

Escola de Pós-Graduação em Economia - EPGE
Fundação Getúlio Vargas

Ensaaios em Macroeconometria

Tese submetida à Escola de Pós-Graduação em Economia da Fundação
Getúlio Vargas como requisito de obtenção do Título de Doutor em
Economia

Aluno: Márcio Antônio Salvato

Orientador: Prof. João Victor Issler

Rio de Janeiro
2003

Escola de Pós-Graduação em Economia - EPGE

Fundação Getúlio Vargas

Ensaaios em Macroeconometria:

- 1 Principais Características do Consumo de Duráveis no Brasil e Testes de Separabilidade entre Duráveis e Não-Duráveis
- 2 Disparidades Regionais ou Educacionais? Um Exercício Contrafactual
- 3 Estimando os Ganhos de Bem-Estar a partir de Funções Impulso-Resposta não Convencionais
- 4 Are Business Cycles All Alike in Europe?

Aluno: Márcio Antônio Salvato

Banca Examinadora:

Professor João Victor Issler (Orientador, EPGE/FGV)

Professor Luiz Renato Lima (EPGE/FGV)

Professor Marcelo Neri (EPGE/FGV)

Professor Naércio Aquino Menezes Filho (FEA/USP)

Professor Paulo Brígido Rocha Macedo (CEDEPLAR/UFGM)

Rio de Janeiro

2003

Agradecimentos

*À Deus pelo direito da Vida
À minha esposa e filhos pelo amor e compreensão
Aos meus pais pela dedicação e orientação
Aos professores e colegas da EPGE pela amizade e respeito
À PUCminas pelo suporte e confiança
A todos que torceram por essa vitória*

Sumário

Lista de Tabelas	i
Lista de Figuras	iii
1 Principais Características do Consumo de Duráveis no Brasil e Testes de Separabilidade entre Duráveis e Não-Duráveis	1
1.1 Introdução	2
1.2 Teoria Econômica e Restrições Testáveis	3
1.2.1 Decisão de Consumo	3
1.2.2 Restrição à Liquidez	7
1.2.3 Evidências Empíricas Prévias para o Brasil	8
1.3 Base de Dados e Análise Preliminar	10
1.4 Testes e Resultados	12
2 Disparidades Regionais ou Educacionais? Um Exercício Contrafactual	26
2.1 Introdução	27
2.2 Fatos Estilizados	30
2.3 Dados	33
2.4 Metodologia	36
2.5 Resultados	39
2.6 Conclusões	49
Referências	50
3 Estimando os Ganhos de Bem-Estar a partir de Funções Impulso-Resposta não Convencionais	58
3.1 Introdução	59
3.2 Funções Impulso-Resposta não Convencionais para o Bem-Estar	61
3.3 Ortogonalização dos Choques	67
3.4 Base de Dados e Resultados Empíricos	70
3.5 Conclusões	79
Referências	81
Apêndice	83
3.A.1 Proposições e Provas	83
3.A.2 Teste de Exogeneidade Fraca	85

3.A.3	Choques Permanentes <i>versus</i> Transitórios	86
3.A.4	Impulso-Resposta Generalizado não Convencional	88
3.A.5	Estimação dos Desvios-padrões das Funções Impulso-Resposta não Convencionais	91
4	Are Business Cycles All Alike in Europe?	92
4.1	Introduction	93
4.2	The Problem	95
4.2.1	Standard Errors of $\lambda(\phi, \beta)$ Estimates	99
4.3	Reduced Form and Long-Run Constraints	100
4.4	Data	102
4.5	Empirical Results	104
4.6	Conclusions	107
	References	108
	Appendix	111
4.A	Convergence of $\lambda(\phi, \beta)$ for $\phi \rightarrow 1$	111
4.B	Deriving Trend and Cycle Formulae from Space-State Form	112
4.C	Computing Conditional Covariances	113
4.D	Tables and Figures	115

Lista de Tabelas

Capítulo 1

Tabela 1	Testes de Raiz Unitária	12
Tabela 2	Estimação por MGM das equações (4) e (5) para $\delta + g = 0.10$	13
Tabela 3	Estimativa do Modelo de Mankiw (1982)	14
Tabela 4	Teste de Cointegração de Johansen - Modelo VAR(1) para cd e cn	15
Tabela 5	Teste de Cointegração de Johansen - Modelo VAR(3) para cd , cn e y	16
Tabela 6	Teste para Ciclos Comuns e Ciclos Codependentes cn_t , cd_t e y_t	17
Tabela 7	Teste de Restrição à liquidez: $\Delta cd_t = \lambda \Delta y_t + (1 - \lambda) \theta r_t + (1 - \lambda) \varepsilon_t$	18
Tabela 8	Teste de Restrição à liquidez: $\Delta cd_t = \lambda \Delta y_t + (1 - \lambda) \varepsilon_t$	19

Capítulo 2

Tabela 1	Índice de Desenvolvimento Humano - IDH	31
Tabela 2	Coeficientes de Theil e de Gini da Renda do Trabalho	31
Tabela 3	Quantitativo por Região/Estado	34
Tabela 4	Estatísticas Descritivas para a Amostra Seleccionada	34
Tabela 5	Pontos Notáveis das Distribuições de Renda do Trabalho Estimadas	42
Tabela 6	Relação da Renda nos Percentis	43
Tabela 7	Distâncias de Kullback-Leibler	47
Tabela 8	Coeficientes de Gini e Theil	48
Tabela 9	Percentual Explicado pela Escolaridade	48
Tabela A.1	Pontos Notáveis das Distribuições de Renda do Trabalho Estimadas	56
Tabela A.2	Relação da Renda nos Percentis	56
Tabela A.3	Coeficientes de Gini e Theil	57
Tabela A.4	Percentual explicado pela Escolaridade	57

Capítulo 3

Tabela 1	Teste de Cointegração de Johansen	72
Tabela 2	Modelo de Correção de Erros (VECM)	72
Tabela 3	Teste de Exogeneidade Fraca sobre os coeficientes de ajustamento do Modelo de Correção de Erros	73
Tabela 4	Decomposição de Choques Permanentes e Transitórios para as FIRNC	75
Tabela 5	Função Impulso-Resposta Convencional	76
Tabela 6	Função Impulso-Resposta não Convencional para o Bem-Estar	77
Tabela 7	Função Impulso-Resposta não Convencional para o Bem-Estar usando Decomposição de Cholesky	79

Capítulo 4

Table 1 Johansen Cointegration Test	103
Table 2 Testing $H_0 : \alpha' = (1, -1)$	104
Table 3 Unit Root Tests	104
Table 4 Welfare Cost of Business Cycles	106
Tabela D.1 Welfare Cost of Business Cycles ($\lambda\%$)	115, 116, 117 and 118

Lista de Figuras e Gráficos

Capítulo 1

Figura 1 Gráfico das Séries em nível	11
--------------------------------------	----

Capítulo 2

Gráfico 1 Renda per capita (em log)	32
Gráfico 2 Renda per capita x Escolaridade Média	32
Gráfico 3.a Escolaridade x Renda	35
Gráfico 3.b Escolaridade x Renda	36
Gráfico 4 Densidades reais para SE1 e NE e contrafactual para NE com escolaridade de SE1	40
Gráfico 5 Densidades reais para SE2 e NE e contrafactual para NE com escolaridade de SE2	40
Gráfico 6 Densidades reais para SP e CE e contrafactual para CE com escolaridade de SP	41
Gráfico 7 Evolução da Renda do Trabalho por Percentis para CE, SP e CE/SP	44
Gráfico 8 Evolução da Renda do Trabalho por Percentis para NE, SE1 e NE/SE1	44
Gráfico 9 Evolução da Renda do Trabalho por Percentis para NE, SE2 e NE/SE2	45
Gráfico 10 Diferença das Distribuições (SE1 x NE)	45
Gráfico 11 Diferença das Distribuições (SE2 x NE)	46
Gráfico 12 Diferença das Distribuições (SP x CE)	46
Gráfico A.1 Densidades reais para SP e NE e contrafactual para NE com escolaridade de SP e Diferenças entre as distribuições de SP e NE (real e contrafactual)	52
Gráfico A.2 Densidades reais para SE1 e CE e contrafactual para CE com escolaridade de SE1 e Diferenças entre as distribuições de SE1 e CE (real e contrafactual)	53
Gráfico A.3 Densidades reais para SE2 e CE e contrafactual para CE com escolaridade de SE2 e Diferenças entre as distribuições de SE2 e CE (real e contrafactual)	54
Gráfico A.4 Evolução da Renda do Trabalho por Percentis para CE, SE1 e CE/SE1	54
Gráfico A.5 Evolução da Renda do Trabalho por Percentis para CE, SE2 e CE/SE2	55
Gráfico A.6 Evolução da Renda do Trabalho por Percentis para NE, SP e NE/SP	55

Capítulo 3

Figura 1	Séries em Nível e 1ª Diferença	71
Figura 2	Resposta a Impulso de 1 Desvio-padrão Generalizado	76
Figura 3	Função Impulso-Resposta Generalizada não Convencional para o Bem-Estar para valores escolhidos de β	78

Capítulo 4

No Figures

Capítulo 1

Principais Características do Consumo de Duráveis no Brasil e Testes de Separabilidade entre Duráveis e Não-Duráveis*

Márcio Antônio Salvato[†]

Resumo

Este trabalho investiga amplamente a evolução do consumo de bens duráveis no Brasil a partir da decisão de consumo individual e da possibilidade de existir restrição ao crédito. A contribuição mais relevante consiste na não rejeição da hipótese de separabilidade nas decisões de consumo de bens duráveis e não duráveis, já que tal hipótese é implicitamente utilizada por vários artigos que trataram a questão do consumo agregado no Brasil. Os resultados, aqui encontrados, sugerem que uma grande parcela dos consumidores está restrita ao crédito, existindo restrições de curto prazo e longo prazo sobre a evolução do consumo de bens duráveis, não duráveis e renda.

JEL Classification Codes: C22 e E21.

*Artigo feito em co-autoria com João Victor Issler (EPGE/FGV) e Fábio Augusto Reis Gomes (EPGE/FGV e CEPE).

[†]EPGE/FGV, PUC.Minas e CEPE, salvato@fgvmail.br.

1.1 Introdução

No Brasil, como na maioria dos países, o consumo agregado é uma das séries macroeconômicas de maior importância, pois, além de representar mais do que 70% do Produto Interno Bruto, é um argumento fundamental na função de bem-estar econômico. A não ser que a função utilidade seja separável, estudos sobre o consumo deveriam tratar de forma integrada as decisões sobre diferentes tipos de gasto. Em particular as decisões sobre o consumo de duráveis e de não-duráveis deveriam ser modeladas de forma integrada, possivelmente incluindo também a renda do consumidor nesta análise. Infelizmente, estudos anteriores sobre o consumo no Brasil não o fizeram, pois consideraram apenas o consumo total, que agrega essas duas categorias — ver Cavalcanti(1993), Reis *et alli* (1998), Issler e Rocha (2000) e Gomes (2003) — sendo importante que se teste a separabilidade da função utilidade para que se cogite validar alguns desses estudos prévios. Como nota Bernanke (1985), o que gera utilidade para o consumidor representativo não é o gasto total em bens duráveis e não-duráveis, mas sim o gasto com bens não-duráveis e o estoque de bens duráveis, respectivamente.

As comparações dos gastos de duráveis e não-duráveis indicam que variações do primeiro têm maior amplitude do que as do segundo, implicando num comportamento cíclico mais pronunciado para o gasto de duráveis, o que poderia ser um elemento importante no entendimento do ciclo de negócios brasileiro, objeto de estudo de uma literatura crescente no Brasil. Dados os resultados prévios da literatura brasileira de consumo, quando se investigou a existência de mercados incompletos de crédito, parte desse comportamento cíclico pode estar associado à restrições à liquidez no financiamento de bens duráveis, assim como se encontrou previamente para o consumo total.

Nesse artigo, investiga-se amplamente as propriedades dos gastos com bens duráveis no Brasil, tendo como ponto de partida um modelo integrado para consumo de bens duráveis, não-duráveis, e renda. A aplicação do método generalizado dos momentos (GMM), às Equações de Euler do problema do consumidor representativo, possibilita testar diretamente a separabilidade da função utilidade entre estes dois tipos de gasto, questão fundamental para a correta avaliação da literatura corrente sobre o consumo no Brasil. Nossos testes de separabilidade são inspirados em Bernanke (1985), que considera a utilidade não-separável no estoque de bens duráveis e no consumo de não duráveis, embora a função de utilidade utilizada aqui seja mais geral que a forma quadrática usada naquele estudo.

A partir dos resultados encontrados nos testes de separabilidade, que não a rejeitaram,

investigou-se as propriedades estocásticas das séries brasileiras de consumo de duráveis e não-duráveis. Estas são estudadas a partir de modelos de otimização intertemporal que consideram o bem-estar associado aos serviços advindos do estoque de bens de consumo duráveis. Tais modelos mostram que o comportamento do consumo de bens duráveis é diferente do consumo de bens não duráveis, encontrando diferentes restrições testáveis para cada tipo de gasto. Ao que sabemos, trata-se de um estudo pioneiro e necessário sobre o assunto no Brasil. Nele investigamos duas questões centrais à essa literatura: a comparação da dinâmica de curto e longo prazos entre os gastos de não-duráveis, duráveis e renda (Mankiw(1982), Bernanke(1985), Galí(1993), Vahid e Engle (1997)) e possível existência de restrições à liquidez (Campbell e Mankiw(1989) e Bacchetta e Stefan(1997)) no consumo de duráveis.

Além do importante resultado de encontrarmos separabilidade entre duráveis e não-duráveis, que validam estudos sobre consumo anteriores ao nosso, concluímos que há uma relação de longo prazo (cointegração) entre o consumo de duráveis, não duráveis, e renda. No curto prazo, essas três séries exibem também uma dinâmica comum (têm um ciclo comum), que pode ser explicada em parte pela existência de restrição à liquidez no Brasil. Essa última evidência sugere que políticas que visem reduzir a restrição de crédito podem ter um impacto grande sobre o consumo, e, por consequência, sobre o bem-estar dos consumidores. Além disso, ao comparar nossos resultados atuais de restrição à liquidez aos resultados prévios de Reis et alli (1998), Issler e Rocha (2000) e Gomes (2003) encontramos evidência, ainda maior, a favor da hipótese de restrição ao crédito.

O resto desse estudo está dividido da seguinte forma: a próxima seção apresenta a teoria e as possíveis proposições testáveis, incluindo resultados empíricos obtidos com dados brasileiros e norte-americanos. A terceira seção apresenta a base de dados e a quarta os resultados empíricos. A quinta seção apresenta as conclusões.

1.2 Teoria Econômica e Restrições Testáveis

1.2.1 Decisão de Consumo

Hall (1978) considerou o problema de otimização no qual o consumidor escolhe o seu consumo de bens não duráveis, cn_t , sujeito a uma restrição orçamentária. A equação de Euler é:

$$u'(cn_t) = E_t \{ \beta u'(cn_{t+1})(1 + r_{t+1}) \}. \quad (1)$$

em que β é a taxa de desconto intertemporal da utilidade futura, e r_{t+1} é a taxa de retôrno real da riqueza. Sob a hipótese de que $\beta(1 + r_{t+1}) = 1, \forall t$:

$$u'(cn_{t+1}) = u'(cn_t) + \epsilon_{t+1},$$

em que $E_t(\epsilon_{t+1}) = 0$, i.e., ϵ_{t+1} é uma inovação. Nesse caso, $u'(cn_{t+1})$ é uma martingala. Se supusermos, adicionalmente, que a utilidade é quadrática em cn_t , então a propriedade de martingala se estende à cn_{t+1} , de modo que $E_t(\Delta cn_{t+1}) = 0$. i.e., é uma Δcn_{t+1} é imprevisível¹, o que é uma hipótese testável a partir do uso de técnicas econométricas.

Mankiw (1982) foi o primeiro a modelar a decisão de gastos em bens duráveis num arcabouço semelhante ao de Hall, no qual o agente decide o gasto em bens duráveis de forma a maximizar sua utilidade² intertemporal, sujeito, além da restrição orçamentária, a uma lei de movimento do estoque de bens duráveis, $k_{t+1} = (1 - \delta)k_t + cd_{t+1}$, onde cd_{t+1} é o gasto em bens duráveis. Considerando-se as hipóteses de separabilidade da utilidade em k_t e cn_t , de utilidade quadrática em k_t , e de $\beta(1 + r_{t+1}) = 1 \forall t$, é possível mostrar que

$$\begin{aligned} k_{t+1} &= k_t + \epsilon_{t+1} & \text{e} \\ cd_{t+1} &= cd_t + \epsilon_{t+1} - (1 - \delta)\epsilon_t, \end{aligned}$$

em que $E_t(\epsilon_{t+1}) = 0$, i.e., ϵ_{t+1} é uma inovação. Neste caso, a variação do estoque de bens de consumo duráveis é um ruído branco e a variação do gasto de duráveis é um processo $MA(1)$, que se constituem em restrições testáveis a partir do uso de técnicas econométricas.

Nos modelos de Hall (1978) e Mankiw (1982) os gastos de não duráveis e de duráveis são modelados separadamente de modo que, necessariamente, uma hipótese de separabilidade

¹Flavin (1981) demonstrou que esse resultado pode ser obtido ao se considerar explicitamente a Teoria da Renda Permanente e a hipótese de expectativas racionais. A renda permanente, y_t^P , é dada por:

$$y_t^P = r \left[A_t + \sum_{s=0}^{\infty} \left(\frac{1}{1+r} \right)^{s+1} E_t(y_{t+s}) \right]$$

em que A_t e y_t são, respectivamente, a riqueza e a renda do trabalho no período t .

²Em Mankiw (1982) a utilidade é função do serviço proporcionado pelo montante de estoque de bens duráveis que o agente possui. Considera-se ainda que o serviço é proporcional ao próprio estoque e portanto a função de utilidade pode ser escrita como função do estoque de bens duráveis.

na função de utilidade foi adotada. Bernanke (1985) foi o primeiro a modelar conjuntamente a decisão de gastos em bens duráveis e não duráveis. Esta forma de tratar o problema torna-se mais relevante quanto maior for a interdependência entre a decisão de gasto em bens duráveis e não duráveis. Seguindo a tradição de Hall e Mankiw, Bernanke considerou a seguinte função utilidade, quadrática no consumo de duráveis e no estoque de não-duráveis,

$$u(cn_t, k_t, k_{t+1}) = -\frac{1}{2}(\bar{cn} - cn_t)^2 - \frac{a}{2}(\bar{k} - k_t)^2 - m(\bar{cn} - cn_t)(\bar{k} - k_t) - \frac{d}{2}(k_{t+1} - k_t)^2,$$

em que o último termo do lado direito modela um custo de ajustamento no estoque de bens de consumo duráveis. A não separabilidade no gasto em bens de consumo não duráveis e no estoque de bens duráveis se deve ao parâmetro m . Quando há separabilidade $m = 0$, o que pode ser testado econometricamente. As equações de Euler desse problema são,

$$\bar{cn} - cn_t + m(\bar{k} - k_t) = \beta E_t \{ (1 + r_t) [\bar{cn} - cn_{t+1} + m(\bar{k} - k_{t+1})] \}, \quad (2)$$

$$\begin{aligned} -d(k_{t+1} - k_t) = & \beta E_t \{ [\bar{cn} - cn_{t+1} + m(\bar{k} - k_{t+1})] [(1 + r_t)p_t^k - (1 - \delta)p_{t+1}^k] \} + \\ & -\beta E_t [a(\bar{k} - k_{t+1}) + m(\bar{cn} - cn_{t+1}) + d(k_{t+2} - k_{t+1})], \end{aligned} \quad (3)$$

em que p_{t+1}^k é o preço relativo de duráveis em unidade de bens não duráveis.

Uma implicação desse modelo, fruto da análise de (2) é que, se $\beta(1 + r) = 1$,

$$(cn_{t+1} - mk_{t+1}) = (cn_t + mk_t) + \epsilon_{t+1},$$

em que $E_t(\epsilon_{t+1}) = 0$, i.e., ϵ_{t+1} é uma inovação. Nesse caso, existe uma combinação linear do gasto de não-duráveis e do estoque de duráveis que é uma martingala. Isto implica a existência de um ciclo comum para Δcn_{t+1} e Δk_{t+1} , com coeficiente $(1, -m)$; ver Vahid e Engle (1993). Note que, quando há separabilidade, $m = 0$, e volta-se ao caso de Hall no qual a primeira diferença do gasto de não-duráveis é imprevisível.

Considerando uma função de utilidade mais geral do que a usada por Bernanke, como, por exemplo, a do tipo CES, $u(cn_t, k_t) = (cn_t^\rho + \alpha k_t^\rho)^{\frac{1}{\rho}}$, as equações de Euler em (2) e (3) tornam-se,

$$E_t \left[\beta (cn_{t+1}^\rho + \alpha k_{t+1}^\rho)^{\frac{1-\rho}{\rho}} cn_{t+1}^{\rho-1} (1 + r_t) \right] = (cn_t^\rho + \alpha k_t^\rho)^{\frac{1-\rho}{\rho}} cn_t^{\rho-1}, \quad (4)$$

$$E_t \left\{ \begin{aligned} & (cn_{t+1}^\rho + \alpha k_{t+1}^\rho)^{\frac{1-\rho}{\rho}} [cn_{t+1}^{\rho-1} p_t^k (1 + r_t) - \alpha k_{t+1}^{\rho-1}] + \\ & -\beta (cn_{t+2}^\rho + \alpha k_{t+2}^\rho)^{\frac{1-\rho}{\rho}} cn_{t+2}^{\rho-1} p_{t+1}^k (1 - \delta) (1 + r_{t+1}) \end{aligned} \right\} = 0. \quad (5)$$

Nesse caso, a hipótese de separabilidade entre a decisão de gasto em bens duráveis e não duráveis pode ser examinada facilmente testando-se a hipótese nula de que $\rho = 1$.

Mais recentemente, autores como Galí (1993) e Vahid e Engle (1997) modelaram conjuntamente a decisão de consumo de bens duráveis, não-duráveis, e renda, impondo a separabilidade *a priori*. Para tanto, partiram de uma função utilidade quadrática que não inclui um termo cruzado entre cn_t e k_t . Vahid e Engle (1997) consideraram a seguinte função de utilidade:

$$u_t = u(cn_t, k_t) = \alpha_1 cn_t - \frac{1}{2} b_1 cn_t^2 + \alpha_2 (\tau k_t) - \frac{1}{2} b_2 (\tau k_t)^2,$$

em que τk_t representa o fluxo de serviços advindo do estoque de bens duráveis em t . Assumindo que $\beta(1+r) = 1$ a solução do problema de otimização é descrita por:

$$cn_t = \frac{\alpha_1}{b_1} - \frac{1}{b_1} \kappa_t, \quad (6)$$

$$k_t = \frac{\alpha_2}{\tau b_2} - \frac{\pi}{\tau^2 b_2} \lambda_t + \frac{\pi \beta (1 - \delta)}{\tau^2 b_2} E_t(\kappa_{t+1}), \quad (7)$$

$$\kappa_t = \kappa_{t-1} + \varepsilon_t, \quad (8)$$

em que ε_t é uma inovação, κ_t é o valor da utilidade marginal da riqueza em t , e π é o preço relativo de duráveis em termos de não duráveis. Usando a lei de movimento para o estoque de bens de consumo duráveis, e a equação de Euler (7), temos:

$$cd_t = \frac{\delta \alpha_2}{\tau b_2} + \rho \kappa_t - (1 - \delta) \rho \kappa_{t-1}, \quad (9)$$

em que $\rho = -\frac{\pi[1-\beta(1-\delta)]}{\tau^2 b_2}$.

Há vários pontos a notar. Primeiro, tanto cn_t quanto cd_t contêm κ_t , que é um processo com raiz unitária (integrado de ordem um). Logo, ambos têm raiz unitária e cointegram. Pode-se mostrar que o vetor de cointegração é $\frac{1}{1-\delta} (b_1 \delta, \rho^{-1})$; ver Galí(1993). Segundo, o consumo de bens não-duráveis é uma martingala, como em Hall (1978) e Flavin (1981): $E_t(cn_{t+1}) = cn_t$ e $\Delta cn_t = -\frac{1}{b_1} \varepsilon_t$. Terceiro, a primeira diferença do consumo de duráveis é um $MA(1)$, como em Mankiw ((1982): $\Delta cd_t = \rho \varepsilon_t - (1 - \delta) \rho \varepsilon_{t-1}$. O fato de que Δcn_t é um ruído branco, e de que Δcd_t é uma $MA(1)$, implica que qualquer combinação linear, não nula, entre a primeira diferença do gasto em bens duráveis e não duráveis resulta em um processo $MA(1)$, isto é, para $\gamma \neq 0$, $\Delta cn_t - \gamma \Delta cd_t \sim MA(1)$ ³. Por fim, pode-se mostrar que

³Na prática, caso as séries de consumo de bens duráveis e não duráveis possuam componente cíclico e ainda assim esta relação persistir, dizemos que há um ciclo codependente entre estas séries. No apêndice apresentamos uma breve revisão do conceito de ciclos codependentes.

se for imposta a condição de que a sequência A_t é limitada⁴, existe uma tendência estocástica comum para consumo de duráveis, não duráveis e renda, que cointegram; ver Vahid e Engle (1997).

1.2.2 Restrição à Liquidez

Campbell e Mankiw (1989) propuseram uma equação de teste para restrição à liquidez quando se supõe uma economia na qual existem dois tipos de consumidores: um consome a totalidade de sua renda corrente, de acordo com uma equação de consumo do tipo Keynesiana (sendo restrito à liquidez), e outro consome a sua renda permanente, de acordo com Hall (1978) e/ou Flavin (1981), ou mesmo levando-se em conta modelos mais elaborados para o crescimento do consumo, baseados na log-linearização da equação de Euler, onde a taxa de crescimento do consumo depende do retôrno real da riqueza, além de sua incerteza.

Considerando-se o modelo de Hall como *benchmark* de otimalidade, a existência de otimalidade (ou a inexistência de restrição à liquidez) pode ser testada através da hipótese nula de que $\lambda = 0$ na equação de teste:

$$\Delta cn_t = \lambda \Delta y_t + \varepsilon_t. \quad (10)$$

obtida quando se considera a heterogeneidade dos dois tipos de agentes na economia. Caso só existam agentes otimizadores, $\lambda = 0$, em que λ é a proporção da renda total pertencente aos consumidores com restrição à liquidez. Caso contrário, $\lambda \neq 0$, pode-se não só verificar a inexistência de optimalidade, como também estimar a sua importância através do termo λ .

Quando consideramos o modelo de Mankiw como *benchmark* de otimização, a primeira diferença do consumo de bens duráveis é um $MA(1)$, $\Delta cd_t = -(1 - \delta)\varepsilon_{t-1} + \varepsilon_t$, e a equação de teste é:

$$\Delta cd_t = \lambda \Delta y_t + (1 - \lambda) [(1 - \delta)\varepsilon_{t-1} + \varepsilon_t] \quad (11)$$

em que λ é a proporção da renda que pertence a consumidores que não tem acesso ao crédito para consumir bens duráveis. Assim, sob a hipótese nula, $\lambda = 0$ e os indivíduos têm pleno acesso ao crédito. Vale notar que esta equação consiste em um teste de ciclos codependentes entre Δcd_t e Δy_t , i.e., a existência de uma combinação linear destas séries que é um processo

⁴Com isso, vale trivialmente uma condição de transversalidade, o que gera uma restrição de longo prazo entre a renda e o consumo.

$MA(\cdot)$ finito. Quando consideramos, adicionalmente, a possibilidade da equação de Euler envolver o retôrno dos ativos, como em Campbell e Mankiw (1989), temos:

$$\Delta cd_t = \lambda \Delta y_t + (1 - \lambda) \theta r_t + (1 - \lambda) \varepsilon_t. \quad (12)$$

que envolve ciclos comuns entre Δcd_t , Δy_t e r_t .

