tiller (2006), para realizar estimação em pesquisas repetidas empregando MEE com erros de observação autocorrelacionados, enfatizando sua utilidade nos casos em que a estrutura desta correlação é complexa. Neste capítulo, vamos abordar sua utilização na estimação simultânea para vários domínios incluindo restrições de benchmark.

O capítulo está organizado da seguinte forma: na Seção 5.1 é feita uma motivação do problema de benchmark no contexto de estimação em pesquisas repetidas; posteriormente, na Seção 5.2 é introduzida uma formulação para considerar séries temporais de vários domínios em um mesmo MEE. Com estes dois elementos, na Seção 5.3, é apresentado o MEE incluindo restrições de benchmark, assim como a forma de calcular as estimativas do vetor de estados e suas covariâncias empregando o filtro GLS. Por fim, na Seção 5.4 são exibidos os resultados de uma aplicação desta metodologia na Pesquisa Mensal de Emprego (PME) do IBGE, utilizando informação de seis regiões metropolitanas do Brasil no período março 2002 - fevereiro 2012.

5.1 As restrições de benchmark

Na implementação de uma pesquisa por amostragem, é comum que sejam requeridas estimativas dos parâmetros de interesse, não só para grandes domínios, para os quais os estimadores diretos do desenho amostral podem ter um nível aceitável de precisão, mas também para subdivisões deles, nas quais este tipo de estimação não é confiável devido ao pequeno tamanho de amostra disponível (Datta, 2009). No caso de levantamentos repetidos, as metodologias apresentadas no Capítulo 3 desta dissertação podem ser empregadas para contornar este problema de estimação, combinando informações dos distintos períodos da pesquisa e de todos os domínios para obter estimadores mais precisos baseados em modelos.

No entanto, se para os domínios considerados *grandes* forem empregadas estimativas baseadas no plano amostral empregado e, para suas subdivisões e outros domínios menores, estimativas baseadas em modelos, como as introduzidas no capítulo 3, seria interessante que estes dois tipos de estimações coincidam quando agregadas, pelo menos ao nível dos principais domínios. Isto não só como uma necessidade de coerência na publicação de resultados, mas também, porque é de utilidade na proteção contra possíveis erros de especificação dos modelos usados na inferência (Pfeffermann e Tiller, 2006).

Na proposta que emprega o filtro GLS, este objetivo é atingido incluindo, no processo de estimação, restrições lineares da forma

$$\sum_{d=1}^D w_{dt} y_{dt} = \sum_{d=1}^D w_{dt} \hat{\theta}_{dt}, \quad \text{para } t = 1, 2, \dots \quad (5.1)$$

onde $d = 1, 2, \dots, D$ indexa o conjunto de domínios ou subdivisões para os quais se deseja gerar estimativas; y_{dt} e $\hat{\theta}_{dt}$ são, respectivamente, as estimativas de planejamento e de modelo para o d -ésimo domínio no t -ésimo período; e w_{dt} é uma ponderação para esse domínio nesse momento¹. Quando se pretende obter coerência entre os dois tipos de estimativas, w_{dt} é escolhida de forma tal que para cada t , o valor empregado como benchmark, $\sum_{d=1}^D w_{dt} y_{dt}$, seja igual à estimativa calculada sobre a união destes domínios,

¹i.e. para cada t , $\sum_{d=1}^D w_{dt} = 1$.

denotada por y_t .

Neste sentido, algumas possíveis escolhas para w_{dt} , sugeridas em Pfeiffermann e Tiller (2006) são:

$$w_{1dt} = \frac{1}{D} \quad \text{ou} \quad w_{2dt} = \frac{N_{dt}}{\sum_{d=1}^D N_{dt}}. \quad (5.2)$$

A primeira, que equivale à restrição $\sum_{d=1}^D y_{dt} = \sum_{d=1}^D \hat{\theta}_{dt}$, resulta apropriada quando $y_t = \sum_{d=1}^D y_{dt}$. A segunda pode ser utilizada nos casos em que y_t é uma proporção com denominador $\sum_{d=1}^D N_{dt}$.

Note que, embora um conjunto de restrições deste tipo seja chamado *de benchmark*, o termo é empregado em forma diferente da habitual e, em particular, de seu uso em trabalhos como Durbin e Quenneville (1997). No caso estudado aqui, tanto a série cujos valores são usados como benchmark, quanto as séries insumo do modelo, provém da mesma pesquisa, o que faz com que a restrição tenha uma variância associada e, mais ainda, que possa existir correlação entre ela e as outras séries. Estes dois fatos devem ser considerados no processo de cálculo da variância dos estimadores.

5.2 Um MEE para todos os domínios

Suponha que foram observadas, com a mesma periodicidade e para o mesmo intervalo de tempo, D séries temporais univariadas, $\{y_{1t}\}, \{y_{2,t}\}, \dots, \{y_{D,t}\}$, correspondentes a igual número de domínios, de forma que para cada período e domínio, y_{dt} é um estimador não viesado do parâmetro de interesse θ_{dt} . Para cada d , assuma que y_{dt} segue um MEE univariado segundo a notação de (1.4) com vetor de estados α_{dt} ; vetor de erros de evolução η_{dt} ; matrizes do sistema \mathbf{Z}_{dt} , \mathbf{T}_{dt} , \mathbf{G}_{dt} e \mathbf{Q}_{dt} e erros de observação e_{dt} , sendo que $\text{Cov}(e_{d\tau}, e_{dt}) = \sigma_{d\tau t}$. Se tanto a evolução do parâmetro de interesse quanto a observação da série são independentes entre domínios, i.e. $\text{Cov}(\alpha_{d\tau}, \alpha_{st}) = \mathbf{0}$ e $\text{Cov}(e_{d\tau}, e_{st}) = 0$ sempre que $d \neq s$ e para todo τ, t , as séries podem ser representadas conjuntamente no MEE multivariado:

O filtro GLS sob restrições de benchmark

$$\begin{aligned} \mathbf{y}_t &= \mathbf{Z}_t \boldsymbol{\alpha}_t + \mathbf{e}_t \\ \boldsymbol{\alpha}_t &= \mathbf{T} \boldsymbol{\alpha}_{t-1} + \mathbf{G} \boldsymbol{\eta}_t, \end{aligned} \tag{5.3}$$

onde

$$\mathbf{y}_t = \begin{pmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{Dt} \end{pmatrix}, \quad \boldsymbol{\alpha}_t = \begin{pmatrix} \alpha_{1t} \\ \alpha_{2t} \\ \vdots \\ \alpha_{Dt} \end{pmatrix}, \quad \mathbf{e}_t = \begin{pmatrix} e_{1t} \\ e_{2t} \\ \vdots \\ e_{Dt} \end{pmatrix}, \quad \boldsymbol{\eta}_t = \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \\ \vdots \\ \eta_{Dt} \end{pmatrix},$$

$\mathbf{Z}_t = \text{diag} \{ \mathbf{Z}_{dt} \}$, $\mathbf{T}_t = \text{diag} \{ \mathbf{T}_{dt} \}$, $\mathbf{G}_t = \text{diag} \{ \mathbf{G}_{dt} \}$, $\mathbf{Q}_t = \text{Cov}(\eta_{d\tau}, \eta'_{dt}) = \text{diag} \{ \mathbf{Q}_{dt} \}$ e $\boldsymbol{\Sigma}_{\tau t} = \text{Cov}(\mathbf{e}_\tau, \mathbf{e}'_t) = \text{diag} \{ \sigma_{d\tau t} \}$. A notação $\text{diag} \{ \mathbf{M}_d \}$ é empregada neste caso para indicar uma matriz diagonal por blocos, com \mathbf{M}_d no d -ésimo bloco, para $\{ \mathbf{M}_d ; d = 1, \dots, D \}$.

5.3 O filtro GLS sob restrições de benchmark

Pela definição de \mathbf{Z}_t em (5.3), a equação (5.1) pode ser reescrita como

$$\sum_{d=1}^D w_{dt} y_{dt} = \sum_{d=1}^D w_{dt} \mathbf{Z}_{dt} \hat{\boldsymbol{\alpha}}_{dt}^{\text{bm}k}, \quad \text{para } t = 1, 2, \dots,$$

onde o sobrescrito *bm**k* é empregado para denotar que as restrições de benchmark são satisfeitas. Por outro lado, para cada t ,

$$\sum_{d=1}^D w_{dt} y_{dt} = \sum_{d=1}^D w_{dt} \mathbf{Z}_{dt} \boldsymbol{\alpha}_{dt} + \sum_{d=1}^D w_{dt} e_{dt}, \quad \text{para } t = 1, 2, \dots, \tag{5.4}$$

também segundo o modelo (5.3). Assim, se incluirmos (5.4) como parte do MEE, o filtro GLS pode ser usado para encontrar estimativas que satisfaçam as restrições de benchmark, sempre que $\sum_{d=1}^D w_{dt} e_{dt}$ seja fixada em zero. Note, no entanto, que o novo modelo incluindo (5.4) é equivalente ao anterior, no sentido de que esta equação não traz consigo nenhuma nova informação.

Sejam então

$$\bar{y}_{wt} = \sum_{d=1}^D w_{dt} y_{dt}, \quad \bar{e}_{wt} = \sum_{d=1}^D w_{dt} e_{dt}, \quad \text{e} \quad \bar{\mathbf{Z}}_{\mathbf{wt}} = \begin{bmatrix} w_{1t} \mathbf{Z}_{1t} & \dots & w_{Dt} \mathbf{Z}_{Dt} \end{bmatrix},$$

a equação (5.4) pode ser escrita como

$$\bar{y}_{wt} = \bar{\mathbf{Z}}_{\mathbf{wt}} \boldsymbol{\alpha}_t + \bar{e}_{wt} \quad \text{para } t = 1, 2, \dots$$

Se definirmos

$$\tilde{\mathbf{y}}'_t = \begin{bmatrix} \mathbf{y}'_t & \bar{y}_{wt} \end{bmatrix}, \quad \tilde{\mathbf{e}}'_t = \begin{bmatrix} \mathbf{e}'_t & \bar{e}_{wt} \end{bmatrix}, \quad \tilde{\mathbf{Z}}'_t = \begin{bmatrix} \mathbf{Z}'_t & \bar{\mathbf{Z}}_{\mathbf{wt}} \end{bmatrix},$$

os D domínios e as restrições de benchmark podem ser considerados conjuntamente no modelo

$$\begin{aligned} \tilde{\mathbf{y}}_t &= \tilde{\mathbf{Z}}_t \boldsymbol{\alpha}_t + \tilde{\mathbf{e}}_t \\ \boldsymbol{\alpha}_t &= \mathbf{T} \boldsymbol{\alpha}_{t-1} + \mathbf{G} \boldsymbol{\eta}_t, \end{aligned} \tag{5.5}$$

com

$$\tilde{\boldsymbol{\Sigma}}_{\tau t} = \text{Cov}(\tilde{\mathbf{e}}_\tau, \tilde{\mathbf{e}}_t) = \begin{bmatrix} \boldsymbol{\Sigma}_{\tau t} & \mathbf{h}_{\tau t} \\ \mathbf{h}'_{t\tau} & \nu_{\tau, t} \end{bmatrix},$$

onde $\mathbf{h}_{\tau t} = \text{Cov}(\mathbf{e}_t, \bar{e}_{wt})$ e $\nu_{\tau t} = \text{Cov}(\bar{e}_{w\tau}, \bar{e}_{wt})$. Veja que, sob a hipótese de independência entre os erros de observação dos domínios,

$$\mathbf{h}'_{\tau t} = \begin{bmatrix} w_{1t} \sigma_{1\tau t} & \dots & w_{Dt} \sigma_{D\tau t} \end{bmatrix} \quad \text{e}$$

$$\nu_{\tau t} = \sum_{d=1}^D w_{d\tau} w_{dt} \sigma_{d\tau t}.$$

Note também que $\mathbf{h}_{\tau t}$ não necessariamente coincide com $\mathbf{h}_{t\tau}$, exceto quando os pesos w_{dt} são mantidos fixos ao longo do tempo. Para cada t , a estimativa do vetor de estados de (5.5) que satisfaz as restrições de benchmark, denotada por $\hat{\boldsymbol{\alpha}}_t^{\text{bm k}}$, pode ser obtida aplicando o filtro GLS a uma modificação deste modelo, onde $\tilde{\mathbf{e}}'_t$ é substituído por $\tilde{\mathbf{e}}'_{0t} = \begin{bmatrix} \mathbf{e}'_t & 0 \end{bmatrix}$. No entanto, embora este procedimento permita o cálculo de $\hat{\boldsymbol{\alpha}}_t^{\text{bm k}}$, ele não pode ser empregado para fazer inferência para este estimador porque o modelo

O filtro GLS sob restrições de benchmark

ajustado não corresponde ao modelo real dado por (5.3). Em particular, a matriz \mathbf{P}_t que poderia ser obtida diretamente ao aplicar o filtro GLS a este modelo restrito, não é a matriz de covariância do estimador $\hat{\alpha}_t^{\text{bmk}}$, pois não considera a variabilidade associada ao benchmark. Apresentaremos a forma de calcular a verdadeira matriz de covariâncias do estimador, $\mathbf{P}_t^{\text{bmk}}$, e a matriz de covariâncias entre o erro de previsão, $\mathbf{u}_{t|t-1}^{\text{bmk}} = \mathbf{T}\hat{\alpha}_{t-1}^{\text{bmk}} - \alpha_t$, e o erro de observação, $\tilde{\mathbf{e}}_t$, denotada por \mathbf{C}_t na Subseção 3.2.1, que será chamada de $\mathbf{C}_t^{\text{bmk}}$.

Inicialmente, note que no modelo utilizando $\tilde{\mathbf{e}}'_{0t}$, as componentes $\nu_{\tau t}$ e $\mathbf{h}_{\tau t}$ de cada matriz $\tilde{\Sigma}_{\tau t}$ são zeros, o que será sinalizado denotando estas como $\tilde{\Sigma}_{\tau t}^*$. Analogamente, para cada t , a matriz \mathbf{C}_t tem sua última coluna igualada a zero, o que será indicado chamando-a de \mathbf{C}_t^* . Com esta nova notação, e empregando a equação (3.14), o estimador sob restrições de benchmark pode ser escrito como:

$$\hat{\alpha}_t^{\text{bmk}} = \mathbf{T}\hat{\alpha}_{t-1}^{\text{bmk}} + (\mathbf{P}_{t|t-1}^{\text{bmk}}\tilde{\mathbf{Z}}'_t - \mathbf{C}_t^*)\mathbf{F}_t^{*-1}(\tilde{\mathbf{y}}_t - \tilde{\mathbf{Z}}_t\mathbf{T}\hat{\alpha}_{t-1}^{\text{bmk}}), \quad (5.6)$$

com $\mathbf{F}_t^* = [\tilde{\mathbf{Z}}_t\mathbf{P}_{t|t-1}^{\text{bmk}}\tilde{\mathbf{Z}}'_t - \tilde{\mathbf{Z}}_t\mathbf{C}_t^* - \mathbf{C}_t^{*'}\tilde{\mathbf{Z}}'_t + \tilde{\Sigma}_{tt}^*]$. Note que a matriz $\mathbf{P}_{t|t-1}^{\text{bmk}}$ aqui utilizada é a verdadeira matriz de covariância da previsão um passo à frente,

$$\mathbf{P}_{t|t-1}^{\text{bmk}} = \mathbf{T}\mathbf{P}_{t-1}^{\text{bmk}}\mathbf{T}' + \mathbf{G}\mathbf{Q}\mathbf{G}'.$$

Abordaremos agora o cálculo da matriz de covariâncias do estimador $\hat{\alpha}_t^{\text{bmk}}$, $\mathbf{P}_t^{\text{bmk}}$. Reordenando os termos na equação (5.6), podemos escrever

$$\hat{\alpha}_t^{\text{bmk}} = \left[\mathbf{I} - (\mathbf{P}_{t|t-1}^{\text{bmk}}\tilde{\mathbf{Z}}'_t - \mathbf{C}_t^*)\mathbf{F}_t^{*-1}\tilde{\mathbf{Z}}_t \right] \mathbf{T}\hat{\alpha}_{t-1}^{\text{bmk}} + (\mathbf{P}_{t|t-1}^{\text{bmk}}\tilde{\mathbf{Z}}'_t - \mathbf{C}_t^*)\mathbf{F}_t^{*-1}\tilde{\mathbf{y}}_t,$$

ou equivalentemente

$$\hat{\alpha}_t^{\text{bmk}} = \left[\mathbf{I} - \mathbf{K}_t^{\text{bmk}}\tilde{\mathbf{Z}}_t \right] \mathbf{T}\hat{\alpha}_{t-1}^{\text{bmk}} + \mathbf{K}_t^{\text{bmk}}\tilde{\mathbf{y}}_t,$$

com $\mathbf{K}_t^{\text{bmk}} = (\mathbf{P}_{t|t-1}^{\text{bmk}}\tilde{\mathbf{Z}}'_t - \mathbf{C}_t^*)\mathbf{F}_t^{*-1}$. Se substituirmos $\tilde{\mathbf{y}}_t$ por $\tilde{\mathbf{Z}}_t\alpha_t + \tilde{\mathbf{e}}_t$, pode-se ver que

$$\hat{\alpha}_t^{\text{bmk}} = \left[\mathbf{I} - \mathbf{K}_t^{\text{bmk}}\tilde{\mathbf{Z}}_t \right] \mathbf{T}\hat{\alpha}_{t-1}^{\text{bmk}} + \mathbf{K}_t^{\text{bmk}}\tilde{\mathbf{Z}}_t\alpha_t + \mathbf{K}_t^{\text{bmk}}\tilde{\mathbf{e}}_t.$$

Para encontrar a esperança de $(\hat{\alpha}_t^{\text{bmk}} - \alpha_t)(\hat{\alpha}_t^{\text{bmk}} - \alpha_t)'$ pode-se escrever α_t como $[\mathbf{I} - \mathbf{K}_t^{\text{bmk}}\tilde{\mathbf{Z}}_t]\alpha_t + \mathbf{K}_t^{\text{bmk}}\tilde{\mathbf{Z}}_t\alpha_t$, para obter

$$\hat{\alpha}_t^{\text{bmk}} - \alpha_t = [\mathbf{I} - \mathbf{K}_t^{\text{bmk}}\tilde{\mathbf{Z}}_t](\mathbf{T}\hat{\alpha}_{t-1}^{\text{bmk}} - \alpha_t) + \mathbf{K}_t^{\text{bmk}}\tilde{\mathbf{e}}_t,$$

e também

$$\begin{aligned} (\hat{\alpha}_t^{\text{bmk}} - \alpha_t)(\hat{\alpha}_t^{\text{bmk}} - \alpha_t)' = \\ [\mathbf{M}_t^{\text{bmk}}(\mathbf{T}\hat{\alpha}_{t-1}^{\text{bmk}} - \alpha_t) + \mathbf{K}_t^{\text{bmk}}\tilde{\mathbf{e}}_t] \left[(\mathbf{T}\hat{\alpha}_{t-1}^{\text{bmk}} - \alpha_t)' \mathbf{M}_t'^{\text{bmk}} + \tilde{\mathbf{e}}_t' \mathbf{K}_t'^{\text{bmk}} \right], \end{aligned}$$

com $\mathbf{M}_t^{\text{bmk}} = \mathbf{I} - \mathbf{K}_t^{\text{bmk}}\tilde{\mathbf{Z}}_t$. Finalmente, para tomar as esperanças nos dois lados desta equação, lembre que

$$(\mathbf{T}\hat{\alpha}_{t-1}^{\text{bmk}} - \alpha_t) = \mathbf{u}_{t|t-1}, \quad \mathbb{E}(\mathbf{u}_{t|t-1}\mathbf{u}_{t|t-1}') = \mathbf{P}_{t|t-1}^{\text{bmk}} \quad \text{e} \quad \mathbb{E}(\mathbf{u}_{t|t-1}\tilde{\mathbf{e}}_t') = \mathbf{C}_t^{\text{bmk}}.$$

Com isto,

$$\mathbf{P}_t^{\text{bmk}} = \mathbf{M}_t^{\text{bmk}}\mathbf{P}_{t|t-1}^{\text{bmk}}\mathbf{M}_t'^{\text{bmk}} + \mathbf{M}_t^{\text{bmk}}\mathbf{C}_t^{\text{bmk}}\mathbf{K}_t'^{\text{bmk}} + \mathbf{K}_t^{\text{bmk}}\mathbf{C}_t'^{\text{bmk}}\mathbf{M}_t'^{\text{bmk}} + \mathbf{K}_t^{\text{bmk}}\Sigma_{tt}\mathbf{K}_t'^{\text{bmk}}. \quad (5.7)$$

O cálculo de $\mathbf{C}_t^{\text{bmk}}$ é feito de forma análoga ao da Subseção 3.2.1. Neste caso,

$$\begin{aligned} \mathbf{C}_t^{\text{bmk}} = \mathbf{A}_{t-1}^{\text{bmk}}\mathbf{A}_{t-2}^{\text{bmk}} \dots \tilde{\mathbf{A}}_1^{\text{bmk}}\Sigma_{1t} + \mathbf{A}_{t-1}^{\text{bmk}}\mathbf{A}_{t-2}^{\text{bmk}} \dots \tilde{\mathbf{A}}_2^{\text{bmk}}\Sigma_{2t} + \dots \\ + \mathbf{A}_{t-1}^{\text{bmk}}\tilde{\mathbf{A}}_{t-2}^{\text{bmk}}\Sigma_{(t-2)t} + \tilde{\mathbf{A}}_{t-1}^{\text{bmk}}\Sigma_{(t-1)t}, \end{aligned}$$

com $\mathbf{A}_j^{\text{bmk}}$ e $\tilde{\mathbf{A}}_j^{\text{bmk}}$ obtidas das equações (3.15) e (3.16), i.e. para cada t ,

$$\mathbf{A}_t^{\text{bmk}} = \mathbf{T}\mathbf{M}_t^{\text{bmk}} \quad \text{e} \quad \tilde{\mathbf{A}}_t^{\text{bmk}} = \mathbf{T}\mathbf{K}_t^{\text{bmk}}.$$

5.4 Aplicação

Nesta seção, será apresentada uma aplicação deste método de estimação sobre a série de taxa de desemprego obtida da Pesquisa Mensal de Emprego - PME feita pelo Instituto

Aplicação

Brasileiro de Geografia e Estatística - IBGE em seis regiões metropolitanas do Brasil. A seguir, são apresentados alguns aspectos desta pesquisa relevantes para a aplicação, segundo indicados no documento metodológico, IBGE(2007). Depois, na Subseção 5.4.2, é apresentado o ajuste das séries empregando o filtro de GLS com restrição de benchmark.

5.4.1 A pesquisa mensal de emprego - PME

A PME é uma pesquisa domiciliar, de periodicidade mensal, que vem sendo desenvolvida pelo IBGE desde 1980 com a finalidade de produzir indicadores sobre a força de trabalho nas Regiões Metropolitanas - RM de seis cidades brasileiras: Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo e Porto Alegre. São objeto de estudo todas as pessoas residentes em domicílios particulares² na área urbana dessas RM.

Nesta pesquisa, todas as pessoas de dez ou mais anos de idade são consideradas como *em idade de trabalhar*. Com relação a uma semana de referência, que é a semana (de domingo a sábado) anterior à de entrevista, as pessoas em idade de trabalhar são classificadas como:

- *Ocupadas*: se trabalharam, remuneradamente ou não, por pelo menos uma hora nessa semana ou, se não tendo trabalhado, tiveram um trabalho remunerado do qual estiveram afastadas por férias, greve ou licença remunerada;
- *Desocupadas*: se não trabalharam nessa semana, mas estiveram disponíveis para fazê-lo e tomaram alguma providência efetiva³ para conseguir um trabalho nos 30 dias anteriores ao início da semana de referência;
- *Inativas*: se não satisfazem nenhum dos critérios anteriores. São incluídos nesta categoria estudantes de tempo completo, pessoas aposentadas, donas de casa, entre outros.

²São excluídas da pesquisa as pessoas que moram em domicílios coletivos institucionais, tais como penitenciárias, hospitais ou instalações militares.

³São consideradas providências efetivas: consultar empregadores, colocar ou responder anúncios, inscrever-se ou fazer concurso, consultar parentes ou amigos, entre outras.

Sob estas definições, a taxa de desemprego é construída como o quociente entre o total de pessoas desocupadas e o total de pessoas economicamente ativas, isto é, ocupadas ou desocupadas, na semana de referência.

Vale a pena notar que a pesquisa é desenvolvida desde 1980 e no ano 2001 foi levada a cabo uma atualização das definições e instrumentos, assim como do plano amostral empregado. Os resultados a serem analisados nesta aplicação correspondem a esta versão revisada da pesquisa, que é a oficial desde março de 2002.

Plano amostral

Como descrito em IBGE(2007), cada uma das RM consideradas na PME está composta, além da cidade principal da qual toma o nome, por um conjunto de municípios próximos geograficamente e relacionados com ela em termos de mercado de trabalho. Alguns destes têm uma baixa população segundo o Censo de 2000, pelo que são juntados a outros para constituir os denominados *pseudo-municípios* e serem tratados como um só município no plano amostral da pesquisa. Com a finalidade de garantir espalhamento da amostra, cada um destes municípios ou pseudo-municípios constitui um estrato.

Assim, em cada estrato foram selecionados setores censitários empregando um desenho sistemático, com probabilidade proporcional à quantidade de domicílios existentes segundo informação do Censo Demográfico 2000. Posteriormente, em cada setor selecionado, foi feita uma operação de listagem, com a finalidade de construir um cadastro de domicílios do qual foi selecionada a amostra a ser entrevistada, empregando um desenho também sistemático. Todas as pessoas residentes nos domicílios selecionados são objeto de entrevista. Desta forma, o plano amostral da PME é probabilístico, estratificado, de conglomerados e em dois estágios de seleção. Um resumo dos tamanhos de amostra utilizados, extraído do documento metodológico da pesquisa, é apresentado na Tabela 5.1.

A amostra desta pesquisa segue um esquema de rotação 4-8(2) como o apresentado na Tabela 1.2. Neste caso, os grupos de rotação estão compostos por grupos de setores censitários e os painéis são conjuntos de domicílios⁴.

⁴Note que por não ser um painel de famílias ou indivíduos, quando a família que mora num domicílio

Aplicação

Tabela 5.1: Tamanho e composição da amostra, segundo Regiões Metropolitanas - 2001.

RM	Municípios	Setores censitários		Domicílios
		Universo	Amostra	
Recife	14	3.068	261	4.715
Salvador	10	4.604	243	4.684
Belo Horizonte	33	14.710	359	6.644
Rio de Janeiro	19	20.612	406	7.576
São Paulo	39	3.023	431	7.820
Porto Alegre	30	4.982	329	5.773

Fonte: IBGE. Documento metodológico da PME.

Estimadores empregados

Em cada repetição da pesquisa e para cada RM, os estimadores *naturais* do desenho de amostra têm a forma

$$\hat{y} = \sum_{h=1}^H \frac{1}{n_h} \sum_{i=1}^{n_h} \frac{1}{p_{hi}} \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hi}} y_{hij},$$

onde o índice h denota os estratos, o índice i os setores censitários ou unidades primárias de amostragem num estrato dado e o índice j os domicílios dentro do setor censitário. Analogamente, H é o número total de estratos em uma RM dada, n_h é o número de setores censitários selecionados no h -ésimo estrato e M_{hi} e m_{hi} são, respectivamente, as quantidades de domicílios no cadastro e a quantidade de domicílios selecionados, no i -ésimo setor do h -ésimo estrato. Finalmente, p_{hi} é o tamanho relativo do i -ésimo setor no h -ésimo estrato, segundo a quantidade de domicílios ocupados referida pelo Censo Demográfico 2000, e y_{hij} é o total da variável de interesse no domicílio correspondente. Assim, para a estimação do total de desempregados numa dada repetição da pesquisa, y_{hij} representa a quantidade de pessoas desocupadas no j -ésimo domicílio do i -ésimo setor censitário no h -ésimo estrato.

Em cada período, estes estimadores recebem um ajuste por não resposta, onde os tamanhos n_h , M_{hi} e m_{hi} são substituídos conforme o número de entrevistas efetivas realizadas. Este estimador ajustado será denotado por \hat{y}^* . Finalmente, o estimador empregado nas publicações oficiais desta pesquisa é um estimador *de razão*, ajustado

selecionado muda de lugar de residência, a informação para as subseqüentes repetições da pesquisa é levantada com a nova família que chegue para morar nele, se a houver.

às projeções de população:

$$\hat{y}^r = \frac{\hat{y}^* P}{\hat{P}^*},$$

onde P é a população projetada para a RM no mês da pesquisa, e \hat{P}^* é a sua estimativa obtida a partir da amostra. Note que como o fator de ajuste P/\hat{P}^* é o mesmo para todas as variáveis da pesquisa, as estimativas obtidas pelos estimadores \hat{y}^* e \hat{y}^r coincidem quando o parâmetro de interesse é do tipo taxa. Em particular para a taxa de desemprego.

Para guardar consistência com a notação empregada ao longo de todo o documento, empregaremos nesta seção o símbolo y_{dt} para denotar a estimativa da taxa de desemprego obtida a partir do estimador \hat{y}^r , para o t -ésimo mês na d -ésima RM.

Precisão das estimativas

Além da estimação dos parâmetros de interesse, rotineiramente são calculadas estimativas da precisão dos estimadores, as quais acompanham as publicações de resultados desta pesquisa na página web do IBGE. Segundo o documento metodológico, estas estimativas são obtidas pelo método *Ultimate Cluster* (Hansen et al., 1953,1993, Seção 6.7). Para exemplificar o nível de precisão observado na prática, são apresentadas na Tabela 5.2 as médias dos coeficientes de variação publicados nas RM para alguns valores específicos da taxa de desemprego durante o período em análise.