Em todas as equações de teste (10), (11) e (12), há a necessidade de se usar técnicas de variáveis instrumentais para estimar os parâmetros de interesse, pois a renda corrente é, em geral, correlacionada com a surpresa no gasto dos consumidores que não estão sujeitos à restrição de liquidez. Devido ao componente $MA(1)$ na equação (11), a escolha de instrumentos tem que ser defasada de pelo menos dois períodos.

Bacchetta e Stefan (1997) modificam levemente a equação de teste proposta por Campbell e Mankiw (1989), usando:

$$\Delta cn_t = \alpha + \theta E_{t-1}(\Delta X_t) + \varepsilon_t, \quad (13)$$

em que X_t é composto pelo logaritmo da renda real disponível, variáveis de crédito imobiliário, variáveis de crédito do consumidor e diferença entre taxa de captação e empréstimo do mercado financeiro (ao invés da própria taxa de juros). Mais uma vez, há a necessidade de se usar técnicas de variáveis instrumentais para se estimar a equação de teste (13).

1.2.3 Evidências Empíricas Prévia para o Brasil

Cavalcanti (1993) testa implicações da Teoria da Renda Permanente e a possibilidade de existência de restrição à liquidez, utilizando dados trimestrais de consumo agregado e PIB, per capita brasileiro, de 1980 a 1989. Esse autor encontrou evidências de que estas séries são não estacionárias, porém não avaliou se há cointegração entre elas. Com respeito ao teste de restrição à liquidez, utilizando estimadores de variáveis instrumentais, encontrou evidência de que 32% da renda pertence a consumidores restritos.

Reis *et alli* (1998) estudaram várias séries brasileiras anuais de consumo e renda agregados, per capita, datadas de 1947 a 1994, bem como séries trimestrais de janeiro de 1975 a abril de 1994. Não rejeitaram a hipótese nula de raiz unitária para as séries. Através da análise de cointegração, investigaram a existência de relações de longo prazo entre consumo e renda, encontrando evidências a favor de um vetor de cointegração. Na maioria dos casos, não rejeitaram a hipótese deste vetor ser $(1, -1)'$ como previsto pela Teoria da Renda

Permanente. Em seguida, utilizando metodologia proposta por Campbell e Mankiw (1989), os autores estimam que aproximadamente 80% da renda no Brasil pertence a consumidores que estão restritos a consumir apenas a sua renda corrente, número bastante superior ao encontrado por Cavalcanti. Essa diferença foi creditada ao conjunto de instrumentos usados na estimativa de (10), sendo que, por problemas de agregação temporal, a escolha de Reis *et alli* deve ser privilegiada⁵. Posteriormente, Issler e Rocha (2000) ratificam os resultados de Reis *et alli* ao estimarem uma proporção de 74%, encontrando também evidências de que consumo e renda cointegram.⁶

Uma estimativa de λ significativamente diferente de zero nas estimativas a partir de variáveis instrumentais e a validade dos testes de sobreidentificação implicam que o consumo e a renda agregados possuem um ciclo comum. Logo, o consumo agregado possui um componente cíclico. Apesar desses autores terem usado o consumo agregado, eles consideraram apenas o resultado de Hall (1978), segundo o qual o consumo não possui um componente cíclico e por isso rejeitam a Teoria da Renda Permanente. De fato, ao proceder a decomposição de Beveridge e Nelson (1981), Gomes (2003) estima que o consumo agregado anual de 1947 a 1999 para o Brasil possui um componente cíclico. Em uma tentativa de compatibilizar esta evidência empírica com o problema de otimização do consumidor, Gomes inclui a hipótese de formação de hábito. Com isso, foi derivada uma equação de teste mais ampla que a proposta por Campbell e Mankiw (1989). Contudo ao proceder a sua estimativa obteve-se um valor para λ próximo de 0.85 e não se obteve evidência a favor da hipótese de formação de hábito. Com isso, Gomes conclui que a existência de um componente cíclico na série de consumo agregada não pode ser compatibilizada com o problema de otimização intertemporal do consumidor, em equilíbrio parcial, através da hipótese de formação de hábito. Ademais, uma grande parcela da renda pertence a consumidores que não têm acesso ao crédito. Não obstante essa evidência, a hipótese de otimização por parte do consumidor representativo ainda poderia ser mantida num arcabouço de equilíbrio geral, como notado por Issler e Vahid (2001).

Uma possível crítica que pode ser feita a todos os estudos prévios sobre o consumo no Brasil é o fato de que estes utilizaram séries de consumo agregado, i.e., que somam o consumo de bens duráveis ao de não-duráveis. Como nota Bernanke (1985), o que gera utilidade para

⁵Ver a discussão em Campbell e Mankiw(1989) e em Deaton(1992).

⁶Reis *et alii* (1998) encontram evidência de que o consumo é pouco sensível ao juro, mas tal investigação não foi conduzida por Issler & Rocha (2000).

o consumidor representativo não é o gasto total em bens duráveis e não-duráveis, mas sim o gasto com bens não-duráveis e o estoque de bens duráveis, respectivamente. Ademais, do ponto de vista teórico, estas séries possuem comportamento diferenciado em modelos de equilíbrio parcial: a série de duráveis possui um componente cíclico, enquanto a de não duráveis não o possui. Dessa forma, a soma das duas tem necessariamente um componente cíclico, o que não foi levado em conta nos estudos anteriores. Nesse sentido, as estimativas significativas de λ podem ter se dado pela existência do componente cíclico da série de consumo de não duráveis e não pela existência de restrição à liquidez.⁷ Os estudos anteriores podem ser validados se for rejeitada a hipótese de que a diferença do consumo de bens duráveis é um processo $MA(1)$, pois neste caso o consumo agregado não possui qualquer componente cíclico.

1.3 Base de Dados e Análise Preliminar

Para testar as implicações do modelo teórico foram obtidos dados anuais para consumo de não-duráveis e de duráveis de 1970 a 1998 construídos por Ellery *et alli* (2002), em termos per capita. Como medida de renda, utilizamos a renda disponível bruta real, deflacionada pelo Deflator Implícito do PIB, em termos per capita, disponibilizada pelo IPEADATA⁸. Para construir a série de preço relativos de duráveis, usou-se o índice IPA-DI para bens de consumo duráveis e de não-duráveis e por definição $p_t^k = \frac{IPA_t^{duráveis}}{IPA_t^{não\ duráveis}}$, disponibilizada pelo FGVDADOS. Utilizou-se ainda a série de juros real do CDB acumulado no ano.

A série de estoque de bens duráveis foi gerada conforme metodologia apresentada no Apêndice, segundo uma função de acumulação dos fluxos de consumo de bens duráveis. Nesta, é necessário fazer uma hipótese sobre o valor da taxa de depreciação acrescida da taxa de crescimento do consumo de bens duráveis, $\delta + g$. Os dados mostram que $g = 0.033$ no período de análise. Considerou-se valor da taxa de depreciação $\delta = 0.067$ de modo que $\delta + g = 0.1$.

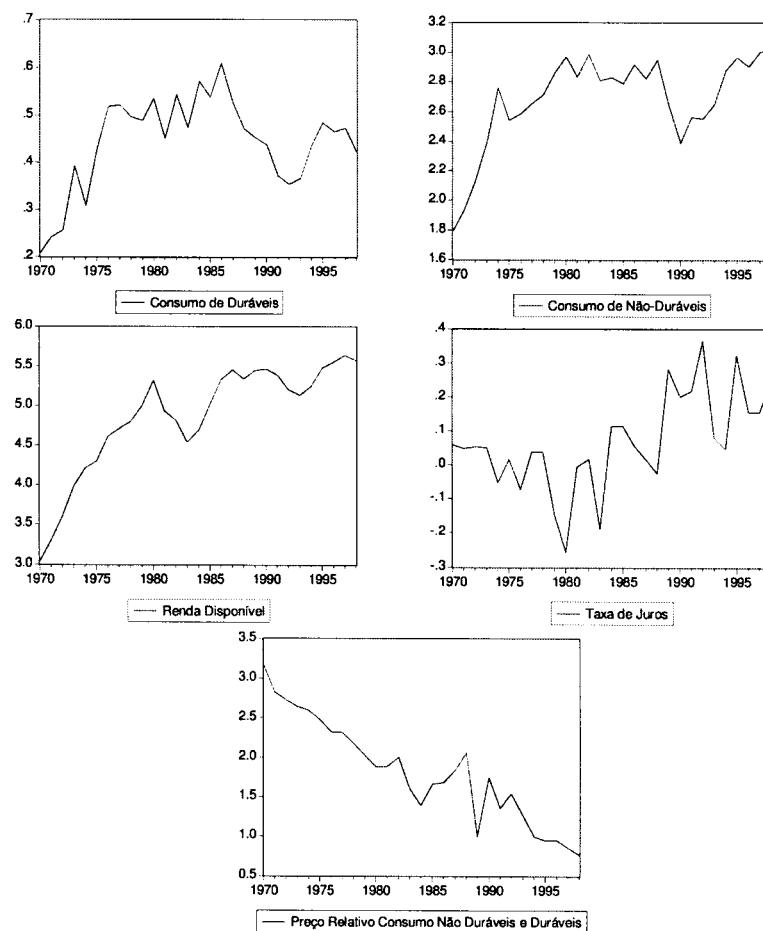
Abaixo na Figura 1 apresentamos os gráficos das séries de consumo duráveis e não-duráveis, renda disponível em termos per capita, além do preço relativo de consumo de

⁷Esta crítica não é pertinente no caso de Gomes (2003), uma vez que este autor considerou um termo $MA(1)$ na equação de teste.

⁸Ver www.ipeadata.gov.br.

duráveis e não-duráveis e a taxa de juros.

Figura 1



Por fim, apresentamos os resultados do teste ADF de raiz unitária das séries de taxa de juros e preço relativo. Como podemos observar na Tabela 1 os resultados sugerem que ambas as séries são estacionárias, no nível de significância de 5%. Como avaliaremos se a renda, o consumo de bens duráveis e não duráveis cointegram, para estas séries o teste de raiz unitária univariado é dispensável. Isto porque o teste de cointegração pode ser visto também como um teste multivariado de raiz unitária, pois se estas séries são estacionárias devemos estimar

3 vetores de cointegração. De fato, se existe uma relação desta natureza entre estas séries utilizar um teste de raiz unitária univariado consiste em um erro de especificação.

Tabela 1 - Testes de Raiz Unitária

Série	Estatística ADF	p-valor	n° de defasagens
Taxa de Juros	-4.14	0.0151	0
Preço Relativo	-5.15	0.0014	0

Nota: Os testes incluíram constante e tendência. O número de defasagens incluídas no teste foi escolhido usando o critério de informação de Akaike.

1.4 Estimações e Resultados

A primeira questão a ser discutida é a interdependência entre a decisão de gasto em bens duráveis e não duráveis: a separabilidade entre o gasto de não-duráveis e o estoque de duráveis na função utilidade do consumidor representativo. Para examinar tal questão, estimamos por GMM as equações (4) e (5), usando $\delta + g = 0.10$ na construção do estoque de duráveis, testando a hipótese nula que $\rho = 1$. Utilizamos 8 conjuntos distintos de instrumentos formados por variáveis que pertencem ao conjunto de informação do consumidor. Usamos o teste de sobreidentificação de Hansen (estatística $T \times J$) para testar a especificação do modelo utilizado. A Tabela 2 mostra os coeficientes estimados para o fator de desconto intertemporal, β , o parâmetro do estoque de bens duráveis da função de utilidade, α e a taxa de substituição intertemporal da função CES, ρ .⁹ Procedemos então o teste de Wald para avaliar a hipótese nula $\rho = 1$.

⁹Usamos $\delta + g = 0.10$. Para testar a robustez dos resultados, considerou-se também outros valores de δ verificando que não há mudança significativa nos resultados.

Tabela 2 -Estimação por MGM das equações (4) e (5) para $\delta + g = 0.10$

Modelo	Instrumentos	Coeficientes Estimados			Teste de Separabilidade ($H_0: \rho=1$)		Teste de Sobreidentificação	
		β	α	ρ	Estatística de Teste	p-valor	Estatística TJ de Hansen	p-valor
1	$cd_t, cd_{t-1}, r_{t-1}, r_t$	0.9847 ** (0.03068)	0.1713 (0.10495)	1.9261 * (0.76001)	1.4847	0.2230	6.7174	0.4589
2	$cd_{t-1}, y_{t-1}, y_{t-2}$	0.9972 ** (0.01859)	0.1548 (0.10310)	1.3665 (0.86991)	0.1775	0.6736	4.6851	0.4555
3	y_{t-1}	0.9924 ** (0.01905)	0.2221 (0.12240)	1.1178 (0.89502)	0.0173	0.8953	3.0488	0.0808
4	$cn_{t-1}, y_t, y_{t-1}, pk_{t-1}, r_{t-1}$	0.9628 ** (0.01530)	0.3088 ** (0.07794)	1.6403 ** (0.35852)	3.1898	0.0741	8.0226	0.5319
5	$cn_{t-1}, y_{t-1}, pk_{t-1}$	0.9918 ** (0.01750)	0.2532 ** (0.09275)	1.3279 ** (0.48153)	0.4637	0.4959	6.5895	0.6798
6	$k_{t-1}, y_{t-1}, y_{t-2}$	0.8990 ** (0.02316)	0.2142 ** (0.08339)	1.5603 * (0.58053)	0.9316	0.3344	5.6613	0.8950
7	$cd_t, cd_{t-1}, cd_{t-2}, y_t$	0.9806 ** (0.01366)	0.0381 (0.06181)	2.1086 (1.89014)	0.3440	0.5575	7.6738	0.1752
8	$k_{t-1}, k_{t-2}, r_{t-1}$	0.9985 ** (0.01935)	0.2098 * (0.08919)	1.6139 ** (0.49010)	1.5693	0.2103	5.7577	0.3305
media		0.9759	0.1965	1.5827				

*(**) denota rejeição de que o coeficiente é significativo ao nível de 5%(1%) de significância; desvios-padrões entre parênteses.

O teste de sobreidentificação de Hansen não rejeita a hipótese nula de que as condições de momento são válidas para todos os grupos de instrumentos utilizados. Os valores estimados para β ficaram próximos a 1, significativamente diferentes de zero para todos os conjuntos de instrumentos, com média 0.976. Isto representa uma taxa de desconto intertemporal de aproximadamente 2.5% ao ano.

Para todos os grupos de instrumentos testados não foi rejeitada a hipótese de que $\rho = 1$, ou seja, não se rejeita a separabilidade entre o gasto de não-duráveis e o estoque de duráveis na função utilidade do consumidor representativo. Esse resultado tem grande importância na medida em que corrobora todo um esforço passado de pesquisa que considerava implicitamente esta hipótese de separabilidade, e que estariam irremediavelmente condenados caso a separabilidade não fosse verificada.

Cabe ressaltar que, para o caso em que $\rho = 1$, o coeficiente α da função CES equivale à taxa marginal de substituição entre consumo de não-duráveis e estoque de duráveis. A taxa marginal de substituição estimada para os grupos de instrumentos utilizados é estatística-

mente inferior a unidade, com média de 0.20, dentre nossas estimativas. Logo, o consumidor brasileiro é indiferente entre 0.2 unidades de estoque de duráveis e uma unidade de bem de consumo não-durável.

O resultado do teste de separabilidade nos permite seguir considerando as restrições testáveis propostas por Vahid e Engle (1997) e Galí (1993), apresentadas na seção 2. A importância de se avaliar tais restrições é o fato de expressarem restrições de curto e longo prazo sobre a dinâmica das séries de gasto de duráveis e não-duráveis.

A Tabela 3 apresenta a estimativa do modelo $MA(1)$ de Mankiw (1982) para a primeira diferença do gasto com duráveis. Apesar do coeficiente ser negativo, como previsto pela teoria, o coeficiente da $MA(1)$ não foi significativamente diferente de zero aos níveis de significância de 5% ou 10%.

Tabela 3 - Estimativa do Modelo de Mankiw (1982)

$$\Delta cd_t = \varepsilon_t - (1 - \delta)\varepsilon_{t-1}$$

Variável Dependente: $\Delta(\text{Consumo de Duráveis})$		
Regressor	(1)	(2)
Constante	0.0078 (0.7558)	-
$-(1-\delta)$	-0.2211 (-1.0504)	-0.2005 (-1.0958)

Nota: Estatística t entre parênteses. Utilizamos o método proposto por Newey & West (1987) para estimar a matriz de covariância dos resíduos.

Os resultados da Tabela 3 sugerem que o consumo de não duráveis é um passeio aleatório. Nesse caso, se a teoria sobre o gasto de não-duráveis fosse válida, a soma do gasto de duráveis e não-duráveis seria um passeio aleatório. Logo, o fato do termo $MA(1)$ ser não significativo pode validar estudos anteriores que usaram o passeio aleatório como *benchmark* para a série de consumo agregado, seguindo Hall (1978). Além disso, a rejeição da hipótese de que o consumo total não é um passeio aleatório, obtida por esses autores, não pode ser atribuída a um problema de viés de agregação.

Para testar as restrições do modelo de Vahid e Engle (1997), de que, para $\gamma \neq 0$, $\Delta cn_t - \gamma \Delta cd_t \sim MA(1)$, e, de que, o gasto de duráveis e não duráveis são cointegrados, estimou-se um VAR contendo o consumo de bens duráveis e não duráveis. Para escolher a ordem deste VAR utilizou-se um teste de Razão de Verossimilhança, aliado ao uso de critérios

de informação. A especificação de um $VAR(1)$ parece ser um modelo razoável. Este modelo foi estimado e, em seguida, foi realizado o teste de cointegração de Johansen. Os resultados são apresentados na Tabela 4 abaixo.

Tabela 4 - Teste de Cointegração de Johansen

Modelo $VAR(1)$ para cd e cn

Teste do Posto de Cointegração sem restrição

Ho: N° de Eq. de Cointegração	Autovalor	Teste do Traço		Teste do max autovalor	
		Estatística de teste	Valor Crítico 5%	Estatística de teste	Valor Crítico 5%
Nenhuma	0.313342	11.5564	12.53	10.5257	11.44
No máximo 1	0.036141	1.0306	3.84	1.0307	3.84

*(**) denota rejeição da hipótese no nível de 5%(1%) de significância

Teste do traço indica não cointegração no nível de 5% de significância

Teste do Máximo autovalor indica não cointegração no nível de 5% de significância

Tanto o teste baseado na estatística do traço quanto na estatística do autovalor máximo não rejeitam a hipótese de não existência de cointegração, no nível de significância de 5%. Há cointegração ao nível de 10%, mas esse resultado não é inequívoco. Incluímos então a renda real no VAR, testando novamente a existência de cointegração, conforme sugere a teoria no caso da riqueza real ser limitada.

Usando o gasto de duráveis, não-duráveis, e renda, a ordem do VAR foi escolhida em 3 defasagens, após consulta a diversos testes de diagnóstico e critérios de informação. Procedeu-se então o teste de cointegração de Johansen (1988), apresentado na Tabela 5.

Tabela 5 - Teste de Cointegração de Johansen(1988)

Modelo VAR(3) para cd , cn e y

Teste do Posto de Cointegração sem restrição

Ho: N° de Eq. de Cointegração	Autovalor	Teste do Traço		Teste do max autovalor	
		Estatística de teste	Valor Crítico 5%	Estatística de teste	Valor Crítico 5%
Nenhuma	0.508875	26.6917	24.31	18.4875	17.89
No máximo 1	0.248983	8.2043 *	12.53	7.4445 *	11.44
No máximo 2	0.028798	0.7597	3.84	0.7597	3.84

*(**) denota rejeição da hipótese no nível de 5%(1%) de significância

Teste do traço indica 1 equação de cointegração no nível de 5% de significância

Teste do Máximo autovalor indica 1 equação de cointegração no nível de 5% de significância

Coeficientes da Equação de Cointegração Normalizados

CN	CD	RENDADISP
1	-2.682945	-0.295274
	(0.40866)	(0.03788)

Desvio-padrão entre parênteses

Tanto o teste baseado na estatística do traço como o teste baseado na estatística do autovalor máximo não rejeitam a hipótese nula de que existe 1 vetor de cointegração, no nível de significância de 5%. Sendo a ordem das variáveis do VAR cn_t , cd_t e y_t , respectivamente, o vetor de cointegração estimado é $(1, -2.6829, -0.2953)$, o que confirma uma das proposições testáveis de Vahid e Engle (1997). Utilizando uma série de consumo agregada, e uma medida da renda, Issler e Rocha (2000) e Reis *et alli* (1998) obtiveram um resultado análogo, não rejeitando a existência de um vetor de cointegração entre consumo e renda.

Uma vez confirmado o comovimento de longo prazo entre as séries de consumo e renda, passamos a analisar a possibilidade de existir comovimentos de curto prazo nas mesmas, i.e., a existência de ciclos comuns ou de ciclos co-dependentes entre elas.¹⁰ A importância de avaliar se as séries possuem alguma característica comum de curto prazo está no fato de que o estudo da dinâmica de curto prazo do conjunto pode ser resumido ao estudo da dinâmica de apenas uma delas. Assim, se este for o caso, o comportamento de curto prazo da renda é decisivo para o entendimento do comportamento das séries de consumo, o que pode indicar a presença de restrição à liquidez no consumo, conforme argumentam Campbell e Mankiw (1989).

¹⁰Estes conceitos são brevemente discutidos no apêndice.

Testa-se a presença de características comuns de curto prazo utilizando os testes de correlação canônica descritos em Vahid e Engle (1993). Usamos 4 defasagens da primeira diferença das séries de consumo de duráveis e não duráveis e renda além da primeira defasagem do termo de correção de erros como instrumentos para calcular as correlações canônicas. Os resultados estão expostos na Tabela 6 abaixo.

Tabela 6 - Teste para Ciclos Comuns e Ciclos Codependentes

cn_t , cd_t e y_t

Teste de Ciclos Comuns

Ho: 1 Vetor de Ciclo Comum		Ho: 2 Vetores de Ciclo Comum		Ho: 3 Vetores de Ciclo Comum	
Estatística de Valor Crítico		Estatística de Valor Crítico		Estatística de Valor Crítico	
Teste	5%	Teste	5%	Teste	5%
9.42	19.68	28.62	36.42	64.17 *	54.57

*(**) denota rejeição de Ho a 5%(1%) de significância

Não rejeita-se a hipótese de 2 ciclos comuns a 5% de significância

Teste de Ciclos Codependentes

Ho: 1 Vetor de Ciclo Codependente		Ho: 2 Vetores de Ciclo Codependente		Ho: 3 Vetores de Ciclo Codependente	
Estatística de Valor Crítico		Estatística de Valor Crítico		Estatística de Valor Crítico	
Teste	5%	Teste	5%	Teste	5%
8.95	19.68	25.98	36.42	49.13	54.57

*(**) denota rejeição de Ho a 5%(1%) de significância

Ao contrário dos resultados encontrados por Vahid e Engle para o caso americano, não encontramos evidência a favor da existência de ciclos codependentes, uma vez que os testes não rejeitaram a existência de ciclos comuns entre o consumo de duráveis, não duráveis e renda, ao nível de significância de 5%. Os testes apontam a existência de dois vetores de ciclos comuns, isto é, existem 2 combinações lineares não-nulas e linearmente independentes entre as primeiras diferenças das séries que geram uma série que é uma inovação.

Os testes de restrição a liquidez para o gasto de duráveis são apresentados abaixo na Tabela 7. Dado que o componente $MA(1)$ foi não significativo em todos os casos, optou-se por reportar apenas a equação de teste quando se elimina o termo $MA(1)$. As estimativas foram feitas utilizando método de variáveis instrumentais, sendo utilizado como instrumentos variáveis defasadas pelo menos em 2 períodos. Na Tabela 7 reportamos os resultados utilizando 5 conjuntos distintos de instrumentos. Em todos os casos não foi rejeitado o teste

de sobre-identificação da validade dos instrumentos, no nível de significância de 5%. As estimativas do coeficiente referente a renda é significativamente diferente de zero, ao nível de significância de 5%, em quase todos os casos. Quanto ao coeficiente da taxa de juros, em todos os casos, sua estimativa é não significativa, ao nível de significância de 5%.

Tabela 7 - Teste de Restrição à liquidez

$$\Delta cd_t = \lambda \Delta y_t + (1 - \lambda) \theta r_t + (1 - \lambda) \varepsilon_t$$

Modelo	Instrumentos	$\lambda^{(a)}$	$\theta^{(a)}$	$(1-\lambda)\theta^{(a)}$	Teste de Sobreidentificação TJ ^(b)
1	$\Delta cd_{t-2}, \Delta Y_{t-2}, \Delta cn_{t-2}, r_{t-2}$	0.3661 (0.0055)	-0.2498 (0.2878)	-0.1583 (0.2604)	2.4308 (0.2966)
2	$\Delta cd_{t-2}, \Delta Y_{t-2}, \Delta cd_{t-3}, \Delta Y_{t-3}, r_{t-2}$	0.0866 (0.4573)	-0.0321 (0.8430)	-0.0293 (0.8399)	3.2877 (0.3494)
3	$\Delta cd_{t-2}, \Delta Y_{t-2}, \Delta cd_{t-3}, \Delta Y_{t-3}, \Delta cn_{t-2}, r_{t-2}$	0.2024 (0.0071)	-0.1346 (0.1773)	-0.1074 (0.1644)	4.3860 (0.3563)
4	$\Delta cd_{t-2}, \Delta Y_{t-2}, \Delta cd_{t-3}, \Delta Y_{t-3}, \Delta cn_{t-2}, \Delta cn_{t-3}, r_{t-2}$	0.1875 (0.0172)	-0.1528 (0.1271)	-0.1241 (0.1094)	4.2271 (0.5172)
5	$\Delta cd_{t-2}, \Delta Y_{t-2}, pk_{t-2}, r_{t-2}$	0.2288 (0.0207)	-0.1120 (0.2164)	-0.0864 (0.2134)	2.7443 (0.2536)

Nota: (a) entre parênteses o p-valor para o teste de significância do respectivo coeficiente; (b) entre parênteses o p-valor para a estatística do teste de Hansen; Nas estimativas por GMM utilizou-se o núcleo de Bartlett com janela de Andrews, além da opção de prewhitening.

A estimativa mediana de λ é aproximadamente 0.20, e significativa para a maioria dos instrumentos utilizados. Para comparar este valor com o valor obtido por Reis *et alli* (1998) de 0.8, por Issler e Rocha (2000) de 0.74 e por Gomes (2003) de 0.85 é preciso levar em conta que ambos trabalharam com uma série de consumo que agrega bens duáveis e não duráveis, $c_t = cd_t + cn_t$. Como a razão cd_t/c_t é bastante estável e possui média aproximadamente igual a 0.14, multiplicando Δc_t por este valor obtemos uma aproximação bastante razoável para Δcd_t . Como sabemos, multiplicar a variável dependente por uma constante implica que os coeficientes estimados também ficam multiplicados por esta constante. Isto posto, multiplicamos 0.8, 0.74 e 0.85 por 0.14 obtendo então as estimativas 0.11, 0.10 e 0.12, respectivamente, valores inferiores a nossa estimativa mediana.¹¹ Logo, apesar de obtermos

¹¹ Como $cd_t/c_t \approx 0.14$, é possível mostrar que $\Delta cd_t \approx 0.14 \Delta c_t$. Reis *et alli* (1998) e Issler & Rocha (2002) consideraram $\Delta c_t = \lambda \Delta y_t + (1 - \lambda) \varepsilon_t$. Multiplicando por 0.14 obtemos, $\Delta cd_t \approx 0.14 \lambda \Delta y_t + 0.14(1 - \lambda) \varepsilon_t$, daí fica claro a necessidade de multiplicar os valores obtidos por estes autores por 0.14.

um coeficiente menor para a renda, obtivemos resultados mais fortes de restrição à liquidez do que os obtidos anteriormente. Isso não nos surpreende, pois usamos a série de gasto de duráveis, que, na maioria dos casos, requer maior crédito para que o gasto seja efetuado.

Dado que o coeficiente da taxa de juros é não-significativo em todas as estimativas da Tabela 7, consideramos reestimar as mesmas equações omitindo o juros real. Os resultados são apresentados abaixo na Tabela 8.

Tabela 8 - Teste de Restrição à liquidez

$$\Delta cd_t = \lambda \Delta y_t + (1 - \lambda) \varepsilon_t$$

Modelo	Instrumentos	$\lambda^{(a)}$	Teste de Sobreidentificação ^(b)
1	$\Delta cd_{t-2}, \Delta Y_{t-2}, \Delta cn_{t-2}, r_{t-2}$	0.2464 (0.0430)	3.9820 (0.2634)
2	$\Delta cd_{t-2}, \Delta Y_{t-2}, \Delta cd_{t-3}, \Delta Y_{t-3}, r_{t-2}$	0.0536 (0.4475)	4.0423 (0.4003)
3	$\Delta cd_{t-2}, \Delta Y_{t-2}, \Delta cd_{t-3}, \Delta Y_{t-3}, \Delta cn_{t-2}, r_{t-2}$	0.1141 (0.1196)	7.9056 (0.1615)
4	$\Delta cd_{t-2}, \Delta Y_{t-2}, \Delta cd_{t-3}, \Delta Y_{t-3}, \Delta cn_{t-2}, \Delta cn_{t-3}, r_{t-2}$	0.1263 (0.0690)	7.8784 (0.2471)
5	$\Delta cd_{t-2}, \Delta Y_{t-2}, pk_{t-2}, r_{t-2}$	0.2070 (0.0448)	2.9833 (0.3942)

Nota: (a) entre parênteses o p-valor para o teste de significância do coeficiente da renda; (b) entre parênteses o p-valor para a estatística de teste do teste de sobreidentificação; Nas regressões utilizou-se o método proposto por Newey & West (1987) para estimar a matriz de covariância dos resíduos.

Observamos na Tabela 8 que em todos os casos não se rejeita o teste de sobreidentificação, no nível de significância de 5%, porém nos casos 2 e 3 não se rejeita que λ seja nulo. A estimativa mediana de λ é aproximadamente igual a 0.14. Se considerarmos apenas os conjuntos de instrumentos 1, 4 e 5 este valor torna-se aproximadamente 0.20.¹² Mais uma vez, apesar de obtermos um coeficiente menor para a renda, obtivemos resultados mais fortes de restrição à liquidez do que os obtidos anteriormente.

Vale notar que a equação estimada na Tabela 8 descreve um ciclo comum entre o gasto de duráveis e a renda, refletindo a possível existência de restrição à liquidez. De acordo com

¹²Estimamos ainda o modelo $\Delta cd_t = \beta r_t + \varepsilon_t$, porém em todos os casos a estimativa de β é não significativa, no nível de 5% de significância. Por isso não reportamos estas estimativas.

essa visão, o comportamento cíclico do gasto de duráveis no Brasil pode ser fruto de restrição à liquidez e não do ajuste ótimo do estoque de duráveis às flutuações da renda permanente. Esse tipo de restrição, se relaxada, resultaria num estoque maior de bens duráveis, e, por consequência, num aumento de bem estar para os consumidores brasileiros.

1.5 Conclusões

Neste artigo investigamos a evolução da renda, consumo de bens duráveis e não duráveis no Brasil, com enfoque especial sobre a decisão de gasto em bens duráveis. Primeiramente, avaliamos se existe separabilidade na função utilidade entre o gasto de não-duráveis e o estoque de duráveis. Esta investigação é muito importante para validar os estudos anteriores, uma vez que todos assumiram implicitamente esta hipótese. O teste de hipótese, realizado com base nas equações de Euler do problema do consumidor, utilizando a função de utilidade do tipo CES, estimada a partir de GMM, não rejeita a separabilidade.

A partir do resultado de separabilidade, foi possível obter ainda os seguintes resultados: *i*) os testes de cointegração não rejeitaram a hipótese da existência de um vetor de cointegração entre consumo de duráveis, não duráveis e renda; *ii*) os testes de ciclo comum não rejeitaram a existência de ciclos comuns entre o consumo de duráveis, não duráveis e renda; *iii*) os testes de restrição à liquidez sugerem que não se pode rejeitar que o consumo de bens duráveis é restrito ao crédito, com restrição superior à encontrada para o consumo total, conforme os resultados de Cavalcanti (1993), Reis et alli (1998), Issler e Rocha (2000) e Gomes (2003).

Os resultados *i*) e *ii*) acima sugerem que, no Brasil, o consumo de duráveis, não duráveis e renda apresentam comovimentos de curto e longo prazo. O resultado *iii*) evidencia que uma fração de consumidores consome a sua renda corrente e outra suaviza seu consumo, de acordo com a Teoria da Renda Permanente. Em ambos os casos, esperamos que o consumo (de bens duráveis e não duráveis) cointegrem com a renda, logo não é possível interpretar o resultado *i*) acima como uma evidência a favor de qualquer um dos casos. No entanto, a existência de ciclo comum entre as séries em questão é uma evidência a favor da hipótese de restrição ao crédito. Tal interpretação é confirmada com base nas estimativas obtidas de λ . Portanto, se o consumo corrente é muito dependente da renda corrente, ao invés da renda permanente, podemos concluir que: *i*) políticas econômicas que afetam a renda têm impacto direto sobre grande parcela do consumo; *ii*) políticas econômicas que visem expandir o crédito têm um

impacto potencial significativo em aumentar o bem-estar dos consumidores.

Referências

- [1] Bacchetta e Stefan, “Consumption and credit constraints: International evidence” , *Journal of Monetary Economics*, 40(2), p.207-238, 1997.
- [2] Bernanke, B., “Adjustment Costs, Durable Goods and Aggregate Consumption,” *Journal of Monetary Economics*, 1985.
- [3] Beveridge, S. e Nelson, C. R., “A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the Business Cycle” *Journal of Monetary Economics* 7(2), p.151-174, 1981.
- [4] Browning, M. e Crossley, T., “The Life-cycle Model of Consumption and Saving,” *Journal Economic Perspectives*, 15, p. 3-22, 2001.
- [5] Bugarin, M. e Ellery, R. Jr. “Liquidity constraints and the behavior of the aggregate consumption over the Brazilian business cycle”, *Estudos Econômicos*, 32(4), p. 551-576, 2002.
- [6] Campbell, J. “Does Saving Anticipate Declining Labor Income? An Alternative Test of the Permanent Income Hypothesis,” *Econometrica*, 55(6), p. 1249-1273, 1987.
- [7] Cavalcanti, Carlos B. “Intertemporal substitution in consumption: an empirical investigation for Brazil”, *Brazilian Review of Econometrics*, 13(2), p. 203-229, 1993.
- [8] Cook, S., “Long-run and Short-run Co-movement in UK Consumption and Income,” *Economics Letters*, 67, p. 11-13, 2000.
- [9] Deaton, Angus S. “Understanding Consumption”. Oxford University Press, New York, 1992.
- [10] Ellery, R. Jr., Gomes, V. e Sachsida, A. “Business cycle fluctuations in Brazil”, *Revista Brasileira de Economia*. 56(2), p. 269-308, 2002.

- [11] Engle, R. F. e Kozicki, S., "Testing for Common Features". *Journal of Business and Economics Statistics* 11, p.369-395, 1993.
- [12] Flavin, M., "The Adjustment of Consumption to Changing Expectations About Future Income", *Journal of Political Economy*, 89(5), 1981.
- [13] Flavin, M., "The Excess Smoothness of Consumption: Identification and Interpretation", *Review of Economic Studies*, 60(204), 1993.
- [14] Hall, Robert., "Stochastic Implications of the Life Cycle Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy*, 1978.
- [15] Issler, J. V. e Rocha, F. P., "Consumo, Restrição a Liquidez e Bem-Estar no Brasil", *Economia Aplicada*, 4(4), p. 637-665, 2000.
- [16] Issler, J.V. e Vahid. F., "Common Cycles and the Importance of Transitory Shocks to Macroeconomic Aggregates," *Journal of Monetary Economics*, 47(3), p. 449-475, 2001.
- [17] Mankiw, G., "Hall's Consumption Hypothesis and Durable Goods", *Journal of Monetary Economics*, 10, p. 417-25, 1981.
- [18] Newey, W. e West, K., "A Simple Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica*, 55, p.703-708, 1987.
- [19] Ogaki, M. e C. Reinhart, "Measuring Intertemporal Substitution: The Role of Durable Goods," *Journal of Political Economy*, 106(5), p. 1078-98, 1998..
- [20] Reis, E., Issler, J. V., Blanco, F. e Carvalho, L., "Renda Permanente e poupança Precaucional: Evidências Empíricas para o Brasil no Passado Recente", *Pesquisa e Planejamento Econômico*, 28(2), p. 233-272, 1998.
- [21] Vahid, F., Engle, R. "Codependent Cycles", *Journal of Econometrics*, 80(2), p.199-121, 1997.

1.A Apêndice

1.A.1 Revisão Econométrica

Considere o modelo de autorregressão vetorial (VAR),

$$x_t = A_1 x_{t-1} + A_2 x_{t-2} + \dots + A_p x_{t-p} + \varepsilon_t,$$

em que $x_t = (x_{1t}, x_{2t}, \dots, x_{nt})'$ é um vetor $n \times 1$ de variáveis do sistema e ε_t é $n \times 1$, os choques de cada equação, tal que $\varepsilon_t \sim iid(0, \Omega)$. Os componentes do vetor x_t serão cointegrados de ordem d , b , $x_t \sim CI(d, b)$, se: *i*) todos os componentes de x_t são $I(d)$; *ii*) existe um vetor $\alpha (\neq 0)$ tal que $\alpha' x_t \sim I(d - b)$, $b > 0$. O vetor α é chamado vetor de cointegração (Engle e Granger, 1987).

A partir do modelo VAR podemos definir também os conceitos de ciclo comum e ciclo codependente. Engle e Kozicki (1993) introduziram o conceito de ciclo comum através de uma medida de comovimento contemporâneo chamada *serial correlation common feature*, SCCF. Esta medida verifica a existência de uma combinação linear entre séries estacionárias que seja imprevisível quando se considera o passado histórico destas variáveis.¹³ Vahid e Engle (1997) por sua vez consideraram a possibilidade de existir uma combinação linear entre as séries que reduz a dependência do passado, mas não a elimina totalmente, sendo este o conceito de ciclos codependentes. Formalmente, considere o seguinte exemplo:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} f_t \\ f_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

$$f_{t+1} = \alpha f_t + \theta' \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

em que ε_{1t} e ε_{2t} são inovações e y_{1t} e y_{2t} são variáveis estacionárias. Pode-se observar que,

$$y_{1t} - \alpha y_{2t} = \begin{pmatrix} 1 & -\alpha \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} + \theta' \begin{pmatrix} \varepsilon_{1t-1} \\ \varepsilon_{2t-1} \end{pmatrix}$$

Assim, a combinação linear $y_{1t} - \alpha y_{2t}$ só tem correlação com sua primeira defasagem, configurando-se a presença de ciclos codependentes. Por outro lado, se existir uma combinação linear entre as séries estacionárias que produz uma série que é uma inovação, caso em que $\theta = 0$, temos a presença de ciclos comuns como definido por Engle e Kozicki (1993).

¹³Note que tais séries devem possuir algum componente autoregressivo, para que estejamos de fato analisando características comuns.

1.A.2 Cálculo do Estoque Inicial de Bens de Consumo Duráveis

Considere a lei de movimento do estoque de bens de consumo duráveis:

$$k_t = cd_{t-1} + (1 - \delta)k_{t-1} \quad (1)$$

Daí,

$$k_t = (1 - \delta)^t k_0 + \sum_{j=1}^t (1 - \delta)^{j-1} cd_{t-j} \quad (2)$$

Através desta lei de movimento é possível obter ainda:

$$\begin{aligned} k_0 &= cd_{-1} + (1 - \delta)[cd_{-2} + (1 - \delta)k_{-2}] \\ &= (1 - \delta)^T k_{-T} + \sum_{j=1}^T (1 - \delta)^{j-1} cd_{-j} \end{aligned} \quad (3)$$

Supondo que o gasto em bens duráveis tenha evoluído à taxa g até o instante zero, isto é, $cd_{-j} = cd_0(1 + g)^{-j}$, isto implica que a equação (14) torna-se:

$$k_0 = (1 - \delta)^T k_{-T} + \left(\frac{cd_0}{1 + g} \right) \sum_{j=0}^{T-1} \left(\frac{1 - \delta}{1 + g} \right)^j \quad (4)$$

Note que $1 - \delta < 1 + g$, logo tomando o limite da equação (4) obtemos:

$$k_0 = \frac{cd_0}{g + \delta} \quad (5)$$

Portanto, obtemos k_0 como função de i_0 , g e δ . Apesar de possuir o valor i_0 , é preciso obter um valor que seja representativo do processo do gasto em duráveis de $-\infty$ a 0. Para tanto considere a seguinte estimação da taxa de crescimento média para dados discretos. No modelo contínuo temos, $cd_t = cd_0 e^{gt}$, onde g é taxa de crescimento. Para estimar g a partir de um modelo de regressão linear fazemos $\ln(cd_t) = \ln(cd_0) + gt$. Assim, ao estimar o valor da taxa de crescimento do caso discreto, g_d , fazemos:

$$g_d = \frac{cd_t - cd_{t-1}}{cd_{t-1}} = \frac{e^{gt} - e^{g(t-1)}}{e^{g(t-1)}} = e^g - 1$$

Definimos então o gasto em bens duráveis estacionário:

$$cd_t^{\text{Estacionário}} \equiv \frac{cd_t}{(1 + g_d)^t}$$

Assim o gasto inicial é dado pela média do gasto estacionário:

$$cd_0 = \frac{1}{T} \sum_{j=1}^T i_t^{Estacionário} \quad (6)$$

Utilizando as equações (5) e (6) encontramos k_0 como função de g e δ . Determinando então estes dois parametros e utilizando a equação (2) obtemos toda série de estoque de capital. Para tanto considere o seguinte:

$$\begin{aligned} k_t &= \int_{-\infty}^t e^{\delta(\tau-t)} cd_{\tau} d\tau \\ &= \int_{-\infty}^t e^{\delta(\tau-t)} (e^{g(\tau-t)} cd_t) d\tau \\ &= \frac{cd_t}{(\delta + g)} \end{aligned} \quad (7)$$

Assim dividindo ambos os lados da equação (14) por y obtemos:

$$\frac{k_t}{y_t} = \frac{1}{(\delta + g)} \frac{cd_t}{y_t}$$

Como g é taxa de crescimento do gasto em duráveiso, conhecendo δ é obtido o capital inicial. Utilizamos então 5 valores distintos para δ , como reportado no texto.

Capítulo 2

Disparidades Regionais ou Educacionais?

Um Exercício Contrafactual*

Márcio Antônio Salvato[†]

Resumo

Este trabalho investiga o impacto da escolaridade sobre a distribuição de renda do trabalho de estados/regiões do Brasil. Usando um método semi-paramétrico, discutido em DiNardo, Fortin & Lemieux (1996), mensuramos o quanto dos diferenciais de renda entre as Regiões Nordeste e Sudeste do Brasil - a mais pobre e a mais rica do país - e entre os Estados do Ceará e São Paulo, podem ser explicados pelas diferenças de escolaridade da população residente. Usando dados da PNAD construímos densidades contrafactuais reponderando a distribuição da região/estado mais pobre pelo perfil de escolaridade da mais rica. Concluímos que: (i) mais de 50% do diferencial de renda é explicado pelo diferencial de escolaridade; (ii) os decis mais elevados da distribuição de renda têm maior ganho com o aumento da escolaridade, se aproximando muito da distribuição de renda do trabalho da região/estado mais rica e; (iii) o aumento da escolaridade, mantendo-se a estrutura de salários, agrava a desigualdade de renda nas regiões/estados mais pobres.

JEL Classification Codes: C14; I20 e J31.

*Artigo feito em co-autoria com Pedro Cavalcante (EPGE/FGV) e Angelo Duarte (EPGE/FGV e BACEN). Os autores agradecem a Luiz Henrique Braidó, Luiz Renato Lima, Carlos Eugênio da Costa, Samuel Pessoa, Marcelo Fernandes e Renato Flôres e participantes dos seminários da EPGE/FGV pelas sugestões e comentários.

[†]EPGE/FGV, PUC.Minas e CEPE, salvato@fgvmail.br.

2.1 Introdução

A desigualdade de renda existente entre as regiões brasileiras, notadamente entre a Região Sudeste e as Regiões Norte e Nordeste, é objeto permanente de discussões entre acadêmicos, elaboradores de políticas governamentais e membros da classe política. Ainda no século XIX o Imperador D. Pedro II já demonstrava sua preocupação ao afirmar que venderia sua coroa para matar a fome dos nordestinos. Ao longo desses quase 150 anos, vários foram os diagnósticos apresentados e as soluções e programas implementados com objetivo de diminuir essa desigualdade. As políticas governamentais se intensificaram e se tornaram sistemáticas a partir da década de 50 quando, sob a liderança do economista Celso Furtado, foi publicado o documento “Uma Política para o Desenvolvimento do Nordeste” que sugeria a industrialização como forma de diminuir a disparidade entre as rendas das regiões nordeste e sudeste. Esse documento inspirou a criação da SUDENE e da SUDAM na qualidade de órgãos planejadores e executores das políticas de combate às desigualdades regionais, e do BNB e BASA na qualidade de agentes financeiros de desenvolvimento regional. A Constituição Federal de 1988 não foi silente em relação à questão e determinou a alocação de 3% das receitas de impostos sobre a renda e produtos industrializados em programas de financiamento ao setor produtivo nas Regiões Norte, Nordeste e Centro-Oeste.

Apesar da elevada quantidade de órgãos e programas governamentais a nível federal e estadual que se prestam ao desenvolvimento econômico regional, as políticas governamentais implementadas nesses últimos 50 anos se utilizaram fortemente dos mesmos instrumentos: concessão de incentivos fiscais e crédito público subsidiado à iniciativa privada e investimento estatal em infra-estrutura. Estima-se que entre 1989 e 2002 os Fundos Constitucionais de desenvolvimento regional tenham aplicado recursos da ordem de US\$ 10 bilhões. Embora bem sucedidos em algumas dimensões, como a de acelerar o processo de industrialização nas Regiões Nordeste e Norte, essas políticas não foram competentes em transformar os indicadores sociais, reduzir o nível de pobreza e modificar a distribuição de renda da região.

Na última década surgiu uma nova abordagem do problema de desigualdade regional no Brasil. Ela baseia-se na tese de que a baixa renda *per capita* nas Regiões Norte e Nordeste está relacionada com a concentração de indivíduos com baixa escolaridade (capital humano) e baixo capital físico, o que fazem com que suas rendas sejam pequenas. Dessa forma a redução da desigualdade regional confunde-se com o combate à pobreza, e para tal necessita-se de uma política de educação e qualificação profissional e programas de acesso ao crédito.

namos de distribuição contrafactual, a qual pode ser comparada com a distribuição original dos salários da população. No nosso caso, trabalharemos com a distribuição da renda do trabalho, uma vez que não se dispõe de dados de salários para o Brasil.

Essa abordagem tem o mesmo fundamento da decomposição de Oaxaca (1973)³, que se baseia na construção de contrafactuais, mas ao contrário deste - que realizou seu trabalho analisando somente as médias das distribuições - analisaremos toda a distribuição. A metodologia que empregamos nos permite visualizar claramente a função de densidade da renda do trabalho, assim como observar as alterações que ocorreriam nessa distribuição caso a população sofresse qualquer alteração nas condições educacionais, mantendo-se a estrutura de remuneração do trabalho originalmente vigente. Neste sentido, ao invés de focarmos o estudo de diferencial de renda e educação em algum índice de desigualdade, como por exemplo os coeficientes de Gini e de Theil ou a variância do logaritmo dos salários, pretendemos extrair nossas conclusões observando toda a distribuição de renda do trabalho, usando medidas de diferenças de distribuições e teste de desigualdade de distribuições de Kolmogorov-Smirnov.

A construção das distribuições contrafactuais se faz reponderando a amostra conforme alguma característica que se pretende estudar. Ou seja, para estudarmos, como no nosso caso, os efeitos da escolaridade sobre a renda do trabalho, estimamos a distribuição da renda do trabalho reponderando-se as amostras disponíveis de tal forma que as mesmas passem a compor uma população com o perfil de escolaridade que se deseja. Dessa forma, podemos obter a distribuição da renda do trabalho de uma região caso ele tivesse o perfil de escolaridade de outra região, mantendo a sua estrutura de remuneração original. Deve ficar claro que existe uma limitação dessa metodologia uma vez que são considerados apenas efeitos parciais, não sendo possível analisar os efeitos em equilíbrio geral. No entanto esta abordagem será bastante útil para responder perguntas do tipo: qual seria a distribuição de renda do trabalho no Estado do Ceará se as condições de educação fossem semelhantes às do Estado de São Paulo?

No presente trabalho mensuramos a mudança na distribuição de renda do trabalho do Estado do Ceará e da Região Nordeste caso a população residente nessas regiões obtivessem

³A abordagem de Oaxaca é geralmente restrita à comparação de médias. Quando as distribuições são unimodais, simétricas e têm variâncias semelhantes este procedimento é bem qualificado para analisar as mudanças na estrutura dos salários. No entanto, não se pode esperar que estas condições sejam válidas quando estamos comparando distribuições de salários de regiões diferentes. Daí a opção por uma estimação de toda a função de densidade.

a mesma escolaridade do Estado de São Paulo e da Região Sudeste⁴.

Na seção seguinte apresentamos alguns fatos estilizados que ilustram o problema de desigualdade no Brasil. Na seção 3 apresentamos os dados utilizados na análise a que se propõe este trabalho. Na seção 4 discutimos detalhadamente a metodologia de estimação de densidade de núcleo ponderada (*weighted kernel density estimator*) e a construção de densidades contrafactuais, além de apresentarmos algumas medidas paramétricas e não-paramétricas utilizadas para comparar as densidades estimadas. Os resultados são apresentados na seção 5, e finalmente, na seção 6, concluímos.

2.2 Fatos Estilizados

A desigualdade entre as regiões e estados brasileiros pode ser verificada tanto nos indicadores de bem-estar da população como no nível de renda dos seus habitantes. A diferença entre os Índices de Desenvolvimento Humano (IDH) das regiões diminuiu entre 1991 e 2000; a distância entre os IDHs das regiões Nordeste e Sudeste, por exemplo, que era 0,16 em 1991, passou para 0,12 em 2000 (Tabela 1)⁵. No entanto a posição relativa das regiões não mudou desde 1970, quando o índice foi computado pela primeira vez. Além disso, a posição relativa dos estados também não apresentou expressiva mudança no mesmo período, haja visto que os nove estados nordestinos sempre se situaram entre as 11 unidades da federação com piores IDH no período compreendido entre 1970 e 2000. A desigualdade de renda, medida pelos coeficientes de Theil e de Gini⁶, se agravou em todas as regiões brasileiras nas décadas de 70 e 80, e apresentou uma leve melhora na década seguinte; no entanto a piora foi mais acentuada nas Regiões Norte e Nordeste, o que as tornou ainda mais desiguais quando comparadas com a Região Sudeste (Tabela 2).

⁴Como será visto na seção seguinte, de fato não consideramos a Região Sudeste integralmente.

⁵Os IDHs das regiões foram calculado pelos autores, seguindo a metodologia do PNUD.

⁶O valor de ambos os coeficientes varia entre 0, quando não há desigualdade, e 1, quando a desigualdade é máxima.

Tabela 1
Índice de Desenvolvimento Humano - IDH

Índice de Desenvolvimento Humano	Ano	Brasil	Regiões				
			Centro-Oeste	Nordeste	Norte	Sudeste	Sul
IDH-Educação	1991	0.745	0.778	0.606	0.705	0.812	0.804
	2000	0.849	0.877	0.762	0.818	0.886	0.895
IDH-Longevidade	1991	0.668	0.682	0.587	0.637	0.709	0.715
	2000	0.731	0.747	0.669	0.706	0.759	0.776
IDH-Renda	1991	0.681	0.699	0.564	0.614	0.732	0.689
	2000	0.723	0.747	0.614	0.640	0.768	0.747
IDH-Total	1991	0.698	0.720	0.586	0.652	0.751	0.736
	2000	0.768	0.790	0.682	0.722	0.805	0.806

Fonte: autores

Tabela 2
Coeficientes de Theil e de Gini da Renda do Trabalho

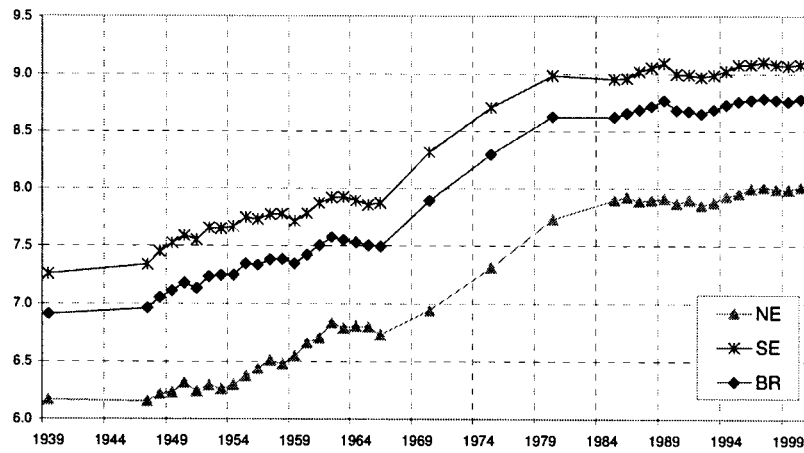
	Theil			Gini		
	1970	1980	1991	1981	1990	2001
Região Norte	0,44	0,56	0,72	0,51	0,58	0,57
Região Nordeste	0,57	0,65	0,78	0,57	0,63	0,60
Região Centro-Oeste	0,55	0,66	0,70	0,58	0,61	0,60
Região Sudeste	0,61	0,60	0,66	0,56	0,58	0,57
Região Sul	0,53	0,58	0,63	0,54	0,58	0,55
Brasil	0,68	0,70	0,78	0,58	0,61	0,60

Fonte: IPEA

Em termos de PIB *per capita*, as diferenças regionais se mantêm desde a década de 40 (gráfico 1); nem mesmo a participação relativa da Região Nordeste no PIB nacional se alterou, de fato, diminuiu de 16,7% em 1939 para 14,8% em 1960, e chegou a 13,1% em 2000⁷.

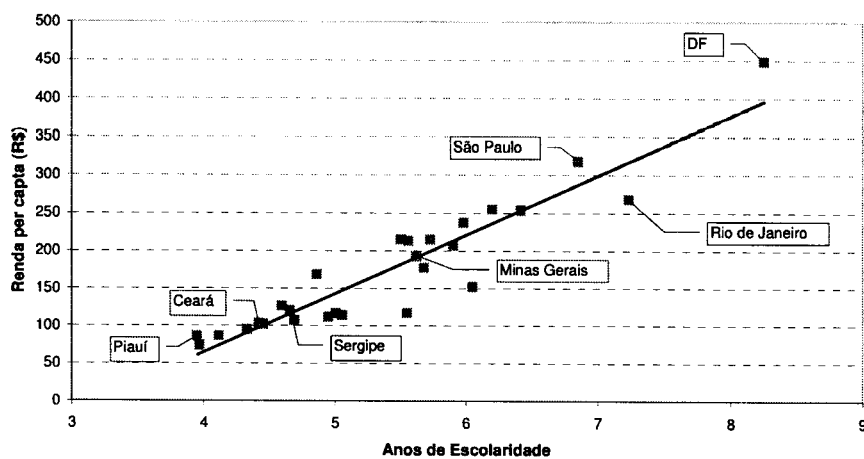
⁷Dados de PIB e população por região do IBGE.

Gráfico 1
Renda *per Capita* (em log)



Os dados do censo de 2000 mostram que há uma relação linear e positiva entre a renda *per capita* nos Estados e a escolaridade média da sua população acima de 25 anos (gráfico 2), o que corroboraria a tese de que as disparidade regionais de renda refletem as diferenças de capital humano de seus habitantes.