Tabela 5.2: Médias dos coeficientes de variação publicados para a taxa de desemprego, segundo RM. PME, março de 2002 - fevereiro 2012.

Intervalo da taxa	Recife	Salvador	Belo Horizonte	Rio de Janeiro	São Paulo	Porto Alegre
4.5 - 5.5	-	-	5.06	5.97	4.8	5.67
7.5 - 8.5	5.6	6	4.3	4.94	3.97	4.8
9.5 - 10.5	5.33	5.56	4.12	4.21	3.75	4.48
11.5 - 12.5	4.57	5.08	3.8	-	3.63	-
14.5 - 15.5	4.02	4.73	-	-	3.02	-

Fonte: Dados publicados pelo IBGE. Cálculos próprios.
O símbolo “-” indica que não se dispõe de dados nesse intervalo.

5.4.2 Ajuste simultâneo das séries de taxa de desemprego por RM

A seguir, serão apresentados os resultados do ajuste simultâneo das séries de taxa de desemprego das seis RM no período de março de 2002 a fevereiro de 2012. Foi empregado um BSM, como o apresentado na equação (1.5), acrescido de uma componente irregular denotada por I_t , o qual foi estimado usando o filtro GLS com uma restrição de benchmark visando garantir coerência entre as estimativas de planejamento e de modelo, para a agregação das seis RM.

Inicialmente, tratando cada RM de forma independente, serão estimadas as autocorrelações das respectivas séries de erros amostrais pelo método dos pseudo-erros, depois, será ajustado um modelo AR a cada uma delas para estimar as variâncias dos erros de evolução do BSM, por máxima verossimilhança empregando a formulação do filtro de Kalman com erros ARMA, apresentada na Seção 3.1. Finalmente, serão apresentados a restrição de benchmark utilizada e os resultados do filtro, assim como alguns elementos de diagnóstico deste ajuste. Tabelas e gráficos adicionais serão apresentados no Apêndice E.

Estimação das autocorrelações dos erros de observação

Considerando que o parâmetro de interesse é uma taxa, é de se esperar que sua variância seja função da magnitude do parâmetro. Por isto, antes de aplicar o método dos pseudo-erros para estimar as autocorrelações de série de erros amostrais, estes foram padronizados empregando as variâncias estimadas obtidas dos coeficientes de variação publicados pelo IBGE. As autocorrelações estimadas para as séries padronizadas, assim como as autocorrelações parciais respectivas, são apresentadas nas Tabelas E.1 e E.2, no Apêndice E. Posteriormente, em cada RM, foram ajustados modelos AR(12) às estimativas de autocorrelação obtidas⁵ via minimização do quadrado das diferenças entre as correlações estimadas e as esperadas sob o modelo⁶. Os parâmetros dos modelos

⁵Mesmo que em algumas RM um modelo AR(2) pareça suficiente para explicar a estrutura de autocorrelação estimada, como nos casos de Salvador ou Belo Horizonte, em todas as RM foram ajustados modelos AR(12) com o objetivo de incluir apropriadamente as autocorrelações esperadas, e de fato presentes em algumas das séries, para os lag 9 e subsequentes.

⁶Para a minimização foram consideradas as correlações desde os lag 1 até 15 e a média dos lag 16 a 24.

ajustados são apresentados na Tabela 5.3, as autocorrelações estimadas e as esperadas sob os modelos ajustados são apresentadas na Figura 5.1.

Tabela 5.3: Ajuste de modelos AR(12) às séries de erros de observação. Aplicação PME.

RM	Parâmetros do AR											
	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8	ϕ_9	ϕ_{10}	ϕ_{11}	ϕ_{12}
Recife	0.2401	0.1238	0.0827	-0.0291	0.0070	0.0478	-0.0242	0.0010	0.0045	0.0780	0.0213	0.1158
Salvador	0.5142	0.1220	0.0477	-0.1561	0.0554	0.0888	-0.0072	-0.0651	0.0316	0.0501	-0.0131	0.0073
Belo Horizonte	0.2393	0.1383	0.0540	-0.0130	0.0420	0.0481	-0.0140	-0.0008	0.0031	0.0174	0.0584	-0.0211
Rio de Janeiro	0.4740	0.1487	-0.0386	-0.0797	0.0517	0.0061	0.0434	-0.0407	0.0529	0.0458	-0.0078	0.0870
São Paulo	0.3829	0.0337	0.1440	-0.1992	0.1519	-0.0973	0.1160	-0.1125	0.1233	-0.0529	0.1045	0.0177
Porto Alegre	0.3149	0.0869	0.0053	0.0850	0.0616	-0.0089	0.0435	-0.0162	0.0409	-0.0432	0.0207	0.0558

Estimação das variâncias dos erros de evolução

Fazendo uso dos modelos apresentados na Tabela 5.3 e das variâncias dos erros amostrais estimadas para cada período a partir dos coeficientes de variação publicados, foi empregado o filtro de Kalman com erros AR(12) para estimar as variâncias dos erros de evolução via máxima verossimilhança. As estimativas obtidas, e seus desvios padrão obtidos a partir da inversa da matriz hessiana, são apresentadas na Tabela 5.4. Note que, empregando uma aproximação normal, só resultam significativamente distintas de zero as variâncias estimadas para a componente de tendência, ao 99% de significância para as áreas de Recife, Belo Horizonte, São Paulo e Porto Alegre, e ao 90% para as de Salvador e Rio de Janeiro. Em cada RM, o filtro foi inicializado de forma difusa.

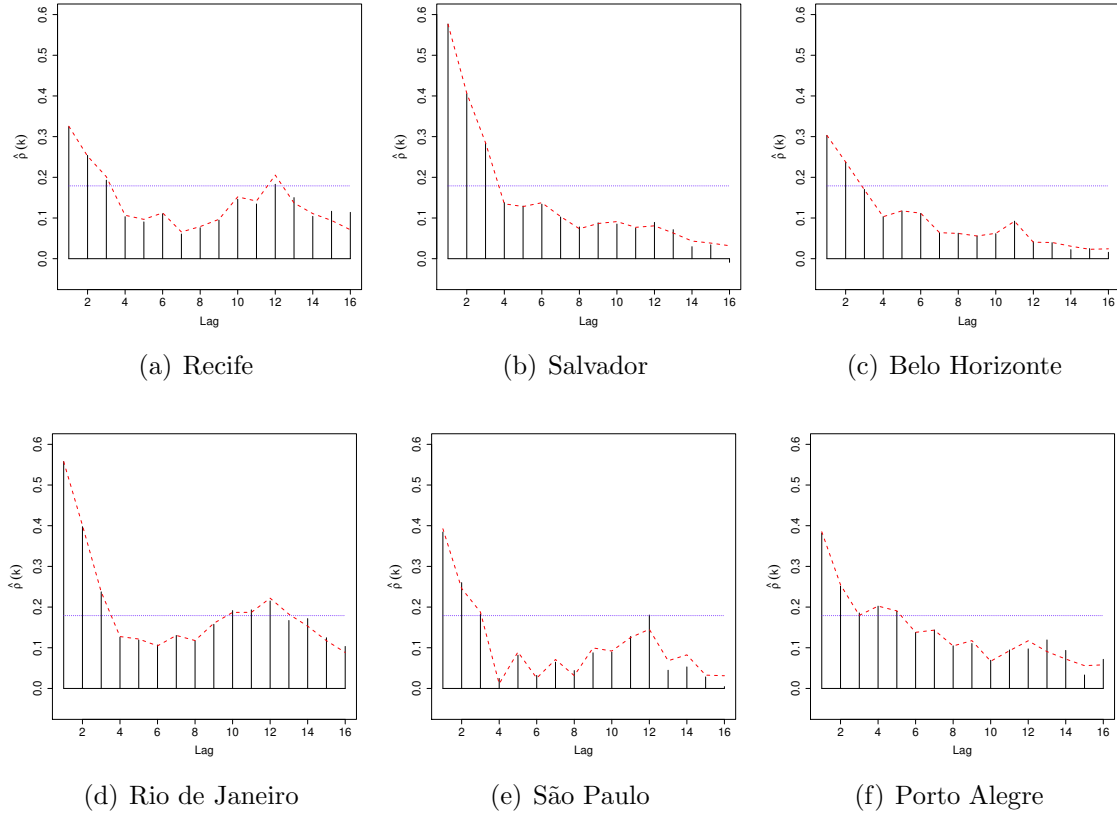
Ajuste simultâneo das séries com restrição de benchmark

Com a finalidade de garantir que a taxa de desemprego calculada para o agregado das seis RM coincida com a média ponderada das estimativas obtidas pelo modelo, foi utilizada uma restrição baseada nos pesos w_{2d} da equação (5.2):

$$w_{2dt} = \frac{N_{dt}}{\sum_{d=1}^D N_{dt}},$$

Aplicação

Figura 5.1: Autocorrelações estimadas para os erros amostrais por RM. Aplicação PME. Linhas: verticais: autocorrelação estimada; vermelha tracejada: autocorrelação ajustada sob o modelo AR(12); azul pontilhada: limite superior de um intervalo com 95% de confiança para a autocorrelação estimada.



com N_{dt} a quantidade de pessoas economicamente ativas estimada pela pesquisa na d -ésima RM no período t , para $d = 1, 2, \dots, 6$. O modelo filtrado foi heteroscedástico, empregando as variâncias obtidas dos coeficientes de variação oficiais publicados e a estrutura de autocorrelação estimada na Subseção 5.4.2.

Resultados

As séries ajustadas são apresentadas na Figura 5.2 e as componentes de nível local L_t e sazonalidade S_t nas Figuras 5.3 e 5.4. As outras componentes do BSM, a inclinação local

Tabela 5.4: Variâncias estimadas para os erros de evolução. Aplicação PME.

Região Metropolitana	σ_L^2	σ_R^2	σ_S^2	σ_I^2
Recife	3.1033×10^{-01} (7.62×10^{-02})	5.6070×10^{-05} (3.97×10^{-04})	5.5835×10^{-07} (4.38×10^{-05})	1.2650×10^{-05} (1.04×10^{-03})
Salvador	7.4107×10^{-02} (4.01×10^{-02})	6.3160×10^{-10} (2.74×10^{-08})	1.0089×10^{-09} (7.99×10^{-07})	1.0741×10^{-08} (3.27×10^{-06})
Belo Horizonte	6.6265×10^{-02} (2.11×10^{-02})	3.9706×10^{-08} (1.81×10^{-06})	2.7786×10^{-10} (4.22×10^{-08})	5.5600×10^{-06} (1.03×10^{-03})
Rio de Janeiro	3.5315×10^{-02} (2.10×10^{-02})	1.3890×10^{-11} (2.15×10^{-09})	1.4427×10^{-08} (1.16×10^{-06})	2.8478×10^{-02} (1.98×10^{-02})
São Paulo	9.3685×10^{-02} (2.29×10^{-02})	9.4535×10^{-09} (1.11×10^{-06})	1.3890×10^{-11} (1.57×10^{-08})	5.6162×10^{-07} (4.09×10^{-05})
Porto Alegre	5.1300×10^{-02} (1.72×10^{-02})	2.6651×10^{-08} (1.12×10^{-06})	1.2174×10^{-10} (2.40×10^{-09})	1.4247×10^{-06} (5.11×10^{-05})

Desvios padrão entre parênteses.

R_t e a componente irregular I_t , são apresentadas nas Figuras E.2 e E.3 no Apêndice E. Os resultados dos estimadores com e sem benchmark são muito similares para a maioria das componentes e períodos, ainda que um pouco diferentes em sua precisão.

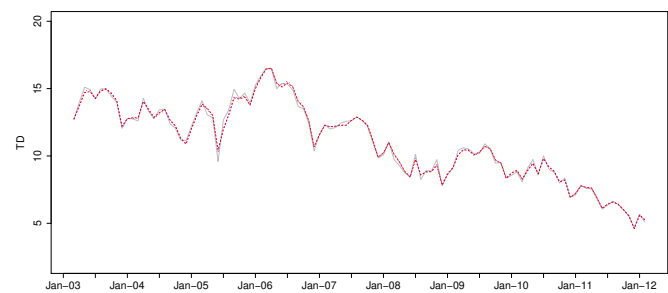
Em geral, pode-se concluir que as séries ajustadas não diferem grandemente das séries originais, exceto por alguns *picos*, o que é desejável pois é um indicio de coerência dos modelos empregados. A tendência estimada, ou nível local, é um pouco menos suave do que a encontrada anteriormente para as RM de Rio de Janeiro e São Paulo no período março 2002 - setembro 2005 em Cruz et al. (2006). Por outro lado, a sazonalidade nessas duas RM é coerente com a encontrada nesse trabalho, se bem no caso de Rio de Janeiro o seu efeito em pontos percentuais pareça menor no nosso ajuste. Estas diferenças nos dois ajustes podem ser devidas, tanto ao menor período de tempo considerado nesse trabalho, quanto às diferenças nas estimativas obtidas para a estrutura de autocorrelação dos erros de observação e para as variâncias dos erros de evolução.

É importante notar que as estimativas obtidas no modelo sem benchmark, satisfazem esta restrição razoavelmente: a maior diferença entre a taxa de desemprego do total das seis RM calculada a partir do desenho e a correspondente média ponderada das estimativas sem benchmark não supera os 0.3 pontos percentuais; já no modelo ajustado com a restrição de benchmark, a maior diferença entre estas estimativas é da ordem de 7×10^{-10} pontos percentuais.

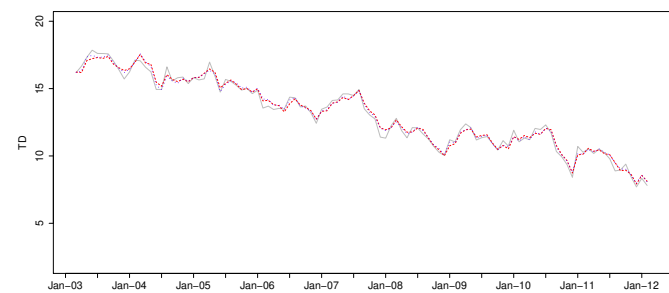
Aplicação

Com relação à precisão dos estimadores note que, em todos os períodos e para todas as RM consideradas, as variâncias dos estimadores obtidos pelo modelo, com ou sem benchmark, foram menores do que as dos estimadores *naturais* de desenho, como pode ser verificado na Figura 5.6. No entanto, nesta aplicação, a inclusão da restrição não parece melhorar a precisão dos estimadores quando comparados com o modelo sem benchmark e, em algumas RM como São Paulo e Rio de Janeiro, os estimadores com benchmark resultam menos precisos do que aqueles sem benchmark. Uma possível explicação para isto deriva-se do fato que a finalidade do benchmark é garantir a coerência entre as estimativas de modelo e desenho, a qual já existe neste caso para estimativas sem benchmark; assim, a inclusão da restrição perderia sua finalidade e só acrescentaria a variabilidade do sistema. Em qualquer caso, resulta de interesse para seu futuro estudo, identificar com maior clareza os fatores associados com um possível ganho em precisão como resultado do processo de benchmark, assunto que não é abordado nos artigos originais onde o método é proposto.

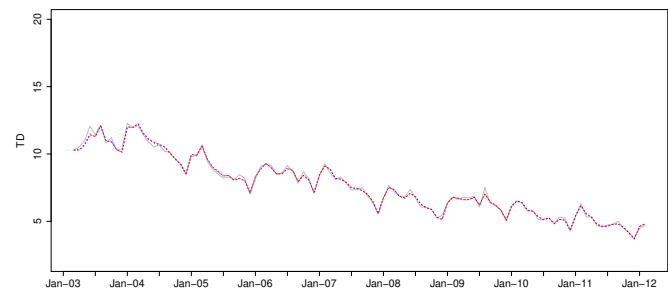
Figura 5.2: Séries ajustadas com e sem benchmark. Aplicação PME. Série filtrada ($\hat{L}_t + \hat{S}_t + \hat{I}_t$).
 Linhas: cinza contínua: série de estimativas de planejamento y_t ; vermelha tracejada: série ajustada sem benchmark; azul pontilhada: série ajustada com benchmark.



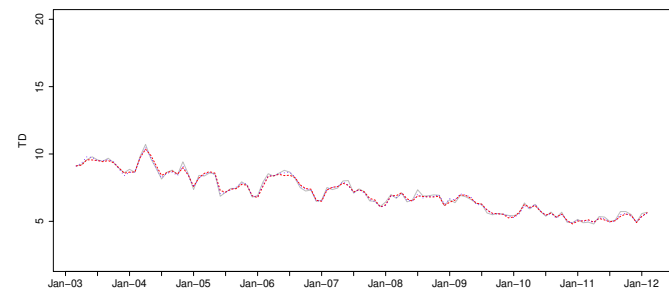
(a) Recife



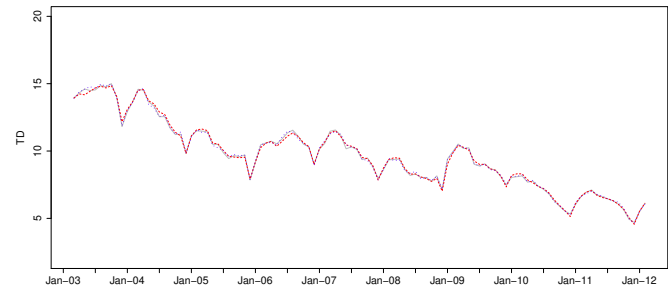
(b) Salvador



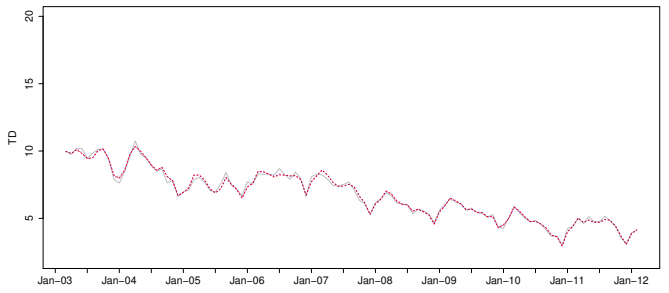
(c) Belo Horizonte



(d) Rio de Janeiro



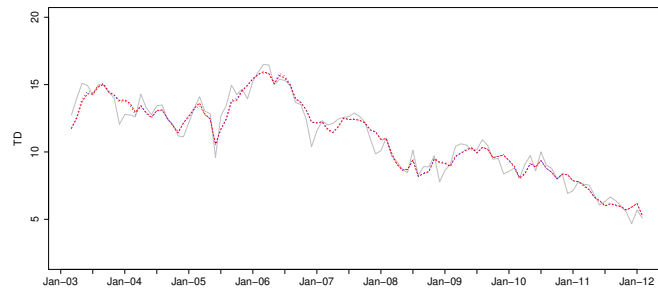
(e) São Paulo



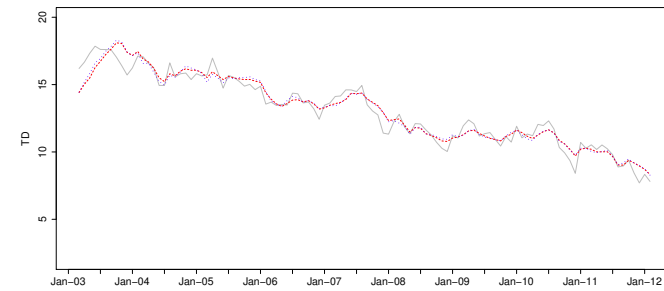
(f) Porto Alegre

Figura 5.3: Séries ajustadas com e sem benchmark. Aplicação PME. Nível Local-Tendência (\hat{L}_t).

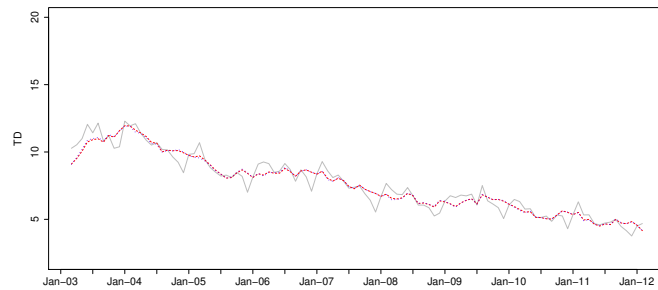
Linhas : cinza contínua: série de estimativas de planejamento y_t ; vermelha tracejada: nível local estimado sem benchmark; azul pontilhada: nível local estimado com benchmark.



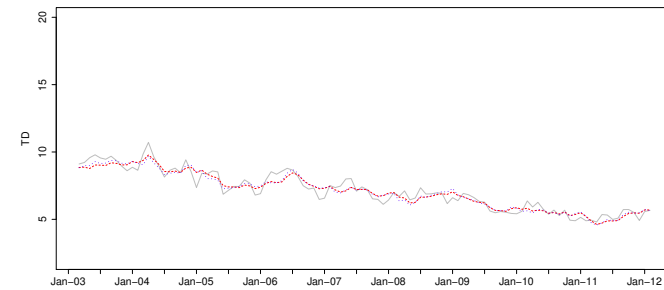
(a) Recife



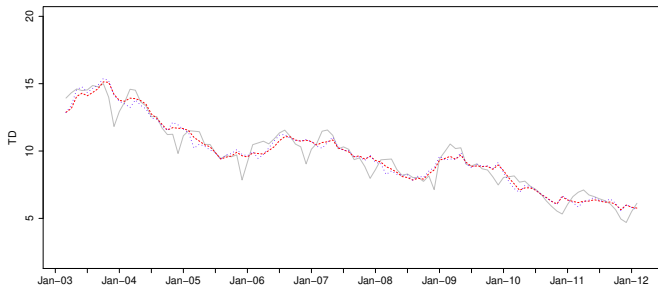
(b) Salvador



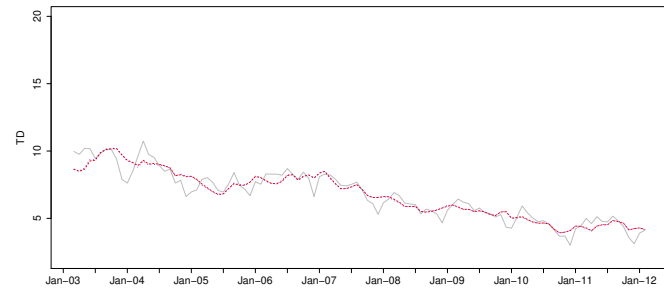
(c) Belo Horizonte



(d) Rio de Janeiro

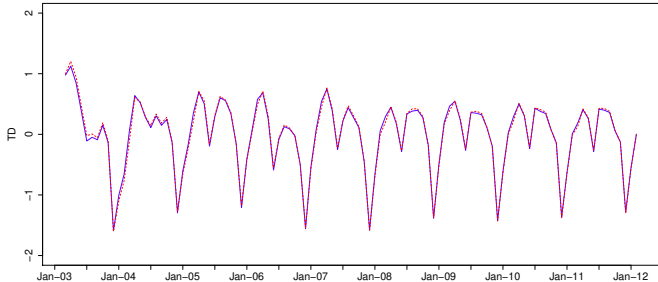


(e) São Paulo

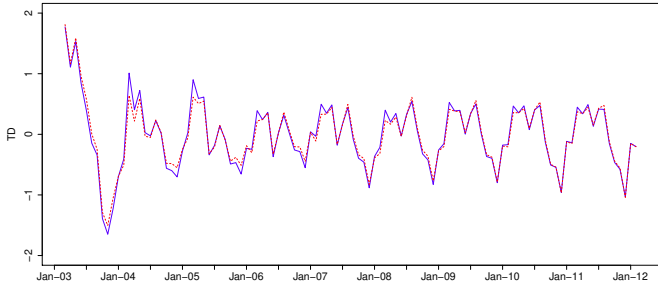


(f) Porto Alegre

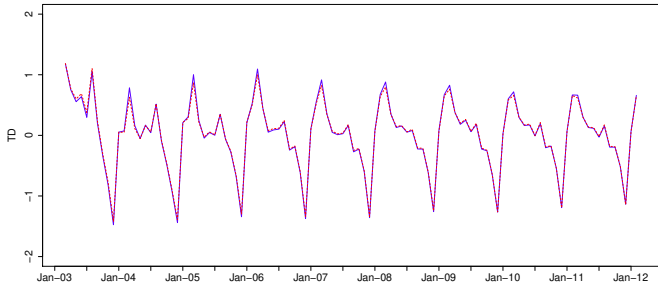
Figura 5.4: Séries ajustadas com e sem benchmark. Aplicação PME. Componente Sazonal (\hat{S}_t).
 Linhas: vermelha tracejada: série ajustada sem benchmark; azul contínua: série ajustada com benchmark.



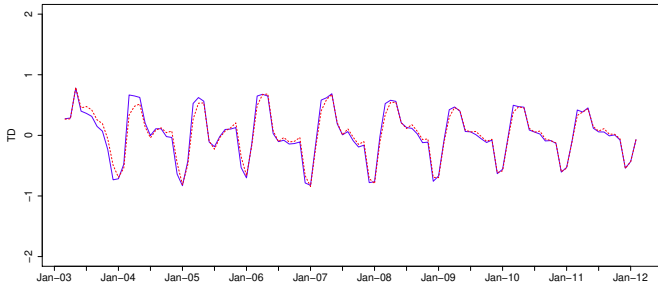
(a) Recife



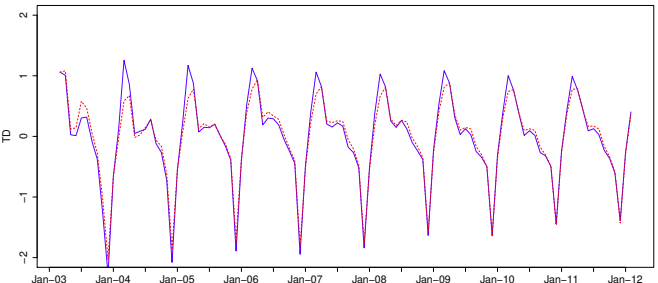
(b) Salvador



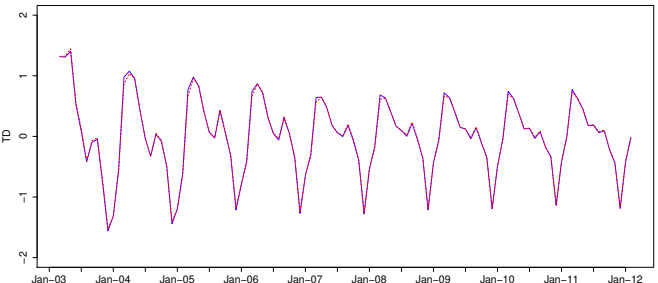
(c) Belo Horizonte



(d) Rio de Janeiro

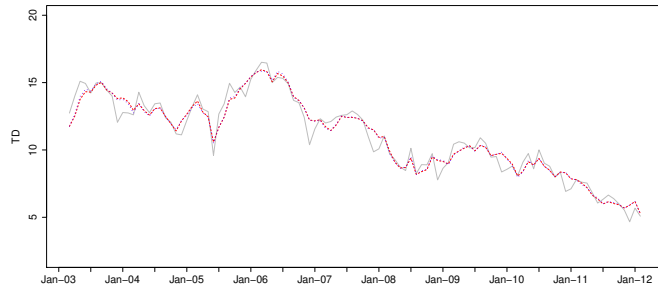


(e) São Paulo

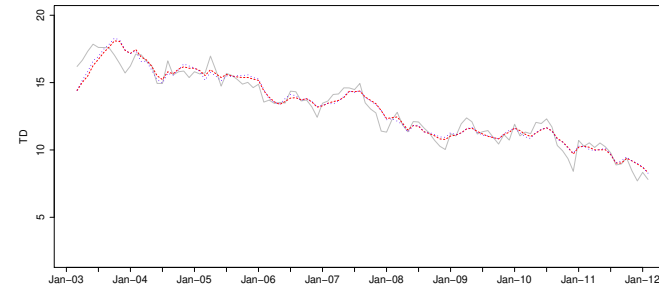


(f) Porto Alegre

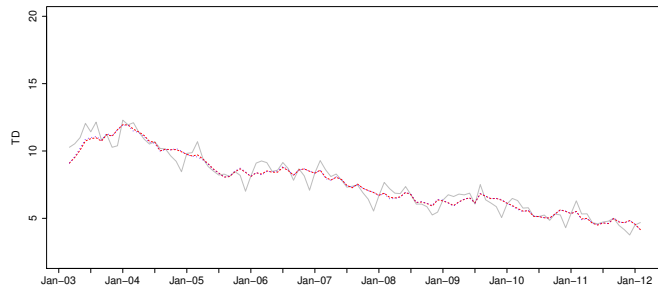
Figura 5.5: Séries ajustadas com e sem benchmark. Aplicação PME. Série Desazonalizada ($\hat{L}_t + \hat{I}_t$).
 Linhas : cinza contínua: série de estimativas de planejamento y_t ; vermelha tracejada: nível local estimado sem benchmark; azul pontilhada: nível local estimado com benchmark.