Gráfico 2
Renda *per Capita* x Escolaridade Média



2.3 Dados

Utilizamos os dados da Pesquisa Nacional por Amostragem Domiciliar (PNAD) do ano de 1999 com as ponderações obtidas a partir dos resultados do censo de 2000. Os resultados da PNAD são apresentados tanto em termos de indivíduos como em termos de domicílios. No caso da pesquisa por indivíduos, que é a de nosso interesse, dispõe-se de informações sobre as características gerais (UF de residência atual, idade, gênero, raça, cor, etc), características de migração (UF onde nasceu, UF onde morava 5 anos antes da pesquisa, etc), educação, trabalho e renda dos pesquisados. A cada indivíduo entrevistado é atribuído um peso, que traduz o quanto as pessoas com as suas características representam em relação à população.

A amostra utilizada consistiu nos indivíduos com renda do trabalho positiva no mês de referência da pesquisa (setembro/1999), com escolaridade conhecida e com jornada de trabalho maior ou igual a 40 horas semanais. Excluiu-se as pessoas com jornada inferior a 40 horas a fim de se tentar uniformizar o tamanho da jornada e aproximar as medidas de renda do trabalho e de salário, uma vez que essa última variável seria a ideal para esse tipo de estudo.

Foram entrevistadas 352.393 pessoas em todo o território nacional, sendo que 67.111 tinham menos de 10 anos de idade a época, e para as quais não se aplica o questionário de trabalho e renda. Das 285.282 pessoas restantes, somente 173.634 eram economicamente ativas⁸, e destas, 3.265 pessoas tinham escolaridade ou renda ignorados, restando portanto 170.369 pessoas. A partir dessa amostra, que engloba indivíduos em todo o Brasil, extraímos 5 sub-amostras de acordo com a UF de residência do entrevistado: Estado do Ceará (CE), Estado de São Paulo (SP), Região Nordeste (NE), Região Sudeste excluindo o Estado do Espírito Santo (SE1), e Região Sudeste excluindo os Estados do Espírito Santo e São Paulo (SE2)⁹.

Escolheu-se as Regiões Nordeste e Sudeste por se tratarem das regiões com menor e maior renda per capita, respectivamente. Os Estados do Ceará e de São Paulo foram escolhidos entre os da Região Nordeste e da Região Sudeste, respectivamente, pelo fato do primeiro, mesmo sendo um dos estados mais pobres da Federação, ser apontado como exemplo de

⁸Compreendem a população economicamente ativa as pessoas ocupadas ou que estão empreendendo esforços a procura de ocupação.

⁹SE2 compreende os Estados do Rio de Janeiro e Minas Gerais, respectivamente a segunda e terceira unidades mais ricas da Federação.

sucesso na implantação de políticas de desenvolvimento baseadas na atração de investimentos industriais privados, e o segundo por se tratar do ente mais rico da Federação. A exclusão do Estado do Espírito Santo se deve ao fato de sua economia ser bem menos desenvolvida e industrializada que a dos demais estados da região a ponto do mesmo ser também alvo das políticas públicas de combate à desigualdade regional, fazendo parte da área de atuação da SUDENE. Optou-se também por considerar uma sub-amostra que não englobasse o Estado de São Paulo, por se tratar de um estado peculiar no que tange à sua renda per capita e ao seu elevado grau de industrialização, que são muito superiores aos dos demais estados da Região Sudeste e da Federação. Ademais, como veremos adiante, São Paulo é o único estado no qual o valor do salário mínimo, que é determinado exogenamente pela União, não se constitui numa forte restrição à distribuição de renda do trabalho. Nos demais estados a moda da distribuição da renda do trabalho é exatamente o valor do salário mínimo.

As tabelas a seguir apresentam os quantitativos da amostra utilizada por região/estado de interesse e as principais estatísticas descritivas, que foram calculadas considerando-se a ponderação de cada indivíduo.

Tabela 3 - Quantitativo por Região/Estado

	Brasil	NE	CE	SE1	SE2	SP
Pessoas pesquisadas	352.393	113.902	22.124	110.558	67.890	42.668
PEA com escolaridade e renda conhecidos	170.369	53.210	10.163	53.162	32.658	20.504
Pessoas com renda do trabalho nula	37.808	14.718	2.588	9.530	6.061	3.469
Pessoas com jornada inferior a 40hr/semana	28.501	10.237	1.867	8.251	5.486	2.765
Tamanho da amostra	104.060	28.255	5.708	35.381	21.111	14.270

Tabela 4 - Estatísticas Descritivas para Amostra Selecionada

(em R\$)	Brasil	NE	CE	SE1	SE2	SP
Renda do trabalho						
média	572	358	338	676	554	775
desvio-padrão	905	704	697	977	866	1.055
Coef. Gini	0,55	0,57	0,58	0,52	0,53	0,50
Log da renda do trabalho						
média	5,81	5,30	5,21	6,04	5,83	6,22
desvio-padrão	0,96	0,95	0,98	0,90	0,90	0,85
Escolaridade (anos)	6,4	4,6	4,5	7,3	6,9	7,8
Coef. Gini	0,32	0,42	0,42	0,29	0,30	0,27

Verifica-se a partir da tabela 4 que quanto maior a renda média, menor a desigualdade de renda e de escolaridade, medidas aqui pelo coeficiente de variação de Pearson e pelo coeficiente de Gini. Isso nos permite levantar a hipótese que a desigualdade de renda e de escolaridade são fortemente correlacionadas, e por sua vez menores nas regiões/estados mais ricos. A escolaridade média nas regiões mais pobres é cerca de 3 anos menor que nas regiões mais ricas. De fato, observa-se que a renda é diretamente proporcional à escolaridade, o que vem a reforçar a hipótese de que o diferencial de renda pode ser explicada pela diferença de escolaridade.

Computamos para cada região/estado a mediana do logaritmo da renda por nível de escolaridade. As curvas escolaridades x renda, também conhecidas como curvas de Mincer, se encontram plotadas nos gráficos 3(a) e (b). Depreende-se que o retorno marginal da escolaridade é crescente para níveis de escolaridade acima de 10 anos, ao contrário do que se observa nos países desenvolvidos com alto nível escolaridade da sua população.

Gráfico 3(a) - Escolaridade x Renda

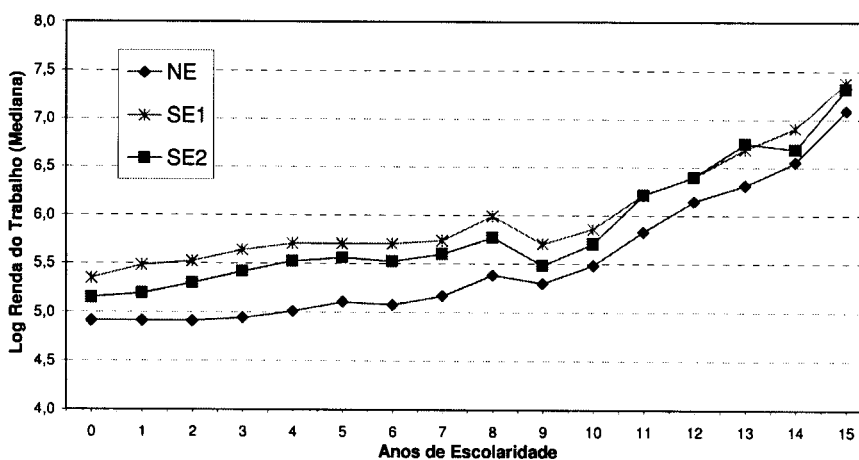
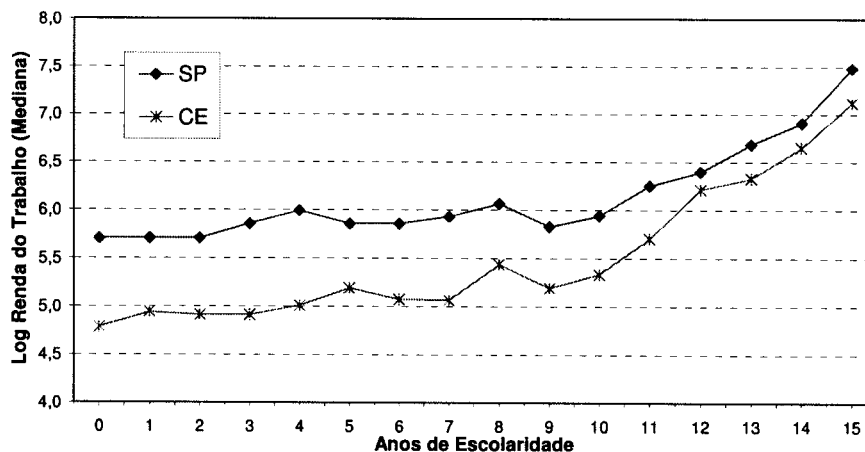


Gráfico 3(b) - Escolaridade x Renda



Apesar de dispormos dos dados de um expressivo número de pessoas entrevistadas, como pode ser observado na última linha da tabela 3, a quantidade de informação contida nessas entrevistas é bem menor do que aparenta. A renda, além de ter sido coletada como um valor inteiro, foi quase sempre reportada em múltiplos de R\$10 ou do salário mínimo (R\$136 à época) e ainda os valores concentram-se bastante em múltiplos de R\$100. A quantidade de níveis de renda do trabalho situam-se entre 480, para CE e 1040, para SP.

2.4 Metodologia

Utilizaremos um modelo semiparamétrico para construir funções de densidade contrafactuais. Essas densidades contrafactuais são estimadas a partir de uma amostra que se gerou tomando por base a amostra original dos dados alterando o atributo do qual se deseja estudar o impacto sobre a distribuição de renda. O método compreende duas etapas: a primeira, paramétrica, que resume-se à construção de funções de reponderação; e a segunda, não paramétrica, que consiste na estimação, baseada em funções núcleo, de funções de densidade, conforme proposto por Rosenblatt (1956) e Parzen (1962).

Seja, \hat{f}_h a estimativa de densidade de núcleo da densidade f , cujo suporte é a variável w , baseada em uma amostra aleatória de tamanho n , $\{W_1, W_2, \dots, W_n\}$, com ponderação $\theta_1, \dots, \theta_n$,

respectivamente, e onde $\sum_i \theta_i = 1$. Temos então:

$$\hat{f}_h(w) = \sum_{i=1}^n \frac{\theta_i}{h} K\left(\frac{w - W_i}{h}\right) \quad (1)$$

onde h é a janela e $K(\cdot)$ é a função núcleo. Os núcleos mais utilizados são o uniforme, o gaussiano e o de Epanechnikov, sendo que a sua escolha é uma decisão ad-hoc do econometrista, que deve levar em conta a natureza da variável cuja densidade está sendo estimada. No presente trabalho, acompanhando as sugestões de DiNardo, Fortin e Lemieux (1996) e Butcher e DiNardo (1998), adotamos o núcleo gaussiano e trabalhamos com o logaritmo da renda do trabalho para reduzir o problema de assimetria.

A escolha da janela é um ponto importante em estimações de densidade de núcleo, pois que há um *trade-off* entre viés (diferença entre a distribuição estimada e a real) e variância: janelas maiores resultam em maior viés e menor variância, e vice-versa. Há diversos métodos de seleção automática da janela, entre os quais destacam-se os métodos de validação cruzada e os métodos de *plug-in*¹⁰. No entanto esses métodos não são adequadas para dados com as características dos que dispomos, visto que são censurados por intervalos (agrupados). Os métodos de validação cruzada, por exemplo, como apontado por Silverman (1986), tendem a gerar resultados inadequados, $h = 0$. Adotamos então o método visual de seleção, conforme procedimento detalhado em Butcher e DiNardo (1998): iniciamos com uma janela extremamente estreita (baixa suavização), $h = 0,05$, e aumentamo-la até obtermos uma distribuição suave, o que veio a acontecer com $h = 0,12$. Justifica-se o procedimento de se iniciar com uma janela pequena e ir aumentando-a pela crença de que o olho humano é mais habilidoso em suavizar do que no sentido contrário (Butcher e DiNardo (1998)). Subjacente a esse método estamos adotando a hipótese que a distribuição de renda do trabalho, e consequentemente a produtividade marginal do trabalho, é suave, o que nos parece bastante razoável haja vista o tamanho da população. A escolha do limite inferior das janelas que produzem distribuições suaves indica que priorizamos o viés em detrimento da variância.

A estimação de densidades contrafactuais é realizada conforme proposto por DiNardo, Fortin e Lemieux (1996), onde se escolhe funções de reponderação da amostra. Pode-se considerar que cada observação da amostra é um vetor (w, z) , onde w representa os salários (uma variável contínua), z os atributos de cada indivíduo (consideraremos apenas a variável educação, medida em anos de escolaridade). Assim temos distribuições de densidade con-

¹⁰Ver Park e Marron (1990) e Sheater e Jones (1991) como exemplos.

juntas $F(w, z)$ de renda do trabalho e educação. Assim, a densidade da renda do trabalho da região 1, $f_{R1}(w)$, pode ser escrita como a integral da densidade de renda do trabalho condicionada no nível de educação dos indivíduos, $f(w/z)$, sobre a distribuição de educação, $F(z)$:

$$\begin{aligned} f_{R1}(w) &= \int_{z \in \Omega_z} dF(w, z | E_{w,z} = R1) \\ &= \int_{z \in \Omega_z} f(w|z, E_w = R1) dF(z | E_z = R1) \\ &\equiv f(w; E_w = R1, E_z = R1) \end{aligned} \quad (2)$$

onde Ω_z é o domínio do conjunto de atributos, E_z representa a região de onde se considera a distribuição de educação e E_w representa a região de onde se considera a distribuição de renda do trabalho. Para fazer o estudo de contrafactuais estaremos interessados em modificar a estrutura de atributos e portanto definiremos como $f(w; E_w = R1, E_z = R1)$ a densidade real de renda do trabalho da região 1 e $f(w; E_w = R1, E_z = R2)$ a densidade de renda do trabalho da região 1 que prevaleceria se a distribuição de educação fosse aquela existente na região 2 no mesmo período.

Sob a hipótese de que a estrutura de renda do trabalho da região 1, $f(w|z, E_w = R1)$, não depende da distribuição de educação na região 2, podemos escrever a densidade hipotética $f(w; E_w = R1, E_z = R2)$ como,

$$\begin{aligned} f(w; E_w = R1, E_z = R2) &= \int f(w|z, E_w = R1) dF(z | E_z = R2) \\ &\equiv \int f(w|z, E_w = R1) \Psi_z(z) dF(z | E_z = R1) \end{aligned} \quad (3)$$

onde $\Psi_z(z)$ é uma função de reponderação definida por

$$\Psi_z(z) \equiv dF(z | E_z = R2) / dF(z | E_z = R1) \quad (4)$$

A equação (3) define a densidade de renda do trabalho na região 1 que prevaleceria se as condições educacionais fossem similares às da região 2 e conforme pode ser observado é idêntica à definição em (2), exceto pela função de reponderação $\Psi_z(z)$. Na verdade, o problema de estimação da função de densidade contrafactual desejada fica reduzido ao cálculo de ponderações apropriadas. Logo, poderemos estimar funções de densidade contrafactuais

usando o método de estimadores de núcleo ponderados onde usamos um novo ponderador que contém uma estimativa para $\Psi_z(z)$. Assim temos,

$$\hat{f}(w; E_w = R1, E_z = R2) = \sum_{i \in S_{R1}} \frac{\theta_i}{h} \hat{\Psi}_z(z_i) K\left(\frac{w - W_i}{h}\right) \quad (5)$$

onde $\sum_{i \in S_{R1}} \theta_i \hat{\Psi}_z(z_i) = 1$ e S_{R1} é o conjunto de índices da amostra de indivíduos da região 1. As diferenças observadas entre a densidade real da região 1 e a densidade hipotética criada representa o efeito de uma mudança na distribuição de educação.

Pode-se observar que aplicando uma regra de Bayes em (4) este quociente pode ser escrito por

$$\Psi_z(z) = \frac{\Pr(E_z = R2|z) \Pr(E_z = R1)}{\Pr(E_z = R1|z) \Pr(E_z = R2)}. \quad (6)$$

Uma vez que o nível de educação é uma variável discreta que assume um número finito de valores, a estimação de $\Psi_z(z)$ por um modelo probit é equivalente a uma simples contagem.

Há diversas formas, paramétricas e não-paramétricas, de comparar as densidades estimadas (reais e contrafactuais) e medir a diferença entre as mesmas. Pode-se simplesmente tomar a diferença entre as densidades contrafactual e real, obtendo-se uma completa descrição das mudanças na distribuição da renda do trabalho oriundas da reponderação pela escolaridade. Outras medidas de comparação entre as densidades estimadas tentam reduzir em um único número as diferenças entre as mesmas: distância de Kullbach-Leibler, de Sibson, de Chernoff, diferença entre os desvios-padrões, diferenças entre os percentis, diferenças entre os diferenciais de percentis (10-90, 10-50, 25-75, 5-95). Todas essas medidas são utilizadas neste trabalho. Além disso, faremos o teste de desigualdade de densidades de Kolmogorov-Smirnov.

2.5 Resultados

Aplicamos a metodologia descrita na seção anterior e estimamos as distribuições reais do logaritmo da renda do trabalho para as 5 sub-amostras e para a amostra completa (Brasil), conforme apresentado na seção 3. Em seguida, estimamos as distribuições contrafactuais para o Nordeste e para o Ceará reponderando as amostras pelas características de escolaridade da Região Sudeste (SE1 e SE2) e do Estado de São Paulo. Relatamos e comentamos nesta seção os resultados para o Estado do Ceará, reponderado pela escolaridade de São

Paulo, e para a Região Nordeste, reponderada pela escolaridade da Região Sudeste (SE1 e SE2). Os resultados para os demais casos estão apresentados no apêndice.

Em cada um dos gráficos abaixo, 4 a 6, apresentamos três densidades: uma densidade contrafactual, a densidade real da qual essa se originou e a densidade real da região usada na reponderação¹¹. O eixo horizontal dos gráficos está em escala logarítmica.

Gráfico 4

Densidades reais para SE1 e NE e contrafactual para NE com escolaridade de SE1

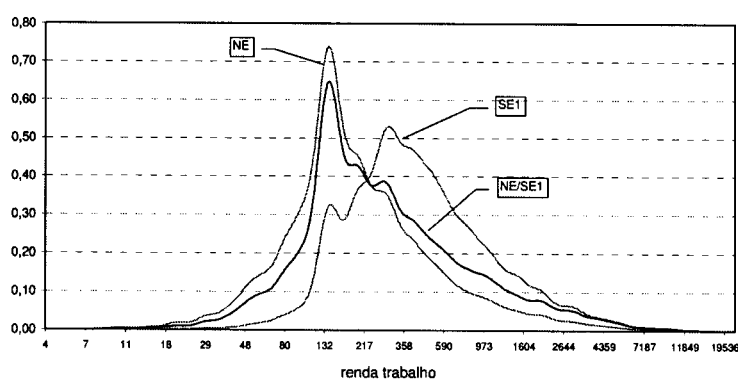
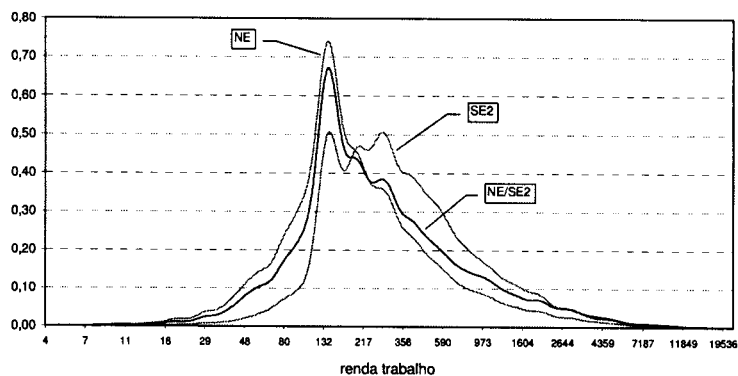


Gráfico 5

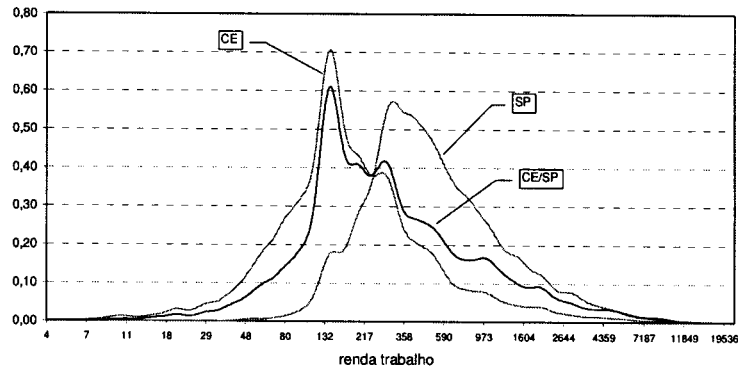
Densidades reais para SE2 e NE e contrafactual para NE com escolaridade de SE2



¹¹Todos as grandezas monetárias estão expressas em Reais.

Gráfico 6

Densidades reais para SP e CE e contrafactual para CE com escolaridade de SP



Testamos a igualdade entre as distribuições empíricas originais e as suas contrafactuais usando o teste de Kolmogorov-Smirnov, no qual a estatística de teste é o máximo da diferença entre as densidades acumuladas: $\max |F_1(w) - F_2(w)|$. Rejeitamos a hipótese nula de que as densidades são iguais a 1% de significância em todos os casos, evidenciando que a mudança do perfil de escolaridade altera a distribuição de renda do trabalho.

Fica evidente que as distribuições ¹² são sensíveis aos diferenciais de escolaridade e que o grau de sensibilidade depende do percentil de renda considerado. O valor do salário mínimo, R\$136 à época da pesquisa, é fator determinante na renda dos trabalhadores do Nordeste, do Ceará e do Sudeste (SE2), haja visto a enorme concentração de massa em torno desse valor na distribuição. A distribuição de renda da Região Sudeste (SE2) é bimodal, sendo que uma das modas corresponde ao valor do salário mínimo e a outra é igual à moda da distribuição de São Paulo, cerca de R\$300. As distribuições de SP e SE1 são bem parecidas o que mostra a significância de São Paulo sobre o agregado da Região Sudeste. Pode-se afirmar então que em todos os estados das Regiões Nordeste e Sudeste, com exceção de São Paulo (SE2), o valor do salário mínimo tem forte impacto sobre a distribuição de renda do trabalho.

Comparando-se as caudas das distribuições contrafactuais e das distribuições originais verifica-se que as mesmas se aproximam bastante a menos de uma translação à direita

¹²Leia-se nos gráficos:

NE/SE1: contrafactual do Nordeste usando escolaridade do Sudeste (MG+RJ+SP).

NE/SE2: contrafactual do Nordeste usando escolaridade do Sudeste (MG+RJ).

CE/SP: contrafactual do Ceará usando escolaridade de S.Paulo.

das primeiras. Isso é decorrente da proporcionalidade direta entre escolaridade e renda em conjunto com a reponderação da amostra, cujos efeitos sobre a distribuição de renda assemelham-se ao que seria obtido caso adicionássemos uma constante à escolaridade de cada indivíduo. Observa-se que as áreas centrais das distribuições de renda, que é exatamente onde se concentram mais indivíduos e aqueles que têm renda próxima a um salário mínimo, pouco se modificam após a reponderação. Justifica-se esse comportamento pelo fato da estrutura salarial não ser alterada nesse processo, sobrepondo-se ao efeito do aumento de escolaridade resultante do mesmo processo. A moda, por exemplo, permanece igual a um salário mínimo em todos os casos.

A tabela 5 apresenta alguns pontos notáveis (percentis, médias, desvios-padrão e diferenças de percentis) das distribuições estimadas, reais e contrafactuais. O suporte das distribuições estimadas é o logaritmo da renda do trabalho, o qual delimitamos para efeitos computacionais ao intervalo [1,10] com passo de 0,01. A partir dessas distribuições construímos distribuições com o suporte sendo o nível da renda do trabalho, simplesmente tomando o exponencial de cada ponto das distribuições estimadas e renormalizando. Apresentamos nas tabelas A.1 e A.2 (no apêndice) os pontos notáveis de todas as distribuições construídas e as razões dos percentis das distribuições estimadas.

Tabela 5 - Pontos Notáveis das Distribuições de Renda do Trabalho Estimadas

Percentil	Distribuições Estimadas(*)								
	Brasil	CE	SP	NE	SE1	SE2	CE/SP	NE/SE1	NE/SE2
10	116	61	187	70	144	125	87	89	84
20	147	90	250	101	196	153	125	124	119
30	189	117	302	124	250	192	147	144	140
40	237	136	358	141	299	235	183	174	164
50	293	158	428	166	365	284	233	219	204
60	365	194	523	202	450	358	296	281	262
70	469	247	672	262	572	459	412	384	351
80	665	334	916	365	796	639	639	578	523
90	1.130	567	1.510	639	1.339	1.086	1.200	1.086	982
Média	572	338	775	358	676	553	584	534	493
DP	905	697	1.056	704	977	855	1.091	687	929
Diferença de Percentis									
10-90	1.014	506	1.323	569	1.195	961	1.113	997	898
10-50	177	97	242	96	221	159	145	130	120
50-90	837	409	1.082	473	974	801	967	867	778
20-80	518	244	666	264	600	486	514	454	404

(*) Os valores apresentados foram extraídos das distribuições estimadas (em log) tomando-se a exponencial de cada ponto.

As médias e os desvios-padrão foram calculados diretamente da amostra reponderada com a renda em nível.

Tabela 6 - Relação da Renda nos Percentis

Percentil	CE x SP	CE/SP x SP	NE x SE1	NE x SE2	NE/SE1 x SE1	NE/SE2 x SE2
10	33%	47%	49%	56%	62%	67%
20	36%	50%	52%	66%	63%	78%
30	39%	49%	50%	64%	58%	73%
40	38%	51%	47%	60%	58%	70%
50	37%	54%	45%	58%	60%	72%
60	37%	57%	45%	57%	63%	73%
70	37%	61%	46%	57%	67%	76%
80	36%	70%	46%	57%	73%	82%
90	38%	79%	48%	59%	81%	90%

Comparando-se as distribuições originais do Ceará e de São Paulo verifica-se que a renda da primeira é de aproximadamente 1/3 da segunda em todos os percentis (tabela 6). Por outro lado, ao compararmos a distribuição do Ceará reponderada pela educação de São Paulo com a distribuição original de São Paulo, observamos dois efeitos: (i) houve um ganho na distribuição de renda do Ceará em todos os percentis, e (ii) o ganho foi maior quanto maior era o nível de renda (percentil)¹³. O retorno marginal crescente da escolaridade, ilustrado no gráfico 3, justifica o segundo efeito, pois os indivíduos mais escolarizados, e conseqüentemente com maior renda, ao receberem um “ganho” de escolaridade por efeito da reponderação terão suas renda elevadas em proporção maior que os menos escolarizados, cuja renda evolui pouco após um aumento de escolaridade.

Comparando-se a Região Nordeste com a Sudeste verificou-se que a renda da primeira se situa entre 45% e 50% da renda do Sudeste para a amostra SE1, e entre 55% e 65% considerando a amostra SE2. Após a reponderação, observaram-se os mesmos efeitos verificados quando comparamos a distribuição do Ceará com a de São Paulo. Observou-se que o efeito da reponderação é diretamente proporcional à diferença entre as distribuições originais, logo sendo maior no caso CE x SP e menor para NE x SE2, para todos os percentis. Uma comparação interessante é que a renda média do Nordeste reponderada pela educação da Região Sudeste (SE1) é 93% da renda média brasileira. Aqui, à semelhança da comparação CE x SP, o fato da escolaridade apresentar retorno marginal crescente justifica um ganho de renda maior nos decis mais altos.

Os ganhos absolutos são significativos em todos os caso: em nenhum decil o ganho de renda com a reponderação é inferior a 13% da renda original, chegando a atingir 70% no nono decil para a comparação CE x SP. O mais relevante é que a renda dos nordestinos

¹³Para o primeiro decil a renda do Ceará evoluiu de 33% para 47% da renda de S. Paulo, e para o nono decil passou de 38% para 79%.

nunca é inferior a cerca de 60% da renda do sudestino, para a amostra SE1 e a 67%, para a amostra SE2, e que nos decis superiores ela se aproxima bastante da renda dos moradores da região sudeste, se aqueles possuísem o mesmo nível de educação destes. A propósito, a renda do nono decil do NE reponderado pela escolaridade da região SE2 atinge 90% da renda de SE2.

Este fato está melhor representado nos gráficos 7 a 9 (e A.4 a A.6 no apêndice), onde a distribuição contrafactual e a distribuição de renda da região/estado que foi utilizado como base para a reponderação se aproximam nos níveis mais elevados de renda.

Gráfico 7 -Evolução da Renda do Trabalho por Percentis para CE, SP e CE/SP

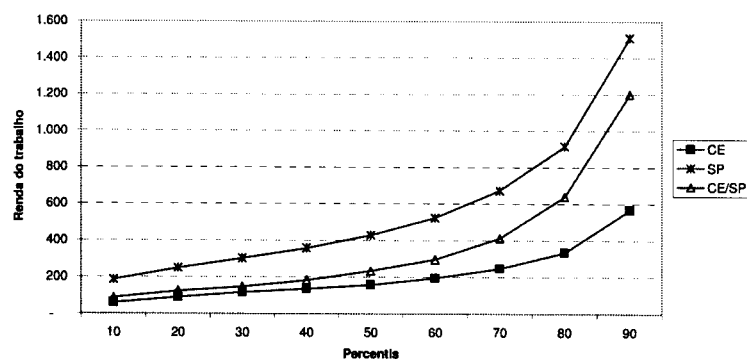


Gráfico 8 -Evolução da Renda do Trabalho por Percentis para NE, SE1 e NE/SE1

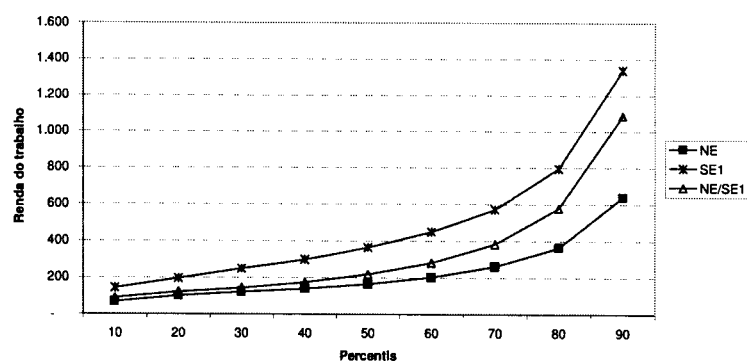
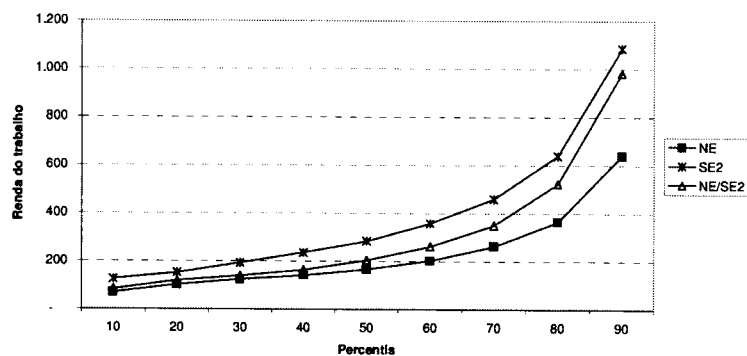
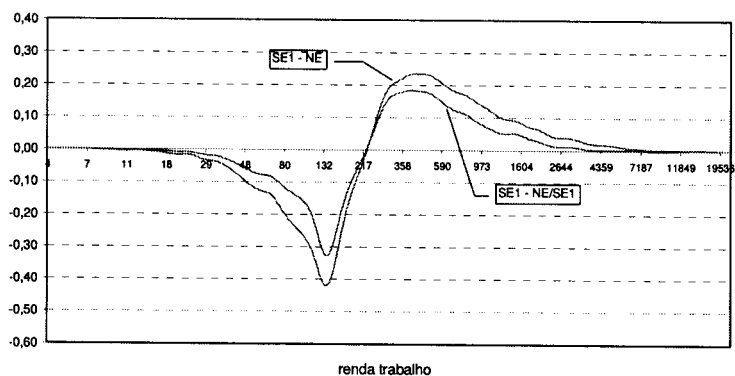


Gráfico 9 -Evolução da Renda do Trabalho por Percentis para NE, SE2 e NE/SE2



Nos gráficos 10 a 12 abaixo apresentamos a diferença entre as duas distribuições reais e a diferença entre a distribuição contrafactual e a real que a originou¹⁴. Esses gráficos nos permitem visualizar o quanto as distribuições de renda se aproximaram após a reponderação pela escolaridade.

Gráfico 10 - Diferença das Distribuições (SE1 x NE)



¹⁴Para cada nível de renda tomou-se a diferença (contrafactual - real), ou seja, $f(w; E_w = R1, E_Z = R2) - f(w; E_w = R1, E_Z = R1)$.

relação não-linear entre a distância das distribuições, para cada w , e sua contribuição para J. Verifica-se que mais de 55% das distâncias entre o Ceará e São Paulo e entre as regiões Nordeste e Sudeste (SE1 e SE2) são explicadas pela escolaridade - 55,3% da distância entre São Paulo e Ceará e 55,0% (57,3%) da distância entre o Nordeste e a região SE1 (SE2). Isto mostra que mais da metade da diferença de renda entre a região/estado mais pobre em relação ao mais rico é devido aos diferenciais de escolaridade da população. Computamos também a distância das distribuições pelas métricas de Chernoff¹⁶ e Sibson¹⁷, sugeridas em Krzanowski (2003), e obtivemos resultados semelhantes, os quais se encontram apresentados na Tabela 9.

Tabela 7 - Distâncias de Kullback-Leibler

Distância entre as Distribuições							
	CE	NE	CE/SP	CE/SE1	CE/SE2	NE/SP	NE/SE1 NE/SE2
SP	1,5901	1,3826	0,7110			0,6700	
SE1	0,9608	0,8031		0,3969			0,3611
SE2	0,5396	0,4127			0,2145		0,1761
% da distância entre as distribuições explicado pela educação							
	SP - CE		55,3%	SP - NE		51,5%	
	SE1 - CE		58,7%	SE1 - NE		55,0%	
	SE2 - CE		60,3%	SE2 - NE		57,3%	

Calculamos os coeficientes de entropia de Theil e de concentração de Gini das distribuições estimadas. Esses coeficientes são medidas de dispersão e são amplamente usadas como medidas de desigualdade quando aplicadas a distribuições de renda. Os resultados estão apresentados na tabela 8 abaixo (e na tabela A.4 no apêndice). Em suma, observou-se um aumento da desigualdade de renda quando reponderamos a amostra. Isto se deve ao fato de que a elevação de renda não é uniforme ao longo de toda a distribuição, observando maiores ganhos para os níveis maiores de renda.

¹⁶ A distância de Chernoff entre as distribuições f_1 e f_2 é definida como $-\log \left(\int \sqrt{f_1(x) f_2(x)} dx \right)$.

¹⁷ A distância de Sibson entre as distribuições f_1 e f_2 é definida como $-\frac{1}{2} \int f_1(x) \ln \frac{f_1(x)}{f(x)} + f_2(x) \ln \frac{f_2(x)}{f(x)}$, onde $f(x) \equiv \frac{1}{2} [f_1(x) + f_2(x)]$.

Tabela 8 - Coeficientes de Gini e Theil

		Coeficiente de Gini	Coeficiente de Theil
Brasil		0,551	0,607
Regiões	CE	0,582	0,768
	NE	0,574	0,731
	SE1	0,523	0,539
	SE2	0,534	0,573
	SP	0,505	0,498
	CE/SP	0,622	0,797
	NE/SE1	0,608	0,766
	NE/SE2	0,604	0,768

Obs.: SE1 = MG + RJ + SP

SE2 = MG + RJ

Mensuramos o efeito da reponderação pela escolaridade computando a variação percentual de algumas métricas, calculadas a partir das distribuições estimadas. Os resultados estão reportados na Tabela 9 abaixo. Em termos de renda média ou considerando a distância de Kullbach-Leibler, a escolaridade foi capaz de explicar mais de 55% da diferença de renda, sendo que no caso da reponderação do Nordeste pela escolaridade da região SE2 esse índice atinge cerca de 70% na primeira métrica. A reponderação ocasionou dois efeitos: (i) as medidas de tendência central se aproximaram e, (ii) as medidas de dispersão se elevaram, sinalizando que a desigualdade de renda nas regiões mais pobres aumentariam caso herdasse o perfil de escolaridade das mais ricas, mantendo-se a estrutura de salários. Em outras palavras, o Nordeste com o perfil de escolaridade do Sudeste seria mais rico, porém mais desigual. O efeito a que nos referimos no item (ii) decorre, em parte, do fato de nosso exercício não levar em conta a mudança na estrutura salarial decorrente da mudança de escolaridade, ou seja, em equilíbrio geral esse fenômeno não seria observado em parte ou totalmente.

Tabela 9 - Percentual Explicado pela Escolaridade

Métrica	SP x CE	SE1 x NE	SE2 x NE
Média	56,3%	55,3%	69,2%
Mediana	27,8%	26,9%	32,6%
Coeficiente de Theil	-10,7%	-18,0%	-23,4%
Coeficiente de Gini	-51,9%	-66,9%	-75,0%
Distância de Kullbach-Leibler	55,3%	55,0%	57,3%
Distância de Chernoff	56,0%	55,5%	57,5%
Distância de Sibson	52,5%	53,6%	56,5%

A robustez dos resultados apresentados ao tamanho da janela foi investigado reestimando-se todas as distribuições com $h = 0,09$ (-25% de h original) e $h = 0,15$ (+25%), e comparando os resultados com os obtidos com o valor original de h . Verificou-se que as alterações nos resultados não os comprometem, tendo ainda não se observado uma relação direta entre a variação do tamanho da janela e a variação dos indicadores apresentados nas tabelas 5 a 9.

Cabe ressaltar que a parcela da diferença de renda explicada pela diferença de escolaridade talvez seja maior que a apresentada nesse trabalho, uma vez que o preço da cesta de consumo dos indivíduos é diferente entre as regiões. Os serviços e bens não-comercializáveis são mais baratos nas regiões mais pobres, refletindo exatamente a diferença de renda do trabalho entre as regiões, vez que eles tendem a ser mais intensivos em mão de obra. O efeito da correção pelo poder de compra da renda em cada região seria menor nas camadas de renda mais baixa (consomem menos serviços) e maior para os indivíduos com renda elevada (consomem mais serviços), acentuando o efeito apresentado nos gráficos 7 a 9.

2.6 Conclusão

Neste trabalho procuramos identificar o quanto do diferencial de renda entre as regiões Nordeste e Sudeste e entre o estado do Ceará e São Paulo é explicado pelo diferencial de escolaridade da população. Usou-se um modelo semiparamétrico para construir funções de densidade contrafactuais, reponderando os indivíduos da região/estado base pela distribuição de educação da região a ser comparada. Estimamos as distribuições de renda do trabalho reais e contrafactuais do Estado do Ceará e da Região Nordeste reponderadas pelas escolaridades da Região Sudeste e do Estado de São Paulo.

Verificou-se que (i) a dispersão de renda é maior nas distribuições com menor média, ou seja, a desigualdade renda, que é enorme em todas as regiões, é maior nas regiões mais pobres; (ii) mais de 55% da diferença de renda do trabalho entre a Região Nordeste e a Região Sudeste e entre os Estados de São Paulo e Ceará, quando medidas pela distância de Kullbach-Leibler ou em termos de renda média, se devem à diferença de escolaridade; (iii) a reponderação pela escolaridade aumentou em cerca de 50% a renda média da Região Nordeste e em mais de 70% a do Estado do Ceará; (iv) a renda do Nordeste reponderada pela escolaridade do Sudeste equivale a 93% da renda média brasileira; (v) quanto mais elevado for o percentil de renda considerado, maior é a contribuição da diferença de escolaridade para

a diferença de renda e, (vi) a dispersão de renda das regiões mais pobres aumenta quando fornecemos a elas o nível de escolaridade das regiões mais ricas, mantendo-se o perfil salarial da região.

Vários são os fatores que podem estar determinando a diferença de renda não explicada pelo diferencial de escolaridade, entre os quais podemos citar a expectativa de vida dos habitantes, fatores étnicos, estrutura etária da população, qualidade da infra-estrutura existente, presença/ausência de estímulos ao desenvolvimento e fatores históricos. Uma extensão natural deste trabalho seria, aplicando a metodologia de contrafactuais, decompor o diferencial de renda em alguns destes fatores, além da escolaridade, de forma a conseguir explicar uma parcela maior do diferencial de renda. A importância relativa dos fatores posta em conjunto com o custo de eliminá-los pode ser de grande valor na orientação de políticas públicas de combate à “desigualdade regional”. Outras extensões interessantes seriam: (i) reponderarmos a distribuição de renda da Região Nordeste pela escolaridade do Brasil sem a Região Norte¹⁸ e; (ii) reponderarmos a distribuição de renda de todas as regiões pelo perfil de escolaridade do Brasil, permitindo que se determine o tamanho das desigualdades regionais controlando-se pela escolaridade.

Referências

- [1] Barros, Ricardo Paes de (1993). “Regional Disparities in Education Within Brazil: the Role of Quality of Education”, Texto para Discussão n° 311, IPEA.
- [2] Barros, Ricardo Paes de & Mendonça, Rosane (1995). “Os Determinantes da Desigualdade no Brasil”, Texto para Discussão n° 377, IPEA.
- [3] Barros, Ricardo Paes de, Camargo, José Márcio & Mendonça, Rosane (1997). “A Estrutura do Desemprego no Brasil”, Texto para Discussão n° 478, IPEA.
- [4] Blinder, Alan S. (1973). “Wage Discrimination: Reduced Form and Structural Estimates”, *Journal of Human Resources*, 8, 436-455.

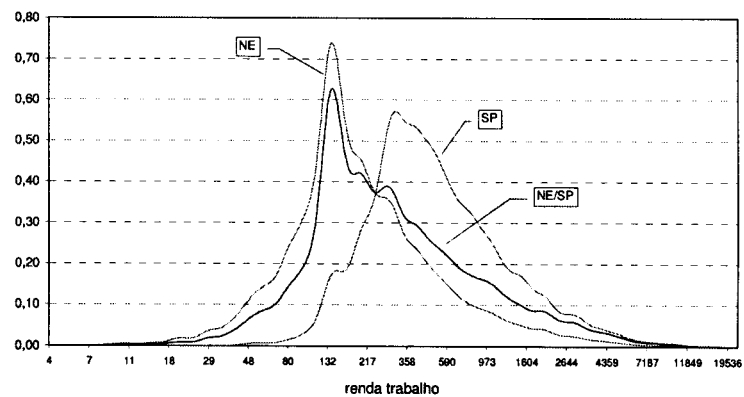
¹⁸Isto porque a PNAD não entrevista a população rural, que é bastante significativa na Região Norte.

- [5] Bourguignon, François, Ferreira, Francisco & Leite, Phillippe (2002). "Beyond Oaxaca-Blinder: Accounting for Differences in Household Income Distributions Across Countries", mimeo.
- [6] DiNardo, John, Fortin, Nicole M. & Lemieux, Thomas (1996). "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach", *Econometrica*, Vol. 64, No. 5, 1001-1044.
- [7] Butcher, Kristin F. & DiNardo, John (1998). "The Immigrant and Native-Born Wage Distributions: Evidence from United States Censuses", *NBER Working Paper Series* 6630.
- [8] Krzanowski, W. J. (2003). "Non-parametric estimation of distance between groups", *Journal of Applied Statistics*, Vol. 30, No. 7, 743-750.
- [9] Oaxaca, R. (1973). "Male-Female Wage Differentials in Urban Labor Markets", *International Economic Review*, 14, 693-709.
- [10] Park, B. U. & Marron, J. S. (1990). "Comparision of Data-driven Bandwidth Selectors", *Journal of American Statistical Association*, 85, 66-72.
- [11] Parzen, E. (1962). "On Estimation of a Probability Density Function and Mode", *The Annals of Mathematical Statistics*, 33, 1065-1076.
- [12] Pessoa, Samuel (2000). "Existe um Problema de Desigualdade Regional no Brasil?", mimeo.
- [13] Rosenblatt, M. (1956). "Remarks on Some Non-parametric Estimates of a Density Function", *The Annals of Mathematical Statistics*, 27, 832-837.
- [14] Sheater, S. J. & Jones, M. C. (1991). "A Reliable Data-based Bandwidth Selection Method for Kernel Density Estimation", *Journal of Royal Statistical Society*, 53, 683-690.
- [15] Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis*, London: Chapman & Hall, 1986.

A Tabelas e Figuras

Gráfico A.1

Densidades reais para SP e NE e contrafactual para NE com escolaridade de SP



Diferenças entre as distribuições de SP e NE (real e contrafactual)

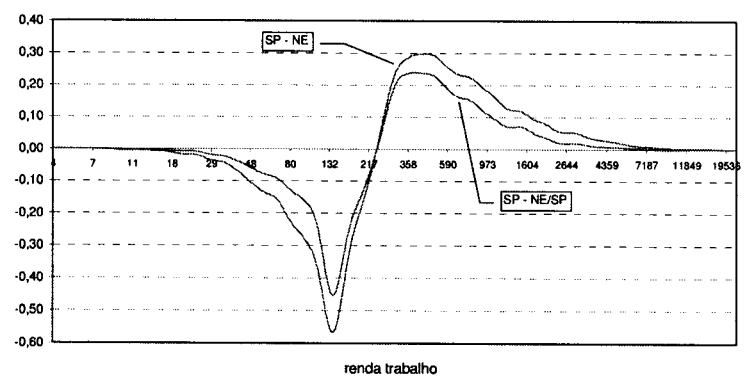
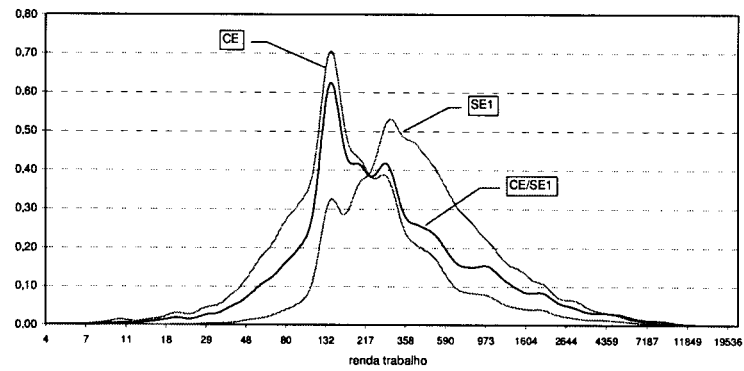


Gráfico A.2

Densidades reais para SE1 e CE e contrafactual para CE com escolaridade de SE1



Diferenças entre as distribuições de SE1 e CE (real e contrafactual)

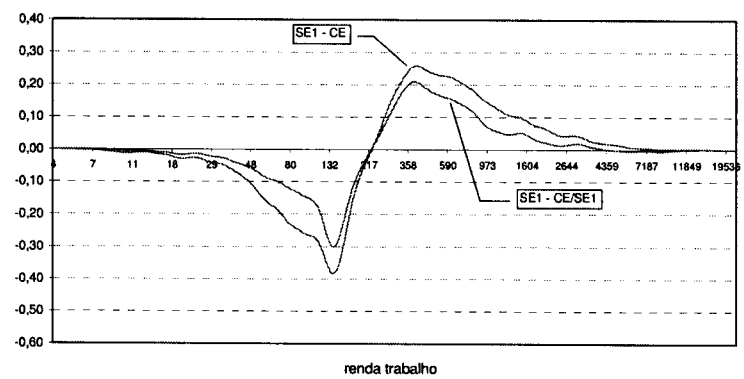
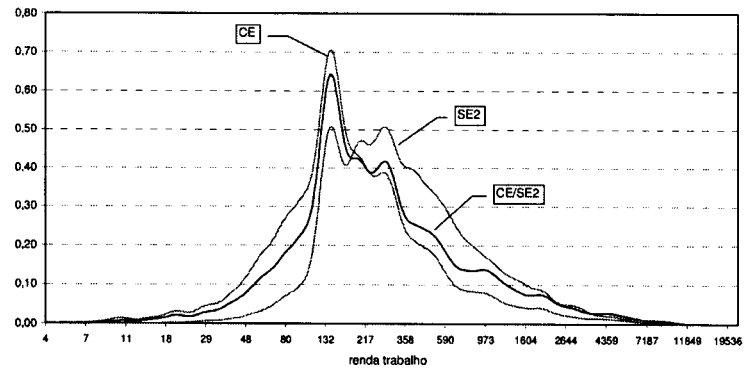


Gráfico A.3

Densidades reais para SE2 e CE e contrafactual para CE com escolaridade de SE2



Diferenças entre as distribuições de SE2 e CE (real e contrafactual)

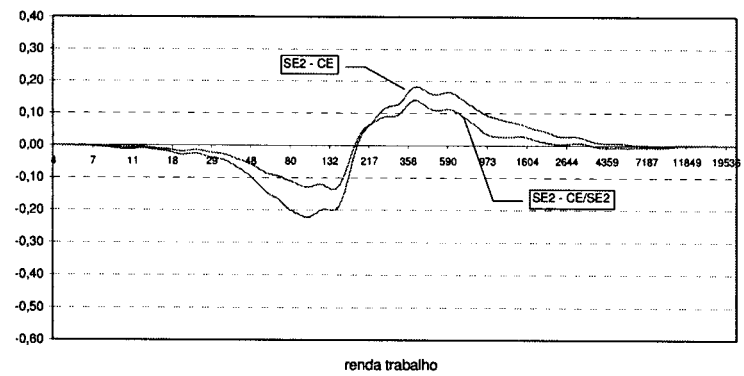


Gráfico A.4 - Evolução da Renda do Trabalho por Percentis para CE, SE1 e CE/SE1

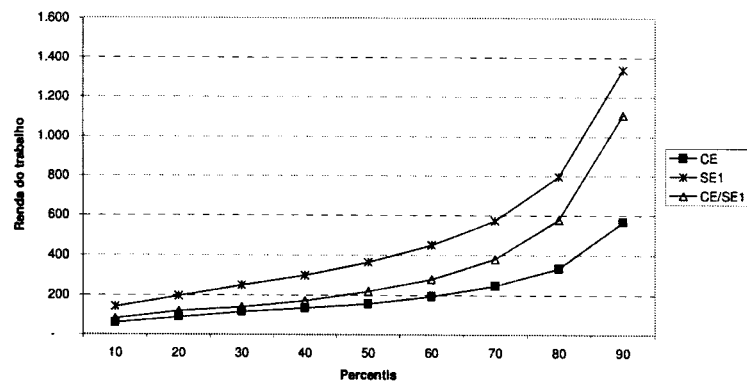


Gráfico A.5 - Evolução da Renda do Trabalho por Percentis para CE, SE2 e CE/SE2

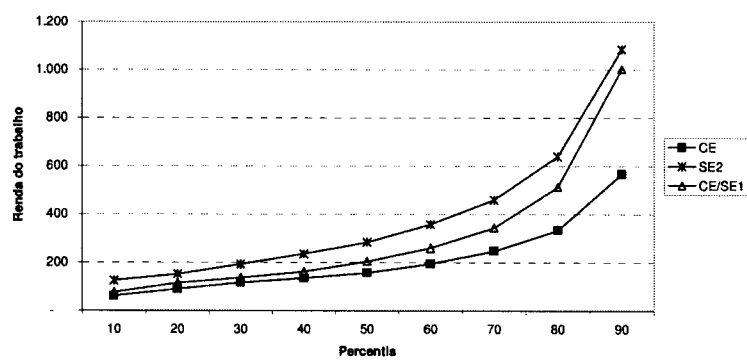


Gráfico A.6 - Evolução da Renda do Trabalho por Percentis para NE, SP e NE/SP

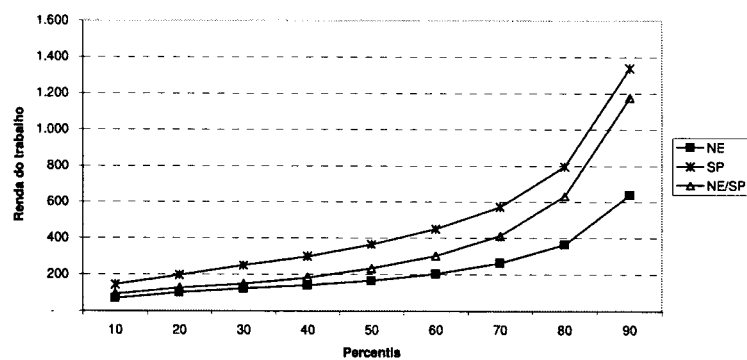


Tabela A.1 - Pontos Notáveis das Distribuições de Renda do Trabalho Estimadas

Percentil	Distribuições Estimadas(*)											
	Brasil	CE	SP	NE	SE1	SE2	CE/SP	CE/SE1	CE/SE2	NE/SP	NE/SE1	NE/SE2
10	116	61	187	70	144	125	87	82	77	94	89	84
20	147	90	250	101	196	153	125	122	116	128	124	119
30	189	117	302	124	250	192	147	143	137	148	144	140
40	237	136	358	141	299	235	183	172	162	183	174	164
50	293	158	428	166	365	284	233	219	204	233	219	204
60	365	194	523	202	450	358	296	279	260	302	281	262
70	469	247	672	262	572	459	412	380	344	412	384	351
80	665	334	916	365	796	639	639	578	513	633	578	523
90	1,130	567	1,510	639	1,339	1,086	1,200	1,108	1,002	1,176	1,086	982
Média	572	338	775	358	676	553	584	547	503	569	534	493
DP	905	697	1,056	704	977	855	1,091	1,041	978	1,033	687	929
Diferença de Percentis												
10-90	1,014	506	1,323	569	1,195	961	1,113	1,025	926	1,082	997	898
10-50	177	97	242	96	221	159	145	137	128	139	130	120
50-90	837	409	1,082	473	974	801	967	888	798	943	867	778
20-80	518	244	666	264	600	486	514	457	397	505	454	404

(*) Os valores apresentados foram extraídos das distribuições estimadas (em log) tomando-se a exponencial de cada ponto.

A média e o desvio-padrão foram calculados diretamente da amostra reponderada com a renda em nível.

Tabela A.2 - Relação da Renda nos Percentis

Percentil	CE x SP	CE/SP x SP	CE x SE1	CE x SE2	CE/SE1 x SE1	CE/SE2 x SE2
10	33%	47%	42%	49%	57%	61%
20	36%	50%	46%	59%	62%	76%
30	39%	49%	47%	61%	57%	71%
40	38%	51%	45%	58%	58%	69%
50	37%	54%	43%	55%	60%	72%
60	37%	57%	43%	54%	62%	73%
70	37%	61%	43%	54%	66%	75%
80	36%	70%	42%	52%	73%	80%
90	38%	79%	42%	52%	83%	92%

Percentil	NE x SP	NE/SP x SP	NE x SE1	NE x SE2	NE/SE1 x SE1	NE/SE2 x SE2
10	38%	50%	49%	56%	62%	67%
20	41%	51%	52%	66%	63%	78%
30	41%	49%	50%	64%	58%	73%
40	39%	51%	47%	60%	58%	70%
50	39%	54%	45%	58%	60%	72%
60	39%	58%	45%	57%	63%	73%
70	39%	61%	46%	57%	67%	76%
80	40%	69%	46%	57%	73%	82%
90	42%	78%	48%	59%	81%	90%

Tabela A.3 - Coeficientes de Gini e Theil

			Coeficiente de Gini	Coeficiente de Theil
Brasil			0,551	0,607
Regiões	Ponderação Original	CE	0,582	0,768
		NE	0,574	0,731
		SE1	0,523	0,539
		SE2	0,534	0,573
		SP	0,505	0,498
	Reponderação pela Educação	CE/SP	0,622	0,797
		CE/SE1	0,621	0,805
		CE/SE2	0,618	0,810
		NE/SP	0,609	0,760
		NE/SE1	0,608	0,766
		NE/SE2	0,604	0,768

Obs.: SE1 = MG + RJ + SP

SE2 = MG + RJ

Tabela A.4 - Percentual explicado pela Escolaridade

Métrica	SP x CE	SE1 x CE	SE2 x CE	SP x NE	SE1 x NE	SE2 x NE
Média	56,3%	61,8%	76,7%	50,6%	55,3%	69,2%
Mediana	27,8%	29,7%	36,9%	25,5%	26,9%	32,6%
Coeficiente de Theil	-10,7%	-16,2%	-21,5%	-12,3%	-18,0%	-23,4%
Coeficiente de Gini	-51,9%	-66,8%	-75,0%	-51,4%	-66,9%	-75,0%
Distância de Kullback-Leibler	55,3%	58,7%	60,3%	51,5%	55,0%	57,3%
Distância de Chernoff	56,0%	59,1%	60,4%	52,3%	55,5%	57,5%
Distância de Sibson	52,5%	56,9%	59,2%	49,2%	53,6%	56,5%

Capítulo 3

Estimando os Ganhos de Bem-Estar a partir de Funções Impulso-Resposta não Convencionais*

Márcio Antônio Salvato [†]

Resumo

Aqui se propõe avaliar o efeito dos diferentes choques econômicos a partir do uso da função de bem-estar. Para tanto, usa-se o conceito de funções impulso-resposta não convencionais, onde o bem-estar é função do valor presente da utilidade do consumo. Essa técnica permite avaliar a importância relativa de diferentes choques sob um novo prisma, o que se constitui em sua maior contribuição. Decompõe-se a função impulso-resposta não convencional em choques transitórios e permanentes. Identifica-se choques “de produtividade” e “de preferências” usando Decomposição de Cholesky e método generalizado para as funções impulso-resposta não convencionais. O resultado permite questionar a adequação da hipótese de identificação de que a única fonte de choques permanentes seja a produtividade: não há um *matching* perfeito entre a decomposição permanente/transitório e a identificação de tipos de choques.