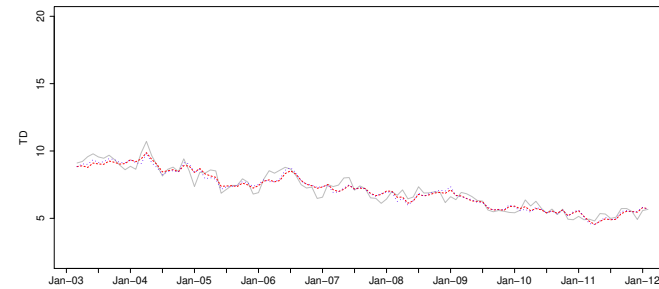
(a) Recife



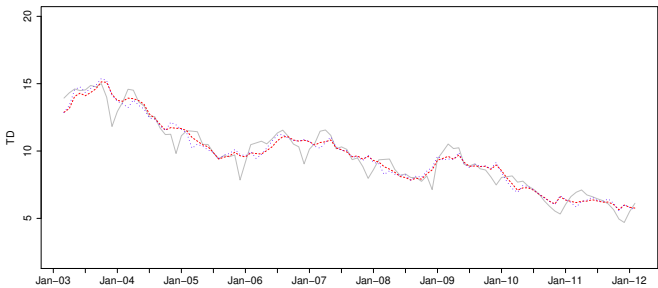
(b) Salvador



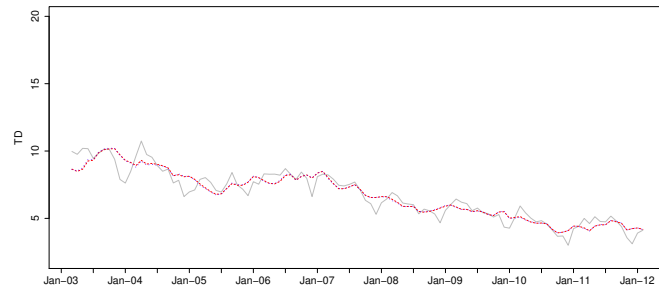
(c) Belo Horizonte



(d) Rio de Janeiro

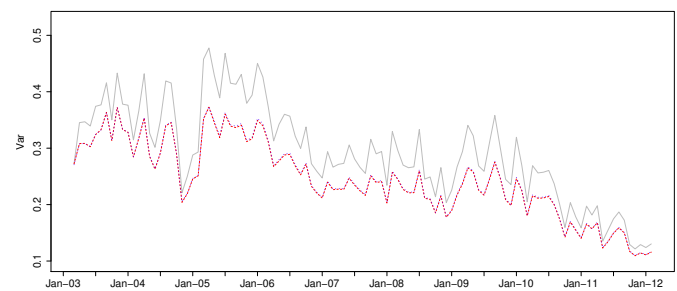


(e) São Paulo

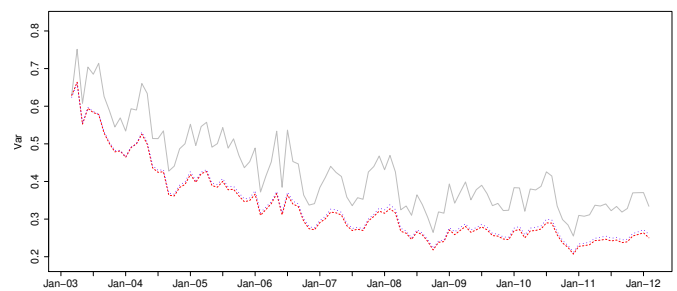


(f) Porto Alegre

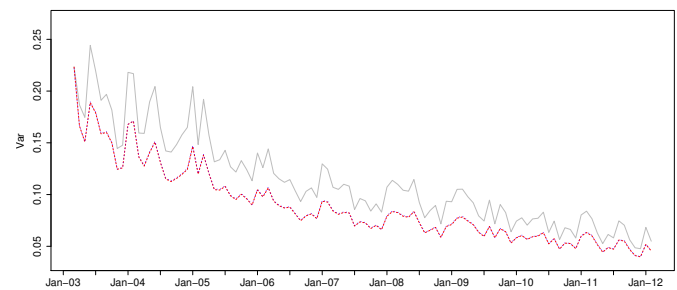
Figura 5.6: Séries ajustadas com e sem benchmark. Aplicação PME. Variâncias estimadas para a série filtrada ($\hat{V}(\hat{L}_t + \hat{S}_t + \hat{I}_t)$). Linhas: cinza contínua: variância das estimativas de planejamento y_t ; vermelha tracejada: variância das estimativas sem benchmark; azul pontilhada: variância das estimativas com benchmark.



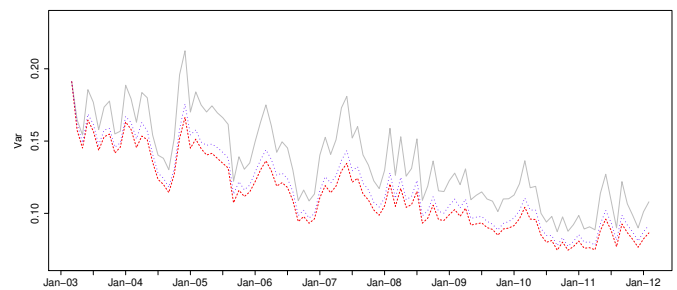
(a) Recife



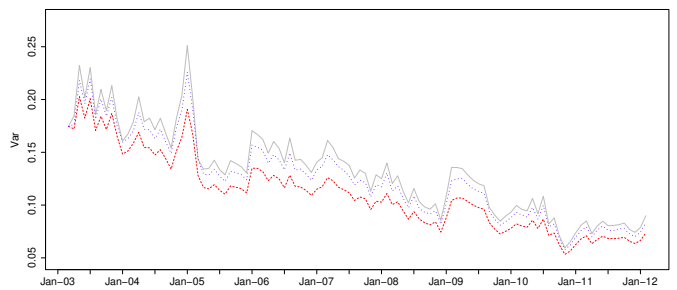
(b) Salvador



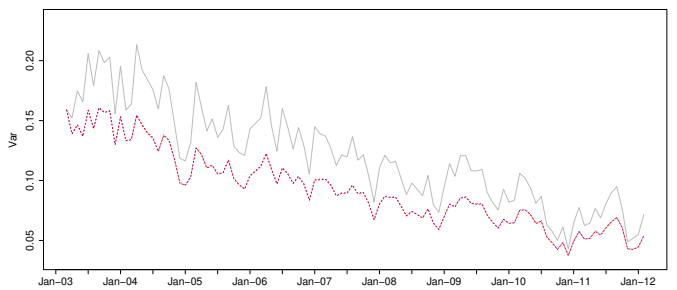
(c) Belo Horizonte



(d) Rio de Janeiro

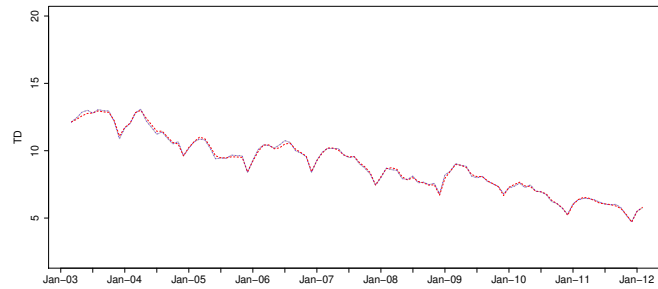


(e) São Paulo

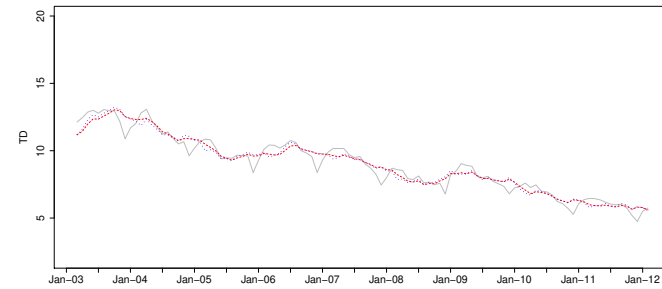


(f) Porto Alegre

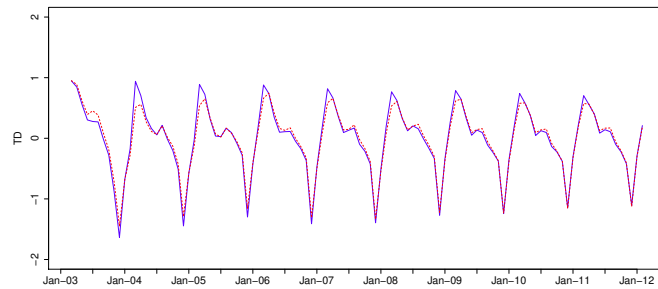
Figura 5.7: Séries ajustadas com e sem benchmark. Aplicação PME. Conjunto 6 Regiões Metropolitanas.
 Linhas: cinza contínua: estimativas de planejamento; vermelha tracejada: estimativas sem benchmark; azul pontilhada ou contínua: estimativas com benchmark.



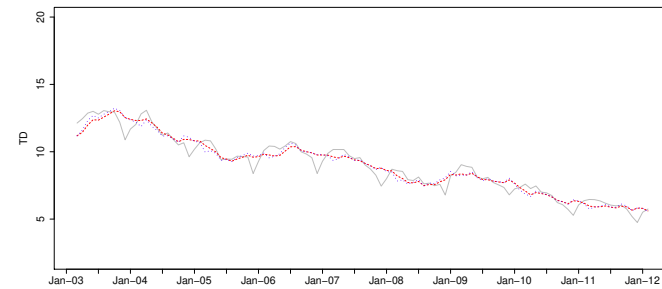
(a) Série Filtrada ($\hat{L}_t + \hat{S}_t + \hat{I}_t$)



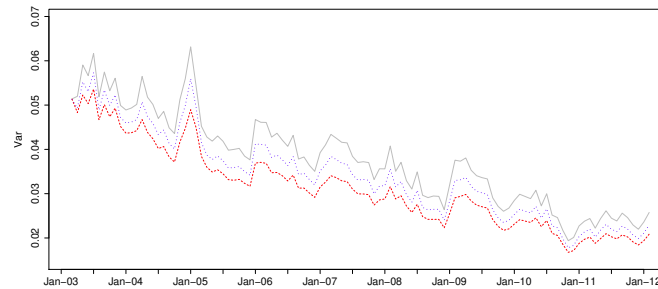
(b) Inclinação Local-Tendência (\hat{L}_t)



(c) Componente Sazonal (\hat{S}_t)



(d) Série Desazonalizada ($\hat{L}_t + \hat{I}_t$)



(e) Variâncias estimadas ($\hat{V}(\hat{L}_t + \hat{S}_t + \hat{I}_t)$)

Capítulo 6

Considerações finais

Neste documento, foi estudada a estimação em pesquisas repetidas utilizando MEE e os filtros GLS e Kalman, para o caso de erros de observação autocorrelacionados. Foi feita uma avaliação de seu desempenho, via simulação, contrastando os dois filtros em dois distintos cenários de autocorrelação. Nesta avaliação foi verificado que a metodologia de estimação utilizando estes tipos de modelos permite obter estimadores sem vício ou com um vício pequeno, e mais precisos do que os estimadores de desenho usualmente empregados no enfoque de inferência clássica, sempre que a autocorrelação presente nos erros amostrais, quando existente, seja considerada no processo de estimação. Comparando os dois filtros incluídos na análise foi observado um melhor desempenho do filtro de Kalman, como esperado, e a diferença em precisão entre eles foi menor no cenário 2, onde o complexo esquema de autocorrelação assumido, faz com que o ajuste dos erros amostrais seja mais difícil de aproximar por um modelo ARMA.

Também, via simulação, foram abordados aspectos prévios ao emprego dos filtros, tais como: i) a estimação da estrutura de autocorrelação dos erros de observação no contexto de pesquisas com painéis rotativos e, ii) a estimação, via máxima verossimilhança, das variâncias dos erros de evolução dos modelos. No primeiro caso, a proposta de Pfeiffermann et al. (1998) mostrou ser apropriada, com os resultados obtidos sugerindo a substituição do denominador T por $T - k$ na estimação da autocovariância com defasagem k , visando corrigir uma possível subestimação das autocovariâncias de lags maiores. Por outra parte, a estimação das variâncias dos erros de evolução não resultou

Trabalhos futuros

suficientemente precisa para nenhum dos dois cenários considerados, como pode ser conferido nas Tabelas D.10 e D.11 no Apêndice D, possivelmente porque estas variâncias costumam ser muito pequenas. Não entanto, como pode ser visto na Seção D.4 do Apêndice D, quando comparados os estimadores (3) e (4) ou os (5) e (7) no cenário 1, ou os (4) e (5), ou (6) e (8) no cenário 2, em geral isto não afeta fortemente a precisão dos estimadores.

Um aspecto interessante do filtro GLS é a possibilidade de incluir uma restrição de benchmark como parte do processo de estimação. Isto foi ilustrado utilizando informações provenientes da Pesquisa Mensal de Emprego do IBGE para seis Regiões Metropolitanas-RM. A nossa aplicação conduz a estimadores mais precisos para as distintas RM e além, faz com que as estimativas guardem coerência com as estimativas de desenho ao nível da agregação das seis RM.

6.1 Trabalhos futuros

Um dos aspectos que resultou problemático na comparação dos filtros GLS e Kalman apresentada no Capítulo 4 foi a estimação das variâncias dos erros de evolução. Como mencionado acima, a metodologia de máxima verossimilhança não resultou satisfatória em nenhum dos cenários considerados. Surge assim o interesse em pesquisar formas mais precisas de estimar estes parâmetros, possivelmente, sob um enfoque Bayesiano.

Também, considerando os resultados obtidos no Capítulo 5, é de interesse um estudo mais detalhado do processo de estimação sob restrições de benchmark que permita identificar os fatores associados ao ganho em precisão dos estimadores, com relação aos estimadores sem benchmark.

Finalmente, com relação à aplicação desta metodologia nas séries de emprego da PME, um encadeamento das séries de taxa de desocupação das pesquisas nova e antiga, i.e., antes e depois de março de 2002, para as seis RM, proporcionaria informações para a inicialização do filtro evitando a perda de observações ao início da série, consequência de uma inicialização difusa. Este trabalho poderia complementar o feito em Cruz et al. (2006) para as RM de Rio de Janeiro e São Paulo.

Referências Bibliográficas

- D. A. Binder e M. A. Hidirolou (1988). Sampling in Time. Em *Handbook of Statistics. 6. Sampling*, editado por P. R. Krishnaiah e C. R. Rao, páginas 187–211. North Holland, Amsterdam.
- B.J.N. Blight e A.J. Scott (1973). A stochastic model for repeated surveys. *Journal of the Royal Statistical Society, Series B*, 35:61–66.
- G.E.P. Box, G.M. Jenkins e G.C. Reinsel (1994). *Time Series Analysis: Forecasting and Control*. Prentice-Hall, Upper Saddle River, NJ, terceira edição.
- J. Chow, P. Zong, D.B.N. Silva, P. Smith e G. Brown (2010). Modelling the UK labour force survey using a structural time series model. Em *Proceedings of the 2010 Joint Statistical Meetings*, páginas 5620–5631, Vancouver. American Statistical Association.
- W.G. Cochran (1977). *Sampling Techniques*. John Wiley & Sons, New York, London.
- M.M. Cruz, M. Pitta e D.B.N. Silva (2006). Uso de modelos de espaço de estados para estimação da taxa de desocupação na PME-IBGE. Em *Memórias do 17o. SINAPE*, Caxambu-MG. ABE - Associação Brasileira de Estatística.
- G.S. Datta (2009). Model-based approach to small area estimation. Em *Handbook of Statistics 29B. Sample Surveys: Inference and Analysis*, editado por D. Pfefferman e C.R. Rao, capítulo 32, páginas 251–288. North Holland, Amsterdam.
- G.J. Duncan e G. Kalton (1987). Issues of design and analysis of surveys across time. *International Statistical Review*, 55(1):97–117.
- J. Durbin e S.J. Koopman (2001). *Time Series Analysis by State Space Methods*. Oxford University Press, Oxford, New York.

Referências Bibliográficas

- J. Durbin e B. Quenneville (1997). Benchmarking by state space models. *International Statistical Review*, 65(1):23–48.
- A.R. Eckler (1955). Rotation sampling. *Annals of Mathematical Statistics*, 26:664–685.
- J. G. Gambino e P.L. do Nascimento Silva (2009). Sampling and estimation in household surveys. Em *Handbook of Statistics 29A. Sample Surveys: Design, Methods and Applications*, editado por D. Pfefferman e C.R. Rao, capítulo 16, páginas 407–439. North Holland, Amsterdam.
- M.H. Hansen, W.N. Hurwitz e W.G. Madow (1953,1993). *Sample Survey Methods and Theory*. John Wiley & Sons, New York, London.
- A. Harvey (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge.
- IBGE. Diretoria de Pesquisas. Coordenação de Trabalho e Rendimento (2007). *Pesquisa Mensal de Emprego. Série Relatórios Metodológicos. Vol. 23*. Rio de Janeiro-RJ, segunda edição.
- R.G. Jones (1980). Best linear unbiased estimators for repeated surveys. *Journal of the Royal Statistical Society, Serie B*, 42:221–226.
- R. E. Kalman (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME - Journal of Basic Engineering*, 82:35–45.
- G. Kalton (2009). Designs for surveys over time. Em *Handbook of Statistics 29A. Sample Surveys: Design, Methods and Applications*, editado por D. Pfefferman e C.R. Rao, capítulo 5, páginas 89–108. North Holland, Amsterdam.
- L. Kish (1965). *Survey Sampling*. John Wiley & Sons, New York, London.
- P.A. Morettin e C.M.C. Toloi (2006). *Análise de Séries Temporais*. Edgard Blucher, São Paulo.
- Y.S. Park, K.W. Kim e J.W. Choi (2001). One-level rotation design balanced on time in monthly sample and in rotation group. *Journal of the American Statistical Association*, 96:1483–1496.
- H. D. Patterson (1950). Sampling on successive occasions with partial replacement of units. *Journal of the Royal Statistical Society. Series B.*, 12(2):241–255.

- G. Petris, S. Petrone e P. Campagnoli (2009). *Dynamic Linear Models with R*. Springer, New York.
- D. Pfeiffermann (2011). New important developments in small area estimation. Em *S3RI Methodology Working Papers, M10/11*. University of Southampton. Southampton Statistical Sciences Research Institute.
- D. Pfeiffermann (2002). Small area estimation: New developments and directions. *International Statistical Review*, 70(1):125–143.
- D. Pfeiffermann e R. Tiller (2006). Small area estimation with state space models subject to benchmark constraints. *Journal of the American Statistical Association*, 101(476): 1387–1397.
- D. Pfeiffermann, P. Bell e D. Signorelli (1996). Labor force trend estimation in small areas. Em *Proceedings of the Annual Research Conference and Technology Interchange*, páginas 407–431. Bureau de Censos dos Estados Unidos, Rosslyn-Virginia.
- D. Pfeiffermann, M. Feder e D. Signorelli (1998). Estimation of autocorrelation of survey errors with application to trend estimation in small areas. *Journal of Business and Economic Statistics*, 16(3):339–348.
- J.N.K. Rao e J.E. Graham (1964). Rotation designs for sampling on repeated occasions. *Journal of the American Statistical Association*, 50:492–509.
- C.E. Särndal, B. Swensson e J. Wretman (1992). *Model Assisted Survey Sampling*. Springer-Verlag, New York.
- A.J. Scott, T.M.F. Smith e R.G. Jones (1974). Analysis of repeated surveys using time series methods. *Journal of the American Statistical Association*, 69:674–678.
- G.A.F. Seber (2008). *A Matrix Handbook for Statisticians*. Ed. John Wiley & Sons, New York, London.
- D.B.N. Silva e M.M. Cruz (2002). Séries temporais de pesquisas amostrais periódicas.
- S. M. Tam (1987). Analysis of repeated surveys using a dynamic linear model. *International Statistical Review*, 55(1):63–73.
- R. Tiller (1992). Time series modelling of sample survey data from the U.S. current population survey. *Journal of Official Statistics*, 8:149–166.

Referências Bibliográficas

- J. Van den Brakel e S. Krieg (2009). Estimation of the monthly unemployment rate through structural time series modelling in a rotating panel design. *Survey Methodology*, 35(2):177–190.
- K. Wang, Y. Li e C. Rizos (2010). The practical approaches to kalman filtering with time-correlated measurement errors. Submitted to IEEE Transactions on Aerospace Electronic Systems. URL http://www.gmat.unsw.edu.au/snap/publications/wangk_etal2010a.pdf.
- W.W.S Wei (1990). *Time Series Analysis. Univariate and Multivariate Methods*. Addison Wesley, Reading, MA.
- M. West e J. Harrison (1997). *Bayesian Forecasting and Dynamic Models*. Springer-Verlag, New York.
- K. Wolter (1979). Composite estimation in finite populations. *Journal of the American Statistical Association*, 74(367):604–613.
- F. Yates (1949). *Sampling Methods for Censuses and Surveys*. Charles Griffin, London.

Apêndice A

Representação de um modelo ARMA em espaço de estados

Seja o modelo ARMA multivariado:

$$\Phi_p(B)\mathbf{y}_t = \Theta_q(B)\mathbf{a}_t, \quad (\text{A.1})$$

onde para cada t , \mathbf{y}_t e \mathbf{a}_t são vetores coluna com k observações cada um, e

$$\Phi_p(B) = \Phi_0 - \Phi_1 B - \Phi_2 B^2 - \dots - \Phi_p B^p$$

e

$$\Theta_q(B) = \Theta_0 - \Theta_1 B - \Theta_2 B^2 - \dots - \Theta_q B^q$$

são dois polinômios de matrizes de dimensão $k \times k$ em B , o operador *backshift*¹, com $\Phi_0 = \Theta_0 = \mathbf{I}_k$. Para garantir a identificabilidade do modelo, seguindo a (Wei 1990, págs. 337 e 347-348), suporemos que os polinômios resultantes de calcular os determinantes $|\Phi_p(B)|$ e $|\Theta_q(B)|$ são não constantes, e que as seguintes condições são satisfeitas:

1. Os únicos divisores comuns a esquerda de $\Phi_p(B)$ e $\Theta_q(B)$ são os unimodulares, i.e., se

$$\Phi_p(B) = \mathbf{C}(B)\mathbf{H}(B) \quad \text{e} \quad \Theta_q(B) = \mathbf{C}(B)\mathbf{K}(B),$$

¹i.e. para $s > 0$, $B^s(\mathbf{y}_t) = \mathbf{y}_{t-s}$

Representação de um modelo ARMA em espaço de estados

então o determinante de $C(B)$ é constante; e

2. Os zeros de $|\Phi_p(B)|$ estão fora do círculo unitário e os zeros de $|\Theta_q(B)|$ estão fora ou no círculo unitário.

Por simplicidade, suporemos também que $E(\mathbf{a}_t) = \mathbf{0}$ e $V(\mathbf{a}_t) = \Sigma_{\mathbf{a}}$, uma matriz diagonal. Uma possível representação de em espaço de estados do modelo A.1 sob estas restrições esta dada por:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{Z}\alpha_t \\ \alpha_t &= \mathbf{T}\alpha_{t-1} + \mathbf{G}\eta_t \quad E(\eta_t) = 0; \quad E(\eta_t\eta_t') = \mathbf{Q}_t, \end{aligned}$$

onde,

$$\mathbf{Z} = \begin{bmatrix} \mathbf{I}_k & | & \mathbf{0}_{(k \times (p+q-1))} \end{bmatrix},$$

$$\alpha'_t = \begin{bmatrix} y'_t & y'_{t-1} & \dots & y'_{t-p+1} & a'_t & a'_{t-1} & \dots & a'_{t-q+1} \end{bmatrix},$$

$$\eta_t = \mathbf{a}_t; \quad \mathbf{Q}_t = \Sigma_{\mathbf{a}},$$

$$\mathbf{T} = \begin{bmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_p & \Theta_1 & \Theta_2 & \dots & \Theta_q \\ \hline \mathbf{I}_{k(p-1)} & \mathbf{0}_{k(p-1) \times k} & \mathbf{0}_{k(p-1) \times k(q-1)} & \mathbf{0}_{k(p-1) \times k} \\ \hline \mathbf{0}_{k \times k(p-1)} & \mathbf{0}_{k \times k} & \mathbf{0}_{k \times k(q-1)} & \mathbf{0}_{k \times k} \\ \hline \mathbf{0}_{k(q-1) \times k(p-1)} & \mathbf{0}_{k(q-1) \times k} & \mathbf{I}_{k(q-1)} & \mathbf{0}_{k(q-1) \times k} \end{bmatrix} \quad e$$

$$\mathbf{G}' = \begin{bmatrix} \mathbf{I}_k & | & \mathbf{0}_{k \times k(p-1)} & | & \mathbf{I}_k & | & \mathbf{0}_{k \times k(q-1)} \end{bmatrix}.$$

As matrizes \mathbf{I}_k e $\mathbf{0}_{k \times l}$ representam a matriz identidade de ordem k e uma matriz de zeros de dimensão $k \times l$, respectivamente. O caso de um modelo ARMA univariado pode ser tratado desta mesma forma, com $k = 1$. Uma outra expressão para este tipo de modelo, equivalente com a apresentada, pode ser encontrada em (Petrís et al., 2009, pág. 140).

Apêndice B

Proposta para a estimação das $C^j(k)$

A estimação das covariâncias considerando a quantidade de vezes que um painel tem sido pesquisado, empregando a metodologia de Pfeffermann et al. (1998), pode ser feita da seguinte forma:

1. Disponha os pseudo erros numa matriz onde as linhas indiquem o período, $t = 1, 2, \dots, T$, e as colunas o número de vezes que o painel tem sido pesquisado, $j = 1, 2, \dots, g$,
2. Centre esta matriz, subtraindo de cada observação a média da coluna,
3. Para cada combinação $\{j, k\}$ e para t fixo, veja que $(\check{e}_t^{(j)} - \check{e}^{(j)})$ corresponde à t -ésima observação da j -ésima coluna, e $(\check{e}_{t-k}^{j,t} - \check{e}^{j,k})$ é a $t - k$ -ésima observação da digamos, m -ésima coluna desta nova matriz, onde m , deve ser estabelecido em função de k , j e o esquema de rotação empregado¹. Assim, identificado m , faça o produto vetorial das colunas j e m , sendo que esta última tem defasagem de k períodos, desprezando os outros termos, e divida por T .

A seguir, é apresentada a sintaxe desenvolvida no R para este fim, no caso de um esquema de rotação 4 – 8(2) balanceado a duas vias.

¹Como exemplo, veja que no esquema 4 – 8(2) segundo apresentado na tabela (1.2), para $k = 1$, $m = j - 1$ se $j \in \{2, 3, \dots, 8\}$, e $m = 8$ para $j = 1$. Neste esquema, em geral, m corresponde a $((j - k - 1) \bmod g) + 1$, onde \bmod indica a operação módulo definida como: $a \bmod b = a - (b \times \text{int}(a/b))$, com int a função parte inteira.

Proposta para a estimação das $C^j(k)$

```
fcovariancias <-function(matriz,nlag){
  T <- nrow(matriz)
  c <-array(0,dim=c(nlag+1,8))
  for (j in 1:8) { #pois um painel pode ser pesquisado até 8 vezes
    temp <- array(0,dim=c(T,nlag+1))
    for (k in 0:nlag){ #para cada lag
      base <- c(0,0)
      for (t in (k+1):T)
        base <- rbind(base,c(matriz[t-k,((j-k-1)%8)+1],matriz[t,j]))
      base<-base[-1,]
      for (m in 1:nrow(base))
        temp[m+k,k+1]<-base[m,1]*base[m,2]
      c[k+1,j] <-sum(temp[,k+1])/(T) #armazena as Cj(k) estimadas
    }
  }
  acf <- array(0,dim=c(nlag+1,1))
  for (i in 1:(nlag+1)) acf[i] <- sum(c[i,])/(8*7) #armazena os gamma(k) estimadas
  acf_j<-list()
  for (j in 1:8) acf_j[[j]] <- t(as.matrix(c[,j]))
  list(acf_j = acf_j, acf= t(as.matrix(acf)))
}
```

Apêndice C

Equações do filtro GLS

Previsão para α_t e y_t :

$$\hat{\alpha}_{t|t-1} = \mathbf{T}\hat{\alpha}_{t-1}; \quad \hat{y}_t = \mathbf{Z}_t\hat{\alpha}_{t-1},$$

com matrizes de covariância

$$\mathbf{P}_{t|t-1} = \mathbf{T}\mathbf{P}_{t-1}\mathbf{T}' + \mathbf{G}\mathbf{Q}\mathbf{G}' \quad \text{e} \quad \mathbf{F}_t = [\mathbf{Z}_t\mathbf{P}_{t|t-1}\mathbf{Z}_t' - \mathbf{C}_t'\mathbf{Z}_t' - \mathbf{Z}_t\mathbf{C}_t + \Sigma_{tt}]$$

respectivamente, onde

$$\mathbf{C}_t = \mathbf{A}_{t-1}\mathbf{A}_{t-2}\dots\tilde{\mathbf{A}}_1\Sigma_{1t} + \mathbf{A}_{t-1}\mathbf{A}_{t-2}\dots\tilde{\mathbf{A}}_2\Sigma_{2t} + \dots + \mathbf{A}_{t-1}\tilde{\mathbf{A}}_{t-2}\Sigma_{(t-2)t} + \tilde{\mathbf{A}}_{t-1}\Sigma_{(t-1)t},$$

para $\mathbf{A}_t = \mathbf{T}\mathbf{M}_t$ e $\tilde{\mathbf{A}}_t = \mathbf{T}\mathbf{K}_t$ se $t = 2, \dots, t-1$, e $\tilde{\mathbf{A}}_1 = \mathbf{T}\mathbf{K}_1$. Por sua vez, as matrizes \mathbf{M}_t e \mathbf{K}_t estão definidas como sendo

$$\mathbf{M}_t = \mathbf{I} - (\mathbf{P}_{t|t-1}\mathbf{Z}_t' - \mathbf{C}_t)\mathbf{F}_t^{-1}\mathbf{Z}_t \quad \text{e} \quad \mathbf{K}_t = (\mathbf{P}_{t|t-1}\mathbf{Z}_t' - \mathbf{C}_t)\mathbf{F}_t^{-1}.$$

Atualização da previsão para α_t :

$$\hat{\alpha}_t = \mathbf{T}\hat{\alpha}_{t-1} + (\mathbf{P}_{t|t-1}\mathbf{Z}_t' - \mathbf{C}_t)\mathbf{F}_t^{-1}(y_t - \mathbf{Z}_t\mathbf{T}\hat{\alpha}_{t-1})$$

Equações do filtro GLS

com matriz de covariâncias

$$\mathbf{P}_t = \mathbf{P}_{t|t-1} - (\mathbf{P}_{t|t-1}\mathbf{Z}'_t - \mathbf{C}_t)\mathbf{F}_t^{-1}(\mathbf{Z}_t\mathbf{P}_{t|t-1} - \mathbf{C}'_t).$$

Apêndice D

Resultados adicionais do exercício de simulação

Neste apêndice são apresentados os resultados em detalhe do exercício de simulação. O apêndice está organizado como segue: na Seção D.1 são entregues os resultados da avaliação do método para estimação da estrutura de autocorrelação dos erros de observação proposta em Pfeiffermann et al. (1998), posteriormente, na Seção D.2 são apresentadas as estatísticas relacionadas com o ajuste de modelos AR a estas estruturas de autocorrelação. Com estes dois elementos, na Seção D.3 são apresentadas as características da estimação das variâncias dos erros de evolução e, finalmente, na Seção sec-apen-filtros são entregues as tabelas ampliadas de resultados da comparação dos filtros de Kalman com erros AR e GLS.

D.1 Avaliação do estimador $\hat{\rho}(k)$

Tabela D.1: Estimação das autocovariâncias dos e_t pelo método dos pseudo-erros. Cenário 1. $\phi = 0.4$. 2000 replicações. $\rho(k)$: autocorrelação teórica. Para o estimador dividindo por T : $\hat{\rho}(k)$: média das estimativas, \sqrt{EQM} raiz quadrada do erro quadrático médio das estimativas, $P_{2.5}$, P_{50} e $P_{97.5}$: percentis de 2.5%, mediana e 97.5% das estimativas. Para o estimador dividindo por $T - k$: $\hat{\rho}^*(k)$: média das estimativas.

Defasagem	$\rho(k)$	$T = 60$							$T = 120$							$T = 240$						
		$\hat{\rho}(k)$	\sqrt{EQM}	$P_{2.5}$	P_{50}	$P_{97.5}$	$\hat{\rho}^*(k)$	$\hat{\rho}(k)$	\sqrt{EQM}	$P_{2.5}$	P_{50}	$P_{97.5}$	$\hat{\rho}^*(k)$	$\hat{\rho}(k)$	\sqrt{EQM}	$P_{2.5}$	P_{50}	$P_{97.5}$	$\hat{\rho}^*(k)$			
1	0.4000	0.3914	0.0462	0.3012	0.3927	0.4775	0.3981	0.3940	0.0326	0.3300	0.3942	0.4555	0.3973	0.3950	0.0227	0.3509	0.3951	0.4373	0.3967			
2	0.1600	0.1573	0.0537	0.0525	0.1574	0.2638	0.1627	0.1583	0.0385	0.0824	0.1580	0.2363	0.1610	0.1589	0.0274	0.1044	0.1590	0.2123	0.1602			
3	0.0640	0.0591	0.0557	-0.0460	0.0586	0.1760	0.0622	0.0580	0.0403	-0.0199	0.0581	0.1391	0.0595	0.0579	0.0294	0.0019	0.0578	0.1150	0.0587			
4	0.0256	0.0298	0.0543	-0.0753	0.0301	0.1362	0.0319	0.0297	0.0393	-0.0459	0.0294	0.1050	0.0308	0.0293	0.0288	-0.0257	0.0283	0.0861	0.0298			
5	0.0102	0.0202	0.0562	-0.0826	0.0185	0.1337	0.0221	0.0206	0.0407	-0.0561	0.0213	0.0995	0.0215	0.0196	0.0297	-0.0373	0.0202	0.0760	0.0200			
6	0.0041	0.0137	0.0564	-0.0943	0.0137	0.1202	0.0153	0.0142	0.0415	-0.0668	0.0147	0.0919	0.0150	0.0143	0.0310	-0.0448	0.0150	0.0705	0.0147			
7	0.0016	0.0080	0.0542	-0.0976	0.0071	0.1126	0.0091	0.0093	0.0398	-0.0631	0.0092	0.0831	0.0098	0.0107	0.0286	-0.0430	0.0106	0.0639	0.0110			
8	0.0007	-0.0061	0.0545	-0.1097	-0.0056	0.0980	-0.0071	0.0007	0.0393	-0.0786	0.0012	0.0774	0.0007	0.0045	0.0278	-0.0501	0.0050	0.0574	0.0046			
9	0.0003	-0.0035	0.0520	-0.1021	-0.0044	0.0984	-0.0041	-0.0018	0.0382	-0.0740	-0.0021	0.0722	-0.0019	-0.0008	0.0283	-0.0564	-0.0012	0.0555	-0.0008			
10	0.0001	-0.0012	0.0515	-0.1056	-0.0000	0.0982	-0.0014	-0.0029	0.0376	-0.0773	-0.0020	0.0710	-0.0032	-0.0030	0.0280	-0.0573	-0.0033	0.0500	-0.0031			
11	0.0000	0.0001	0.0522	-0.0980	-0.0002	0.1002	0.0001	-0.0020	0.0392	-0.0777	-0.0011	0.0731	-0.0022	-0.0027	0.0282	-0.0581	-0.0025	0.0506	-0.0028			
12	0.0000	0.0064	0.0525	-0.0955	0.0053	0.1056	0.0081	0.0046	0.0385	-0.0699	0.0048	0.0780	0.0051	0.0032	0.0279	-0.0505	0.0033	0.0576	0.0034			
13	0.0000	0.0049	0.0513	-0.0936	0.0067	0.0989	0.0063	0.0038	0.0383	-0.0728	0.0043	0.0770	0.0043	0.0012	0.0272	-0.0522	0.0009	0.0555	0.0013			
14	0.0000	-0.0023	0.0503	-0.0970	-0.0017	0.0953	-0.0030	-0.0041	0.0381	-0.0785	-0.0036	0.0685	-0.0047	-0.0064	0.0287	-0.0611	-0.0065	0.0507	-0.0068			
15	0.0000	-0.0050	0.0498	-0.1040	-0.0057	0.0944	-0.0067	-0.0049	0.0376	-0.0773	-0.0053	0.0701	-0.0057	-0.0051	0.0281	-0.0596	-0.0055	0.0475	-0.0055			
16	0.0000	-0.0118	0.0500	-0.1032	-0.0120	0.0910	-0.0161	-0.0074	0.0371	-0.0766	-0.0088	0.0645	-0.0085	-0.0043	0.0274	-0.0568	-0.0046	0.0481	-0.0046			
17	0.0000	-0.0021	0.0492	-0.1009	-0.0036	0.0943	-0.0029	-0.0014	0.0383	-0.0743	-0.0019	0.0779	-0.0016	-0.0001	0.0284	-0.0563	-0.0001	0.0546	-0.0002			
18	0.0000	0.0020	0.0494	-0.0992	0.0031	0.0961	0.0028	0.0021	0.0389	-0.0754	0.0029	0.0790	0.0025	0.0033	0.0282	-0.0527	0.0027	0.0572	0.0035			
19	0.0000	0.0064	0.0498	-0.0890	0.0068	0.1039	0.0094	0.0058	0.0401	-0.0697	0.0053	0.0867	0.0069	0.0065	0.0285	-0.0444	0.0059	0.0622	0.0070			
20	0.0000	0.0026	0.0475	-0.0890	0.0018	0.0959	0.0039	0.0010	0.0373	-0.0712	0.0006	0.0742	0.0012	0.0003	0.0268	-0.0517	-0.0010	0.0533	0.0003			

Tabela D.2: Estimação das autocovariâncias dos e_t pelo método dos pseudo-erros. Cenário 1. $\phi = 0.6$. 2000 replicações. $\rho(k)$: autocorrelação teórica. Para o estimador dividindo por T : $\hat{\rho}(k)$: média das estimativas, \sqrt{EQM} raiz quadrada do erro quadrático médio das estimativas, $P_{2.5}$, P_{50} e $P_{97.5}$: percentis de 2.5%, mediana e 97.5% das estimativas. Para o estimador dividindo por $T - k$: $\hat{\rho}^*(k)$: média das estimativas.

Defasagem	$\rho(k)$	$T = 60$						$T = 120$						$T = 240$					
		$\hat{\rho}(k)$	\sqrt{EQM}	$P_{2.5}$	P_{50}	$P_{97.5}$	$\hat{\rho}^*(k)$	$\hat{\rho}(k)$	\sqrt{EQM}	$P_{2.5}$	P_{50}	$P_{97.5}$	$\hat{\rho}^*(k)$	$\hat{\rho}(k)$	\sqrt{EQM}	$P_{2.5}$	P_{50}	$P_{97.5}$	$\hat{\rho}^*(k)$
1	0.6000	0.5880	0.0414	0.5088	0.5900	0.6639	0.5980	0.5929	0.0290	0.5357	0.5934	0.6465	0.5979	0.5952	0.0202	0.5570	0.5952	0.6333	0.5977
2	0.3600	0.3491	0.0566	0.2384	0.3496	0.4567	0.3611	0.3544	0.0405	0.2763	0.3551	0.4326	0.3604	0.3571	0.0287	0.2997	0.3578	0.4121	0.3601
3	0.2160	0.2051	0.0638	0.0836	0.2048	0.3332	0.2159	0.2090	0.0461	0.1181	0.2095	0.2966	0.2144	0.2112	0.0332	0.1455	0.2113	0.2761	0.2139
4	0.1296	0.1272	0.0651	0.0002	0.1273	0.2492	0.1363	0.1313	0.0470	0.0413	0.1314	0.2221	0.1358	0.1329	0.0343	0.0660	0.1324	0.2020	0.1351
5	0.0778	0.0820	0.0672	-0.0459	0.0803	0.2156	0.0894	0.0857	0.0486	-0.0061	0.0856	0.1778	0.0894	0.0866	0.0357	0.0188	0.0870	0.1547	0.0885
6	0.0467	0.0521	0.0681	-0.0759	0.0522	0.1841	0.0579	0.0556	0.0497	-0.0393	0.0565	0.1499	0.0585	0.0572	0.0369	-0.0129	0.0575	0.1245	0.0587
7	0.0280	0.0312	0.0668	-0.0985	0.0320	0.1580	0.0354	0.0350	0.0489	-0.0561	0.0356	0.1264	0.0372	0.0376	0.0353	-0.0280	0.0372	0.1064	0.0387
8	0.0168	0.0112	0.0668	-0.1205	0.0123	0.1427	0.0129	0.0181	0.0485	-0.0777	0.0180	0.1095	0.0193	0.0219	0.0345	-0.0461	0.0219	0.0914	0.0227
9	0.0101	0.0060	0.0648	-0.1170	0.0049	0.1353	0.0070	0.0090	0.0477	-0.0824	0.0074	0.1016	0.0097	0.0107	0.0347	-0.0582	0.0108	0.0790	0.0111
10	0.0060	0.0041	0.0639	-0.1215	0.0038	0.1317	0.0049	0.0039	0.0472	-0.0881	0.0046	0.0940	0.0042	0.0042	0.0346	-0.0634	0.0044	0.0685	0.0044
11	0.0036	0.0035	0.0640	-0.1198	0.0043	0.1287	0.0043	0.0022	0.0480	-0.0909	0.0034	0.0941	0.0024	0.0016	0.0348	-0.0684	0.0015	0.0673	0.0016
12	0.0022	0.0065	0.0639	-0.1189	0.0061	0.1271	0.0082	0.0049	0.0474	-0.0882	0.0053	0.0978	0.0055	0.0034	0.0344	-0.0630	0.0044	0.0699	0.0036
13	0.0013	0.0043	0.0629	-0.1213	0.0047	0.1203	0.0055	0.0032	0.0470	-0.0861	0.0045	0.0959	0.0035	0.0008	0.0339	-0.0633	0.0001	0.0691	0.0008
14	0.0008	-0.0019	0.0620	-0.1240	-0.0006	0.1203	-0.0025	-0.0031	0.0466	-0.0915	-0.0022	0.0838	-0.0036	-0.0052	0.0347	-0.0723	-0.0050	0.0630	-0.0055
15	0.0005	-0.0053	0.0611	-0.1250	-0.0040	0.1163	-0.0070	-0.0048	0.0460	-0.0946	-0.0040	0.0846	-0.0055	-0.0049	0.0343	-0.0717	-0.0046	0.0608	-0.0053
16	0.0003	-0.0097	0.0608	-0.1245	-0.0099	0.1132	-0.0132	-0.0063	0.0458	-0.0909	-0.0069	0.0847	-0.0073	-0.0039	0.0338	-0.0704	-0.0048	0.0602	-0.0042
17	0.0002	-0.0028	0.0606	-0.1232	-0.0040	0.1191	-0.0039	-0.0018	0.0470	-0.0921	-0.0025	0.0927	-0.0021	-0.0003	0.0345	-0.0658	-0.0008	0.0690	-0.0003
18	0.0001	0.0013	0.0605	-0.1191	0.0010	0.1180	0.0019	0.0016	0.0478	-0.0894	0.0014	0.0932	0.0018	0.0030	0.0346	-0.0639	0.0028	0.0701	0.0033
19	0.0001	0.0052	0.0606	-0.1153	0.0046	0.1251	0.0076	0.0045	0.0486	-0.0895	0.0042	0.0991	0.0053	0.0055	0.0346	-0.0610	0.0050	0.0747	0.0060
20	0.0000	0.0033	0.0589	-0.1132	0.0017	0.1209	0.0049	0.0016	0.0467	-0.0905	0.0005	0.0944	0.0020	0.0018	0.0334	-0.0631	0.0008	0.0685	0.0020

Tabela D.3: Estimação das autocovariâncias dos e_t pelo método dos pseudo-erros. Cenário 1. $\phi = 0.8$. 2000 replicações. $\rho(k)$: autocorrelação teórica. Para o estimador dividindo por T : $\hat{\rho}(k)$: média das estimativas, \sqrt{EQM} raiz quadrada do erro quadrático médio das estimativas, $P_{2.5}$, P_{50} e $P_{97.5}$: percentis de 2.5%, mediana e 97.5% das estimativas. Para o estimador dividindo por $T - k$: $\hat{\rho}^*(k)$: média das estimativas.

Defasagem	$\rho(k)$	$T = 60$						$T = 120$						$T = 240$					
		$\hat{\rho}(k)$	\sqrt{EQM}	$P_{2.5}$	P_{50}	$P_{97.5}$	$\hat{\rho}^*(k)$	$\hat{\rho}(k)$	\sqrt{EQM}	$P_{2.5}$	P_{50}	$P_{97.5}$	$\hat{\rho}^*(k)$	$\hat{\rho}(k)$	\sqrt{EQM}	$P_{2.5}$	P_{50}	$P_{97.5}$	$\hat{\rho}^*(k)$
1	0.8000	0.7843	0.0341	0.7196	0.7868	0.8382	0.7976	0.7922	0.0228	0.7495	0.7932	0.8315	0.7989	0.7960	0.0154	0.7657	0.7964	0.8235	0.7993
2	0.6400	0.6171	0.0554	0.5159	0.6196	0.7109	0.6384	0.6295	0.0375	0.5562	0.6314	0.6977	0.6402	0.6356	0.0258	0.5859	0.6364	0.6835	0.6409
3	0.5120	0.4846	0.0705	0.3514	0.4868	0.6066	0.5101	0.4992	0.0481	0.4057	0.5009	0.5864	0.5120	0.5065	0.0335	0.4402	0.5074	0.5688	0.5129
4	0.4096	0.3840	0.0787	0.2347	0.3868	0.5239	0.4115	0.4004	0.0538	0.2936	0.4013	0.4983	0.4142	0.4082	0.0381	0.3308	0.4082	0.4813	0.4151
5	0.3277	0.3047	0.0848	0.1384	0.3056	0.4628	0.3324	0.3217	0.0584	0.2044	0.3225	0.4310	0.3356	0.3296	0.0417	0.2468	0.3298	0.4115	0.3366
6	0.2621	0.2400	0.0893	0.0684	0.2414	0.4045	0.2667	0.2570	0.0621	0.1363	0.2571	0.3741	0.2705	0.2654	0.0445	0.1790	0.2648	0.3518	0.2722
7	0.2097	0.1871	0.0919	0.0041	0.1893	0.3615	0.2118	0.2039	0.0646	0.0795	0.2041	0.3253	0.2165	0.2126	0.0458	0.1228	0.2121	0.3009	0.2190
8	0.1678	0.1421	0.0945	-0.0433	0.1446	0.3188	0.1640	0.1594	0.0667	0.0332	0.1588	0.2858	0.1708	0.1683	0.0471	0.0760	0.1686	0.2583	0.1741
9	0.1342	0.1108	0.0938	-0.0657	0.1124	0.2897	0.1304	0.1250	0.0674	-0.0001	0.1250	0.2567	0.1351	0.1322	0.0485	0.0359	0.1324	0.2241	0.1373
10	0.1074	0.0870	0.0927	-0.0874	0.0886	0.2660	0.1044	0.0981	0.0679	-0.0292	0.0976	0.2289	0.1070	0.1038	0.0494	0.0061	0.1032	0.1973	0.1083
11	0.0859	0.0685	0.0918	-0.1006	0.0683	0.2443	0.0839	0.0776	0.0684	-0.0548	0.0785	0.2079	0.0854	0.0819	0.0499	-0.0172	0.0814	0.1781	0.0859
12	0.0687	0.0554	0.0907	-0.1163	0.0555	0.2284	0.0692	0.0629	0.0679	-0.0679	0.0646	0.1957	0.0699	0.0661	0.0496	-0.0325	0.0659	0.1604	0.0696
13	0.0550	0.0425	0.0899	-0.1245	0.0414	0.2166	0.0543	0.0491	0.0675	-0.0784	0.0496	0.1820	0.0551	0.0515	0.0495	-0.0446	0.0511	0.1478	0.0544
14	0.0440	0.0301	0.0894	-0.1388	0.0302	0.2066	0.0393	0.0358	0.0674	-0.0939	0.0362	0.1659	0.0406	0.0381	0.0498	-0.0569	0.0373	0.1346	0.0404
15	0.0352	0.0211	0.0884	-0.1491	0.0222	0.1959	0.0282	0.0268	0.0670	-0.1002	0.0266	0.1562	0.0306	0.0298	0.0495	-0.0641	0.0300	0.1280	0.0318
16	0.0281	0.0137	0.0877	-0.1546	0.0119	0.1886	0.0186	0.0199	0.0669	-0.1080	0.0192	0.1526	0.0230	0.0241	0.0492	-0.0687	0.0235	0.1203	0.0259
17	0.0225	0.0131	0.0866	-0.1525	0.0130	0.1861	0.0183	0.0174	0.0672	-0.1128	0.0174	0.1471	0.0203	0.0212	0.0493	-0.0733	0.0209	0.1181	0.0229
18	0.0180	0.0125	0.0856	-0.1528	0.0120	0.1846	0.0178	0.0156	0.0675	-0.1169	0.0159	0.1499	0.0183	0.0194	0.0493	-0.0710	0.0187	0.1165	0.0209
19	0.0144	0.0124	0.0848	-0.1600	0.0105	0.1809	0.0181	0.0142	0.0677	-0.1155	0.0152	0.1445	0.0169	0.0178	0.0492	-0.0763	0.0175	0.1162	0.0193
20	0.0115	0.0100	0.0835	-0.1563	0.0093	0.1763	0.0150	0.0107	0.0668	-0.1206	0.0105	0.1382	0.0129	0.0139	0.0486	-0.0822	0.0140	0.1127	0.0152

Tabela D.4: Estimação das autocovariâncias dos e_t pelo método dos pseudo-erros. Cenário 2. $\phi = 0.4$. 2000 replicações. $\rho(k)$: autocorrelação teórica. Para o estimador dividindo por T : $\hat{\rho}(k)$: média das estimativas, \sqrt{EQM} raiz quadrada do erro quadrático médio das estimativas, $P_{2.5}$, P_{50} e $P_{97.5}$: percentis de 2.5%, mediana e 97.5% das estimativas. Para o estimador dividindo por $T - k$: $\hat{\rho}^*(k)$: média das estimativas.

Defasagem	$\rho(k)$	$T = 60$						$T = 120$						$T = 240$					
		$\hat{\rho}(k)$	\sqrt{EQM}	$P_{2.5}$	P_{50}	$P_{97.5}$	$\hat{\rho}^*(k)$	$\hat{\rho}(k)$	\sqrt{EQM}	$P_{2.5}$	P_{50}	$P_{97.5}$	$\hat{\rho}^*(k)$	$\hat{\rho}(k)$	\sqrt{EQM}	$P_{2.5}$	P_{50}	$P_{97.5}$	$\hat{\rho}^*(k)$
1	0.3010	0.2977	0.2203	0.2040	0.2998	0.3839	0.3027	0.2997	0.2200	0.2348	0.3010	0.3609	0.3023	0.3006	0.2198	0.2557	0.3007	0.3447	0.3019
2	0.0820	0.0817	0.0805	-0.0165	0.0821	0.1818	0.0845	0.0829	0.0734	0.0121	0.0837	0.1516	0.0843	0.0822	0.0685	0.0287	0.0827	0.1326	0.0829
3	0.0191	0.0197	0.0558	-0.0831	0.0186	0.1295	0.0208	0.0211	0.0407	-0.0520	0.0212	0.0944	0.0217	0.0195	0.0309	-0.0346	0.0193	0.0697	0.0197
4	0.0041	0.0087	0.0548	-0.1013	0.0085	0.1175	0.0093	0.0057	0.0379	-0.0661	0.0051	0.0767	0.0059	0.0038	0.0264	-0.0475	0.0030	0.0543	0.0039
5	0.0041	0.0093	0.0522	-0.0874	0.0099	0.1116	0.0101	0.0063	0.0387	-0.0665	0.0057	0.0812	0.0066	0.0046	0.0270	-0.0487	0.0044	0.0593	0.0047
6	0.0041	0.0091	0.0521	-0.0920	0.0103	0.1116	0.0101	0.0057	0.0386	-0.0694	0.0064	0.0840	0.0060	0.0045	0.0271	-0.0482	0.0053	0.0565	0.0046
7	0.0041	0.0061	0.0502	-0.0922	0.0065	0.1026	0.0069	0.0039	0.0376	-0.0716	0.0045	0.0785	0.0042	0.0036	0.0268	-0.0505	0.0035	0.0551	0.0038
8	0.0041	-0.0068	0.0538	-0.1075	-0.0068	0.0878	-0.0079	-0.0013	0.0403	-0.0760	-0.0012	0.0754	-0.0014	0.0007	0.0291	-0.0535	0.0005	0.0537	0.0007
9	0.0116	0.0097	0.0492	-0.0841	0.0088	0.1085	0.0115	0.0115	0.0368	-0.0605	0.0115	0.0831	0.0124	0.0112	0.0275	-0.0424	0.0112	0.0623	0.0116
10	0.0191	0.0182	0.0498	-0.0766	0.0184	0.1163	0.0219	0.0202	0.0356	-0.0482	0.0204	0.0863	0.0221	0.0194	0.0272	-0.0319	0.0195	0.0714	0.0202
11	0.0266	0.0244	0.0503	-0.0705	0.0250	0.1208	0.0299	0.0256	0.0363	-0.0439	0.0255	0.0947	0.0281	0.0267	0.0272	-0.0238	0.0261	0.0780	0.0279
12	0.0340	0.0302	0.0492	-0.0658	0.0314	0.1275	0.0378	0.0311	0.0364	-0.0391	0.0310	0.1031	0.0346	0.0338	0.0278	-0.0204	0.0339	0.0848	0.0356
13	0.0266	0.0233	0.0483	-0.0703	0.0230	0.1185	0.0297	0.0242	0.0361	-0.0455	0.0247	0.0921	0.0271	0.0261	0.0267	-0.0247	0.0253	0.0775	0.0276
14	0.0191	0.0166	0.0463	-0.0712	0.0169	0.1072	0.0217	0.0180	0.0363	-0.0520	0.0189	0.0892	0.0204	0.0189	0.0264	-0.0311	0.0186	0.0690	0.0201
15	0.0116	0.0074	0.0457	-0.0799	0.0073	0.0968	0.0098	0.0111	0.0363	-0.0577	0.0107	0.0777	0.0126	0.0107	0.0261	-0.0363	0.0107	0.0593	0.0114
16	0.0041	-0.0073	0.0477	-0.0979	-0.0078	0.0826	-0.0100	-0.0012	0.0358	-0.0681	-0.0018	0.0706	-0.0014	-0.0002	0.0277	-0.0550	-0.0005	0.0539	-0.0002
17	0.0041	0.0020	0.0445	-0.0867	0.0012	0.0899	0.0028	0.0027	0.0346	-0.0652	0.0022	0.0686	0.0032	0.0031	0.0264	-0.0481	0.0031	0.0559	0.0034
18	0.0041	0.0037	0.0445	-0.0851	0.0044	0.0884	0.0053	0.0044	0.0342	-0.0641	0.0047	0.0712	0.0052	0.0039	0.0254	-0.0470	0.0039	0.0522	0.0042
19	0.0041	0.0043	0.0429	-0.0839	0.0046	0.0873	0.0063	0.0058	0.0351	-0.0629	0.0056	0.0727	0.0069	0.0051	0.0259	-0.0445	0.0053	0.0594	0.0055
20	0.0041	0.0047	0.0454	-0.0862	0.0046	0.0913	0.0070	0.0060	0.0350	-0.0613	0.0071	0.0754	0.0072	0.0045	0.0271	-0.0488	0.0044	0.0564	0.0049

Tabela D.5: Estimação das autocovariâncias dos e_t pelo método dos pseudo-erros. Cenário 2. $\phi = 0.6$. 2000 replicações. $\rho(k)$: autocorrelação teórica. Para o estimador dividindo por T : $\hat{\rho}(k)$: média das estimativas, \sqrt{EQM} raiz quadrada do erro quadrático médio das estimativas, $P_{2.5}$, P_{50} e $P_{97.5}$: percentis de 2.5%, mediana e 97.5% das estimativas. Para o estimador dividindo por $T - k$: $\hat{\rho}^*(k)$: média das estimativas.

Defasagem	$\rho(k)$	$T = 60$						$T = 120$						$T = 240$					
		$\hat{\rho}(k)$	\sqrt{EQM}	$P_{2.5}$	P_{50}	$P_{97.5}$	$\hat{\rho}^*(k)$	$\hat{\rho}(k)$	\sqrt{EQM}	$P_{2.5}$	P_{50}	$P_{97.5}$	$\hat{\rho}^*(k)$	$\hat{\rho}(k)$	\sqrt{EQM}	$P_{2.5}$	P_{50}	$P_{97.5}$	$\hat{\rho}^*(k)$
1	0.4617	0.4546	0.2552	0.3666	0.4569	0.5414	0.4623	0.4587	0.2575	0.3934	0.4596	0.5254	0.4625	0.4605	0.2584	0.4111	0.4610	0.5099	0.4624
2	0.2033	0.1985	0.1233	0.0857	0.1993	0.3088	0.2053	0.2015	0.1207	0.1145	0.2009	0.2857	0.2050	0.2024	0.1183	0.1393	0.2017	0.2715	0.2041
3	0.0890	0.0867	0.0732	-0.0301	0.0874	0.2109	0.0912	0.0883	0.0641	-0.0014	0.0869	0.1872	0.0906	0.0882	0.0559	0.0194	0.0872	0.1663	0.0893
4	0.0467	0.0451	0.0656	-0.0796	0.0437	0.1768	0.0484	0.0458	0.0516	-0.0518	0.0457	0.1520	0.0474	0.0468	0.0403	-0.0262	0.0448	0.1263	0.0476
5	0.0467	0.0436	0.0681	-0.0876	0.0435	0.1754	0.0476	0.0448	0.0511	-0.0523	0.0434	0.1527	0.0468	0.0472	0.0398	-0.0260	0.0445	0.1290	0.0482
6	0.0467	0.0426	0.0685	-0.0883	0.0431	0.1745	0.0474	0.0438	0.0519	-0.0571	0.0443	0.1508	0.0461	0.0459	0.0407	-0.0294	0.0451	0.1276	0.0471
7	0.0467	0.0379	0.0670	-0.0873	0.0369	0.1697	0.0429	0.0420	0.0520	-0.0574	0.0415	0.1485	0.0446	0.0447	0.0404	-0.0330	0.0444	0.1247	0.0461
8	0.0467	0.0284	0.0769	-0.0959	0.0284	0.1599	0.0327	0.0387	0.0585	-0.0635	0.0394	0.1348	0.0415	0.0426	0.0477	-0.0325	0.0426	0.1271	0.0440
9	0.0678	0.0547	0.0706	-0.0689	0.0544	0.1762	0.0643	0.0615	0.0548	-0.0282	0.0612	0.1584	0.0665	0.0649	0.0449	-0.0055	0.0638	0.1447	0.0674
10	0.0890	0.0757	0.0688	-0.0386	0.0756	0.1896	0.0909	0.0822	0.0538	-0.0046	0.0823	0.1745	0.0896	0.0860	0.0435	0.0176	0.0840	0.1605	0.0897
11	0.1102	0.0925	0.0699	-0.0179	0.0912	0.2056	0.1133	0.1008	0.0548	0.0136	0.1000	0.1902	0.1110	0.1061	0.0442	0.0398	0.1050	0.1799	0.1112
12	0.1313	0.1080	0.0583	-0.0040	0.1064	0.2251	0.1350	0.1200	0.0474	0.0313	0.1191	0.2130	0.1333	0.1256	0.0396	0.0544	0.1241	0.2015	0.1322
13	0.1102	0.0891	0.0551	-0.0150	0.0894	0.1968	0.1138	0.1000	0.0461	0.0116	0.0985	0.1881	0.1121	0.1055	0.0396	0.0378	0.1037	0.1802	0.1116
14	0.0890	0.0705	0.0547	-0.0320	0.0692	0.1747	0.0919	0.0793	0.0465	-0.0067	0.0775	0.1729	0.0897	0.0848	0.0397	0.0198	0.0830	0.1594	0.0901
15	0.0678	0.0502	0.0560	-0.0550	0.0501	0.1567	0.0670	0.0579	0.0476	-0.0275	0.0568	0.1552	0.0662	0.0631	0.0410	-0.0042	0.0617	0.1392	0.0673
16	0.0467	0.0260	0.0616	-0.0875	0.0256	0.1394	0.0354	0.0345	0.0499	-0.0542	0.0332	0.1325	0.0398	0.0402	0.0403	-0.0331	0.0370	0.1225	0.0431
17	0.0467	0.0333	0.0562	-0.0705	0.0309	0.1413	0.0464	0.0387	0.0471	-0.0468	0.0380	0.1312	0.0451	0.0419	0.0378	-0.0263	0.0390	0.1219	0.0451
18	0.0467	0.0359	0.0538	-0.0640	0.0349	0.1426	0.0513	0.0403	0.0454	-0.0450	0.0402	0.1289	0.0475	0.0432	0.0361	-0.0204	0.0416	0.1174	0.0467
19	0.0467	0.0362	0.0543	-0.0664	0.0368	0.1404	0.0530	0.0402	0.0460	-0.0475	0.0408	0.1323	0.0478	0.0436	0.0365	-0.0228	0.0423	0.1169	0.0473
20	0.0467	0.0342	0.0557	-0.0760	0.0345	0.1363	0.0512	0.0407	0.0475	-0.0489	0.0393	0.1319	0.0488	0.0433	0.0378	-0.0266	0.0416	0.1199	0.0472

Tabela D.6: Estimação das autocovariâncias dos e_t pelo método dos pseudo-erros. Cenário 2. $\phi = 0.8$. 2000 replicações. $\rho(k)$: autocorrelação teórica. Para o estimador dividindo por T : $\hat{\rho}(k)$: média das estimativas, \sqrt{EQM} raiz quadrada do erro quadrático médio das estimativas, $P_{2.5}$, P_{50} e $P_{97.5}$: percentis de 2.5%, mediana e 97.5% das estimativas. Para o estimador dividindo por $T - k$: $\hat{\rho}^*(k)$: média das estimativas.