JEL Classification Codes: C32; D60 e E21

*Artigo feito em co-autoria com João Victor Issler (EPGE/FGV). Os autores agradecem a Luiz Renato Lima, Marcelo Fernandes e participantes dos seminários da EPGE/FGV pelas sugestões e comentários.

[†]EPGE/FGV, PUC.Minas e CEPE, salvato@fgvmail.br.

3.1 Introdução

Desde que Nelson e Plosser (1982) propuseram a idéia de que as séries macroeconômicas possuem uma raiz unitária, o que abre a possibilidade de que essas séries tenham componentes permanentes e transitórios, um grande esforço tem sido feito nas últimas duas décadas para investigar os efeitos de choques econômicos sobre flutuações agregadas; ver Nelson e Plosser (1982), Watson (1986), Campbell e Mankiw (1987), Cochrane (1988, 1994), Blanchard e Quah (1989), King *et alli* (1991), Galí (1999), Issler e Vahid (2001), Francis e Ramey (2001, 2003), Hartley e Whitt (2002), Galí *et alli* (2003), Christiano, Eichenbaum e Vigfusson (2003). Nessa crescente literatura, muitas vezes é praxe “identificar” choques econômicos, no sentido econométrico, a partir de seus efeitos permanentes sobre os agregados macroeconômicos. Por exemplo, a partir do uso de modelos Autoregressivos Vetoriais (VAR), Blanchard e Quah denominam de “choque de oferta” o único choque que tem efeitos permanentes sobre o produto. King *et alli* denominam de “choque de produtividade” o choque que tem impactos permanentes idênticos sobre o produto, o consumo e o investimento. Para Galí, esse é o choque que tem efeitos permanentes sobre o produto por trabalhador e as horas trabalhadas.

Como notaram Issler e Vahid, os modelos teóricos em macroeconomia não são construídos usando como base choques com efeitos transitórios ou permanentes. Ao contrário, são construídos em termos de choques reais (e.g., choques de produtividade) e nominais (e.g., choques monetários). Por essa razão, para testar essas teorias, alguns autores lançam mão de hipóteses de identificação um tanto restritivas, onde a apenas um dos choques é facultada a possibilidade de ter efeitos permanentes. Esse é o caso de King *et alli* e de Galí, que postulam que a única fonte de choques permanentes é a produtividade. Ora, pode-se facilmente imaginar outros tipos de choque que tenham efeitos permanentes nas variáveis macroeconômicas: mudanças permanentes de preferências, por exemplo. Na presença de mais de uma fonte de choques permanentes, modelos econométricos que “identifiquem” apenas um choque permanente estão necessariamente mesclando os efeitos de várias fontes de flutuação numa só. Como consequência, superestimam a importância do choque que é identificado como sendo esse choque permanente.

A visão de que modelos econométricos com hipóteses de identificação

restritivas podem estimar erroneamente a importância de alguns choques econômicos encontra eco nos trabalhos recentes de Francis e Ramey (2001), Hartley e Whitt(2002); ver também a discussão em Christiano, Eichenbaum e Vigfusson (2003). Os primeiros argumentam que os choques identificados econometricamente são potencialmente uma mistura de vários tipos de choques. Uma explicação para esse fenômeno é encontrada em Hartley e Whitt (2002), que acreditam que o número de choques que governam as flutuações agregadas excedem o número de variáveis endógenas.

Esse artigo tem potencialmente duas contribuições relevantes para a mensuração da importância dos choques econômicos sobre as variáveis agregadas, principalmente na mensuração da importância relativa dos choques de produtividade. A primeira diz respeito à grande diferença de resultados empíricos que se verifica quando se implementa a mensuração da importância de choques permanentes e transitórios, quando os choques de produtividade são identificados como os únicos a ter efeitos permanentes, *vis-à-vis* ao que se verifica numa decomposição estrutural onde choques de produtividade e de preferências podem ambos ter efeitos permanentes e transitórios. Ora, sob a hipótese de correta identificação, seria de se esperar resultados similares em ambos os exercícios. Grandes diferenças podem então indicar a inadequação da hipótese de identificação de que a única fonte de choques permanentes seja a produtividade. A segunda contribuição diz respeito ao uso da função de bem-estar econômico na mensuração da importância relativa dos choques econômicos. Durlauf (1989) lembra que, acima de tudo, devemos estar interessados em saber como os choques afetam o nível de bem-estar do indivíduo. Entretanto, do ponto de vista econométrico, a literatura se interessa pelas funções de resposta a impulso das variáveis macroeconômicas, e não pela resposta do bem-estar econômico a esses mesmos choques. Como o bem-estar econômico é uma função do consumo, uma das formas de se investigar como os choques afetam o nível de bem-estar seria usar o próprio consumo na análise econométrica, filtrado pela forma funcional da utilidade instantânea, trazendo os impactos dos diferentes choques a valor presente. Tal exercício é feito aqui seguindo a idéia de Bohn(1991), que considera o que denominamos de função impulso-resposta não-convencional, i.e., a resposta ao impacto dos choques em valor presente.

Após apresentada a metodologia de construção de funções impulso-resposta não convencionais para o bem-estar, investigamos a importância relativa de diferentes tipos de choques. Para isso, aplicaremos a metodologia originalmente proposta aqui a um VAR contendo duas variáveis: o logaritmo do consumo e o da renda. Para fins de comparação, usaremos a mesma base de dados de Blinder e Deaton (1985), também utilizada por Campbell (1987).

A partir do uso desses dados, fazemos uma decomposição dos choques em permanentes (produtividade) e transitórios seguindo King *et alli* (1991); ver também Warne (1993) para uma explicação mais detalhada dessa técnica. Além disso, implementamos as funções impulso-resposta generalizadas de Koop *et alli* (1996) e Pesaran e Shin (1998), onde não há a necessidade de ortogonalizar choques, estimando a importância de choques de preferências e de produtividade. Nesse caso, ambos podem ter tanto efeitos permanentes quanto transitórios. Para as duas formas de identificação de choques propostas, estimaremos através do Método Delta os erros padrões para as estimativas pontuais de modo avaliar a significância das mesmas. Os resultados das duas técnicas são então comparados de forma a validar as hipóteses de identificação. Nesses exercícios, a função de utilidade do consumidor utilizada para compor a função de resposta a impulso não-convencional é a que tem a propriedade CRRA – Constant Relative Risk Aversion – para o coeficiente de aversão relativa ao risco.

Os principais resultados obtidos sugerem que: *i)* a grande importância relativa de choques permanente *vis-à-vis* a dos choques transitórios; *ii)* os choques de produtividade têm um impacto sobre o bem-estar maior que choques de preferências; entretanto, *iii)* não é possível fazer um *matching* entre choques de preferências e de produtividade com as definições de choques permanentes e transitórios. Este último resultado corrobora a discussão acima em Issler e Vahid (2001), Francis e Ramey (2001) e Hartley e Whitt (2002).

Na seção 2 apresentamos a metodologia desenvolvida para a função impulso-resposta não convencional para o bem-estar. Na seção 3 serão discutidas algumas formas de ortogonalização dos erros, usando funções impulso-resposta generalizada de Koop *et alli* (1996) e Pesaran e Shin (1998) e decomposição de choques permanentes e transitórios de Warne (1993). Na Seção 4 apresentamos os resultados empíricos do uso da metodologia proposta na Seção 3. Por fim, na seção 5 serão apresentadas as principais conclusões deste trabalho.

3.2 Funções Impulso-Resposta Não Convencionais para o Bem-Estar

Para avaliar como o nível de bem-estar do agente econômico responde a choques sobre variáveis econômicas devemos estudar como os choques afetam a sua trajetória ótima de consumo, solução de seu problema de otimiza-

ção intertemporal. Neste sentido, o impacto no bem-estar deve ser uma soma descontada dos impactos na utilidade do consumo a cada período $t + i$, $i = 0, 1, \dots$. Considere então um problema de otimização intertemporal em que um consumidor representativo *forward-looking* escolhe a trajetória de consumo que seja solução do seguinte problema:

$$\begin{aligned} \underset{\{C_{t+i}\}_{i=0}^{\infty}}{\text{Max}} \quad & E_t \left[\sum_{i=0}^{\infty} (\beta^i u(C_{t+i})) \right] \\ \text{s.a.} \quad & W_{t+1+i} = (W_{t+i} + Y_{t+i} - C_{t+i}) (1 + r_{t+i}), \forall i \geq 0 \\ & W_0 \text{ dado e } \lim_{j \rightarrow \infty} \beta^j E_t W_{t+j} = 0 \end{aligned} \quad (1)$$

em que W_t , Y_t , C_t e r_t são, respectivamente a riqueza, a renda do trabalho, o consumo e a taxa de juros no período t , todos em termos reais, ressaltando que W_0 é a riqueza inicial.

O problema do consumidor descrito em (1) pode ser entendido como a maximização de uma função valor presente de modo que podemos utilizar a metodologia proposta por Bohn (1991) para mensurar o efeito de choques no período t sobre o bem-estar do consumidor, utilizando o conceito de função impulso-resposta não convencional.

Seja U_t o valor presente descontado da utilidade na trajetória do consumo $\{C_{t+i}\}_{i=0}^{\infty}$ solução do problema de otimização do consumidor.

$$\begin{aligned} U_t &= \sum_{i=0}^{\infty} \beta^i E[u(C_{t+i}) | \Omega_t] \\ &= u(C_t) + \sum_{i \geq 1} \beta^i E[u(C_{t+i}) | \Omega_t] \end{aligned} \quad (2)$$

em que $E[u(C_{t+i}) | \Omega_t]$ é o valor esperado da utilidade do consumo no período $t + i$ condicionado ao conjunto de informações Ω_t disponível no período t .

Defina o valor presente de uma variável Z avaliado no período t como a soma descontada de todos os valores futuros desta, ou seja, $PV(Z)_t = \sum_{j \geq 1} \beta^j Z_{t+j}$. A inovação no valor presente de uma variável pode ser definida como $\widehat{PV}(Z)_t = E_t[PV(Z)_t] - E_{t-1}[PV(Z)_t]$. Logo, a equação (2) pode ser reescrita como,

$$U_t = u(C_t) + E_t PV[u(C)]_t \quad (3)$$

Considerando a diferença entre o valor esperado condicionado ao conjunto de informação em t e $t - 1$ da equação (3), podemos escrever a inovação

na utilidade indireta do consumidor advinda de um choque no período t como,

$$\widehat{U}_t = \widehat{u(C_t)} + \widehat{PV}[u(C)]_t \quad (4)$$

Proposição 1 *Para qualquer variável aleatória Z , onde $\widehat{Z}_t = E_t Z_t - E_{t-1} Z_t$ é a inovação em Z_t e β é um fator de desconto intertemporal, valem as seguintes relações:*

- i) $(1 - \beta) [Z_t + PV(Z)_t] = Z_t + PV(\Delta Z)_t$
- ii) $(1 - \beta) [\widehat{Z}_t + \widehat{PV}(Z)_t] = \widehat{Z}_t + \widehat{PV}(\Delta Z)_t$
- iii) $\widehat{\Delta Z}_t = \widehat{Z}_t$

Prova.

$$\begin{aligned} i) \quad (1 - \beta) [Z_t + PV(Z)_t] &= Z_t + \sum_{j \geq 1} \beta^j Z_{t+j} - \beta Z_t - \sum_{j \geq 1} \beta^{j+1} Z_{t+j} \\ &= Z_t + \sum_{j \geq 1} \beta^j Z_{t+j} - \sum_{j \geq 0} \beta^{j+1} Z_{t+j} \\ &= Z_t + \sum_{j \geq 1} \beta^j Z_{t+j} - \sum_{j \geq 1} \beta^j Z_{t+j-1} \\ &= Z_t + \sum_{j \geq 1} \beta^j (Z_{t+j} - Z_{t+j-1}) \\ &= Z_t + PV(\Delta Z)_t \end{aligned}$$

ii) Basta tomar a diferença entre os valores esperados condicionados aos conjuntos de informação em t e $t - 1$ da expressão do item i).

$$\begin{aligned} iii) \quad \widehat{Z}_t &= E_t Z_t - E_{t-1} Z_t \\ \widehat{\Delta Z}_t &= E_t \Delta Z_t - E_{t-1} \Delta Z_t = E_t (Z_t - Z_{t-1}) - E_{t-1} (Z_t - Z_{t-1}) \\ &= E_t Z_t - E_{t-1} Z_t - Z_{t-1} + Z_{t-1} = \widehat{Z}_t \end{aligned}$$

o último passo vem do fato de que $Z_{t-1} \in \Omega_{t-1}$, e portanto, $E_t Z_{t-1} = E_{t-1} Z_{t-1} = Z_{t-1}$. ■

Multiplicando a equação (4) por $(1 - \beta)$ e aplicando a proposição 1 temos,

$$\begin{aligned} (1 - \beta) \widehat{U}_t &= (1 - \beta) \left\{ \widehat{u(C_t)} + \widehat{PV}[u(C)]_t \right\} \\ &= \widehat{\Delta u(C_t)} + \widehat{PV}[\Delta u(C)]_t \end{aligned}$$

Assim a inovação na utilidade indireta do indivíduo pode ser expressa por,

$$\widehat{U}_t = \frac{1}{1-\beta} \left[\widehat{\Delta u(C_t)} + \widehat{PV}[\Delta u(C)]_t \right] \quad (5)$$

Como o objetivo final é estimar a resposta de um choque de variáveis que estão no conjunto de informações do indivíduo no período t sobre a equação (5), temos que obter,

$$\frac{\partial \widehat{U}_t}{\partial \mu_t^i} = \frac{1}{1-\beta} \left[\frac{\partial \widehat{\Delta u(c_t)}}{\partial \mu_t^i} + \frac{\partial \widehat{PV}[\Delta u(c)]_t}{\partial \mu_t^i} \right] \quad (6)$$

em que μ_t^i representa o choque na i -ésima variável que está no conjunto de informação do indivíduo no período t .¹

Para estimar a equação (6) utilizaremos as funções impulso-resposta de um modelo autorregressivo vetorial (VAR)². Evidências apontam para o fato de que existem choques que têm efeito persistente sobre as variáveis econômicas como por exemplo um avanço tecnológico que afetam as decisões de consumo ou a produtividade do trabalho. Tais efeitos permanentes está diretamente ligado à hipótese de que as variáveis são integradas de ordem 1, $I(1)$. Logo, estamos supondo que as variáveis do conjunto de informação do consumidor são estacionárias em diferença e que podem existir relações de longo prazo, de modo que podemos escrever o VAR na sua representação de correção de erros (VECM).

Seja um VECM de ordem p , incluindo a variável $\Delta u(C_t)$ e m outras variáveis³ que estão no conjunto de informações do consumidor (Ω_t).

$$\begin{pmatrix} \Delta u(C_t) \\ \Delta \Psi_t \end{pmatrix} = \delta - \gamma \alpha' \begin{pmatrix} u(C_{t-1}) \\ \Psi_{t-1} \end{pmatrix} + \begin{pmatrix} a(L) & B(L) \\ C(L) & D(L) \end{pmatrix} \begin{pmatrix} \Delta u(C_{t-1}) \\ \Delta \Psi_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ \varsigma_t \end{pmatrix} \quad (7)$$

em que supõe-se que Ψ_t seja um vetor de m variáveis $I(1)$ que estão no conjunto de informação (Ω_t) do consumidor no período t , $a(L)$ é um polinômio autorregressivo de ordem $p-1$, $B(L)$, $C(L)$ e $D(L)$ são polinômios autorregressivos matriciais também de ordem $p-1$, sendo que B_i é $(1 \times m)$, C_i é $(m \times 1)$ e D_i é $(m \times m)$.⁴ ε_t e ς_t são ruídos brancos. δ é um vetor

¹A interpretação do choque μ_t^i está vinculada ao modelo estrutural considerado e à hipótese de ortogonalização da matriz de covariância dos choques.

²de ordem $p+1$.

³ $n = m + 1 = n^\circ$ de variáveis do VAR

⁴ $a(L) = a_1 + a_2 L + \dots + a_p L^{p-1}$

com n constantes. Além disso, o termo do lado direito da equação (7) com variáveis em nível representa a relação estável de longo prazo, onde γ é uma matriz de coeficientes ($n \times r$) denotando a sensibilidade à desvios desta condição, α é uma matriz ($n \times r$), contendo os r vetores de cointegração. Fazendo $\Delta X_t = \begin{pmatrix} \Delta u(C_t) & (\Delta \Psi_t)' \end{pmatrix}'$ um vetor [$n \times 1$], $A(L)$ um polinômio autorregressivo matricial⁵ de ordem p , $RLP_t = -\alpha' \begin{pmatrix} u(C_{t-1}) \\ \Psi_{t-1} \end{pmatrix}$ a relação estável de longo prazo e $\eta_t = \begin{pmatrix} \varepsilon_t & \varsigma_t' \end{pmatrix}'$ podemos escrever (7) como,

$$\Delta X_t = \delta + A_1 \Delta X_{t-1} + \dots + A_p \Delta X_{t-p} + \gamma RLP_t + \eta_t \quad (8)$$

Reescrevendo a equação (8) na forma de um VAR(1) e lembrando que $RLP_t = -\alpha' \begin{pmatrix} \Delta u(C_{t-1}) \\ \Delta \Psi_{t-1} \end{pmatrix} + RLP_{t-1}$ temos,

$$\begin{pmatrix} \Delta X_t \\ \vdots \\ \Delta X_{t-p+1} \\ RLP_{t-p} \end{pmatrix} = \begin{pmatrix} \delta \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} A_1^* & \dots & A_{p-1}^* & A_p^* & \gamma \\ I_{(n+1)} & \dots & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & I_{(n+1)} & 0 & 0 \\ 0 & \dots & 0 & -\alpha' & 1 \end{pmatrix} \begin{pmatrix} \Delta X_{t-1} \\ \vdots \\ \Delta X_{t-p} \\ RLP_{t-p-1} \end{pmatrix} + \begin{pmatrix} \eta_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$\Delta X_t^* = \delta^* + A^* \Delta X_{t-1}^* + \eta_t^* \quad (9)$$

em que $A_i^* = A_i - \gamma \alpha'$ são matrizes ($n \times n$), $\Delta X_t^* = \begin{pmatrix} \Delta X_t' & \dots & \Delta X_{t-p+1}' & RLP_{t-p}' \end{pmatrix}'$, $\delta^* = \begin{pmatrix} \delta' & 0 & \dots & 0 \end{pmatrix}'$ e $\eta_t^* = \begin{pmatrix} \eta_t' & 0 & \dots & 0 \end{pmatrix}'$ são vetores com $np + 1$ elementos. Veremos a seguir que a matriz A^* , $(np + 1) \times (np + 1)$, será útil para definir a função impulso-resposta não-convencional. A princípio, pode-se observar que a resposta j -períodos a frente pode ser escrita por⁶,

$$\widehat{\Delta X_{t+j}^*} = E_t \Delta X_{t+j}^* - E_{t-1} \Delta X_{t+j}^* = (A^*)^j \eta_t^*$$

De modo que,

$$\widehat{PV}(\Delta X^*)_t = \sum_{j \geq 1} (\beta A^*)^j \eta_t^*$$

$$\begin{aligned} B(L) &= B_1 + B_2 L + \dots + B_p L^{p-1} \\ C(L) &= C_1 + C_2 L + \dots + C_p L^{p-1} \\ D(L) &= D_1 + D_2 L + \dots + D_p L^{p-1} \end{aligned}$$

$$^5 A(L) = A_1 L + A_2 L + \dots + A_p L$$

onde A_i é ($n \times n$)

⁶Desde que $E_t[\eta_{t+j}^*] = 0$, para todo $j \geq 1$

$$\begin{aligned}
\frac{\partial \widehat{PV}(\Delta X^*)_t}{\partial \eta_t^*} &= \sum_{j \geq 1} (\beta A^*)^j = \beta A^* + (\beta A^*)^2 + (\beta A^*)^3 + \dots \\
&= \beta A^* [I + \beta A^* + (\beta A^*)^2 + (\beta A^*)^3 + \dots] = (\beta A^*)(I - \beta A^*)^{-1} \\
&= (I - \beta A^*)^{-1} - I
\end{aligned} \tag{10}$$

Além disso,

$$\frac{\partial \widehat{\Delta X_t^*}}{\partial \eta_t^*} = I \tag{11}$$

Como visto anteriormente, queremos estimar a equação (6), onde precisamos da inovação no incremento de utilidade, $\widehat{\Delta u(c_t)}$, e de seu valor presente, $\widehat{PV}(\Delta u(c))_t$. Deste modo podemos usar um vetor linha canônico, $e'_1 = (1 \ 0 \ \dots \ 0)$, de modo a selecionar o primeiro elemento do vetor coluna ΔX_t^* e assim obter a função impulso-resposta não convencional sobre a utilidade indireta do indivíduo, como escrita na equação abaixo.

$$\begin{aligned}
\frac{\partial \widehat{U}_t}{\partial \eta_t^*} &= \frac{1}{1 - \beta} e'_1 [I + \beta A^*(I - \beta A^*)^{-1}] \\
&= \frac{1}{1 - \beta} e'_1 [I + (I - \beta A^*)^{-1} - I] \\
&= \frac{1}{1 - \beta} e'_1 (I - \beta A^*)^{-1}
\end{aligned} \tag{12}$$

em que $e'_1(I - \beta A^*)^{-1}$ é um vetor linha $np + 1$, tal que o i -ésimo elemento representa a resposta sobre a utilidade indireta advinda de um choque na i -ésima variável do vetor ΔX_t^* . A equação (12) é o que temos chamado de função impulso-resposta não convencional. É considerada não convencional por calcular uma resposta sobre o valor-presente de uma inovação a partir de choques no sistema. A equação (12) depende apenas dos valores de β e de A^* . O último pode ser consistentemente estimado através do VAR e o primeiro pode ser estimado ou calibrado.

Cabe ressaltar que o objetivo principal é estimar o impacto de um choque na i -ésima variável no conjunto de informação do indivíduo sobre o nível da utilidade do mesmo, ou seja, a resposta em termos de ganhos de bem-estar advindos de choques no sistema. Até agora vimos o efeito sobre a inovação na utilidade do indivíduo e não sobre o nível de utilidade. No entanto, é possível mostrar que $\frac{\partial U_t}{\partial \eta_t} = \frac{\partial \widehat{U}_t}{\partial \eta_t}$, uma vez que U_t é uma combinação linear de variáveis que são tratadas dentro um modelo autorregressivo linear. No *Apêndice A.1* temos uma prova para esta argumentação.

Observe ainda que ao escrever o VAR na forma de um VAR(1), geramos um vetor de choques $\eta_t^* = (\eta_t' \ 0 \ \cdots \ 0)'$, onde apenas os n primeiros elementos representam os choques do sistema e os demais são todos nulos. Logo, estamos interessados apenas em $\frac{\partial U}{\partial \eta_t}$, que são exatamente os n primeiros elementos de $\frac{1}{1-\beta} e_1' (I - \beta A^*)^{-1}$.

3.3 Ortogonalização dos Choques

Uma discussão que merece destaque é se a matriz de variância-covariância de η_t é ou não diagonal, já que a interpretação de uma função impulso-resposta pode ficar comprometida caso os choques em η_t não sejam ortogonais. Isso requer que a covariância entre os elementos de η_t seja nula, ou seja, que $E(\eta_t \eta_t') = \Sigma$ seja diagonal. A seguir, descrevemos três estratégias para interpretar os choques sobre o bem-estar: *i)* testes de exogeneidade fraca [Johansen (1992,1995)]^{7 8}; *ii)* decomposição de choques transitórios e permanentes [Warne (1993)] e *iii)* impulso-resposta generalizado [Koop *et alli* (1996) e Pesaran e Shin (1998)] e Decomposição de Cholesky.

A não rejeição de uma hipótese de exogeneidade fraca para todas as variáveis de um VAR implica em uma propriedade de separação, tal que a função impulso-resposta não requer uma ortogonalização dos choques. Considerando a representação VECM definida na equação (8) para $n = 2$ e válida a hipótese de que $u(C_t)$ é fracamente exógeno para determinar os parâmetros de interesse do modelo condicional de Ψ_t ,⁹ então choques na média condicional de Ψ_t não afetam $\bar{P}V(\Delta u(C))_t$ e portanto, não teria impacto sobre o bem-estar. Tal hipótese pode ser testada usando procedimento proposto por Johansen; ver *Apêndice A.2*.

Podemos ainda considerar uma abordagem de decomposição de choques em permanente e transitória que é completamente consistente com o modelo de tendência comuns, conforme descrita em Warne (1993); ver *Apêndice A.3*. Tal decomposição está ligada à hipótese de que temos um conjunto de variáveis que tem uma raiz unitária, $X_t \sim I(1)$, e r vetores de cointegração.

Aplicando a decomposição permanente/transitório às funções impulso-resposta não convencionais podemos construir combinações lineares de el-

⁷Ver Engle, Hendry e Richard (1983), Maddala (1992) para uma introdução ao assunto; Ericsson e Irons (1994) para uma revisão mais completa e Ericsson, Irons e Johansen (1992, 1995) para procedimentos de testes.

⁸O procedimento só será útil quando o vetor de variáveis for 2×1 , pois neste caso, se existir exogeneidade fraca, a matriz de variância dos choques é diagonal.