Defasagem	$\rho(k)$	$T = 60$						$T = 120$						$T = 240$					
		$\hat{\rho}(k)$	\sqrt{EQM}	$P_{2.5}$	P_{50}	$P_{97.5}$	$\hat{\rho}^*(k)$	$\hat{\rho}(k)$	\sqrt{EQM}	$P_{2.5}$	P_{50}	$P_{97.5}$	$\hat{\rho}^*(k)$	$\hat{\rho}(k)$	\sqrt{EQM}	$P_{2.5}$	P_{50}	$P_{97.5}$	$\hat{\rho}^*(k)$
1	0.6655	0.6496	0.2062	0.5381	0.6503	0.7505	0.6606	0.6558	0.2108	0.5607	0.6545	0.7505	0.6613	0.6582	0.2122	0.5747	0.6569	0.7467	0.6610
2	0.4511	0.4291	0.1366	0.2612	0.4312	0.5910	0.4439	0.4362	0.1378	0.2814	0.4349	0.5914	0.4436	0.4399	0.1373	0.3021	0.4371	0.5852	0.4436
3	0.3246	0.3007	0.1130	0.1011	0.3011	0.5003	0.3165	0.3069	0.1083	0.1225	0.3058	0.4962	0.3147	0.3117	0.1042	0.1439	0.3112	0.4916	0.3157
4	0.2621	0.2373	0.1180	0.0170	0.2371	0.4520	0.2543	0.2426	0.1101	0.0483	0.2412	0.4497	0.2510	0.2477	0.1016	0.0592	0.2458	0.4424	0.2519
5	0.2621	0.2333	0.1176	0.0107	0.2333	0.4440	0.2545	0.2410	0.1091	0.0466	0.2398	0.4497	0.2515	0.2465	0.1019	0.0586	0.2437	0.4429	0.2518
6	0.2621	0.2282	0.1184	0.0152	0.2282	0.4344	0.2536	0.2388	0.1086	0.0373	0.2388	0.4408	0.2513	0.2456	0.1018	0.0604	0.2446	0.4423	0.2519
7	0.2621	0.2216	0.1193	0.0065	0.2229	0.4322	0.2508	0.2362	0.1083	0.0334	0.2355	0.4351	0.2508	0.2437	0.1016	0.0609	0.2417	0.4392	0.2510
8	0.2621	0.2119	0.1383	-0.0022	0.2127	0.4223	0.2446	0.2327	0.1212	0.0283	0.2325	0.4303	0.2494	0.2418	0.1120	0.0600	0.2386	0.4356	0.2501
9	0.2934	0.2390	0.1339	0.0398	0.2380	0.4326	0.2812	0.2620	0.1166	0.0775	0.2643	0.4537	0.2833	0.2731	0.1069	0.1043	0.2697	0.4538	0.2837
10	0.3246	0.2632	0.1334	0.0826	0.2609	0.4451	0.3158	0.2894	0.1141	0.1165	0.2884	0.4665	0.3157	0.3033	0.1029	0.1416	0.3007	0.4750	0.3165
11	0.3558	0.2853	0.1359	0.1160	0.2847	0.4548	0.3494	0.3153	0.1139	0.1514	0.3132	0.4814	0.3471	0.3324	0.1007	0.1770	0.3312	0.4957	0.3484
12	0.3871	0.3052	0.1007	0.1345	0.3052	0.4666	0.3815	0.3411	0.0862	0.1845	0.3383	0.5016	0.3790	0.3608	0.0815	0.2070	0.3607	0.5171	0.3798
13	0.3558	0.2731	0.1016	0.1001	0.2737	0.4391	0.3486	0.3098	0.0880	0.1498	0.3076	0.4773	0.3475	0.3285	0.0849	0.1723	0.3273	0.4929	0.3473
14	0.3246	0.2419	0.1036	0.0748	0.2426	0.4116	0.3156	0.2779	0.0911	0.1147	0.2745	0.4553	0.3146	0.2963	0.0886	0.1363	0.2918	0.4695	0.3147
15	0.2934	0.2113	0.1056	0.0382	0.2078	0.3866	0.2817	0.2458	0.0953	0.0754	0.2441	0.4336	0.2809	0.2645	0.0929	0.0954	0.2611	0.4478	0.2821
16	0.2621	0.1794	0.1278	-0.0147	0.1788	0.3633	0.2447	0.2134	0.1106	0.0298	0.2108	0.4101	0.2463	0.2325	0.1021	0.0517	0.2306	0.4249	0.2491
17	0.2621	0.1792	0.1239	0.0058	0.1778	0.3547	0.2501	0.2136	0.1086	0.0374	0.2109	0.4081	0.2488	0.2330	0.1005	0.0576	0.2320	0.4225	0.2507
18	0.2621	0.1772	0.1231	0.0103	0.1772	0.3442	0.2532	0.2127	0.1076	0.0414	0.2105	0.4013	0.2503	0.2326	0.0995	0.0567	0.2293	0.4219	0.2514
19	0.2621	0.1742	0.1240	0.0065	0.1760	0.3408	0.2550	0.2119	0.1068	0.0412	0.2091	0.3922	0.2518	0.2320	0.0994	0.0598	0.2300	0.4167	0.2519
20	0.2621	0.1700	0.1266	0.0024	0.1735	0.3355	0.2550	0.2102	0.1070	0.0366	0.2077	0.3896	0.2523	0.2311	0.1001	0.0569	0.2316	0.4150	0.2521

Ajuste das funções de autocorrelação dos erros com modelos AR.

D.2 Ajuste das funções de autocorrelação dos erros com modelos AR.

Tabela D.7: Estimação da componente autoregressiva dos e_t no cenário 1. 2000 replicações. $p_{2.5}$ e $p_{97.5}$: percentis de 2.5% e 97.5% das estimativas.

ϕ	T	Média	Desv.Padr.	$p_{2.5}$	Mediana	$p_{97.5}$
0.4	60	0.3954	0.0566	0.2906	0.3930	0.5132
	120	0.3968	0.0406	0.3205	0.3956	0.4806
	240	0.3973	0.0288	0.3416	0.3969	0.4547
0.6	60	0.5927	0.0563	0.4845	0.5923	0.7065
	120	0.5981	0.0410	0.5193	0.5981	0.6784
	240	0.6006	0.0293	0.5444	0.6003	0.6598
0.8	60	0.7828	0.0455	0.6866	0.7859	0.8649
	120	0.7932	0.0321	0.7286	0.7950	0.8520
	240	0.7981	0.0229	0.7528	0.7984	0.8408

Tabela D.8: Ajuste da componente autoregressiva dos e_t no cenário 2 empregando um modelo AR(1). 2000 replicações.

$p_{2.5}$ e $p_{97.5}$: percentis de 2.5% e 97.5% das estimativas.

ϕ	T	Média	Desv.Padr.	$p_{2.5}$	Mediana	$p_{97.5}$
0.4	60	0.2951	0.0531	0.1955	0.2914	0.4051
	120	0.2963	0.0367	0.2260	0.2965	0.3689
	240	0.2959	0.0261	0.2465	0.2950	0.3487
0.6	60	0.4603	0.0707	0.3431	0.4549	0.6164
	120	0.4632	0.0559	0.3671	0.4579	0.5881
	240	0.4641	0.0448	0.3903	0.4607	0.5619
0.8	60	0.7644	0.1128	0.5164	0.7969	0.9070
	120	0.7827	0.1064	0.5445	0.8159	0.9133
	240	0.7905	0.0991	0.5669	0.8212	0.9142

Estimação das variâncias dos erros de evolução

Tabela D.9: Ajuste da componente autoregressiva dos e_t no cenário 2 empregando um modelo AR(12). 2000 replicações.

ϕ	T	Estatística	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8	ϕ_9	ϕ_{10}	ϕ_{11}	ϕ_{12}
0.4	60	Média	0.3036	-0.0098	-0.0034	-0.0002	0.0061	0.0012	0.0054	-0.0184	0.0139	0.0055	0.0077	0.0231
		Desv.Padr.	0.0507	0.0531	0.0539	0.0562	0.0531	0.0512	0.0525	0.0529	0.0511	0.0505	0.0540	0.0550
	120	Média	0.3034	-0.0085	-0.0016	-0.0025	0.0050	0.0010	0.0027	-0.0088	0.0102	0.0086	0.0067	0.0254
		Desv.Padr.	0.0345	0.0361	0.0366	0.0381	0.0375	0.0357	0.0363	0.0375	0.0357	0.0349	0.0372	0.0405
	240	Média	0.3036	-0.0086	-0.0028	-0.0025	0.0039	0.0011	0.0021	-0.0052	0.0085	0.0085	0.0074	0.0286
		Desv.Padr.	0.0244	0.0260	0.0260	0.0264	0.0258	0.0248	0.0260	0.0263	0.0249	0.0254	0.0266	0.0290
0.6	60	Média	0.4585	-0.0151	-0.0068	-0.0081	0.0230	0.0054	0.0074	-0.0200	0.0330	0.0185	0.0045	0.0680
		Desv.Padr.	0.0545	0.0596	0.0602	0.0664	0.0642	0.0503	0.0614	0.0620	0.0596	0.0566	0.0675	0.0635
	120	Média	0.4578	-0.0138	-0.0071	-0.0080	0.0223	0.0061	0.0058	-0.0076	0.0272	0.0222	0.0033	0.0799
		Desv.Padr.	0.0369	0.0396	0.0393	0.0437	0.0430	0.0351	0.0425	0.0409	0.0411	0.0404	0.0481	0.0486
	240	Média	0.4575	-0.0139	-0.0082	-0.0065	0.0229	0.0059	0.0064	-0.0045	0.0261	0.0245	0.0025	0.0857
		Desv.Padr.	0.0262	0.0276	0.0273	0.0307	0.0290	0.0242	0.0290	0.0290	0.0285	0.0274	0.0343	0.0359
0.8	60	Média	0.6022	-0.0030	-0.0244	-0.0162	0.0611	0.0168	0.0062	-0.0207	0.0601	0.0270	0.0090	0.0943
		Desv.Padr.	0.0500	0.0548	0.0526	0.0609	0.0568	0.0512	0.0568	0.0583	0.0559	0.0509	0.0755	0.0666
	120	Média	0.5945	-0.0028	-0.0311	-0.0196	0.0637	0.0176	-0.0002	-0.0091	0.0631	0.0306	-0.0041	0.1277
		Desv.Padr.	0.0358	0.0382	0.0378	0.0440	0.0441	0.0325	0.0433	0.0400	0.0423	0.0354	0.0591	0.0512
	240	Média	0.5862	-0.0031	-0.0303	-0.0202	0.0611	0.0192	-0.0015	-0.0069	0.0643	0.0319	-0.0057	0.1422
		Desv.Padr.	0.0261	0.0281	0.0272	0.0299	0.0297	0.0225	0.0295	0.0283	0.0299	0.0242	0.0415	0.0389

D.3 Estimação das variâncias dos erros de evolução

Tabela D.10: Resumo das estimativas para σ_L^2 , σ_R^2 , σ_S^2 , σ_I^2 , pelo método de máxima verossimilhança. Cenário 1. Erros ajustados com um AR(1). 2000 replicações.

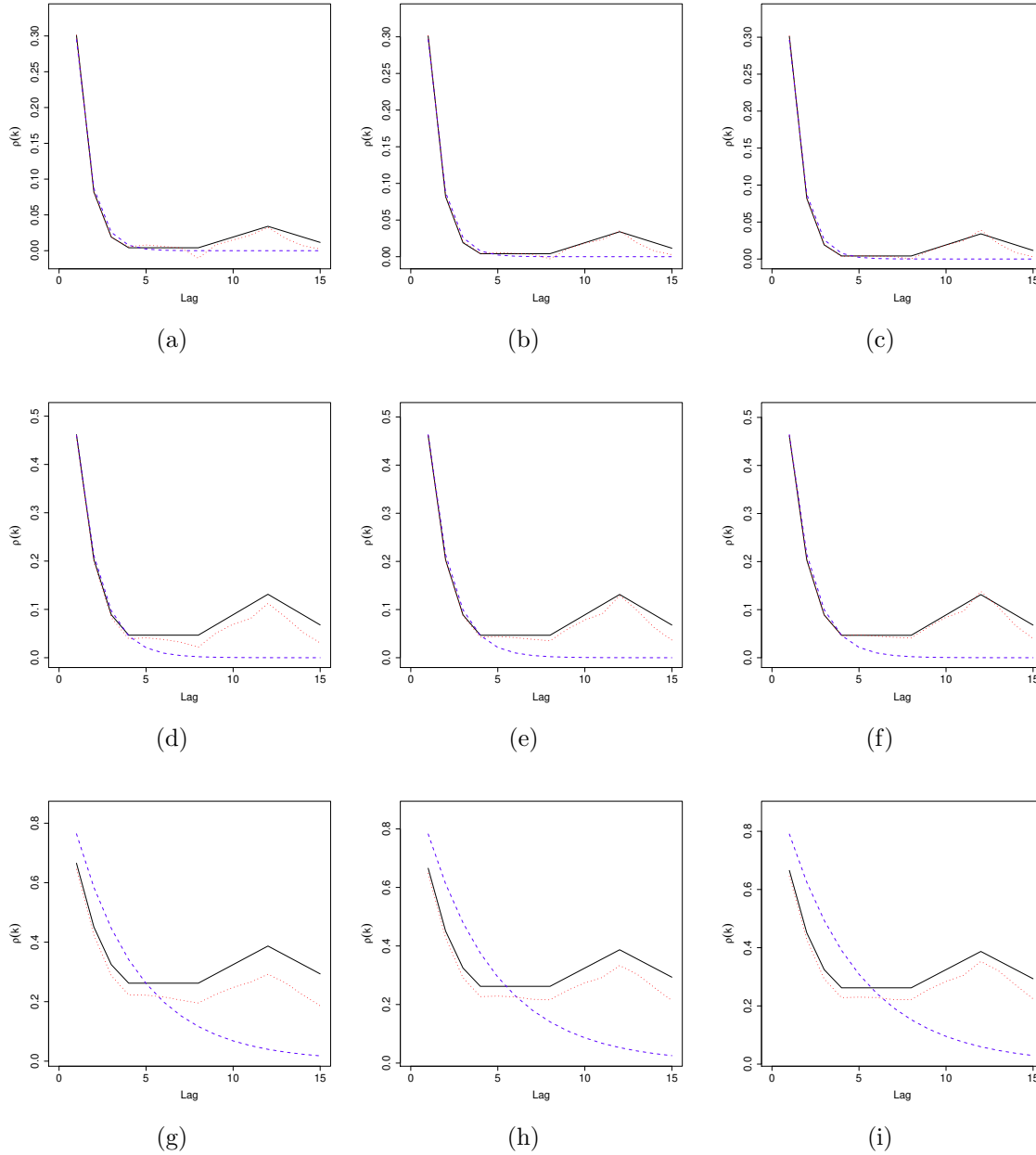
ϕ	T	Estatística	σ_L^2	σ_R^2	σ_S^2	σ_I^2
-	-	Valores Teóricos	0.317000	1.13×10^{-05}	6.98×10^{-08}	0.000129
0.4	60	Média	0.224051	0.001846	0.005451	0.039707
		Desv. Padrão	0.142846	0.005386	0.011460	0.065938
	120	Média	0.265447	0.000543	0.000674	0.031683
		Desv. Padrão	0.101040	0.002715	0.002170	0.048934
	240	Média	0.274046	0.001406	0.000313	0.015793
		Desv. Padrão	0.091252	0.006775	0.000742	0.025632
0.6	60	Média	0.219742	0.002039	0.004334	0.034572
		Desv. Padrão	0.136997	0.006064	0.009181	0.054968
	120	Média	0.248521	0.001558	0.000240	0.034315
		Desv. Padrão	0.115902	0.006076	0.001138	0.048713
	240	Média	0.240677	0.012551	0.000170	0.008731
		Desv. Padrão	0.130789	0.043319	0.000508	0.020996
0.8	60	Média	0.220431	0.002947	0.003255	0.027948
		Desv. Padrão	0.119552	0.008851	0.006987	0.043125
	120	Média	0.241602	0.002822	0.000253	0.031503
		Desv. Padrão	0.119580	0.009324	0.000993	0.046670
	240	Média	0.201148	0.090997	0.000070	0.007696
		Desv. Padrão	0.150725	0.149533	0.000294	0.023979

Estimação das variâncias dos erros de evolução

Figura D.1: Ajuste das funções de autocorrelação dos erros com modelos AR. Cenário 2. 2000 replicações.

Nas linhas: $\phi = 0.4, 0.6$ e 0.8 . Nas colunas: $T = 60, 120$ e 240 .

Cor das linhas: preta contínua: valor teórico, azul tracejada: ajuste modelo AR(1), vermelha pontilhada: ajuste modelo AR(12).



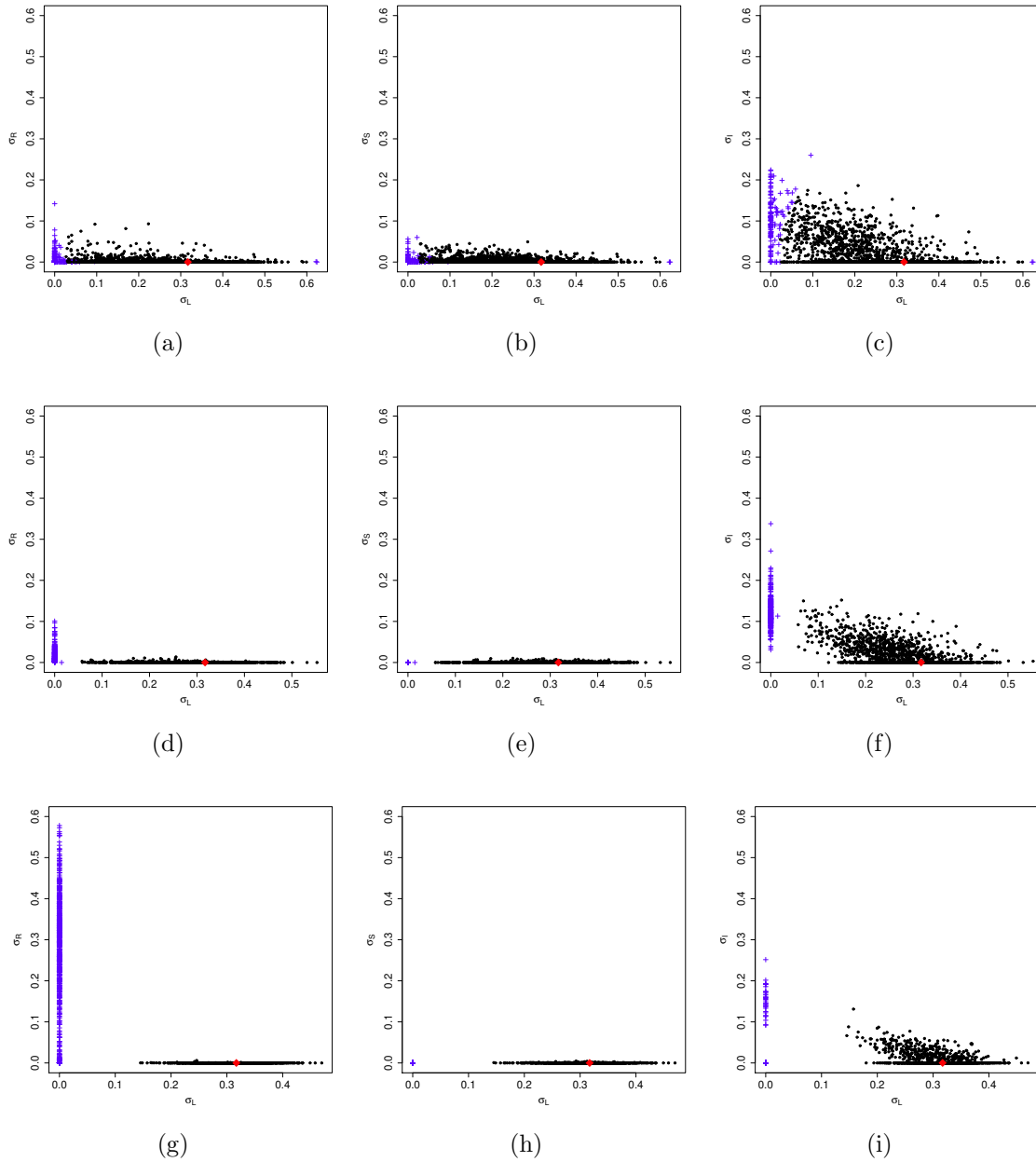
Estimação das variâncias dos erros de evolução

Tabela D.11: Resumo das estimativas para σ_L^2 , σ_R^2 , σ_S^2 , σ_I^2 , pelo método de máxima verossimilhança. Cenário 2. Erros ajustados com modelos AR(1) e AR(12). 2000 replicações.

ϕ	T	Estatística	AR(1)				AR(12)			
			σ_L^2	σ_R^2	σ_S^2	σ_I^2	σ_L^2	σ_R^2	σ_S^2	σ_I^2
Valores Teóricos			0.317000	1.13×10^{-05}	6.98×10^{-08}	0.000129	0.317000	1.13×10^{-05}	6.98×10^{-08}	0.000129
0.4	60	Média	0.229865	0.001763	0.005267	0.040820	0.208380	0.001881	0.005280	0.039854
		Desv.Padr.	0.142408	0.006588	0.011244	0.070519	0.144547	0.006510	0.012820	0.070117
	120	Média	0.274645	0.000485	0.000869	0.029792	0.261379	0.000418	0.000715	0.026626
		Desv.Padr.	0.098437	0.003278	0.002617	0.050033	0.096031	0.002096	0.002536	0.045651
	240	Média	0.282153	0.000782	0.000373	0.017947	0.267723	0.000959	0.000302	0.013532
		Desv.Padr.	0.078510	0.005195	0.000841	0.028935	0.083555	0.005067	0.000754	0.024445
0.6	60	Média	0.218750	0.001851	0.005044	0.039540	0.187901	0.002183	0.004767	0.035802
		Desv.Padr.	0.137856	0.006371	0.010315	0.064924	0.133543	0.006948	0.010830	0.061212
	120	Média	0.259376	0.001000	0.000522	0.033110	0.240763	0.000749	0.000361	0.024008
		Desv.Padr.	0.108494	0.004514	0.001954	0.051932	0.098263	0.003564	0.001679	0.041559
	240	Média	0.263243	0.003081	0.000306	0.014042	0.221914	0.003425	0.000139	0.006824
		Desv.Padr.	0.104910	0.013312	0.000691	0.025412	0.120276	0.008673	0.000459	0.019661
0.8	60	Média	0.254741	0.002925	0.004296	0.041049	0.144419	0.002634	0.003499	0.020133
		Desv.Padr.	0.145310	0.008692	0.009165	0.059261	0.108559	0.006105	0.008680	0.038508
	120	Média	0.293710	0.001396	0.000878	0.042719	0.191981	0.001138	0.000171	0.015309
		Desv.Padr.	0.130692	0.007099	0.002259	0.053766	0.089127	0.003311	0.000950	0.028682
	240	Média	0.294224	0.032357	0.000244	0.026375	0.175721	0.004993	0.000032	0.003890
		Desv.Padr.	0.134620	0.111576	0.000616	0.034772	0.120813	0.010786	0.000180	0.019161

Estimação das variâncias dos erros de evolução

Figura D.2: Gráficos de dispersão. Valores estimados para σ_L^2 , σ_R^2 , σ_S^2 e σ_I^2 segundo distância aos valores teóricos e tamanho da série. Cenário 1, $\phi = 0.8$. 2000 replicações. Nas linhas: $T = 60, 120$ e 240 . Nas colunas: $\hat{\sigma}_L^2 \times \hat{\sigma}_R^2$, $\hat{\sigma}_L^2 \times \hat{\sigma}_S^2$, $\hat{\sigma}_L^2 \times \hat{\sigma}_I^2$. Tipo de ponto: distância D do vetor de estimativas ao vetor de valores teóricos. Bola preta: $D \leq 0.3$. Cruz azul: $D > 0.3$. Losângulo vermelho: valor real.



D.4 Filtros de Kalman com erros ARMA e GLS

Tabela D.12: Resumo das estimativas para $\{\theta_t\}$. Cenário 1 com $\phi = 0.4$ e $T = 60$. 2000 replicações.

Modelos considerados: (1) Estimativa de desenho; Filtro de Kalman: (2) com e_t independentes, (3) com e_t $AR(1)$, (4) com e_t $AR(1)$ com hiperparâmetros conhecidos; Filtro GLS, estrutura de autocorrelação: (5) estimativas iniciais (6), estimativas iniciais ajustadas $AR(1)$, (7) hiperparâmetros conhecidos.

$p_{2.5}$ e $p_{97.5}$: percentis de 2.5% e 97.5% das estimativas.

Estimador		(1)	(2)	(3)	(4)	(5)	(6)	(7)
EQM total	Média	0.3257	0.2726	0.2693	0.2641	0.2725	0.2724	0.2709
	Desv.Padr.	0.0877	0.0801	0.0788	0.0771	0.0795	0.0795	0.0782
	$p_{2.5}$	0.1836	0.1455	0.1446	0.1426	0.1459	0.1458	0.1477
	Mediana	0.3162	0.2611	0.2580	0.2533	0.2619	0.2618	0.2606
	$p_{97.5}$	0.5182	0.4562	0.4511	0.4408	0.4525	0.4536	0.4478
Vicio médio total	Média	0.0011	-0.0053	-0.0116	-0.0038	-0.0074	-0.0074	-0.0017
	Desv.Padr.	0.1260	0.1308	0.1354	0.1349	0.1299	0.1298	0.1297
	$p_{2.5}$	-0.2549	-0.2641	-0.2751	-0.2735	-0.2690	-0.2684	-0.2597
	Mediana	0.0051	-0.0042	-0.0094	-0.0037	-0.0052	-0.0055	-0.0012
	$p_{97.5}$	0.2462	0.2488	0.2567	0.2645	0.2463	0.2465	0.2502
DW (5%)	Rejeita (%)	-	35.65	4.10	4.70	20.90	21.05	14.15
EQM ult.	Média	0.3373	0.2717	0.2689	0.2608	0.2718	0.2715	0.2676
	Desv.Padr.	0.4866	0.4009	0.3943	0.3844	0.4023	0.4019	0.3945
	$p_{2.5}$	0.0003	0.0003	0.0003	0.0002	0.0004	0.0004	0.0003
	Mediana	0.1465	0.1234	0.1205	0.1142	0.1200	0.1216	0.1183
	$p_{97.5}$	1.7242	1.3585	1.3029	1.2933	1.3014	1.2855	1.2883
Vicio médio ult.	Média	0.0171	0.0063	-0.0033	0.0059	0.0101	0.0101	0.0167
	Desv.Padr.	0.5806	0.5213	0.5186	0.5108	0.5214	0.5211	0.5172
	$p_{2.5}$	-1.1024	-0.9644	-1.0041	-0.9685	-0.9943	-0.9932	-0.9473
	Mediana	-0.0093	-0.0074	-0.0257	-0.0098	-0.0186	-0.0182	-0.0007
	$p_{97.5}$	1.1834	1.0518	1.0368	0.9984	1.0416	1.0431	1.0335
$V(\hat{\theta}_t)$ ult.	Média	0.3211	0.2112	0.2476	0.2560	0.2553	0.2554	0.2636
	Desv.Padr.	0.0584	0.0355	0.0436	0.0381	0.0448	0.0450	0.0399
	$p_{2.5}$	0.2039	0.1401	0.1604	0.1759	0.1636	0.1628	0.1796
	Mediana	0.3244	0.2117	0.2490	0.2592	0.2566	0.2568	0.2669
	$p_{97.5}$	0.4276	0.2807	0.3299	0.3224	0.3384	0.3411	0.3330
Prob. Cober. ult.	IC(95%)	94.90	91.40	94.20	95.40	94.60	94.65	95.30

Filtros de Kalman com erros ARMA e GLS

Tabela D.13: Resumo das estimativas para $\{\theta_t\}$. Cenário 1 com $\phi = 0.4$ e $T = 120$. 2000 replicações.

Modelos considerados: (1) Estimativa de desenho; Filtro de Kalman: (2) com e_t independentes, (3) com e_t $AR(1)$, (4) com e_t $AR(1)$ com hiperparâmetros conhecidos; Filtro GLS, estrutura de autocorrelação: (5) estimativas iniciais (6), estimativas iniciais ajustadas $AR(1)$, (7) hiperparâmetros conhecidos.

$p_{2.5}$ e $p_{97.5}$: percentis de 2.5% e 97.5% das estimativas.

Estimador		(1)	(2)	(3)	(4)	(5)	(6)	(7)
EQM total	Média	0.3241	0.2644	0.2593	0.2566	0.2643	0.2643	0.2643
	Desv.Padr.	0.0647	0.0563	0.0556	0.0547	0.0561	0.0561	0.0556
	$p_{2.5}$	0.2104	0.1680	0.1635	0.1647	0.1703	0.1701	0.1703
	Mediana	0.3194	0.2599	0.2555	0.2519	0.2599	0.2598	0.2593
	$p_{97.5}$	0.4608	0.3877	0.3835	0.3778	0.3880	0.3880	0.3883
Vicio médio total	Média	-0.0007	-0.0098	-0.0171	-0.0079	-0.0115	-0.0115	-0.0051
	Desv.Padr.	0.0801	0.0844	0.0863	0.0859	0.0830	0.0830	0.0828
	$p_{2.5}$	-0.1563	-0.1765	-0.1871	-0.1792	-0.1729	-0.1729	-0.1665
	Mediana	-0.0009	-0.0113	-0.0177	-0.0093	-0.0127	-0.0129	-0.0073
	$p_{97.5}$	0.1526	0.1570	0.1518	0.1600	0.1515	0.1519	0.1588
DW (5%)	Rejeita (%)	-	69.05	5.60	4.05	38.35	38.35	25.50
EQM ult.	Média	0.3194	0.2478	0.2432	0.2375	0.2495	0.2495	0.2478
	Desv.Padr.	0.4748	0.3738	0.3604	0.3511	0.3748	0.3746	0.3693
	$p_{2.5}$	0.0002	0.0003	0.0002	0.0002	0.0002	0.0002	0.0002
	Mediana	0.1298	0.1031	0.0998	0.1009	0.1027	0.1027	0.1040
	$p_{97.5}$	1.6816	1.3090	1.2164	1.1866	1.2870	1.2612	1.2741
Vicio médio ult.	Média	0.0085	-0.0118	-0.0189	-0.0089	-0.0080	-0.0081	-0.0019
	Desv.Padr.	0.5652	0.4978	0.4929	0.4874	0.4995	0.4996	0.4979
	$p_{2.5}$	-1.1198	-0.9699	-0.9827	-0.9783	-0.9963	-0.9982	-0.9667
	Mediana	0.0099	-0.0160	-0.0125	-0.0086	-0.0046	-0.0047	0.0009
	$p_{97.5}$	1.1786	0.9873	0.9979	0.9984	1.0004	0.9975	1.0045
$V(\hat{\theta}_t)$ ult.	Média	0.3163	0.2008	0.2418	0.2467	0.2495	0.2496	0.2545
	Desv.Padr.	0.0655	0.0342	0.0433	0.0416	0.0451	0.0453	0.0437
	$p_{2.5}$	0.1788	0.1275	0.1488	0.1548	0.1533	0.1527	0.1582
	Mediana	0.3201	0.2028	0.2442	0.2501	0.2524	0.2526	0.2579
	$p_{97.5}$	0.4395	0.2637	0.3201	0.3190	0.3308	0.3308	0.3307
Prob. Cober. ult.	IC(95%)	94.20	91.55	94.45	95.15	94.50	94.55	95.15

Tabela D.14: Resumo das estimativas para $\{\theta_t\}$. Cenário 1 com $\phi = 0.4$ e $T = 240$. 2000 replicações.

Modelos considerados: (1) Estimativa de desenho; Filtro de Kalman: (2) com e_t independentes, (3) com e_t $AR(1)$, (4) com e_t $AR(1)$ com hiperparâmetros conhecidos; Filtro GLS, estrutura de autocorrelação: (5) estimativas iniciais (6), estimativas iniciais ajustadas $AR(1)$, (7) hiperparâmetros conhecidos.

$p_{2.5}$ e $p_{97.5}$: percentis de 2.5% e 97.5% das estimativas.

Estimador		(1)	(2)	(3)	(4)	(5)	(6)	(7)
EQM total	Média	0.3211	0.2573	0.2537	0.2496	0.2603	0.2603	0.2580
	Desv.Padr.	0.0506	0.0410	0.0509	0.0400	0.0498	0.0497	0.0409
	$p_{2.5}$	0.2250	0.1840	0.1790	0.1785	0.1841	0.1843	0.1839
	Mediana	0.3198	0.2556	0.2502	0.2475	0.2568	0.2567	0.2560
	$p_{97.5}$	0.4252	0.3424	0.3415	0.3313	0.3514	0.3513	0.3419
Vicio médio total	Média	0.0000	-0.0113	-0.0197	-0.0088	-0.0129	-0.0129	-0.0054
	Desv.Padr.	0.0528	0.0552	0.0559	0.0550	0.0543	0.0543	0.0538
	$p_{2.5}$	-0.1096	-0.1232	-0.1329	-0.1216	-0.1270	-0.1271	-0.1143
	Mediana	-0.0002	-0.0118	-0.0195	-0.0096	-0.0135	-0.0134	-0.0065
	$p_{97.5}$	0.1031	0.1006	0.0921	0.1003	0.0947	0.0948	0.1008
DW (5%)	Rejeita (%)	-	95.70	4.70	3.20	73.45	73.40	45.10
EQM ult.	Média	0.3162	0.2420	0.2436	0.2368	0.2484	0.2483	0.2440
	Desv.Padr.	0.4734	0.3530	0.3900	0.3494	0.3928	0.3940	0.3585
	$p_{2.5}$	0.0005	0.0002	0.0002	0.0002	0.0001	0.0002	0.0002
	Mediana	0.1379	0.1062	0.1078	0.1081	0.1078	0.1062	0.1066
	$p_{97.5}$	1.7397	1.3045	1.2460	1.2095	1.2721	1.2781	1.2729
Vicio médio ult.	Média	-0.0190	-0.0355	-0.0398	-0.0306	-0.0330	-0.0331	-0.0273
	Desv.Padr.	0.5622	0.4907	0.4920	0.4858	0.4974	0.4973	0.4934
	$p_{2.5}$	-1.1216	-1.0174	-1.0370	-0.9978	-1.0264	-1.0293	-1.0291
	Mediana	-0.0338	-0.0286	-0.0249	-0.0194	-0.0195	-0.0193	-0.0166
	$p_{97.5}$	1.1181	0.9545	0.9171	0.9227	0.9622	0.9597	0.9759
$V(\hat{\theta}_t)$ ult.	Média	0.3091	0.1919	0.2342	0.2391	0.2419	0.2421	0.2468
	Desv.Padr.	0.0766	0.0374	0.0498	0.0488	0.0520	0.0521	0.0512
	$p_{2.5}$	0.1473	0.1099	0.1261	0.1285	0.1287	0.1286	0.1311
	Mediana	0.3158	0.1957	0.2390	0.2458	0.2470	0.2471	0.2538
	$p_{97.5}$	0.4406	0.2537	0.3179	0.3154	0.3279	0.3295	0.3269
Prob. Cober. ult.	IC(95%)	94.00	91.35	94.15	95.00	94.45	94.45	95.40

Filtros de Kalman com erros ARMA e GLS

Tabela D.15: Resumo das estimativas para $\{\theta_t\}$. Cenário 1 com $\phi = 0.6$ e $T = 60$. 2000 replicações.

Modelos considerados: (1) Estimativa de desenho; Filtro de Kalman: (2) com e_t independentes, (3) com e_t $AR(1)$, (4) com e_t $AR(1)$ com hiperparâmetros conhecidos; Filtro GLS, estrutura de autocorrelação: (5) estimativas iniciais (6), estimativas iniciais ajustadas $AR(1)$, (7) hiperparâmetros conhecidos.

$p_{2.5}$ e $p_{97.5}$: percentis de 2.5% e 97.5% das estimativas.

Estimador		(1)	(2)	(3)	(4)	(5)	(6)	(7)
EQM total	Média	0.3242	0.3037	0.2888	0.2822	0.2958	0.2958	0.2945
	Desv.Padr.	0.1037	0.1046	0.1015	0.0982	0.1011	0.1011	0.0999
	$p_{2.5}$	0.1647	0.1455	0.1389	0.1368	0.1441	0.1440	0.1457
	Mediana	0.3094	0.2871	0.2723	0.2655	0.2809	0.2806	0.2795
	$p_{97.5}$	0.5714	0.5517	0.5329	0.5224	0.5309	0.5321	0.5230
Vicio médio total	Média	0.0009	-0.0044	-0.0183	-0.0089	-0.0079	-0.0079	-0.0032
	Desv.Padr.	0.1633	0.1676	0.1752	0.1721	0.1659	0.1658	0.1653
	$p_{2.5}$	-0.3379	-0.3392	-0.3631	-0.3510	-0.3445	-0.3452	-0.3357
	Mediana	0.0060	-0.0042	-0.0183	-0.0085	-0.0058	-0.0059	0.0004
	$p_{97.5}$	0.3172	0.3218	0.3336	0.3390	0.3087	0.3083	0.3142
DW (5%)	Rejeita (%)	-	65.85	3.30	4.35	27.25	27.10	16.90
EQM ult.	Média	0.3399	0.3108	0.2924	0.2855	0.3005	0.3005	0.2983
	Desv.Padr.	0.4875	0.4498	0.4224	0.4143	0.4373	0.4379	0.4315
	$p_{2.5}$	0.0003	0.0004	0.0003	0.0003	0.0003	0.0003	0.0003
	Mediana	0.1467	0.1410	0.1375	0.1296	0.1385	0.1389	0.1304
	$p_{97.5}$	1.7088	1.6195	1.4499	1.3866	1.4850	1.4887	1.4322
Vicio médio ult.	Média	0.0233	0.0145	-0.0025	0.0062	0.0183	0.0182	0.0223
	Desv.Padr.	0.5827	0.5574	0.5408	0.5344	0.5480	0.5480	0.5459
	$p_{2.5}$	-1.0855	-1.0409	-1.0083	-0.9854	-1.0131	-1.0083	-0.9991
	Mediana	0.0050	0.0089	-0.0217	-0.0085	0.0040	0.0035	0.0190
	$p_{97.5}$	1.1809	1.1302	1.0828	1.0827	1.0703	1.0813	1.0694
$V(\hat{\theta}_t)$ ult.	Média	0.3213	0.2062	0.2665	0.2757	0.2812	0.2815	0.2897
	Desv.Padr.	0.0585	0.0349	0.0481	0.0430	0.0501	0.0504	0.0467
	$p_{2.5}$	0.2030	0.1373	0.1661	0.1860	0.1760	0.1767	0.1924
	Mediana	0.3246	0.2064	0.2682	0.2791	0.2838	0.2832	0.2933
	$p_{97.5}$	0.4289	0.2725	0.3569	0.3516	0.3742	0.3740	0.3719
Prob. Cober. ult.	IC(95%)	95.05	88.65	94.10	94.80	94.45	94.80	95.30

Tabela D.16: Resumo das estimativas para $\{\theta_t\}$. Cenário 1 com $\phi = 0.6$ e $T = 120$. 2000 replicações.

Modelos considerados: (1) Estimativa de desenho; Filtro de Kalman: (2) com e_t independentes, (3) com e_t $AR(1)$, (4) com e_t $AR(1)$ com hiperparâmetros conhecidos; Filtro GLS, estrutura de autocorrelação: (5) estimativas iniciais (6), estimativas iniciais ajustadas $AR(1)$, (7) hiperparâmetros conhecidos.

$p_{2.5}$ e $p_{97.5}$: percentis de 2.5% e 97.5% das estimativas.

Estimador		(1)	(2)	(3)	(4)	(5)	(6)	(7)
EQM total	Média	0.3236	0.2973	0.2811	0.2769	0.2913	0.2914	0.2908
	Desv.Padr.	0.0756	0.0743	0.0718	0.0697	0.0723	0.0723	0.0714
	$p_{2.5}$	0.1957	0.1771	0.1647	0.1647	0.1732	0.1737	0.1736
	Mediana	0.3162	0.2900	0.2730	0.2684	0.2838	0.2836	0.2834
	$p_{97.5}$	0.4939	0.4760	0.4493	0.4378	0.4647	0.4634	0.4580
Vicio médio total	Média	-0.0010	-0.0088	-0.0247	-0.0139	-0.0116	-0.0116	-0.0065
	Desv.Padr.	0.1045	0.1084	0.1112	0.1104	0.1063	0.1063	0.1061
	$p_{2.5}$	-0.2025	-0.2187	-0.2432	-0.2297	-0.2184	-0.2181	-0.2137
	Mediana	-0.0018	-0.0104	-0.0249	-0.0151	-0.0131	-0.0130	-0.0071
	$p_{97.5}$	0.2004	0.2012	0.1909	0.2005	0.1934	0.1927	0.1999
DW (5%)	Rejeita (%)	-	96.90	7.70	3.60	54.50	54.30	31.90
EQM ult.	Média	0.3164	0.2850	0.2661	0.2587	0.2792	0.2795	0.2770
	Desv.Padr.	0.4766	0.4317	0.3962	0.3854	0.4252	0.4250	0.4191
	$p_{2.5}$	0.0002	0.0006	0.0002	0.0003	0.0002	0.0003	0.0003
	Mediana	0.1278	0.1206	0.1111	0.1117	0.1212	0.1213	0.1196
	$p_{97.5}$	1.7762	1.4875	1.3948	1.3575	1.4594	1.4621	1.4195
Vicio médio ult.	Média	0.0119	-0.0085	-0.0226	-0.0109	-0.0026	-0.0029	0.0018
	Desv.Padr.	0.5625	0.5339	0.5155	0.5087	0.5286	0.5288	0.5265
	$p_{2.5}$	-1.0800	-1.0299	-1.0349	-1.0132	-1.0530	-1.0626	-1.0378
	Mediana	0.0160	-0.0119	-0.0152	-0.0097	0.0048	0.0039	0.0066
	$p_{97.5}$	1.1474	1.0229	1.0143	1.0044	1.0630	1.0575	1.0654
$V(\hat{\theta}_t)$ ult.	Média	0.3164	0.1963	0.2632	0.2679	0.2780	0.2784	0.2826
	Desv.Padr.	0.0655	0.0334	0.0491	0.0476	0.0527	0.0530	0.0519
	$p_{2.5}$	0.1794	0.1238	0.1570	0.1640	0.1649	0.1652	0.1698
	Mediana	0.3196	0.1985	0.2670	0.2714	0.2822	0.2828	0.2861
	$p_{97.5}$	0.4371	0.2572	0.3527	0.3510	0.3736	0.3754	0.3746
Prob. Cober. ult.	IC(95%)	94.45	90.00	94.55	94.95	94.30	94.45	95.15

Filtros de Kalman com erros ARMA e GLS

Tabela D.17: Resumo das estimativas para $\{\theta_t\}$. Cenário 1 com $\phi = 0.6$ e $T = 240$. 2000 replicações.

Modelos considerados: (1) Estimativa de desenho; Filtro de Kalman: (2) com e_t independentes, (3) com e_t $AR(1)$, (4) com e_t $AR(1)$ com hiperparâmetros conhecidos; Filtro GLS, estrutura de autocorrelação: (5) estimativas iniciais (6), estimativas iniciais ajustadas $AR(1)$, (7) hiperparâmetros conhecidos.

$p_{2.5}$ e $p_{97.5}$: percentis de 2.5% e 97.5% das estimativas.

Estimador		(1)	(2)	(3)	(4)	(5)	(6)	(7)
EQM total	Média	0.3210	0.2900	0.2806	0.2709	0.2935	0.2935	0.2862
	Desv.Padr.	0.0573	0.0529	0.0776	0.0502	0.0686	0.0684	0.0523
	$p_{2.5}$	0.2188	0.1980	0.1854	0.1839	0.1967	0.1966	0.1954
	Mediana	0.3191	0.2863	0.2736	0.2676	0.2891	0.2892	0.2843
	$p_{97.5}$	0.4403	0.4011	0.3982	0.3753	0.4112	0.4094	0.3975
Vicio médio total	Média	0.0001	-0.0100	-0.0295	-0.0150	-0.0128	-0.0129	-0.0063
	Desv.Padr.	0.0690	0.0713	0.0733	0.0713	0.0705	0.0705	0.0697
	$p_{2.5}$	-0.1456	-0.1575	-0.1793	-0.1602	-0.1627	-0.1630	-0.1520
	Mediana	-0.0008	-0.0116	-0.0303	-0.0167	-0.0141	-0.0143	-0.0081
	$p_{97.5}$	0.1374	0.1321	0.1151	0.1268	0.1281	0.1282	0.1340
DW (5%)	Rejeita (%)	-	99.95	4.05	3.10	74.00	73.25	58.70
EQM ult.	Média	0.3137	0.2779	0.2746	0.2603	0.2828	0.2828	0.2744
	Desv.Padr.	0.4642	0.3999	0.4326	0.3776	0.4351	0.4360	0.3983
	$p_{2.5}$	0.0002	0.0002	0.0004	0.0002	0.0002	0.0002	0.0002
	Mediana	0.1352	0.1257	0.1235	0.1213	0.1203	0.1208	0.1180
	$p_{97.5}$	1.6382	1.5198	1.4725	1.3960	1.4809	1.4846	1.4385
Vicio médio ult.	Média	-0.0196	-0.0338	-0.0497	-0.0353	-0.0332	-0.0333	-0.0277
	Desv.Padr.	0.5599	0.5262	0.5218	0.5091	0.5309	0.5309	0.5232
	$p_{2.5}$	-1.1055	-1.0736	-1.0856	-1.0299	-1.0878	-1.0871	-1.0619
	Mediana	-0.0229	-0.0230	-0.0452	-0.0290	-0.0289	-0.0260	-0.0224
	$p_{97.5}$	1.1003	1.0076	0.9477	0.9728	1.0241	1.0244	1.0402
$V(\hat{\theta}_t)$ ult.	Média	0.3090	0.1876	0.2596	0.2608	0.2734	0.2740	0.2755
	Desv.Padr.	0.0766	0.0363	0.0597	0.0560	0.0637	0.0639	0.0612
	$p_{2.5}$	0.1465	0.1082	0.1327	0.1356	0.1381	0.1382	0.1398
	Mediana	0.3148	0.1917	0.2648	0.2682	0.2803	0.2804	0.2832
	$p_{97.5}$	0.4435	0.2472	0.3609	0.3497	0.3814	0.3829	0.3728
Prob. Cober. ult.	IC(95%)	94.40	88.35	94.35	95.00	94.15	94.15	94.75

Tabela D.18: Resumo das estimativas para $\{\theta_t\}$. Cenário 1 com $\phi = 0.8$ e $T = 60$. 2000 replicações.

Modelos considerados: (1) Estimativa de desenho; Filtro de Kalman: (2) com e_t independentes, (3) com e_t $AR(1)$, (4) com e_t $AR(1)$ com hiperparâmetros conhecidos; Filtro GLS, estrutura de autocorrelação: (5) estimativas iniciais (6), estimativas iniciais ajustadas $AR(1)$, (7) hiperparâmetros conhecidos.

$p_{2.5}$ e $p_{97.5}$: percentis de 2.5% e 97.5% das estimativas.

Estimador		(1)	(2)	(3)	(4)	(5)	(6)	(7)
EQM total	Média	0.3226	0.3498	0.3119	0.3040	0.3163	0.3165	0.3149
	Desv.Padr.	0.1445	0.1553	0.1498	0.1437	0.1449	0.1449	0.1444
	$p_{2.5}$	0.1255	0.1361	0.1148	0.1142	0.1208	0.1217	0.1197
	Mediana	0.2943	0.3200	0.2797	0.2732	0.2865	0.2870	0.2859
	$p_{97.5}$	0.6750	0.7255	0.6813	0.6608	0.6746	0.6741	0.6609
Vicio médio total	Média	-0.0005	-0.0043	-0.0377	-0.0267	-0.0096	-0.0096	-0.0064
	Desv.Padr.	0.2382	0.2420	0.2502	0.2462	0.2396	0.2395	0.2392
	$p_{2.5}$	-0.4926	-0.4966	-0.5310	-0.5200	-0.5014	-0.5014	-0.5035
	Mediana	0.0005	-0.0025	-0.0383	-0.0267	-0.0085	-0.0080	-0.0060
	$p_{97.5}$	0.4600	0.4618	0.4545	0.4542	0.4583	0.4576	0.4586
DW (5%)	Rejeita (%)	-	88.75	2.60	3.60	16.90	18.95	12.95
EQM ult.	Média	0.3443	0.3723	0.3230	0.3163	0.3331	0.3331	0.3310
	Desv.Padr.	0.4914	0.5255	0.4611	0.4491	0.4733	0.4743	0.4690
	$p_{2.5}$	0.0003	0.0003	0.0004	0.0004	0.0002	0.0002	0.0003
	Mediana	0.1568	0.1714	0.1526	0.1446	0.1535	0.1498	0.1524
	$p_{97.5}$	1.7441	1.8061	1.6697	1.5769	1.7010	1.6921	1.6749
Vicio médio ult.	Média	0.0273	0.0207	-0.0195	-0.0087	0.0206	0.0207	0.0227
	Desv.Padr.	0.5863	0.6100	0.5682	0.5625	0.5769	0.5769	0.5750
	$p_{2.5}$	-1.0584	-1.1067	-1.1093	-1.0800	-1.0801	-1.0824	-1.0427
	Mediana	0.0270	0.0173	-0.0317	-0.0231	0.0159	0.0149	0.0147
	$p_{97.5}$	1.2391	1.2691	1.1296	1.1272	1.1975	1.2017	1.2063
$V(\hat{\theta}_t)$ ult.	Média	0.3213	0.1981	0.2913	0.2979	0.3086	0.3079	0.3125
	Desv.Padr.	0.0585	0.0337	0.0526	0.0493	0.0550	0.0551	0.0537
	$p_{2.5}$	0.2022	0.1306	0.1821	0.1966	0.1956	0.1949	0.2026
	Mediana	0.3245	0.1986	0.2935	0.3012	0.3111	0.3102	0.3160
	$p_{97.5}$	0.4280	0.2622	0.3899	0.3869	0.4089	0.4090	0.4087
Prob. Cober. ult.	IC(95%)	94.10	84.45	93.30	94.25	93.70	93.70	93.80

Filtros de Kalman com erros ARMA e GLS

Tabela D.19: Resumo das estimativas para $\{\theta_t\}$. Cenário 1 com $\phi = 0.8$ e $T = 120$. 2000 replicações.

Modelos considerados: (1) Estimativa de desenho; Filtro de Kalman: (2) com e_t independentes, (3) com e_t $AR(1)$, (4) com e_t $AR(1)$ com hiperparâmetros conhecidos; Filtro GLS, estrutura de autocorrelação: (5) estimativas iniciais (6), estimativas iniciais ajustadas $AR(1)$, (7) hiperparâmetros conhecidos.

$p_{2.5}$ e $p_{97.5}$: percentis de 2.5% e 97.5% das estimativas.

Estimador		(1)	(2)	(3)	(4)	(5)	(6)	(7)
EQM total	Média	0.3248	0.3498	0.3072	0.3025	0.3166	0.3167	0.3158
	Desv.Padr.	0.1037	0.1119	0.1043	0.1019	0.1037	0.1037	0.1033
	$p_{2.5}$	0.1645	0.1802	0.1526	0.1513	0.1565	0.1570	0.1558
	Mediana	0.3093	0.3330	0.2890	0.2865	0.3014	0.3019	0.3004
	$p_{97.5}$	0.5749	0.6180	0.5603	0.5490	0.5624	0.5635	0.5614
Vicio médio total	Média	-0.0021	-0.0081	-0.0471	-0.0341	-0.0123	-0.0123	-0.0091
	Desv.Padr.	0.1552	0.1586	0.1619	0.1610	0.1560	0.1560	0.1560
	$p_{2.5}$	-0.3106	-0.3162	-0.3633	-0.3449	-0.3202	-0.3200	-0.3167
	Mediana	0.0008	-0.0114	-0.0490	-0.0361	-0.0130	-0.0131	-0.0095
	$p_{97.5}$	0.2975	0.3042	0.2654	0.2699	0.2891	0.2892	0.2909
DW (5%)	Rejeita (%)	-	99.90	8.95	3.35	38.45	39.55	23.05
EQM ult.	Média	0.3160	0.3381	0.2915	0.2866	0.3054	0.3056	0.3044
	Desv.Padr.	0.4722	0.5017	0.4304	0.4252	0.4577	0.4578	0.4568
	$p_{2.5}$	0.0002	0.0003	0.0005	0.0003	0.0003	0.0003	0.0002
	Mediana	0.1383	0.1426	0.1261	0.1220	0.1304	0.1325	0.1314
	$p_{97.5}$	1.6268	1.6894	1.5008	1.4818	1.5535	1.5572	1.5868
Vicio médio ult.	Média	0.0127	-0.0059	-0.0449	-0.0301	-0.0006	-0.0008	0.0023
	Desv.Padr.	0.5621	0.5815	0.5382	0.5346	0.5528	0.5529	0.5519
	$p_{2.5}$	-1.1180	-1.1588	-1.1166	-1.1077	-1.1095	-1.1067	-1.1053
	Mediana	0.0082	-0.0099	-0.0511	-0.0353	0.0012	-0.0029	-0.0045
	$p_{97.5}$	1.1169	1.1855	1.0398	1.0383	1.0780	1.0728	1.0689
$V(\hat{\theta}_t)$ ult.	Média	0.3164	0.1886	0.2888	0.2913	0.3055	0.3053	0.3071
	Desv.Padr.	0.0656	0.0321	0.0564	0.0549	0.0612	0.0611	0.0602
	$p_{2.5}$	0.1781	0.1185	0.1689	0.1731	0.1747	0.1749	0.1788
	Mediana	0.3208	0.1907	0.2921	0.2948	0.3099	0.3097	0.3104
	$p_{97.5}$	0.4352	0.2459	0.3913	0.3885	0.4160	0.4163	0.4156
Prob. Cober. ult.	IC(95%)	94.70	85.20	94.45	94.60	94.70	94.80	94.85

Tabela D.20: Resumo das estimativas para $\{\theta_t\}$. Cenário 1 com $\phi = 0.8$ e $T = 240$. 2000 replicações.

Modelos considerados: (1) Estimativa de desenho; Filtro de Kalman: (2) com e_t independentes, (3) com e_t $AR(1)$, (4) com e_t $AR(1)$ com hiperparâmetros conhecidos; Filtro GLS, estrutura de autocorrelação: (5) estimativas iniciais (6), estimativas iniciais ajustadas $AR(1)$, (7) hiperparâmetros conhecidos.

$p_{2.5}$ e $p_{97.5}$: percentis de 2.5% e 97.5% das estimativas.