⁹para $n = 2$, Ψ_t é um conjunto de uma única variável.

ementos do vetor de choques η_t de tal forma que sejam consideradas as seguintes restrições: $(n - r)$ inovações permanentes, φ_t , são independentes das r inovações transitórias, ψ_t ; as r inovações transitórias são mutuamente independentes. Assim, $v_t = \begin{bmatrix} \varphi_t \\ \psi_t \end{bmatrix} \eta_t = \Gamma \eta_t$, em que,¹⁰

$$\Gamma = \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \end{bmatrix} = \begin{bmatrix} (\Upsilon' \Upsilon)^{-1} \Upsilon' E_n (I - A^*)^{-1} E_n' \\ Q^{-1} \zeta' \Sigma^{-1} \end{bmatrix} \quad (13)$$

onde $E_n = \begin{pmatrix} I_n & 0 \end{pmatrix}$ é uma matriz que seleciona as n primeiras linhas de X_t^* . Desta forma decomponemos a série em uma parte tendência (não estacionária) e outra cíclica (estacionária). Assim, a Representação de Wold da equação (9) pode ser reescrita por,

$$\begin{pmatrix} \Delta u(C_t) \\ \Delta \Psi_t \end{pmatrix} = \rho + E_n (I - A^* L)^{-1} E_n' \Gamma^{-1} \begin{pmatrix} \varphi_t \\ \psi_t \end{pmatrix} \quad (14)$$

em que $\rho = E_n \rho^* = E_n (I - A^*)^{-1} E_n' \delta$. Portanto, a decomposição de choques permanentes/transitórios para a função impulso-resposta não-convencional é,

$$\begin{aligned} \frac{\partial U}{\partial \varphi_{jt}} &= \frac{1}{1 - \beta} e_1' (I - \beta A^*)^{-1} E_n' \Gamma^{-1} e_j, & 1 \leq j \leq n - r \\ \frac{\partial U}{\partial \psi_{jt}} &= \frac{1}{1 - \beta} e_1' (I - \beta A^*)^{-1} E_n' \Gamma^{-1} e_j, & r + 1 \leq j \leq n \end{aligned} \quad (15)$$

em que $\frac{\partial U}{\partial \varphi_{jt}}$ e $\frac{\partial U}{\partial \psi_{jt}}$ são respectivamente a resposta sobre a utilidade de um choque permanente e um choque transitório.

Pode-se ainda fazer uma identificação do tipo do choque, adotando outras restrições sobre a ortogonalização dos choques. Primeiramente considere uma fatorização de Cholesky tal que a matriz de variância dos resíduos do VAR pode ser escrita por $\Sigma = PP'$, sendo P uma matriz $n \times n$ triangular inferior. Assim, pode-se ortogonalizar os choques escrevendo,

$$X_t = \sum_{i \geq 0} A_i P P^{-1} \eta_{t-i} = \sum_{i \geq 0} A_i P \xi_{t-i}$$

em que $\xi_t = P^{-1} \eta_t$ é tal que $E(\xi_t \xi_t') = I_n$. Assim, o vetor $n \times 1$ da função impulso-resposta ortogonalizado de um choque na j -ésima equação sobre X_{t+k} é dado por,

$$\psi_j^{ortog}(k) = A_k P e_j, \quad k = 0, 1, 2, \dots \quad (16)$$

¹⁰Definições de Υ, Q e ζ ver no Apêndice A.3.

em que e_j é um vetor $n \times 1$ com j -ésimo elemento unitário e zero para os demais. Neste caso, a função impulso-resposta não convencional ortogonalizada usando a Decomposição de Cholesky pode ser escrita por,

$$[\psi_j^{ortog}]_U = \frac{1}{1-\beta} e'_1 [I - \beta A^*]^{-1} E'_n P e_j \quad (17)$$

O problema desta abordagem é que depende da ordem das variáveis apresentadas no VAR. Neste trabalho a ordem do VAR para construção da função impulso-resposta ortogonal é definida de acordo com o resultado do teste de exogeneidade discutido acima.

Koop *et alli* (1996) e Pesaran e Shin (1997) descrevem uma abordagem alternativa que é invariante à ordem do VAR; ver *Apêndice A.4*. Neste caso, a função impulso-resposta generalizada para o efeito de um choque de um desvio-padrão ocorrido na j -ésima equação do VAR para k períodos a frente pode ser escrita como,

$$GIR_X \left(k, \frac{\delta_j}{\sqrt{\sigma_{jj}}}, \Omega_{t-1} \right) = \psi_j^{gen}(k) = \sigma_{jj}^{-\frac{1}{2}} A_k \Sigma e_j, \quad k = 0, 1, 2, \dots \quad (18)$$

Assim, podemos escrever a função impulso-resposta generalizada não convencional para o bem-estar, U , para um choque de um desvio-padrão na j -ésima equação como,

$$\begin{aligned} GIR_U \left(\frac{\delta_j}{\sqrt{\sigma_{22}}}, \Omega_{t-1} \right) &= \frac{1}{1-\beta} \frac{1}{\sqrt{\sigma_{22}}} e'_1 \left[\sum_{i \geq 0} \beta^i (A^*)^i \right] \Sigma^* e_j \\ &= \frac{1}{1-\beta} \frac{1}{\sqrt{\sigma_{22}}} e'_1 [I - \beta A^*]^{-1} \Sigma^* e_j \end{aligned} \quad (19)$$

em que $E[\eta_t^* \eta_t^{*'}] = \Sigma^* = \begin{bmatrix} \Sigma_{n \times n} & 0_{n \times (np-1)} \\ 0_{(np-1) \times n} & 0_{(np-1) \times (np-1)} \end{bmatrix}_{(np+1) \times (np+1)}$ ¹¹.

Para proceder testes de hipóteses para as funções impulso-resposta não convencionais é necessário estimar os desvios-padrões destas estimativas pontuais. Para tanto, pode-se construir experimentos de Monte Carlo usando um grande número de réplicas a partir da geração de vetores de erros aleatórios, ou ainda fazer reamostragem do vetor de resíduos do modelo segundo o método de bootstrap. No entanto, optou-se aqui utilizar o Método Delta em que os desvios-padrões são construídos a partir dos desvios-padrões

¹¹Deve ficar claro que só faz sentido tomar impulso na j -ésima variável para $j = 1, 2, \dots, n$.

estimados das estimativas do modelo VEC considerado. Tal exercício é possível dado que as funções impulso-resposta não convencionais como definidas nas equações (15), (17) e (19) são funções contínuas em $vec(\hat{A}^*)$ e $vech(\hat{\Sigma})$; ver *Apêndice A.5*.

3.4 Base de Dados e Resultados Empíricos

Neste trabalho consideramos uma função de utilidade do tipo CRRA, $u(C) = \frac{C^{1-\sigma}-1}{1-\sigma}$, na qual ¹² $\sigma = 1$, ou seja, $u(C) = \ln(C) = c$,¹³ e que o conjunto Ψ_t está restrito a uma única variável, o logaritmo da renda do trabalho ($\ln(Y) = y$).¹⁴ Neste caso, o vetor X_t é (2×1) definido pelas variáveis $X_t = (c_t \ y_t)'$. Considere a representação VECM descrita pela equação (8), ou sua forma VAR(1) descrita pela equação (9), sendo $m = 1$. Desta maneira, podemos definir a resposta no bem-estar, conforme a equação (12), $\frac{\partial U}{\partial \eta_t^*} = \frac{1}{1-\beta} e_1' (I - \beta A^*)^{-1}$, dependendo apenas dos valores de β e A^* . A matriz A^* pode ser estimada consistentemente usando a forma reduzida do VAR e β pode ser calibrado.

Usaremos a mesma base de dados de Blinder e Deaton (1985) e Campbell (1987) para as séries trimestrais de renda do trabalho e consumo de bens não duráveis e serviços¹⁵, ambas em valores reais per capita para o período de 1953:2 a 1985:4. A *Figura 1* mostra as séries em nível e primeiras diferenças.

Serão considerados os valores do fator de desconto intertemporal (base trimestral) $\beta = \{0.997, 0.995, 0.992, 0.990, 0.987, 0.985, 0.983, 0.980, 0.978, 0.976\}$ para a estimação da função impulso-resposta não convencional da utilidade. Cabe lembrar que os valores escolhidos são condizentes com taxa anuais de juros no conjunto $r = \{0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.10\}$.

¹² σ é o coeficiente de aversão relativa ao risco ou o inverso da elasticidade substituição intertemporal.

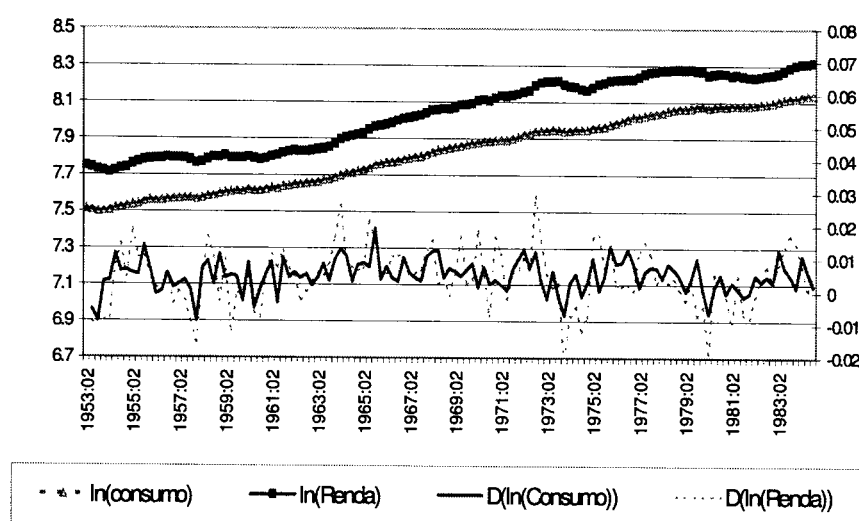
¹³ Aplicando o teorema de L'Hopital, $\lim_{\sigma \rightarrow 1} \frac{C^{1-\sigma}-1}{1-\sigma} = \lim_{\sigma \rightarrow 1} \frac{-C^{1-\sigma} \ln(C)}{-1} = \ln(C) = c$

¹⁴ Considerando o problema (1), basta fazer a hipótese de Hall (1978) que a taxa de juros é constante, satisfazendo a equação $1 = (1+r)\beta$. Assim o conjunto de informações do consumidor fica restrito à $\{u(C_{t-i}), \ln(Y_{t-i})\}_{i=1}^{\infty}$.

¹⁵ A construção das séries de renda do trabalho (Y) e consumo de não-duráveis e serviços (C) seguiu as seguintes restrições: a) O pagamento de juros dos consumidores para as firmas são subtraídos do computo da renda disponível; b) gastos com vestuário e calçados são considerados como consumo durável. Ver Campbell (1987, p. 1259).

Observamos na *Figura 1* que todas as séries são muito suaves, o que indicam um alto grau de persistência e suas primeiras diferenças elimina esta característica. Tal comportamento pode ser justificado pela presença de uma raiz unitária, comum a ambas as séries, o que é corroborado pelo teste de cointegração, conforme veremos abaixo.

Figura 1
Séries em Nível e 1ª Diferença



Realizamos o teste de cointegração entre $u(C_t)$ e $\ln(Y_t)$ para verificar se existe alguma relação de longo prazo estacionária. É de conhecimento que os resultados do teste de cointegração usando Johansen (1988) depende dos componentes determinísticos incluídos no VAR e da escolha das defasagens. Assim, pré-testes são realizados para proceder a escolha do modelo a ser usado no teste de cointegração. O número de defasagens usado no VAR foi selecionado usando os critérios de informação de Schwarz e Hannan-Quinn. A escolha dos componentes determinísticos usados no teste se deu com base no teste de razão de verossimilhança proposto por Johansen. Os testes indicaram a escolha de um VAR com duas defasagens e sem constante.

Os resultados para o teste de cointegração estão apresentados na *Tabela 1*. Não rejeitamos a hipótese nula de um vetor de cointegração ao nível de 1% de significância. O vetor de cointegração estimado é (1, -1.023) (normalização sobre $u(C) = \ln(C)$). É natural testar a hipótese de que o vetor

de cointegração é (1,-1), uma vez que este formato da função de utilidade sugere também que existe uma relação $\ln(C/Y)$ estacionária. Conclui-se que a hipótese não pode ser rejeitada ao nível de 5% de significância.

Tabela 1
Teste de Cointegração de Johansen

Ho: No. Eq de Cointegração	Estatística do Traço	Val. Críticos		Estatística do Max. Autovalor	Val. Críticos	
k = 0	46.275 **	5%	12.53	43.356 **	5%	11.44
		1%	16.31		1%	15.69
k ≤ 1	2.918	5%	3.84	2.918	5%	3.84
		1%	6.51		1%	6.51

*(**) indica rejeição de Ho a 5%(1%) de significância

Vetor de Cointegração:

$u(C)=\ln(Cons)$ $\ln(Renda)$
1.000 -1.023

Teste de Hipótese para vetor de cointegração (1,-1)

Estatística de Teste ($\chi^2(1)$) 0.369
p-valor 54.4%

A *Tabela 2* mostra os resultados das estimativas do modelo de correção de erros, construído a partir de um VAR(2) sem tendência determinística e com a restrição de 1 vetor de cointegração igual a (1, -1). A equação 1 mostra que Δy_{t-1} e a equação de cointegração são significativos para estimar $\Delta u(C_t)$, enquanto que a equação 2 mostra que $\Delta u(C_{t-1})$ e Δy_{t-1} são significativos para estimar Δy_t .

Tabela 2
Modelo de Correção de Erros (VECM)

Regressores	Coeficiente	Desvio-Padrão	Estatística t	p-valor
Eq.1 - D(u(C)=ln(C))				
D(u(C),-1)	0.08528	0.10020	0.85115	0.39576
D(ln(Y),-1)	0.14115	0.05576	2.53145	0.01217
RLP(-1)	-0.01014	0.00150	-6.76366	0.00000
Eq.2 - D(ln(Y))				
D(u(C),-1)	0.36478	0.16909	2.15730	0.03224
D(ln(Y),-1)	0.33696	0.09409	3.58106	0.00043
RLP(-1)	-0.00303	0.00253	-1.19794	0.23243

A *Tabela 3* apresenta os resultados do teste de exogeneidade fraca conforme discutido na seção anterior. Rejeita-se a hipótese de que $u(C) = \ln(C)$ seja fracamente exógena para os parâmetros de interesse do

modelo condicional de $\ln(Y)$, o que mostra que os choques na média condicional de $\ln(Y)$ não podem ser desconsiderados ao construir $\widehat{PV}(\Delta u(C))_t$ e portanto espera-se que tenham impacto sobre o bem-estar do indivíduo. Entretanto, não se rejeita a hipótese de que $\ln(Y)$ seja fracamente exógeno para os parâmetros de interesse do modelo condicional de $u(C)$. Isso implica que a ortogonalização dos choques da forma reduzida pode ser feita colocando-se primeiramente o choque da equação da renda e, posteriormente, uma combinação linear dos dois choques da forma reduzida. Logo, na Decomposição de Cholesky, o choque da renda aparece hierarquicamente no vetor de choques da forma estrutural, afetando a equação do consumo. Já o choque do consumo, não afeta a equação da renda.

Tabela 3
Teste de Exogeneidade Fraca sobre os
coeficientes de ajustamento do Modelo de Correção de Erros

Hipótese Nula	Estatística do Teste	p-valor
$\ln(Y)$ é fracamente exógeno para os parâmetros de interesse do modelo condicional de $u(C)=\ln(C)$	1.4351	0.2309
$u(C)=\ln(C)$ é fracamente exógeno para os parâmetros de interesse do modelo condicional de $\ln(Y)$	45.7471	0.0000

Quando ortogonalizamos os choques seguindo a técnica de Cholesky, herdamos dois choques estruturais interpretáveis. O primeiro afeta $u(C)$ e o segundo afeta $\ln(Y)$ sendo ortogonal ao primeiro. Denominaremos o primeiro choque de “choque de preferências”, por razões óbvias. Denominaremos o segundo de “choque de produtividade”, pois, além de ser ortogonal ao choque de preferências, é o choque da equação da renda. Note que, nesse contexto, ambos podem ter efeitos permanentes e transitórios, o que nos será útil para comparar os resultados aqui obtidos aos da decomposição permanente/transitório feita abaixo.

Conforme discutido na seção anterior, os resultados de uma função impulso-resposta só são interpretáveis quando a matriz de variância-covariância dos erros do modelo de correção de erros é diagonal. O coeficiente de correlação estimado entre as duas formas reduzidas dos resíduos estimados no modelo é 0.46494. O teste de razão de verossimilhança para uma matriz de variância-covariância diagonal é 53.91 com uma distribuição $\chi^2(1)$. Os valores críticos para 5% e 1% são 3.84 e 6.63 respectivamente. Assim, rejeitamos a hipótese de ortogonalidade das inovações, requerendo

alguma técnica de ortogonização sobre a forma reduzida dos resíduos para a construção das funções impulso-resposta convencional e não-convencional. Utilizamos assim três técnicas para avaliar o impulso-resposta: *i*) decomposição de choques transitórios e permanentes construindo combinações lineares dos resíduos que são ortogonais com matriz de variância-covariância I_2 ; *ii*) construção de funções impulso-resposta generalizadas não convencionais que não requerem a necessidade da ortogonalização dos resíduos, mas usam a informação da matriz de variância-covariância estimada para os resíduos; e *iii*) ortogonalização dos resíduos usando decomposição de Cholesky.

A decomposição em choques permanentes e transitórios da FIRNC foi realizada utilizando a equação (15) e os resultados são apresentados na *Tabela 4*. Estimamos os desvios-padrões das estimativas pontuais usando o Método Delta discutido na seção anterior. Todas as estimativas pontuais são significativamente diferentes de zero. Com a decomposição de Warne observa-se que os choques permanentes têm um impacto no bem-estar de magnitude considerável. Por outro lado, os choques transitórios são desprezíveis em afetar o bem-estar do consumidor *forward looking*. Este resultado é condizente com a proposição de Lucas (1987) em afirmar que a perda de bem-estar produzida pelo ciclo econômico é desprezível e portanto políticas de suavização do ciclo não são eficientes. Observa-se ainda que a resposta de choque permanentes depende do valor de β : quanto mais paciente for o indivíduo (maior β), maior o efeito sobre o bem-estar.

Tabela 4
Decomposição de choques Permanentes e Transitórios para as FIRNC

β	Tipo de Choque	
	Permanente	Transitório
0.997	164.2640 (13.47200)	0.0042 (0.00037)
0.995	74.8187 (7.58255)	0.0031 (0.00022)
0.992	34.8960 (4.08087)	0.0022 (0.00012)
0.99	24.0763 (2.96926)	0.0019 (0.00009)
0.987	15.4970 (2.01370)	0.0015 (0.00007)
0.985	12.1759 (1.62128)	0.0014 (0.00006)
0.983	9.8601 (1.33850)	0.0012 (0.00005)
0.98	7.5016 (1.04148)	0.0011 (0.00004)
0.978	6.3938 (0.89836)	0.0010 (0.00003)
0.976	5.5286 (0.78476)	0.0009 (0.00003)

Nota: Desvio-padrão entre parênteses

Para identificar diferentes tipos de choques, construiu-se funções impulso-resposta não convencionais usando a Decomposição de Cholesky e o método generalizado, apresentando choques “de preferência” e “de produtividade”. Primeiramente, discutiremos os resultados para o método generalizado.

Na *Tabela 5* e *Figura 2* apresentamos a função impulso-resposta generalizada convencional para $u(C) = \ln(C)$ e $\Delta u(C)$. Para o período inicial a função impulso-resposta generalizada para $\Delta u(C)$ dado um choque de 1 desvio-padrão em $\Delta u(C_t)$ é 0.0046 e 0.00214.¹⁶ Para um período a frente os valores são 0.00094 e 0.00138; 10 períodos a frente são 0.000065 e 0.00023. Com isso, verifica-se que a resposta de $\Delta u(C)$ a choques no consumo é maior somente no primeiro período, tendendo mais rapidamente a zero que choques ocorridos na renda do trabalho.

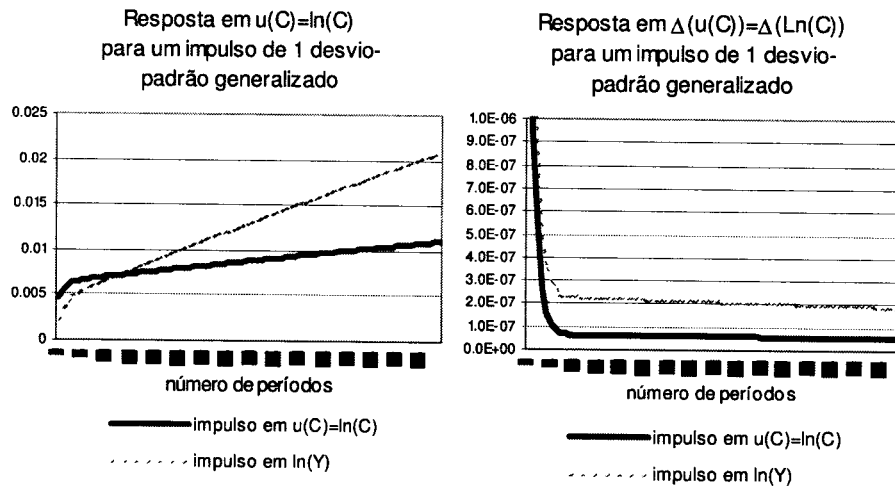
¹⁶Exatamente igual a 1 $DP(\Delta u(C))$ e respectivo efeito contemporâneo para o choque em $\Delta \ln(Y)$, i.e., $Cov(\Delta u(C), \Delta \ln(Y))/DP(\Delta \ln(Y))$.

Tabela 5
Função impulso-resposta convencional

Período	Resposta em $\Delta u(C)$	
	Impulso: $\Delta u(C)$	Impulso: $\Delta \ln(Y)$
t	0.004600	0.002139
t+1	0.000939	0.001380
t+2	0.000521	0.000754
t+3	0.000275	0.000474
t+4	0.000162	0.000343
t+5	0.000110	0.000282
t+10	0.000065	0.000227
t+20	0.000062	0.000220
t+50	0.000057	0.000203
t+100	0.000050	0.000177

Obs.: $DP(\Delta u(C))=0.0046$; $DP(\Delta \ln(Y))=0.00776$ e
 $Cov(Du(C), D\ln(Y))=0.0000166$

Figura 2



Usando a equação (19), construímos a função impulso-resposta generalizada não-convencional (FIRNC) para os valores do fator de desconto intertemporal, β , definidos no início desta seção. Novamente, usamos o Método Delta para estimar os desvios-padrões das estimativas pontuais. Os resultados estão apresentados na *Tabela 6*. Observa-se que todas as estimativas pontuais são significativamente diferentes de zero. Analisando a evolução da FIRNC e da razão $\frac{GIR_U(\ln(Y))}{GIR_U(u(C))}$ para diferentes valores do fator de desconto β , conforme reportadas na *Figura 3*, pode-se concluir que: i) a FIRNC aumenta monotonicamente com o aumento de β , ou seja, consumidores que descontam menos o futuro (β maior) terão maior impacto

no seu bem-estar a partir de choques no seu conjunto de informações; *ii*) $\frac{GIR_U(\ln(Y))}{GIR_U(u(C))} > 1$ evidencia que choques sobre variáveis reais têm um impacto consideravelmente maior que choques de preferência; *iii*) efeito diferenciado segundo a origem do choque de modo que a razão $\frac{GIR_U(\ln(Y))}{GIR_U(u(C))}$ cresce monotonicamente, mostrando que os choques sobre preferências têm um impacto relativo cada vez menor para os indivíduos que descontam menos o futuro.

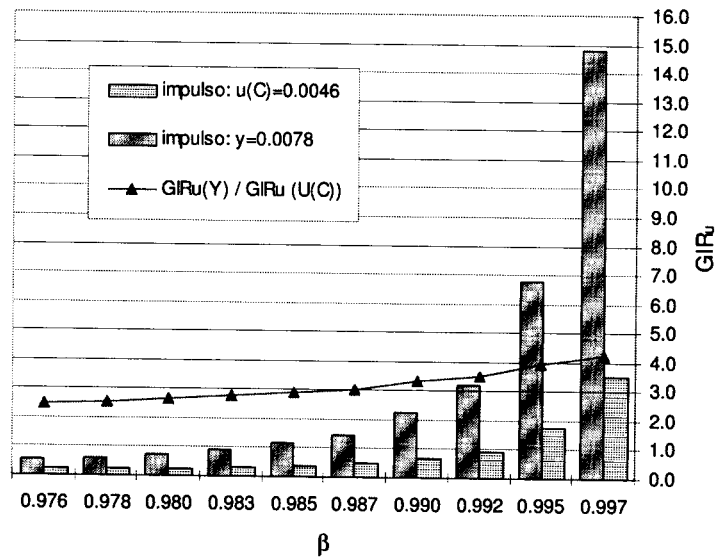
Tabela 6

Função Impulso-Resposta Generalizada não-Convencional para o Bem-estar

Variável de Impulso	u(C)=ln(C)	y=ln(Y)	
Choque Inicial = 1DP	0.00462	0.00777	
β	$GIR_U(u(C))$	$GIR_U(y)$	$GIR_U(y) / GIR_U(u(C))$
0.997	3.4868 (0.20427)	14.7976 (1.01141)	4.24
0.995	1.7390 (0.07662)	6.7798 (0.33964)	3.90
0.992	0.9102 (0.03048)	3.1884 (0.11479)	3.50
0.99	0.6708 (0.01980)	2.2111 (0.06755)	3.30
0.987	0.4708 (0.01206)	1.4335 (0.03618)	3.04
0.985	0.3892 (0.00927)	1.1314 (0.02584)	2.91
0.983	0.3302 (0.00740)	0.9202 (0.01933)	2.79
0.98	0.2674 (0.00556)	0.7044 (0.01339)	2.63
0.978	0.2366 (0.00472)	0.6026 (0.01085)	2.55
0.976	0.2119 (0.00408)	0.5230 (0.00900)	2.47

Nota: Desvio-padrão entre parênteses

Figura 3
 Função Impulso-Resposta Generalizada não-Convencional
 para o Bem-estar para valores escolhidos de β



Estes resultados são corroborados ao se usar a Decomposição de Cholesky para ortogonalizar os choques, conforme apresentados na *Tabela 7*, onde a resposta a choques de preferências se mostraram um pouco maiores aos resultados do método generalizado, reduzindo a razão entre choques de produtividade e de preferências.

Tabela 7
Função Impulso-Resposta não-Convencional para o Bem-estar
usando Decomposição de Cholesky

β	$\Psi_U^{ortog}(u(C))$	$\Psi_U^{ortog}(y)$	$\Psi_U^{ortog}(y) / \Psi_U^{ortog}(u(C))$
0.997	5.8691 (0.34502)	13.6417 (0.98865)	2.32
0.995	2.9271 (0.12936)	6.1230 (0.33052)	2.09
0.992	1.5320 (0.05125)	2.7964 (0.11071)	1.83
0.99	1.1292 (0.03314)	1.9036 (0.06469)	1.69
0.987	0.7925 (0.02001)	1.2019 (0.03423)	1.52
0.985	0.6552 (0.01528)	0.9327 (0.02423)	1.42
0.983	0.5557 (0.01211)	0.7463 (0.01796)	1.34
0.98	0.4501 (0.00900)	0.5581 (0.01226)	1.24
0.978	0.3983 (0.00758)	0.4704 (0.00984)	1.18
0.976	0.3566 (0.00650)	0.4024 (0.00808)	1.13

Nota: Desvio-padrão entre parênteses

Contudo, ambas as razões resultantes, $\frac{GIR_U(\ln(Y))}{GIR_U(u(C))}$ e $\frac{[\psi_y^{ortog}]_U}{[\psi_{u(C)}^{ortog}]_U}$, não são grandes o suficiente quando comparadas com a razão entre choques permanentes e transitórios – cuja razão é de, pelo menos, 6000 vezes! Ora, à luz desses resultados díspares, não se pode crer que os choques com efeitos permanentes sejam apenas de produtividade (ou apenas de preferências). Somos obrigados a concordar com a tendência da literatura recente, e.g., Issler e Vahid(2001), Francis e Ramey (2001) e Hartley e Whitt (2002), que advogam que choques permanentes são de fato uma mistura de choques de diferentes fontes.

3.5 Conclusão

Esse artigo discute a mensuração da importância dos choques econômicos sobre as variáveis agregadas, principalmente a mensuração da importância relativa dos choques de produtividade. Para tanto, compara-se os resultados

empíricos da importância de choques permanentes e transitórios, quando os choques de produtividade são identificados como os únicos a ter efeitos permanentes, *vis-à-vis* aos de uma decomposição estrutural, onde choques de produtividade e de preferências podem ambos ter efeitos permanentes e transitórios. Sob a hipótese de correta identificação, seria de se esperar resultados similares em ambos os exercícios.

Na comparação da importância dos choques econômicos, tratou-se a questão de forma inovadora, avaliando o impacto desses choques sobre a função de bem-estar econômico. Tal exercício é feito aqui seguindo a idéia de Bohn (1991), que considera o que denominamos de função impulso-resposta não-convencional, i.e., a resposta ao impacto dos choques em valor presente, na função utilidade do agente representativo.