Estimador		(1)	(2)	(3)	(4)	(5)	(6)	(7)
EQM total	Média	0.3222	0.3401	0.3399	0.2972	0.3313	0.3312	0.3124
	Desv.Padr.	0.0761	0.0795	0.2449	0.0721	0.1294	0.1288	0.0748
	$p_{2.5}$	0.1968	0.2095	0.1856	0.1765	0.1920	0.1918	0.1883
	Mediana	0.3163	0.3325	0.3019	0.2897	0.3118	0.3116	0.3074
	$p_{97.5}$	0.4865	0.5176	0.8466	0.4574	0.6299	0.6299	0.4772
Vicio médio total	Média	0.0000	-0.0082	-0.0604	-0.0365	-0.0142	-0.0142	-0.0079
	Desv.Padr.	0.1031	0.1053	0.1306	0.1055	0.1067	0.1067	0.1034
	$p_{2.5}$	-0.2157	-0.2245	-0.3053	-0.2530	-0.2404	-0.2400	-0.2267
	Mediana	-0.0003	-0.0084	-0.0564	-0.0377	-0.0142	-0.0138	-0.0088
	$p_{97.5}$	0.2034	0.2059	0.1664	0.1728	0.1932	0.1935	0.1968
DW (5%)	Rejeita (%)	-	100.00	5.95	3.10	44.35	43.25	42.35
EQM ult.	Média	0.3157	0.3419	0.3531	0.2913	0.3440	0.3443	0.3067
	Desv.Padr.	0.4631	0.5044	0.7386	0.4174	0.6062	0.6071	0.4475
	$p_{2.5}$	0.0003	0.0003	0.0002	0.0003	0.0004	0.0003	0.0003
	Mediana	0.1406	0.1499	0.1363	0.1305	0.1383	0.1379	0.1334
	$p_{97.5}$	1.5784	1.6775	1.9298	1.4425	1.8519	1.8674	1.5546
Vicio médio ult.	Média	-0.0128	-0.0242	-0.0825	-0.0499	-0.0345	-0.0344	-0.0216
	Desv.Padr.	0.5618	0.5844	0.5886	0.5376	0.5857	0.5859	0.5535
	$p_{2.5}$	-1.0813	-1.1522	-1.1991	-1.0905	-1.1634	-1.1612	-1.0783
	Mediana	-0.0183	-0.0351	-0.0728	-0.0486	-0.0265	-0.0251	-0.0252
	$p_{97.5}$	1.0933	1.1279	1.0302	0.9964	1.1339	1.1282	1.0852
$V(\hat{\theta}_t)$ ult.	Média	0.3092	0.1803	0.2835	0.2845	0.2963	0.2963	0.3004
	Desv.Padr.	0.0766	0.0346	0.0796	0.0649	0.0797	0.0796	0.0711
	$p_{2.5}$	0.1477	0.1054	0.0671	0.1420	0.1104	0.1109	0.1459
	Mediana	0.3150	0.1836	0.2934	0.2928	0.3053	0.3053	0.3088
	$p_{97.5}$	0.4394	0.2381	0.4110	0.3891	0.4265	0.4256	0.4165
Prob. Cober. ult.	IC(95%)	94.45	84.70	92.00	94.65	92.75	92.70	94.70

Filtros de Kalman com erros ARMA e GLS

Tabela D.21: Resumo das estimativas para $\{\theta_t\}$. Cenário 2 com $\phi = 0.4$ e $T = 60$. 2000 replicações.

Modelos considerados: (1) Estimativa de desenho; Filtro de Kalman: (2) com e_t independentes, (3) com e_t $AR(1)$, (4) com e_t $AR(12)$ com hiperparâmetros estimados, (5) com e_t $AR(12)$ com hiperparâmetros conhecidos; Filtro GLS, estrutura de autocorrelação: (6) estimativas iniciais (7), estimativas iniciais ajustadas $AR(12)$, (8) hiperparâmetros conhecidos.

$p_{2.5}$ e $p_{97.5}$: percentis de 2.5% e 97.5% das estimativas.

Estimador		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
EQM total	Média	0.3260	0.2614	0.2601	0.2635	0.2560	0.2634	0.2634	0.2601
	Desv.Padr.	0.0835	0.0717	0.0711	0.0727	0.0694	0.0722	0.0721	0.0700
	$p_{2.5}$	0.1925	0.1486	0.1482	0.1482	0.1438	0.1510	0.1512	0.1458
	Mediana	0.3176	0.2524	0.2505	0.2533	0.2463	0.2535	0.2538	0.2504
	$p_{97.5}$	0.5161	0.4269	0.4185	0.4289	0.4158	0.4259	0.4271	0.4209
Vicio médio total	Média	0.0002	-0.0066	-0.0107	-0.0133	-0.0038	-0.0083	-0.0083	-0.0018
	Desv.Padr.	0.1157	0.1215	0.1248	0.1282	0.1268	0.1220	0.1219	0.1214
	$p_{2.5}$	-0.2236	-0.2407	-0.2500	-0.2537	-0.2488	-0.2395	-0.2390	-0.2339
	Mediana	0.0007	-0.0070	-0.0110	-0.0149	-0.0048	-0.0085	-0.0083	-0.0024
	$p_{97.5}$	0.2218	0.2276	0.2267	0.2338	0.2405	0.2250	0.2255	0.2385
DW (5%)	Rejeita (%)	-	24.65	5.55	7.25	5.65	20.45	20.25	11.65
EQM ult.	Média	0.3229	0.2709	0.2665	0.2706	0.2555	0.2712	0.2709	0.2597
	Desv.Padr.	0.4504	0.3897	0.3830	0.3840	0.3615	0.3872	0.3868	0.3682
	$p_{2.5}$	0.0004	0.0003	0.0003	0.0002	0.0002	0.0002	0.0003	0.0003
	Mediana	0.1457	0.1218	0.1172	0.1187	0.1146	0.1235	0.1224	0.1163
	$p_{97.5}$	1.6169	1.4218	1.3862	1.3846	1.2720	1.3952	1.4014	1.3221
Vicio médio ult.	Média	-0.0143	-0.0240	-0.0324	-0.0350	-0.0258	-0.0212	-0.0214	-0.0157
	Desv.Padr.	0.5682	0.5200	0.5154	0.5191	0.5050	0.5204	0.5202	0.5095
	$p_{2.5}$	-1.1397	-1.0538	-1.0765	-1.0782	-1.0490	-1.0831	-1.0881	-1.0452
	Mediana	-0.0085	-0.0204	-0.0229	-0.0175	-0.0162	-0.0125	-0.0121	0.0020
	$p_{97.5}$	1.0579	0.9591	0.9440	0.9914	0.9328	0.9763	0.9720	0.9465
$V(\hat{\theta}_t)$ ult.	Média	0.3204	0.2128	0.2391	0.2526	0.2650	0.2400	0.2407	0.2513
	Desv.Padr.	0.0578	0.0359	0.0415	0.0462	0.0396	0.0431	0.0431	0.0369
	$p_{2.5}$	0.2001	0.1355	0.1514	0.1559	0.1835	0.1494	0.1500	0.1732
	Mediana	0.3232	0.2137	0.2406	0.2534	0.2659	0.2417	0.2422	0.2545
	$p_{97.5}$	0.4288	0.2805	0.3166	0.3388	0.3372	0.3188	0.3194	0.3155
Prob. Cober. ult.	IC(95%)	94.50	91.25	93.40	93.70	95.20	92.90	93.10	94.15

Tabela D.22: Resumo das estimativas para $\{\theta_t\}$. Cenário 2 com $\phi = 0.4$ e $T = 120$. 2000 replicações.

Modelos considerados: (1) Estimativa de desenho; Filtro de Kalman: (2) com e_t independentes, (3) com e_t $AR(1)$, (4) com e_t $AR(12)$ com hiperparâmetros estimados, (5) com e_t $AR(12)$ com hiperparâmetros conhecidos; Filtro GLS, estrutura de autocorrelação: (6) estimativas iniciais (7), estimativas iniciais ajustadas $AR(12)$, (8) hiperparâmetros conhecidos.

$p_{2.5}$ e $p_{97.5}$: percentis de 2.5% e 97.5% das estimativas.

Estimador		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
EQM total	Média	0.3241	0.2516	0.2489	0.2499	0.2469	0.2523	0.2523	0.2518
	Desv.Padr.	0.0619	0.0511	0.0506	0.0510	0.0500	0.0513	0.0513	0.0504
	$p_{2.5}$	0.2173	0.1658	0.1651	0.1659	0.1648	0.1664	0.1665	0.1683
	Mediana	0.3189	0.2468	0.2434	0.2451	0.2417	0.2474	0.2475	0.2469
	$p_{97.5}$	0.4586	0.3661	0.3610	0.3609	0.3588	0.3708	0.3701	0.3661
Vicio médio total	Média	0.0005	-0.0091	-0.0136	-0.0166	-0.0063	-0.0107	-0.0107	-0.0036
	Desv.Padr.	0.0823	0.0856	0.0868	0.0877	0.0872	0.0850	0.0850	0.0849
	$p_{2.5}$	-0.1584	-0.1669	-0.1761	-0.1805	-0.1713	-0.1688	-0.1688	-0.1613
	Mediana	-0.0012	-0.0101	-0.0158	-0.0202	-0.0067	-0.0120	-0.0120	-0.0040
	$p_{97.5}$	0.1633	0.1573	0.1566	0.1547	0.1659	0.1562	0.1562	0.1630
DW (5%)	Rejeita (%)	-	48.05	5.60	7.10	5.75	32.60	31.85	21.05
EQM ult.	Média	0.3226	0.2380	0.2348	0.2350	0.2318	0.2387	0.2386	0.2385
	Desv.Padr.	0.4797	0.3448	0.3429	0.3424	0.3350	0.3476	0.3473	0.3443
	$p_{2.5}$	0.0003	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
	Mediana	0.1376	0.1032	0.0995	0.1006	0.0994	0.1000	0.0993	0.1045
	$p_{97.5}$	1.8028	1.2077	1.2635	1.2534	1.2129	1.2117	1.2099	1.1934
Vicio médio ult.	Média	-0.0057	-0.0249	-0.0307	-0.0349	-0.0238	-0.0238	-0.0239	-0.0168
	Desv.Padr.	0.5681	0.4874	0.4837	0.4837	0.4810	0.4881	0.4880	0.4882
	$p_{2.5}$	-1.1322	-1.0220	-1.0292	-1.0254	-0.9991	-1.0240	-1.0197	-1.0005
	Mediana	-0.0160	-0.0236	-0.0345	-0.0292	-0.0288	-0.0196	-0.0193	-0.0162
	$p_{97.5}$	1.1002	0.9425	0.8879	0.9088	0.9152	0.9250	0.9254	0.9267
$V(\hat{\theta}_t)$ ult.	Média	0.3160	0.2028	0.2328	0.2450	0.2518	0.2349	0.2353	0.2411
	Desv.Padr.	0.0658	0.0349	0.0415	0.0436	0.0416	0.0420	0.0421	0.0401
	$p_{2.5}$	0.1772	0.1307	0.1463	0.1537	0.1615	0.1476	0.1479	0.1522
	Mediana	0.3189	0.2046	0.2352	0.2475	0.2554	0.2379	0.2379	0.2443
	$p_{97.5}$	0.4369	0.2666	0.3083	0.3220	0.3243	0.3085	0.3089	0.3107
Prob. Cober. ult.	IC(95%)	94.50	92.20	94.70	95.35	95.95	94.20	94.25	95.05

Filtros de Kalman com erros ARMA e GLS

Tabela D.23: Resumo das estimativas para $\{\theta_t\}$. Cenário 2 com $\phi = 0.4$ e $T = 240$. 2000 replicações.

Modelos considerados: (1) Estimativa de desenho; Filtro de Kalman: (2) com e_t independentes, (3) com e_t $AR(1)$, (4) com e_t $AR(12)$ com hiperparâmetros estimados, (5) com e_t $AR(12)$ com hiperparâmetros conhecidos; Filtro GLS, estrutura de autocorrelação: (6) estimativas iniciais (7), estimativas iniciais ajustadas $AR(12)$, (8) hiperparâmetros conhecidos.

$p_{2.5}$ e $p_{97.5}$: percentis de 2.5% e 97.5% das estimativas.

Estimador		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
EQM total	Média	0.3217	0.2441	0.2419	0.2429	0.2397	0.2462	0.2462	0.2450
	Desv.Padr.	0.0479	0.0364	0.0418	0.0374	0.0358	0.0379	0.0379	0.0363
	$p_{2.5}$	0.2333	0.1786	0.1763	0.1771	0.1759	0.1785	0.1784	0.1798
	Mediana	0.3207	0.2427	0.2395	0.2406	0.2377	0.2441	0.2444	0.2437
	$p_{97.5}$	0.4219	0.3216	0.3236	0.3256	0.3174	0.3264	0.3263	0.3227
Vicio médio total	Média	0.0003	-0.0122	-0.0176	-0.0209	-0.0087	-0.0138	-0.0138	-0.0054
	Desv.Padr.	0.0606	0.0625	0.0631	0.0633	0.0623	0.0620	0.0620	0.0615
	$p_{2.5}$	-0.1175	-0.1339	-0.1410	-0.1443	-0.1307	-0.1348	-0.1349	-0.1249
	Mediana	0.0008	-0.0124	-0.0171	-0.0209	-0.0088	-0.0140	-0.0138	-0.0056
	$p_{97.5}$	0.1179	0.1089	0.1044	0.1012	0.1119	0.1075	0.1075	0.1150
DW (5%)	Rejeita (%)	-	78.15	5.80	7.20	5.55	56.45	56.10	34.80
EQM ult.	Média	0.3072	0.2319	0.2291	0.2307	0.2280	0.2321	0.2321	0.2318
	Desv.Padr.	0.4641	0.3423	0.3405	0.3437	0.3413	0.3415	0.3415	0.3418
	$p_{2.5}$	0.0002	0.0002	0.0003	0.0001	0.0002	0.0002	0.0002	0.0003
	Mediana	0.1324	0.1047	0.1011	0.1027	0.1035	0.1036	0.1033	0.1034
	$p_{97.5}$	1.5524	1.1365	1.1384	1.1487	1.1223	1.1609	1.1592	1.1443
Vicio médio ult.	Média	-0.0131	-0.0272	-0.0307	-0.0329	-0.0229	-0.0245	-0.0245	-0.0185
	Desv.Padr.	0.5543	0.4809	0.4778	0.4793	0.4771	0.4813	0.4813	0.4812
	$p_{2.5}$	-1.0996	-0.9954	-1.0014	-1.0003	-0.9970	-0.9892	-0.9927	-0.9823
	Mediana	-0.0026	-0.0216	-0.0251	-0.0234	-0.0196	-0.0203	-0.0207	-0.0181
	$p_{97.5}$	1.0583	0.9082	0.9178	0.9176	0.9275	0.9096	0.9081	0.9101
$V(\hat{\theta}_t)$ ult.	Média	0.3093	0.1937	0.2247	0.2357	0.2424	0.2270	0.2273	0.2330
	Desv.Padr.	0.0762	0.0378	0.0466	0.0486	0.0484	0.0472	0.0473	0.0469
	$p_{2.5}$	0.1455	0.1119	0.1232	0.1322	0.1318	0.1249	0.1249	0.1264
	Mediana	0.3150	0.1986	0.2301	0.2413	0.2497	0.2323	0.2326	0.2397
	$p_{97.5}$	0.4361	0.2550	0.3017	0.3156	0.3198	0.3047	0.3049	0.3058
Prob. Cober. ult.	IC(95%)	94.65	92.75	94.80	95.05	95.80	94.65	94.70	95.15

Tabela D.24: Resumo das estimativas para $\{\theta_t\}$. Cenário 2 com $\phi = 0.6$ e $T = 60$. 2000 replicações.

Modelos considerados: (1) Estimativa de desenho; Filtro de Kalman: (2) com e_t independentes, (3) com e_t $AR(1)$, (4) com e_t $AR(12)$ com hiperparâmetros estimados, (5) com e_t $AR(12)$ com hiperparâmetros conhecidos; Filtro GLS, estrutura de autocorrelação: (6) estimativas iniciais (7), estimativas iniciais ajustadas $AR(12)$, (8) hiperparâmetros conhecidos.

$p_{2.5}$ e $p_{97.5}$: percentis de 2.5% e 97.5% das estimativas.

Estimador		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
EQM total	Média	0.3256	0.2798	0.2760	0.2793	0.2700	0.2798	0.2799	0.2781
	Desv.Padr.	0.0919	0.0861	0.0864	0.0903	0.0838	0.0860	0.0859	0.0837
	$p_{2.5}$	0.1712	0.1419	0.1418	0.1413	0.1393	0.1449	0.1444	0.1417
	Mediana	0.3166	0.2681	0.2654	0.2687	0.2602	0.2686	0.2689	0.2674
	$p_{97.5}$	0.5296	0.4697	0.4716	0.4755	0.4610	0.4741	0.4736	0.4644
Vicio médio total	Média	-0.0011	-0.0076	-0.0161	-0.0287	-0.0116	-0.0117	-0.0118	-0.0041
	Desv.Padr.	0.1715	0.1766	0.1824	0.1852	0.1815	0.1765	0.1764	0.1753
	$p_{2.5}$	-0.3384	-0.3615	-0.3846	-0.3936	-0.3717	-0.3615	-0.3617	-0.3515
	Mediana	0.0003	-0.0078	-0.0174	-0.0308	-0.0156	-0.0128	-0.0130	-0.0045
	$p_{97.5}$	0.3303	0.3349	0.3421	0.3355	0.3511	0.3314	0.3296	0.3393
DW (5%)	Rejeita (%)	-	45.40	4.85	6.80	6.20	30.85	29.80	16.00
EQM ult.	Média	0.3156	0.2780	0.2699	0.2758	0.2567	0.2772	0.2771	0.2680
	Desv.Padr.	0.4592	0.4067	0.3896	0.3992	0.3754	0.4018	0.4017	0.3931
	$p_{2.5}$	0.0004	0.0003	0.0002	0.0004	0.0003	0.0003	0.0003	0.0002
	Mediana	0.1443	0.1222	0.1213	0.1230	0.1133	0.1253	0.1245	0.1173
	$p_{97.5}$	1.5989	1.4061	1.3539	1.4458	1.3269	1.4068	1.4034	1.3989
Vicio médio ult.	Média	-0.0041	-0.0187	-0.0315	-0.0523	-0.0298	-0.0185	-0.0186	-0.0081
	Desv.Padr.	0.5619	0.5270	0.5187	0.5227	0.5059	0.5263	0.5262	0.5178
	$p_{2.5}$	-1.1089	-1.0410	-1.0185	-1.0743	-1.0028	-1.0324	-1.0343	-1.0174
	Mediana	-0.0084	-0.0361	-0.0423	-0.0600	-0.0376	-0.0289	-0.0275	-0.0141
	$p_{97.5}$	1.0882	1.0442	1.0032	0.9773	0.9518	1.0388	1.0394	1.0182
$V(\hat{\theta}_t)$ ult.	Média	0.3207	0.2091	0.2521	0.2838	0.3000	0.2579	0.2589	0.2721
	Desv.Padr.	0.0580	0.0350	0.0447	0.0511	0.0467	0.0453	0.0453	0.0420
	$p_{2.5}$	0.2023	0.1391	0.1603	0.1824	0.2034	0.1672	0.1675	0.1839
	Mediana	0.3249	0.2090	0.2534	0.2848	0.3022	0.2596	0.2608	0.2756
	$p_{97.5}$	0.4265	0.2760	0.3379	0.3820	0.3862	0.3437	0.3452	0.3456
Prob. Cober. ult.	IC(95%)	95.10	91.00	94.35	94.75	96.40	93.80	93.85	95.25

Filtros de Kalman com erros ARMA e GLS

Tabela D.25: Resumo das estimativas para $\{\theta_t\}$. Cenário 2 com $\phi = 0.6$ e $T = 120$. 2000 replicações.

Modelos considerados: (1) Estimativa de desenho; Filtro de Kalman: (2) com e_t independentes, (3) com e_t $AR(1)$, (4) com e_t $AR(12)$ com hiperparâmetros estimados, (5) com e_t $AR(12)$ com hiperparâmetros conhecidos; Filtro GLS, estrutura de autocorrelação: (6) estimativas iniciais (7), estimativas iniciais ajustadas $AR(12)$, (8) hiperparâmetros conhecidos.

$p_{2.5}$ e $p_{97.5}$: percentis de 2.5% e 97.5% das estimativas.

Estimador		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
EQM total	Média	0.3228	0.2717	0.2644	0.2663	0.2613	0.2711	0.2712	0.2712
	Desv.Padr.	0.0697	0.0635	0.0619	0.0629	0.0610	0.0629	0.0630	0.0622
	$p_{2.5}$	0.2079	0.1699	0.1646	0.1646	0.1619	0.1704	0.1704	0.1719
	Mediana	0.3146	0.2643	0.2560	0.2576	0.2530	0.2630	0.2633	0.2639
	$p_{97.5}$	0.4767	0.4160	0.4080	0.4128	0.4015	0.4145	0.4140	0.4140
Vicio médio total	Média	0.0002	-0.0085	-0.0176	-0.0327	-0.0148	-0.0116	-0.0116	-0.0045
	Desv.Padr.	0.1430	0.1455	0.1470	0.1485	0.1470	0.1448	0.1448	0.1444
	$p_{2.5}$	-0.2803	-0.2970	-0.3075	-0.3261	-0.3069	-0.2985	-0.2985	-0.2938
	Mediana	-0.0023	-0.0106	-0.0198	-0.0351	-0.0164	-0.0133	-0.0133	-0.0074
	$p_{97.5}$	0.2851	0.2840	0.2708	0.2602	0.2725	0.2767	0.2768	0.2831
DW (5%)	Rejeita (%)	-	83.90	7.50	7.35	5.45	52.85	51.55	28.95
EQM ult.	Média	0.3113	0.2591	0.2491	0.2535	0.2485	0.2581	0.2581	0.2581
	Desv.Padr.	0.4719	0.3793	0.3653	0.3686	0.3661	0.3822	0.3826	0.3846
	$p_{2.5}$	0.0001	0.0003	0.0002	0.0003	0.0002	0.0003	0.0003	0.0005
	Mediana	0.1422	0.1195	0.1138	0.1165	0.1134	0.1231	0.1234	0.1170
	$p_{97.5}$	1.5208	1.3171	1.2156	1.2209	1.2399	1.3198	1.3306	1.3104
Vicio médio ult.	Média	0.0178	-0.0117	-0.0163	-0.0397	-0.0193	-0.0077	-0.0077	0.0001
	Desv.Padr.	0.5578	0.5090	0.4990	0.5020	0.4983	0.5081	0.5081	0.5082
	$p_{2.5}$	-1.0468	-0.9881	-0.9853	-0.9926	-0.9687	-0.9777	-0.9767	-0.9804
	Mediana	0.0040	-0.0043	-0.0216	-0.0392	-0.0238	-0.0013	-0.0005	-0.0022
	$p_{97.5}$	1.1305	0.9992	0.9505	0.9256	0.9529	0.9695	0.9705	0.9933
$V(\hat{\theta}_t)$ ult.	Média	0.3166	0.1999	0.2488	0.2769	0.2873	0.2546	0.2553	0.2632
	Desv.Padr.	0.0654	0.0340	0.0456	0.0502	0.0493	0.0465	0.0467	0.0461
	$p_{2.5}$	0.1846	0.1314	0.1569	0.1753	0.1818	0.1594	0.1596	0.1618
	Mediana	0.3199	0.2023	0.2509	0.2782	0.2904	0.2569	0.2577	0.2667
	$p_{97.5}$	0.4391	0.2617	0.3336	0.3719	0.3762	0.3396	0.3406	0.3441
Prob. Cober. ult.	IC(95%)	95.15	92.10	95.25	96.05	96.70	95.15	95.25	95.50

Tabela D.26: Resumo das estimativas para $\{\theta_t\}$. Cenário 2 com $\phi = 0.6$ e $T = 240$. 2000 replicações.

Modelos considerados: (1) Estimativa de desenho; Filtro de Kalman: (2) com e_t independentes, (3) com e_t $AR(1)$, (4) com e_t $AR(12)$ com hiperparâmetros estimados, (5) com e_t $AR(12)$ com hiperparâmetros conhecidos; Filtro GLS, estrutura de autocorrelação: (6) estimativas iniciais (7), estimativas iniciais ajustadas $AR(12)$, (8) hiperparâmetros conhecidos.

$p_{2.5}$ e $p_{97.5}$: percentis de 2.5% e 97.5% das estimativas.

Estimador		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
EQM total	Média	0.3195	0.2649	0.2602	0.2685	0.2551	0.2742	0.2741	0.2651
	Desv.Padr.	0.0526	0.0455	0.0532	0.0770	0.0446	0.0701	0.0699	0.0453
	$p_{2.5}$	0.2292	0.1885	0.1809	0.1842	0.1815	0.1883	0.1884	0.1884
	Mediana	0.3161	0.2606	0.2552	0.2608	0.2505	0.2673	0.2674	0.2610
	$p_{97.5}$	0.4306	0.3659	0.3687	0.3897	0.3548	0.3946	0.3948	0.3637
Vicio médio total	Média	-0.0005	-0.0115	-0.0227	-0.0411	-0.0186	-0.0146	-0.0147	-0.0062
	Desv.Padr.	0.1274	0.1287	0.1291	0.1307	0.1287	0.1284	0.1283	0.1278
	$p_{2.5}$	-0.2445	-0.2607	-0.2736	-0.2991	-0.2674	-0.2633	-0.2632	-0.2515
	Mediana	-0.0030	-0.0143	-0.0259	-0.0441	-0.0212	-0.0176	-0.0177	-0.0096
	$p_{97.5}$	0.2560	0.2442	0.2334	0.2154	0.2365	0.2430	0.2427	0.2487
DW (5%)	Rejeita (%)	-	98.80	5.40	8.45	5.50	78.20	76.90	53.10
EQM ult.	Média	0.3021	0.2552	0.2474	0.2572	0.2452	0.2618	0.2620	0.2539
	Desv.Padr.	0.4278	0.3709	0.3655	0.3829	0.3584	0.3894	0.3896	0.3658
	$p_{2.5}$	0.0003	0.0005	0.0004	0.0003	0.0004	0.0002	0.0002	0.0002
	Mediana	0.1422	0.1179	0.1097	0.1167	0.1141	0.1177	0.1180	0.1184
	$p_{97.5}$	1.5181	1.3070	1.2238	1.3025	1.2732	1.2941	1.3043	1.2878
Vicio médio ult.	Média	-0.0025	-0.0246	-0.0319	-0.0539	-0.0306	-0.0215	-0.0216	-0.0150
	Desv.Padr.	0.5498	0.5047	0.4965	0.5044	0.4944	0.5114	0.5115	0.5038
	$p_{2.5}$	-1.0497	-0.9876	-0.9719	-1.0414	-0.9981	-1.0045	-1.0194	-0.9851
	Mediana	-0.0118	-0.0223	-0.0292	-0.0491	-0.0345	-0.0158	-0.0163	-0.0091
	$p_{97.5}$	1.0479	0.9430	0.9460	0.9513	0.9552	0.9478	0.9470	0.9378
$V(\hat{\theta}_t)$ ult.	Média	0.3095	0.1911	0.2420	0.2668	0.2775	0.2469	0.2475	0.2554
	Desv.Padr.	0.0765	0.0372	0.0527	0.0583	0.0575	0.0548	0.0550	0.0541
	$p_{2.5}$	0.1489	0.1100	0.1272	0.1418	0.1470	0.1268	0.1269	0.1339
	Mediana	0.3165	0.1947	0.2469	0.2733	0.2842	0.2521	0.2525	0.2627
	$p_{97.5}$	0.4367	0.2517	0.3316	0.3642	0.3703	0.3355	0.3368	0.3404
Prob. Cober. ult.	IC(95%)	95.95	91.30	94.85	95.25	96.50	94.65	94.65	95.55

Filtros de Kalman com erros ARMA e GLS

Tabela D.27: Resumo das estimativas para $\{\theta_t\}$. Cenário 2 com $\phi = 0.8$ e $T = 60$. 2000 replicações.

Modelos considerados: (1) Estimativa de desenho; Filtro de Kalman: (2) com e_t independentes, (3) com e_t $AR(1)$, (4) com e_t $AR(12)$ com hiperparâmetros estimados, (5) com e_t $AR(12)$ com hiperparâmetros conhecidos; Filtro GLS, estrutura de autocorrelação: (6) estimativas iniciais (7), estimativas iniciais ajustadas $AR(12)$, (8) hiperparâmetros conhecidos.

$p_{2.5}$ e $p_{97.5}$: percentis de 2.5% e 97.5% das estimativas.

Estimador		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
EQM total	Média	0.3150	0.3088	0.3014	0.3120	0.2877	0.3003	0.3002	0.2954
	Desv.Padr.	0.1449	0.1479	0.1522	0.1664	0.1470	0.1460	0.1459	0.1444
	$p_{2.5}$	0.1412	0.1297	0.1275	0.1215	0.1192	0.1287	0.1289	0.1256
	Mediana	0.2823	0.2737	0.2631	0.2715	0.2492	0.2667	0.2663	0.2602
	$p_{97.5}$	0.7006	0.6866	0.7037	0.7494	0.6824	0.6745	0.6751	0.6725
Vicio médio total	Média	-0.0017	-0.0068	-0.0430	-0.0927	-0.0400	-0.0190	-0.0190	-0.0067
	Desv.Padr.	0.3009	0.3052	0.3110	0.3139	0.3056	0.3042	0.3043	0.3029
	$p_{2.5}$	-0.5905	-0.5973	-0.6495	-0.7036	-0.6343	-0.6064	-0.6063	-0.5988
	Mediana	-0.0050	-0.0074	-0.0430	-0.0932	-0.0407	-0.0200	-0.0204	-0.0056
	$p_{97.5}$	0.5902	0.5834	0.5567	0.5181	0.5520	0.5745	0.5745	0.5828
DW (5%)	Rejeita (%)	-	76.18	2.85	7.46	6.31	33.53	33.28	13.81
EQM ult.	Média	0.2976	0.3091	0.2868	0.3171	0.2756	0.2871	0.2871	0.2793
	Desv.Padr.	0.4258	0.4208	0.4068	0.4517	0.3884	0.3943	0.3940	0.3898
	$p_{2.5}$	0.0003	0.0003	0.0002	0.0003	0.0003	0.0001	0.0002	0.0002
	Mediana	0.1348	0.1406	0.1296	0.1463	0.1258	0.1303	0.1316	0.1252
	$p_{97.5}$	1.5474	1.5128	1.4200	1.6276	1.373	1.4452	1.4436	1.4010
Vicio médio ult.	Média	-0.0013	-0.0088	-0.0549	-0.1502	-0.0683	-0.0154	-0.0157	-0.0024
	Desv.Padr.sd	0.5457	0.5560	0.5328	0.5428	0.5207	0.5358	0.5357	0.5286
	$p_{2.5}$	-1.0977	-1.1027	-1.1372	-1.2603	-1.1272	-1.0762	-1.0801	-1.0320
	Mediana	0.0171	-0.0087	-0.0406	-0.1288	-0.0649	-0.0055	-0.0061	-0.0060
	$p_{97.5}$	1.0574	1.0992	0.9428	0.8543	0.9108	0.9925	1.0043	0.9951
$V(\hat{\theta}_t)$ ult.	Média	0.3209	0.2019	0.2897	0.3325	0.3515	0.2849	0.2850	0.2997
	Desv.Padr.	0.0576	0.0337	0.0542	0.0588	0.0583	0.0492	0.0492	0.0496
	$p_{2.5}$	0.2059	0.1350	0.1809	0.2129	0.2311	0.1824	0.1846	0.1972
	Mediana	0.3250	0.2029	0.2902	0.3337	0.3543	0.2864	0.2870	0.3032
	$p_{97.5}$	0.4283	0.2662	0.3924	0.4438	0.4588	0.3801	0.3797	0.3879
Prob. Cober. ult.	IC(95%)	95.65	88.39	95.20	94.64	97.60	95.15	95.15	95.95

Tabela D.28: Resumo das estimativas para $\{\theta_t\}$. Cenário 2 com $\phi = 0.8$ e $T = 120$. 2000 replicações.

Modelos considerados: (1) Estimativa de desenho; Filtro de Kalman: (2) com e_t independentes, (3) com e_t $AR(1)$, (4) com e_t $AR(12)$ com hiperparâmetros estimados, (5) com e_t $AR(12)$ com hiperparâmetros conhecidos; Filtro GLS, estrutura de autocorrelação: (6) estimativas iniciais (7), estimativas iniciais ajustadas $AR(12)$, (8) hiperparâmetros conhecidos.

$p_{2.5}$ e $p_{97.5}$: percentis de 2.5% e 97.5% das estimativas.

Estimador		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
EQM total	Média	0.3153	0.3039	0.2968	0.3149	0.2872	0.2955	0.2945	0.2923
	Desv.Padr.	0.1288	0.1298	0.1333	0.1568	0.1314	0.1296	0.1281	0.1275
	$p_{2.5}$	0.1684	0.1563	0.1503	0.1507	0.1450	0.1519	0.1517	0.1518
	Mediana	0.2830	0.2698	0.2605	0.2708	0.2514	0.2623	0.2620	0.2606
	$p_{97.5}$	0.6474	0.6429	0.6479	0.7563	0.6456	0.6255	0.6216	0.6111
Vicio médio total	Média	0.0002	-0.0072	-0.0520	-0.1292	-0.0631	-0.0151	-0.0149	-0.0060
	Desv.Padr.	0.2847	0.2868	0.2891	0.2957	0.2867	0.2858	0.2858	0.2854
	$p_{2.5}$	-0.5403	-0.5601	-0.6133	-0.7269	-0.6266	-0.5650	-0.5652	-0.5526
	Mediana	-0.0023	-0.0096	-0.0572	-0.1329	-0.0726	-0.0203	-0.0200	-0.0110
	$p_{97.5}$	0.5723	0.5724	0.5225	0.4593	0.5035	0.5633	0.5631	0.5703
DW (5%)	Rejeita (%)	-	98.64	5.54	9.83	6.75	56.85	57.51	27.97
EQM ult.	Média	0.3068	0.2969	0.2893	0.3317	0.2860	0.2836	0.2828	0.2816
	Desv.Padr.	0.4248	0.4105	0.3993	0.4682	0.3942	0.3913	0.3895	0.3888
	$p_{2.5}$	0.0004	0.0003	0.0005	0.0003	0.0003	0.0003	0.0003	0.0003
	Mediana	0.1385	0.1394	0.1350	0.1538	0.1340	0.1327	0.1326	0.1231
	$p_{97.5}$	1.6102	1.4954	1.4834	1.7385	1.4168	1.4319	1.4218	1.4552
Vicio médio ult.	Média	-0.0022	-0.0270	-0.0695	-0.1992	-0.1055	-0.0271	-0.0261	-0.0153
	Desv.Padr.	0.5540	0.5444	0.5335	0.5406	0.5244	0.5320	0.5313	0.5305
	$p_{2.5}$	-1.1095	-1.1196	-1.1490	-1.2816	-1.1515	-1.0953	-1.0969	-1.0961
	Mediana	0.0009	-0.0158	-0.0753	-0.1990	-0.1108	-0.0259	-0.0260	-0.0150
	$p_{97.5}$	1.1026	1.0571	0.9793	0.8406	0.9107	1.0427	1.0371	1.0319
$V(\hat{\theta}_t)$ ult.	Média	0.3162	0.1926	0.2879	0.3265	0.3404	0.2822	0.2823	0.2927
	Desv.Padr.	0.0651	0.0328	0.0578	0.0614	0.0628	0.0535	0.0535	0.0551
	$p_{2.5}$	0.1808	0.1224	0.1684	0.1967	0.2077	0.1673	0.1674	0.1730
	Mediana	0.3195	0.1944	0.2895	0.3285	0.3441	0.2851	0.2849	0.2961
	$p_{97.5}$	0.4385	0.2528	0.3955	0.4414	0.4558	0.3792	0.3791	0.3912
Prob. Cober. ult.	IC(95%)	95.06	88.31	94.30	94.56	96.67	94.46	94.35	95.01

Filtros de Kalman com erros ARMA e GLS

Tabela D.29: Resumo das estimativas para $\{\theta_t\}$. Cenário 2 com $\phi = 0.8$ e $T = 240$. 2000 replicações.

Modelos considerados: (1) Estimativa de desenho; Filtro de Kalman: (2) com e_t independentes, (3) com e_t $AR(1)$, (4) com e_t $AR(12)$ com hiperparâmetros estimados, (5) com e_t $AR(12)$ com hiperparâmetros conhecidos; Filtro GLS, estrutura de autocorrelação: (6) estimativas iniciais (7), estimativas iniciais ajustadas $AR(12)$, (8) hiperparâmetros conhecidos.

$p_{2.5}$ e $p_{97.5}$: percentis de 2.5% e 97.5% das estimativas.

Estimador		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
EQM total	Média	0.3095	0.2951	0.3018	0.3484	0.2836	0.3079	0.3011	0.2848
	Desv.Padr.	0.1134	0.1118	0.1635	0.2473	0.1185	0.2129	0.1316	0.1114
	$p_{2.5}$	0.1794	0.1701	0.1604	0.1653	0.1554	0.1648	0.1650	0.1606
	Mediana	0.2805	0.2643	0.2594	0.2829	0.2489	0.2648	0.2642	0.2550
	$p_{97.5}$	0.6050	0.5822	0.6575	0.9439	0.6179	0.6649	0.6601	0.5675
Vicio médio total	Média	0.0008	-0.0088	-0.0623	-0.1804	-0.0839	-0.0162	-0.0163	-0.0061
	Desv.Padr.	0.2738	0.2753	0.2783	0.2997	0.2758	0.2753	0.2749	0.2741
	$p_{2.5}$	-0.5263	-0.5365	-0.6031	-0.8029	-0.6340	-0.5440	-0.5440	-0.5345
	Mediana	-0.0037	-0.0153	-0.0684	-0.1841	-0.0942	-0.0218	-0.0209	-0.0116
	$p_{97.5}$	0.5505	0.5406	0.4873	0.3930	0.4597	0.5346	0.5346	0.5439
DW (5%)	Rejeita (%)	-	100.00	2.10	10.73	6.52	71.23	72.80	51.87
EQM ult.	Média	0.2861	0.2779	0.2837	0.3394	0.2633	0.2834	0.2805	0.2643
	Desv.Padr.	0.4245	0.4127	0.4712	0.5323	0.3864	0.427	0.4149	0.3895
	$p_{2.5}$	0.0004	0.0002	0.0003	0.0005	0.0002	0.0002	0.0002	0.0003
	Mediana	0.1238	0.1161	0.1217	0.1381	0.1086	0.1174	0.1165	0.1134
	$p_{97.5}$	1.5331	1.4441	1.4462	1.8184	1.4259	1.5147	1.5342	1.3962
Vicio médio ult.	Média	0.0042	-0.0102	-0.0659	-0.2146	-0.0999	-0.0101	-0.0109	-0.0039
	Desv.Padr.	0.5350	0.5272	0.5287	0.5418	0.5035	0.5324	0.5296	0.5142
	$p_{2.5}$	-1.0205	-1.0952	-1.1197	-1.3168	-1.1130	-1.0911	-1.0853	-1.0264
	Mediana	0.0076	-0.0058	-0.0588	-0.2012	-0.0913	0.0040	0.0043	0.0152
	$p_{97.5}$	1.1012	1.0398	0.9690	0.8238	0.8937	1.0333	1.0302	1.0132
$V(\hat{\theta}_t)$ ult.	Média	0.3087	0.1840	0.2821	0.3141	0.3282	0.2744	0.2744	0.2846
	Desv.Padr.	0.0771	0.0355	0.0732	0.0763	0.0739	0.0675	0.0670	0.0653
	$p_{2.5}$	0.1473	0.1059	0.1295	0.1450	0.1654	0.1274	0.1281	0.1413
	Mediana	0.3152	0.1871	0.2879	0.3217	0.3351	0.2813	0.2815	0.2930
	$p_{97.5}$	0.4389	0.2451	0.4051	0.4403	0.4530	0.3846	0.3835	0.3888
Prob. Cober. ult.	IC(95%)	95.95	88.48	94.69	93.00	96.74	93.63	93.79	95.79

Apêndice E

Resultados adicionais da aplicação

Tabela E.1: Estimativas da autocorrelação das séries de erros amostrais por RM. Método dos pseudo-erros. PME, março de 2002 - fevereiro 2012.

Defasagem	Recife	Salvador	Belo Horizonte	Rio de Janeiro	São Paulo	Porto Alegre
1	0.3246***	0.5766***	0.3031***	0.5577***	0.3844***	0.3819***
2	0.2539**	0.4054***	0.2374**	0.3976***	0.2601**	0.2514**
3	0.1929*	0.2843***	0.1703	0.2375**	0.1827*	0.1855*
4	0.1033	0.1378	0.1033	0.1259	0.0245	0.2023*
5	0.0904	0.1294	0.1174	0.1192	0.0818	0.1898*
6	0.1105	0.1341	0.1113	0.1068	0.0320	0.1376
7	0.0607	0.1025	0.0636	0.1303	0.0646	0.1437
8	0.0761	0.0781	0.0626	0.1170	0.0440	0.1048
9	0.0947	0.0881	0.0558	0.1576	0.0877	0.1105
10	0.1461	0.0851	0.0614	0.1914*	0.0894	0.0692
11	0.1345	0.0750	0.0922	0.1931*	0.1276	0.0947
12	0.1829*	0.0892	0.0417	0.2150**	0.1804*	0.0972
13	0.1501	0.0713	0.0394	0.1670	0.0448	0.1193
14	0.1041	0.0296	0.0222	0.1717	0.0529	0.0935
15	0.1166	0.0341	0.0249	0.1246	0.0283	0.0331
16	0.1139	-0.0089	0.0157	0.1033	0.0045	0.0716
17	0.0990	-0.0042	0.0563	0.0529	0.0870	0.0640
18	0.1212	0.0079	0.0309	0.0343	0.0541	0.1135
19	0.1603	0.0283	0.0629	0.0764	0.0145	0.0432
20	0.0706	0.0486	0.0663	0.0478	0.0277	0.0573

(***, ** e *) indicam significância ao 99%, 95% e 90% respectivamente, empregando como limite $z_{\alpha/2}/\sqrt{120}$.

Resultados adicionais da aplicação

Figura E.1: Boxplots dos Pseudo-erros observados por RM. Aplicação PME.
Ponto vermelho: Média.

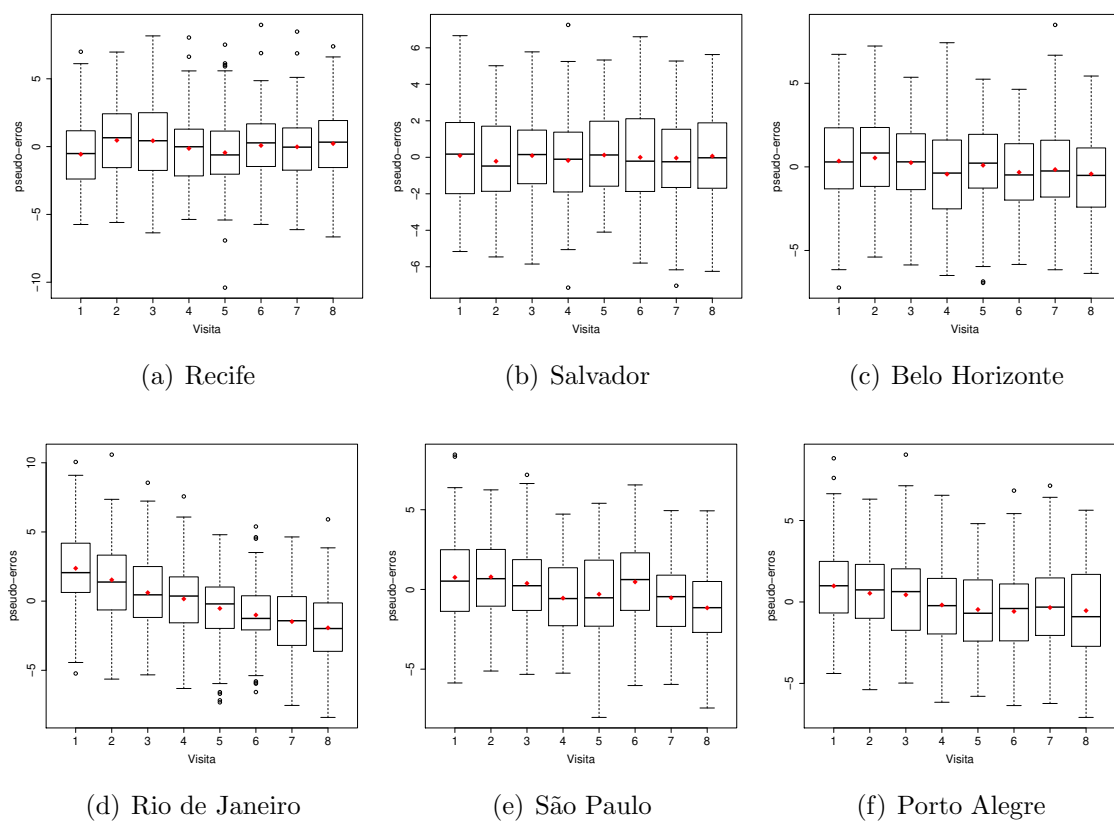


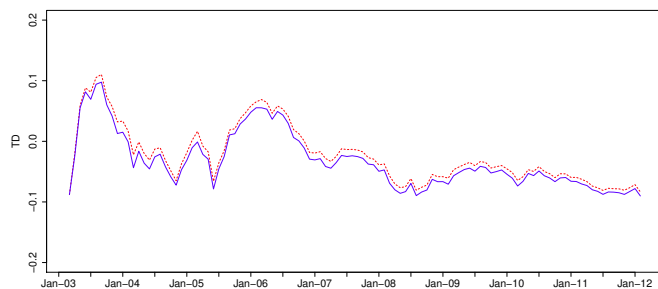
Tabela E.2: Estimativas da autocorrelação parcial das séries de erros amostrais por RM. PME, março de 2002 - fevereiro 2012.

Defasagem	Recife	Salvador	Belo Horizonte	Rio de Janeiro	São Paulo	Porto Alegre
1	0.3246***	0.5766***	0.3031***	0.5577***	0.3844***	0.3819***
2	0.1661	0.1092	0.1603	0.1257	0.1318	0.1236
3	0.0808	0.0198	0.0693	-0.0391	0.0541	0.0644
4	-0.0106	-0.0897	0.0079	-0.0396	-0.1024	0.1092
5	0.0207	0.0812	0.0566	0.0762	0.0815	0.0718
6	0.0626	0.0660	0.0487	0.0362	-0.0128	0.0087
7	-0.0061	-0.0095	-0.0096	0.0586	0.0588	0.0525
8	0.0288	-0.0223	0.0109	0.0045	-0.0171	-0.0024
9	0.0493	0.0492	0.0156	0.0901	0.0821	0.0282
10	0.1012	0.0308	0.0262	0.0862	0.0206	-0.0187
11	0.0454	-0.0015	0.0542	0.0368	0.0896	0.0398
12	0.0938	0.0286	-0.0195	0.0664	0.0916	0.0287
13	0.0327	-0.0027	-0.0012	-0.0138	-0.0896	0.0528
14	-0.0103	-0.0410	-0.0101	0.0571	-0.0015	0.0061
15	0.0319	0.0184	0.0070	-0.0175	0.0028	-0.0491
16	0.0367	-0.0482	-0.0084	0.0026	0.0049	0.0372
17	0.0238	0.0195	0.0453	-0.0528	0.0744	0.0051
18	0.0435	0.0071	0.0010	-0.0048	0.0021	0.0639
19	0.0884	0.0374	0.0403	0.0624	-0.0549	-0.0419
20	-0.0485	0.0215	0.0299	-0.0442	0.0075	0.0161

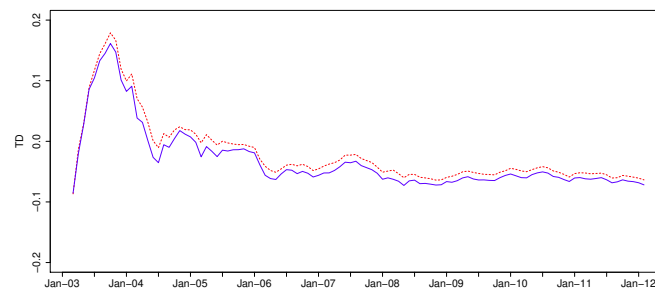
(***, ** e *) indicam significância ao 99%, 95% e 90% respectivamente, empregando como limite $z_{\alpha/2}/\sqrt{120}$.

Figura E.2: Séries ajustadas com e sem benchmark. Inclinação local (\hat{R}_t).

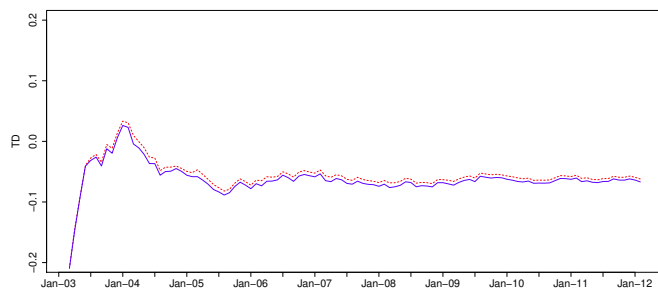
Linhas: vermelha tracejada: série ajustada sem benchmark; azul contínua: série ajustada com benchmark.



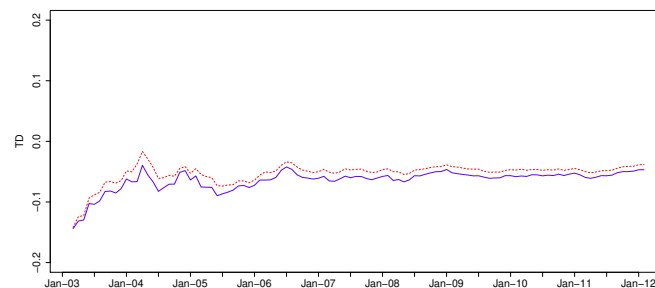
(a) Recife



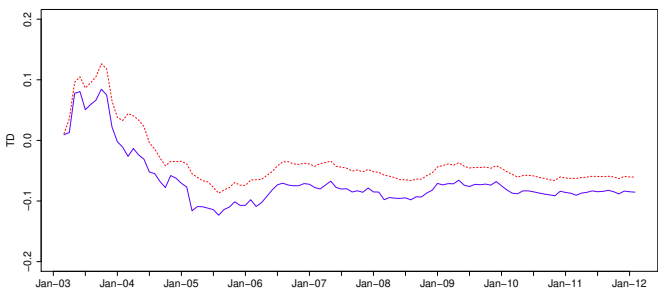
(b) Salvador



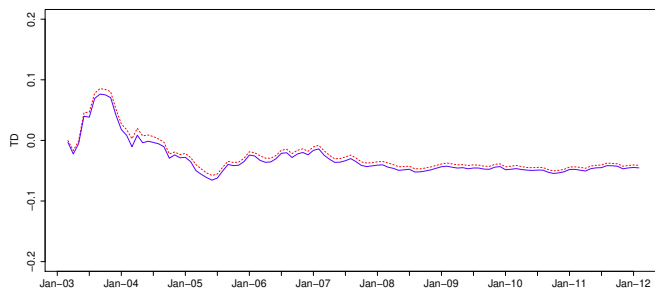
(c) Belo Horizonte



(d) Rio de Janeiro

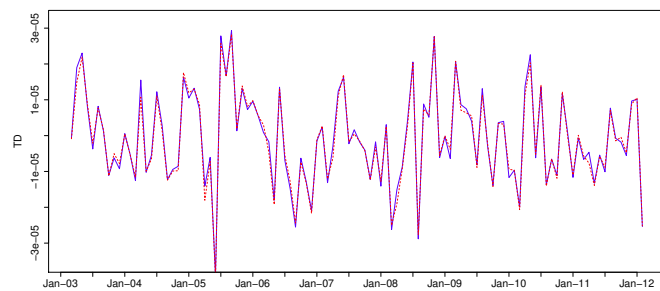


(e) São Paulo

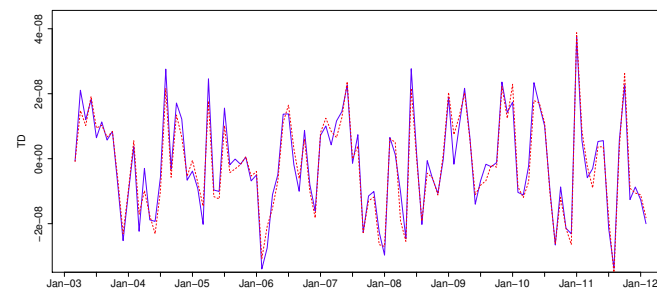


(f) Porto Alegre

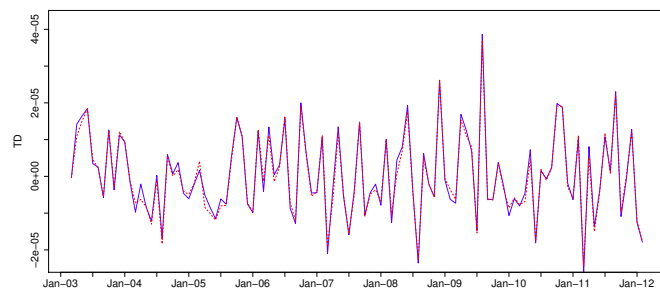
Figura E.3: Séries ajustadas com e sem benchmark. Componente Irregular (\hat{I}_t).
Linhas: vermelha tracejada: série ajustada sem benchmark; azul contínua: série ajustada com benchmark.



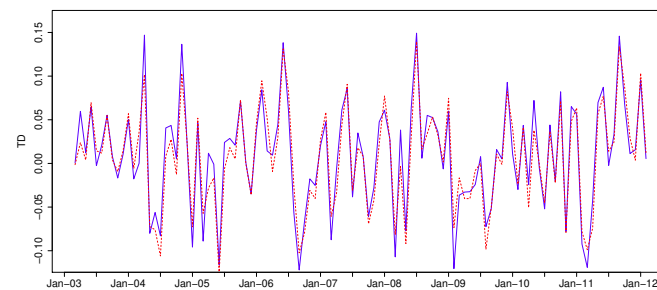
(a) Recife



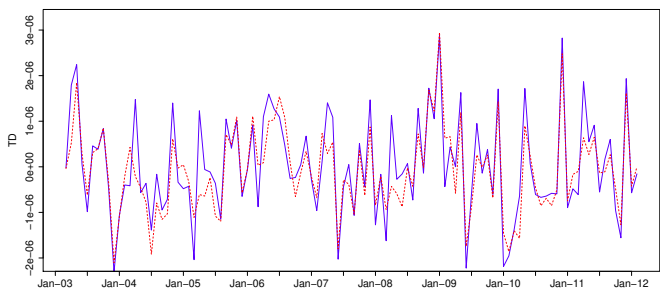
(b) Salvador



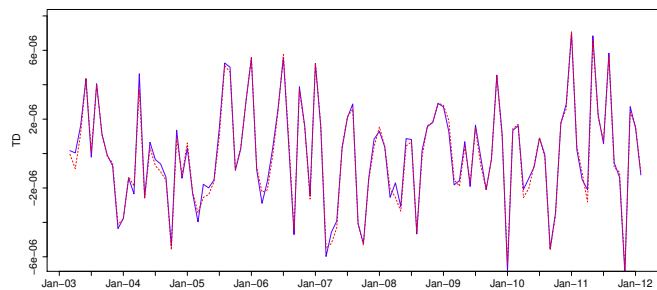
(c) Belo Horizonte



(d) Rio de Janeiro



(e) São Paulo

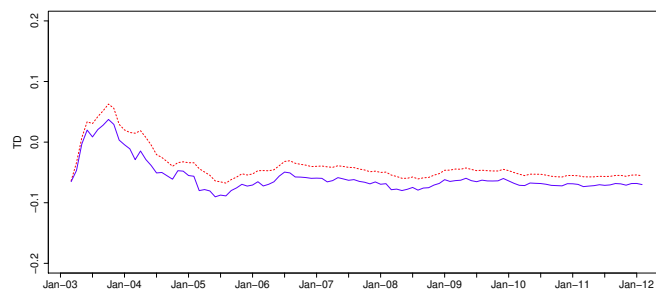


(f) Porto Alegre

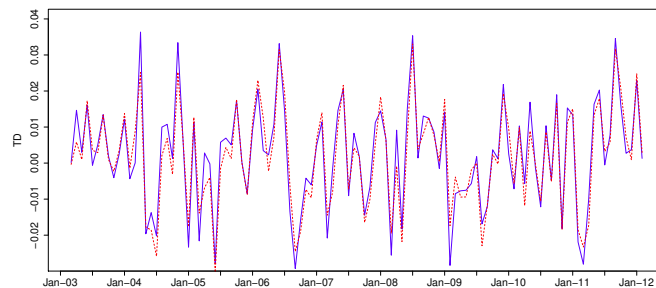
Resultados adicionais da aplicação

Figura E.4: Séries ajustadas com e sem benchmark. Adicionais Conjunto 6 Regiões Metropolitanas.

Linhas: vermelha tracejada: série ajustada sem benchmark; azul contínua: série ajustada com benchmark.



(a) Inclinação Local (\hat{R}_t)



(b) Componente Irregular (\hat{I}_t)

Tabela E.3: P-valores do teste de Jarque-Bera para normalidade dos erros de previsão padronizados no modelo ajustado, segundo RM.

RM	Estatística	P-valor
Recife	0.5124	0.7740
Salvador	0.9115	0.6340
Belo Horizonte	4.1742	0.1240
Rio de Janeiro	0.1985	0.9055
São Paulo	0.0483	0.9762
Porto Alegre	0.5388	0.7638

Tabela E.4: P-valores do teste de Ljung-Box para a não autocorrelação dos erros de previsão padronizados no modelo ajustado, segundo RM. Teste para os primeiros 16 lag, com 12 graus de liberdade.

RM	Estatística	P-valor
Recife	17.5114	0.1314
Salvador	22.5426	0.0319
Belo Horizonte	5.0645	0.9558
Rio de Janeiro	9.3033	0.6768
São Paulo	19.6965	0.0730
Porto Alegre	22.7033	0.0304

Tabela E.5: Autocorrelações estimadas dos erros de previsão por RM. Filtro GLS com restrição de benchmark. PME, março de 2002 - fevereiro 2012.

Defasagem	Recife	Salvador	Belo Horizonte	Rio de Janeiro	São Paulo	Porto Alegre
1	0.1120	0.2745**	-0.0005	0.0775	0.0304	0.2039*
2	0.0891	0.0226	-0.0184	-0.0853	-0.0188	-0.1017
3	-0.1142	0.0831	0.0654	0.0310	-0.1381	-0.0782
4	0.0766	-0.0663	0.0151	-0.1126	0.0400	-0.0743
5	0.0164	-0.2027*	-0.0783	0.0109	0.1259	0.0553
6	-0.0841	-0.0548	-0.0026	-0.0911	-0.0081	0.0257
7	0.0213	-0.1385	0.0795	-0.0322	0.1197	0.0773
8	-0.0618	-0.0099	0.0152	-0.0006	-0.0038	-0.1133
9	-0.0635	0.0108	-0.0416	-0.1851*	-0.0190	-0.1158
10	-0.2017*	-0.0840	-0.1072	-0.0382	-0.1383	-0.0012
11	0.0241	-0.1137	0.0335	0.0032	-0.0662	-0.0673
12	-0.1940*	-0.0025	-0.0420	0.0115	0.1089	0.0667
13	0.0427	-0.0813	-0.0696	-0.0555	0.0061	0.1363
14	-0.0984	0.0463	-0.0093	-0.0594	-0.2029*	-0.0196
15	0.0163	-0.0593	-0.0532	0.0026	-0.0393	-0.1659
16	-0.0270	-0.1070	-0.0286	-0.0136	-0.1743	-0.1835*

(***, ** e *) indicam significância ao 99%, 95% e 90% respectivamente, empregando como limite $z_{\alpha/2}/\sqrt{120}$.