Uma decomposição de choques permanentes e transitórios foi estimada obtendo-se os seguintes resultados: *i*) choques permanentes têm um impacto no bem-estar de magnitude considerável; *ii*) choques transitórios são desprezíveis; *iii*) quanto mais paciente for o indivíduo (maior β) maior o efeito de choques permanentes sobre o bem-estar do indivíduo. Este último resultado é condizente com a proposição de Lucas (1987) em afirmar que a perda de bem-estar produzida pelo ciclo econômico é desprezível e portanto políticas de suavização do ciclo não são eficientes.

Foram identificados choques “de preferência” e “de produtividade” sobre o bem-estar usando funções impulso-resposta não convencionais pelo método generalizado e por ortogonalização via Decomposição de Cholesky. Os resultados apontam que: *i*) os choques de preferência e de produtividade têm impacto significativo sobre o nível de bem-estar; *ii*) a função impulso-resposta não-convencional aumenta monotonicamente com o aumento de β , ou seja, consumidores que descontam menos o futuro (β maior) terão maior impacto no seu bem-estar; *iii*) choques de produtividade tem um impacto consideravelmente maior que choques de preferência; *iv*) choques de preferências têm um impacto relativo cada vez menor para os indivíduos que descontam menos o futuro (β maior).

Contudo, observando as razões entre os choques de produtividade e de preferências na decomposição permanente/transitório com aquela obtida na Decomposição de Cholesky (ou generalizada), nota-se que estas são incompatíveis entre si. À luz desses resultados díspares, não se pode crer que os choques com efeitos permanentes sejam apenas de produtividade (ou apenas de preferências). Somos obrigados a concordar com a tendência da literatura recente, e.g., Issler e Vahid(2001), Francis e Ramey (2001) e Hartley e Whitt (2002), que advogam que choques permanentes são de fato uma mistura de choques de diferentes fontes.

Referências

- [1] Blinder, A., Deaton, A., 1985. The time series consumption function revisited, *Brookings Papers on Economic Activity*, 465-511.
- [2] Blanchard, O.J., Quah, D., 1989. The Dynamic Effects of Aggregate Supply and Demand Disturbances. *American Economic Review* 79, 151-174.
- [3] Bohn, H., 1991. Budget balance through revenue or spending adjustments? Some historical evidence for the United States. *Journal of Monetary Economics* 27, 333-359.
- [4] Campbell, J., 1987. Does saving anticipate declining labor income? An alternative test of the permanent income hypothesis. *Econometrica* 55(6), 1249-1273.
- [5] Campbell, J.Y., Mankiw, N.G., 1987. Permanent and transitory components in macroeconomic fluctuations. *American Economic Review* 77, 111-117.
- [6] Christiano, L., Eichenbaum, M., Vigfusson, R., 2003, "What Happens After a Technology Shock?" NBER Working Paper, Julho de 2003.
- [7] Cochrane, J.H., 1988. How big is the random walk in GNP? *Journal of Political Economy* 96, 893-920.
- [8] Cochrane, J.H., 1994. Permanent and transitory components of GNP and stock prices. *Quarterly Journal of Economics* 109, 241-265.
- [9] Durlauf, S.N., 1989. Output Persistence, Economic Structure, and the Choice of Stabilization Policy. *Brookings Papers on Economic Activity* 1989(2), 69-136, com comentários.
- [10] Engle, R.F., Hendry, D.F., Richard, J.F. 1983. Exogeneity. *Econometrica* 55, 277-304.
- [11] Engle, R.F., Granger, C.W.J., 1987. Co-integration and error correction: representation, estimation and testing. *Econometrica* 55, 251-276.
- [12] Ericsson, N.R., 1994. Testing exogeneity: an introduction. In: Ericsson, N.R., Iron, J.S. (Eds.), Testing Exogeneity. *Oxford Univ. Press*, Oxford.
- [13] Ericsson, N.R., Irons, J.S., 1994. Testing exogeneity. *Oxford Univ. Press*, Oxford.

- [26] Johansen, S., 1995. Likelihood-Based Inference in Cointegrated Vector Autoregressive Models. *Oxford University Press*, Oxford.
- [27] King, R.G., Plosser, C.I., Stock, J.H., Watson, M.W., 1991. Stochastic trends and economic fluctuations. *American Economic Review*, 81, 819-840.
- [28] Koop, G., Pesaran, M.H., Potter, S.M., 1996. Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics* 74, 119-47.
- [29] Lucas, R. “*Models of Business Cycles*”, Oxford: Blackwell, 1987.
- [30] Maddala, G.S., 1992. Introduction to Econometrics. 2nd Edition Prentice-Hall, Englewood Cliffs, NJ.
- [31] Nelson, C.R., Plosser, C.I., 1982. Trends and random walks in macroeconomic time series. *Journal of Monetary Economics* 10, 139-162.
- [32] Pesaran, M.H., Shin, Y., 1998. Generalized impulse response analysis in linear multivariate models. *Economics Letters*, 1998, Vol.58, pp.17-29.
- [33] Sims, C., 1980. Comparison of Interwar and Postwar Business Cycles: Monetarism Reconsidered. *The American Economic Review* 70 (2), 250-257.
- [34] Warne, A., 1993. A common trends model: identification, estimation and asymptotics. Institute for International Economic Studies, University of Stockholm, Seminar Paper No. 555.
- [35] Watson, M.W., 1986. Univariate detrending methods with stochastic trends. *Journal of Monetary Economics* 18, 49-75

3.A Apêndice

3.A.1 Proposições e Provas

Proposição 2 *Todo ARIMA pode ser escrito como uma função linear da história de choques.*

Prova. Seja $X_t \sim ARIMA(\cdot, 0, \cdot)$ estacionário. Logo \exists uma representação de Wold, $X_t = C(L)\eta_t = f(H_t)$, onde f é uma função linear e $H_t = (\eta_t \ \eta_{t-1} \ \cdots \ \eta_0)'$ é a história dos choques até o período t . A igualdade $C(L)\eta_t = f(H_t)$ só é válida sob a hipótese de que $\eta_i = 0, \forall i \leq 0$.

Para $X_t \sim ARIMA(\cdot, 1, \cdot)$, \exists uma representação de Wold para a primeira diferença, logo $\Delta X_t = f(H_t)$ e $X_t = X_0 + \sum_{i=0}^{t-1} f(H_{t-i}) = g(H_t)$, onde g é uma função linear.

Para mostrar que vale para todo processo ARIMA basta provar por indução: $ARIMA(\cdot, N, \cdot) = f_N(H_t)$, com f_N função linear $\implies ARIMA(\cdot, N+1, \cdot) = f_{N+1}(H_t)$, com f_{N+1} função linear.

Se $Y_t \sim ARIMA(\cdot, N+1, \cdot)$, então $\Delta Y_t \sim ARIMA(\cdot, N, \cdot) = f_N(H_t)$ e $Y_t = Y_0 + \sum_{i=0}^{t-1} f_N(H_{t-i}) = f_{N+1}(H_t)$.

Se vale para $ARIMA(\cdot, k, \cdot)$ com $k = 0, 1, N$ e $N+1$ então vale para todo k . ■

Proposição 3 *Seja $X_t \sim ARIMA$ qualquer e $\widehat{X}_t = (E_t - E_{t-1})X_t$. Então,*

$$\frac{\partial X_t}{\partial \eta_t} = \frac{\partial \widehat{X}_t}{\partial \eta_t}$$

Prova. $X_t \sim ARIMA$ implica que $X_t = f(H_t)$, com f linear. X_t pode ser escrito como $X_t = h_1(\varepsilon_t) + h_2(H_{t-1})$, com h_1 e h_2 funções linear. Logo, $\frac{\partial X_t}{\partial \eta_t} = \frac{\partial f(H_t)}{\partial \eta_t} = \frac{dh_1(\eta_t)}{d\eta_t}$. Por sua vez, $\widehat{X}_t = h_1(\eta_t)$ e assim $\frac{\partial \widehat{X}_t}{\partial \eta_t} = \frac{dh_1(\eta_t)}{d\eta_t} = \frac{\partial X_t}{\partial \eta_t}$. ■

Proposição 4 *Seja $X_t \sim ARIMA$ qualquer e $Y_t = E_t \sum_{i \geq 0} \beta^i X_{t+i}$, onde $\beta \in (0, 1)$ é um fator de desconto. Então $\frac{\partial Y_t}{\partial \eta_t} = \frac{\partial \widehat{Y}_t}{\partial \eta_t}$.*

Prova. $Y_t = E_t \sum_{i \geq 0} \beta^i f(H_{t+i})$, onde f é função linear. Y_t é o valor esperado de uma função linear de H_∞ . Portanto, todos os termos de choques não contidos no conjunto de informação Ω_t terão $E_t(\cdot) = 0$. Assim, Y_t pode ser escrito como $Y_t = h(H_t)$, onde h é uma função linear. Usando a proposição anterior temos o resultado. ■

Corolário 5 $\frac{\partial U_t}{\partial \eta_t} = \frac{\partial \widehat{U}_t}{\partial \eta_t}$, onde $U_t = E_t \sum_{i \geq 0} \beta^i u(C_{t+i})$

3.A.2 Teste de exogeneidade fraca

Considerando o caso em que o vetor de variáveis X_t seja composto por apenas duas variáveis,^{17 18} podemos decompor a densidade conjunta do vetor $X_t = (X_{1t}, X_{2t})'$ no produto das densidades condicional e marginal, $f(X_{1t}, X_{2t}) = f(X_{1t}|X_{2t}) \cdot f(X_{2t})$, pode-se dizer que X_{2t} é fracamente exógeno para o parâmetro β (parâmetros da densidade condicional de X_{1t}) se $f(X_{2t})$ contém somente parâmetros de estorvo (“*nuisance parameters*”) que são irrelevantes para a inferência de β ¹⁹. Desta maneira, poderíamos estimar β considerando na maximização da verossimilhança apenas a função de densidade condicional, $f(X_{1t}|X_{2t})$. Isto implica uma propriedade de separação, onde o vetor de parâmetros associados à densidade conjunta pode ser decomposto em parâmetros associados às densidades marginais e condicionais. Desta maneira, a construção da função impulso-resposta para choques originados na variável X_2 não requer uma ortogonalização dos choques.

Considere a representação VECM definida na equação (8) para $n = 2$. Decompondo a representação VECM,

$$\begin{aligned}\Delta X_{1t} &= \gamma_1 \alpha' X_{t-1} + \sum_{i=1}^p A_{1i} \Delta X_{t-i} + \eta_{1t} \\ \Delta X_{2t} &= \gamma_2 \alpha' X_{t-1} + \sum_{i=1}^p A_{2i} \Delta X_{t-i} + \eta_{2t}\end{aligned}\quad (20)$$

em que $X_t = \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix}$, $\gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}$, $A_i = \begin{bmatrix} A_{1i} \\ A_{2i} \end{bmatrix}$ e $\eta_t = \begin{bmatrix} \eta_{1t} \\ \eta_{2t} \end{bmatrix} \sim iidN(0, \Sigma)$, com $\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$, positiva definida não diagonal.

Decompondo a função de densidade conjunta no produto das densidades condicional de X_{1t} condicionado em X_{2t} (dado o passado de X), $f(\Delta X_{1t}|\Delta X_{2t}; X_{t-1}, \Delta X_{t-1}, \dots, \Delta X_{t-p})$, e a marginal de X_{2t} (dado todo o passado de X), $f(\Delta X_{2t}|X_{t-1}, \Delta X_{t-1}, \dots, \Delta X_{t-p})$, podemos reescrever o

¹⁷ $X_t = (u(C_t), \Psi_1)'$, onde Ψ_1 é o conjunto de informações do indivíduo, composto por apenas 1 variável.

¹⁸Nesse caso, o consumo e a renda, em logaritmo.

¹⁹Diz-se que X_{2t} é fortemente exógeno se, além da condição para exogeneidade fraca, ainda tem a condição de que X_1 não granger-causa X_2 .

sistema,

$$\begin{aligned}\Delta X_{1t} &= \omega \Delta X_{2t} + \tilde{\gamma}_1 \alpha' X_{t-1} + \sum_{i=1}^p \tilde{A}_{1i} \Delta X_{t-i} + \tilde{\eta}_{1t} \\ \Delta X_{2t} &= \gamma_2 \alpha' X_{t-1} + \sum_{i=1}^p A_{2i} \Delta X_{t-i} + \eta_{2t}\end{aligned}\quad (21)$$

em que $\omega = \sigma_{12}/\sigma_{22}$, $\tilde{\gamma}_1 = \gamma_1 - \omega\gamma_2$, $\tilde{A}_{1i} = A_{1i} - \omega A_{2i}$ e $\tilde{\eta}_{1t} = \eta_{1t} - \omega\eta_{2t}$. O sistema (21) é tal que a variância do modelo condicional é $Var(\tilde{\eta}_{1t}) = \sigma_{11} - \sigma_{12}\sigma_{22}^{-1}\sigma_{21} = \sigma_{11.2}$ e os choques $\tilde{\eta}_{1t}$ e η_{2t} são independentes, o que não é verdade para os choques da forma reduzida descrita no sistema (20). Os parâmetros dos modelos condicional e marginal no sistema (21) são respectivamente $(\omega, \tilde{\gamma}_1, \alpha, \tilde{A}_{1i}, \sigma_{11.2})$ e $(\gamma_2, \alpha, A_{2i}, \sigma_{22})$. Percebe-se que α está presente nas duas equações de (21) e isto pode ser considerado um problema, ou uma perda de informação, se formos analisar a primeira equação - modelo condicional - sem levar em conta a segunda - modelo marginal. Johansen (1995) apresenta o seguinte teorema:

Teorema 6 (Johansen, 1995) *Se $\gamma_2 = 0$, então X_{2t} é fracamente exógeno para os parâmetros (α, γ_1) e o estimador de máxima verossimilhança de α e γ_1 podem ser calculados do modelo condicional.*

3.A.3 Choques Permanentes *versus* Transitórios

Warne (1993) apresenta a forma de se proceder uma decomposição dos choques em parte permanente e transitória que é completamente consistente com o modelo de tendência comuns. Necessariamente, devemos considerar um vetor de variáveis $X_t \sim I(1)$, sendo $\alpha' X_t$ estacionário com $rank(\alpha) = r$, ou seja, temos r vetores de cointegração e $(n - r)$ tendências estocásticas comuns. Decompondo a representação de Wold e integrando temos,

$$\begin{aligned}X_t &= E_n X_t^* = X_0 + E_n C(1) E_n' \xi_t + E_n C^*(L) E_n' \eta_t \\ \xi_t &= \delta + \xi_{t-1} + \eta_t \\ C(1) &= (I - A^*)^{-1}, \quad C_i^* = - \sum_{j=i+1}^{\infty} (A^*)^j, \quad \forall i \geq 0\end{aligned}\quad (22)$$

em que $E_n = \begin{pmatrix} I_n & 0 \end{pmatrix}$ é uma matriz que seleciona as n primeiras linhas de X_t^* . Desta forma decomporomos a série em uma parte tendência (não estacionária) e outra cíclica (estacionária), $X_t = X_t^p + X_t^s$, onde

$X_t^p = E_n(I - A^*)^{-1}E_n' \left[\xi_0 + \delta t + \sum_{j=1}^t \eta_j \right]$ e $X_t^s = X_0 + E_n C^*(L)E_n' \eta_t$, tal que $\alpha' X_t = \alpha' X_t^s$, pois existem r combinações lineares das séries não estacionárias que são estacionárias, i.e., $\alpha' X_t^p = 0$. No entanto, a matriz de variância-covariância entre os choques, não é necessariamente diagonal, $E[\eta_t \eta_t'] = \Sigma$.

Por outro lado, se considerarmos o modelo na forma estrutural, $\Delta X_t = \rho + R(L)v_t$, onde a $R(L)$ é um polinômio de defasagens matricial, $v_t = \begin{bmatrix} \varphi_t & \psi_t \end{bmatrix}'$, φ_t é o choque permanente, ψ_t é o choque transitório e $E[v_t v_t'] = I_n$, podemos então decompor X_t da seguinte forma,

$$\begin{aligned} X_t &= X_0 + \Upsilon \tau_t + \Phi(L)v_t \\ \tau_t &= \mu + \tau_{t-1} + \varphi_t \end{aligned} \quad (23)$$

em que Υ é $(n \times (n - r))$ e deve ser tal que $\alpha' \Upsilon = 0$ ²⁰. Comparando (22) e (23) temos a seguinte relação:

$$\Upsilon \varphi_t = E_n(I - A^*)^{-1}E_n' \eta_t \quad (24)$$

Seja Γ uma matriz $(n \times n)$, não singular tal que $\Gamma \Sigma \Gamma'$ é diagonal. Assim, podemos escrever $R(1) = E_n(I - A^*(1))^{-1}E_n' \Gamma^{-1}$ e $v_t = \Gamma \eta_t$.

Seguindo a definição de Warne (1993) pode-se afirmar:

Definição 7 - Γ identifica o modelo de tendências comuns se,

- i) é unicamente determinado pelos parâmetros do VECM
- ii) a matriz de variância-covariância de $\Gamma \eta_t$ é diagonal com elementos não nulos na diagonal
- iii) $R(1) = \begin{bmatrix} \Upsilon & 0 \end{bmatrix}$ é a matriz de impacto total.

Definição 8 - ν_{it} é um choque permanente se a i -ésima coluna de $R(1)$ é não nula. ν_{it} é um choque transitório se a i -ésima coluna de $R(1)$ é nula.

Logo, para escolher Γ deve-se considerar as seguintes condições: as $(n - r)$ inovações permanentes, φ_t , são independentes das r inovações transitórias, ψ_t ; as r inovações transitórias são mutuamente independentes.

²⁰Warne (1993) considera o seguinte procedimento: $\Upsilon = \Upsilon_0 \pi$ onde Υ_0 deve ser da mesma dimensão de Υ , tal que $\alpha' \Upsilon_0 = 0$ e π é um conjunto de parâmetros livres. Para identificação exata é necessário impor $(n - r)(n - r - 1)/2$ restrições adicionais, onde r é o número de restrições de cointegração. Para o caso em que, $n = 2$ e $r = 1$, temos identificação exata e basta fazer $\Upsilon = \begin{pmatrix} \alpha_1 & 1 \end{pmatrix}'$, considerando que o vetor de cointegração é $\alpha = \begin{pmatrix} 1 & -\alpha_1 \end{pmatrix}'$.

Assim, $v_t = \begin{bmatrix} \varphi_t \\ \psi_t \end{bmatrix} = \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \end{bmatrix} \eta_t$. Considerando a equação (24) temos

$\Gamma_1 = (\Upsilon' \Upsilon)^{-1} \Upsilon' E_n (I - A^*)^{-1} E_n'$ de dimensão $((n - r) \times n)$.

Para satisfazer a condição de independência entre os choques, $E[\varphi_t \psi_t'] = \Gamma_1 \Sigma \Gamma_2' = (\Upsilon' \Upsilon)^{-1} \Upsilon' E_n (I - A^*)^{-1} E_n' \Sigma \Gamma_2' = 0$. Como $C(1)$ é de posto reduzido por causa da restrição de cointegração, basta fazer $\Gamma_2 = H_1 \Sigma^{-1}$, onde H_1 é um vetor $(r \times n)$ linearmente independente das linhas de $C(1)$. Considere $H_1 = Q^{-1} \zeta'$, onde Q é um escalar e $\zeta = \gamma(U\gamma)^{-1}$.²¹

²² Assim, podemos definir,

$$\Gamma = \begin{bmatrix} (\Upsilon' \Upsilon)^{-1} \Upsilon' E_n (I - A^*)^{-1} E_n' \\ Q^{-1} \zeta' \Sigma^{-1} \end{bmatrix} \quad (25)$$

e ainda,

$$\begin{pmatrix} \Delta u(C_t) \\ \Delta \Psi_t \end{pmatrix} = \rho + E_n (I - A^* L)^{-1} E_n' \Gamma^{-1} \begin{pmatrix} \varphi_t \\ \psi_t \end{pmatrix}$$

em que $\rho = E_n \rho^* = E_n (I - A^*)^{-1} E_n' \delta$.

Para calcular a função impulso-resposta não-convencional, equação (12), basta usar Γ para separar os choques permanentes dos transitórios. Assim,

$$\begin{aligned} \frac{\partial U}{\partial \varphi_{jt}} &= \frac{1}{1 - \beta} e_1' (I - \beta A^*)^{-1} E_n' \Gamma^{-1} e_j, & 1 \leq j \leq n - r \\ \frac{\partial U}{\partial \psi_{jt}} &= \frac{1}{1 - \beta} e_1' (I - \beta A^*)^{-1} E_n' \Gamma^{-1} e_j, & r + 1 \leq j \leq n \end{aligned} \quad (26)$$

em que $\frac{\partial U}{\partial \varphi_{jt}}$ e $\frac{\partial U}{\partial \psi_{jt}}$ são respectivamente a resposta sobre a utilidade de um choque permanente e um choque transitório.

3.A.4 Impulso-Resposta Generalizado Não Convencional

Koop *et alli* (1996) e Pesaran e Shin (1998) consideram uma definição de função impulso-resposta que não requer a ortogonalização dos choques,

²¹ U é $(r \times n)$ tal que $U\gamma$ é inversível. Construir H_1 como função de γ é ideal pois sabemos que $C(1)\gamma = 0$.

²² Para o caso em que $n = 2$ e $r = 1$, $U\gamma$ é um escalar e portanto basta escolher U tal que o produto seja não nulo. Como temos apenas um choque transitório, a segunda condição é trivialmente satisfeita e uma escolha conveniente de Q é tal que $E[\psi_t \psi_t'] = \frac{1}{Q^2} \zeta' \Sigma^{-1} \zeta = 1$. Assim, basta fazer $Q = \sqrt{\frac{\gamma' \gamma}{(U\gamma)^2}}$ e escolher U tal que $UA(1)\alpha \neq 0$.

chamada de função impulso-resposta generalizado (GIR). Considere um VAR estacionário para n variáveis, escrito em sua representação de Wold, $X_t = \sum_{i \geq 0} A_i \eta_{t-i}$, onde $E[\eta_t \eta_t'] = \Sigma$. Uma função impulso-resposta generalizado descreve a trajetória de efeitos esperados sobre X_{t+k} advindos de um vetor de choques, $\eta_t = \delta$, que atinge a economia no período t , considerando que não ocorreram outros choques entre t e $t+k$. Ou seja,

$$\begin{aligned} GI_X(k, \eta_t = \delta, \Omega_{t-1}) &= E(X_{t+k} | \eta_t = \delta, \Omega_{t-1}) - E(X_{t+k} | \Omega_{t-1}) \\ &= A_k \delta \end{aligned}$$

em que Ω_{t-1} é o conjunto de informação em $t-1$. Neste caso, $GI_X(k, \eta_t = \delta, \Omega_{t-1})$ independe de Ω_{t-1} ²³, mas depende da composição de choques definido pelo vetor δ . Claramente, a escolha de δ é fundamental para definir as propriedades da função impulso-resposta.

Sims (1980) escolhe δ usando uma decomposição de Cholesky da matriz de covariância de choques, tal que $PP' = \Sigma$ e P é $n \times n$ triangular inferior. Assim, pode-se ortogonalizar os choques escrevendo,

$$X_t = \sum_{i \geq 0} A_i P P^{-1} \eta_{t-i} = \sum_{i \geq 0} A_i P \xi_{t-i}$$

em que $\xi_t = P^{-1} \eta_t$ é tal que $E(\xi_t \xi_t') = I_n$. Assim, o vetor $n \times 1$ da função impulso-resposta ortogonalizado de um choque na j -ésima equação sobre X_{t+k} é dado por,

$$\psi_j^{ortog}(k) = A_k P e_j, \quad k = 0, 1, 2, \dots \quad (27)$$

em que e_j é um vetor $n \times 1$ com j -ésimo elemento unitário e zero para os demais. Neste caso, a função impulso-resposta não convencional ortogonalizada usando a Decomposição de Cholesky pode ser escrita por,

$$[\psi_j^{ortog}]_U = \frac{1}{1-\beta} e_j' [I - \beta A^*]^{-1} E_n' P e_j \quad (28)$$

O problema desta abordagem é que depende da ordem das variáveis apresentadas no VAR. Neste trabalho a ordem do VAR para construção da função impulso-resposta ortogonal é definida de acordo com o resultado do teste de exogeneidade discutido acima.

Koop *et alli* (1996) e Pesaran e Shin (1997) descrevem uma abordagem alternativa para escolher δ , que ao invés de produzir um choque em todos

²³Para sistemas lineares, a função impulso-resposta possui a propriedade de ser invariante com a história, o que não ocorre para modelos não lineares.

os elementos de η_t , produz-se um choque apenas sobre o j -ésimo elemento de η_t e constrói-se δ usando a distribuição de η_t assumida ou observada historicamente.

Asumindo que η_t tem uma distribuição multivariada normal, $\eta_t \sim iidN(0, \Sigma)$, sabemos que $E[\eta_t | \eta_{jt} = \delta_j]$ é linear em δ_j , i.e., $E[\eta_t | \eta_{jt} = \delta_j] = \frac{1}{\sigma_{jj}} \Sigma e_j \delta_j$. Assim a função impulso-resposta generalizada para o efeito de um choque de um desvio-padrão ocorrido na j -ésima equação do VAR para k períodos a frente pode ser escrita como,

$$GIR_X \left(k, \frac{\delta_j}{\sqrt{\sigma_{jj}}}, \Omega_{t-1} \right) = \psi_j^{gen}(k) = \sigma_{jj}^{-\frac{1}{2}} A_k \Sigma e_j, \quad k = 0, 1, 2, \dots \quad (29)$$

Ao contrário da ortogonalização usando decomposição de Cholesky, uma função impulso-resposta generalizada é invariante com a ordem das variável que se apresentam no VAR.

Para o caso em que $X_t = \begin{pmatrix} u(C_t) & \Psi_t \end{pmatrix}'$ e $U = E_t \sum_{i \geq 0} \beta^i u(C_{t+i})$, conforme modelo definido na equação (9), sua Representação de Wold pode ser escrita por,

$$\begin{aligned} \Delta X_t^* &= \rho^* + (I - A^* L)^{-1} \eta_t^* \\ &= \rho^* + \sum_{i \geq 0} (A^*)^i \eta_{t+i}^* \end{aligned} \quad (30)$$

e inovação na utilidade descrita pela equação (5). Podemos escrever a função impulso-resposta generalizada para U para um choque de um desvio-padrão na j -ésima equação como,

$$\begin{aligned} GIR_U \left(\frac{\delta_j}{\sqrt{\sigma_{22}}}, \Omega_{t-1} \right) &= \frac{1}{1 - \beta} \frac{1}{\sqrt{\sigma_{22}}} e_1' \left[\sum_{i \geq 0} \beta^i (A^*)^i \right] \Sigma^* e_j \\ &= \frac{1}{1 - \beta} \frac{1}{\sqrt{\sigma_{22}}} e_1' [I - \beta A^*]^{-1} \Sigma^* e_j \end{aligned} \quad (31)$$

em que $E[\eta_t^* \eta_t^{*'}] = \Sigma^* = \begin{bmatrix} \Sigma_{n \times n} & 0_{n \times (np-1)} \\ 0_{(np-1) \times n} & 0_{(np-1) \times (np-1)} \end{bmatrix}_{(np+1) \times (np+1)}$.²⁴

²⁴Deve ficar claro que só faz sentido tomar impulso na j -ésima variável para $j = 1, 2, \dots, n$.

3.A.5 Estimação dos Desvios-padrões das Funções Impulso-resposta não Convencionais

Pelo Método Delta, pode-se construir os desvios padrões das funções impulso-resposta não convencionais, usando desvios-padrões estimados das estimativas do modelo VEC considerado.

Nas equações (26), (28) e (31) podemos observar que $\widehat{\frac{\partial U}{\partial \varphi_{jt}}} = h_1 \left(\text{vec}(\widehat{A^*}), \text{vech}(\widehat{\Sigma}); \beta \right)$, $\widehat{\frac{\partial U}{\partial \psi_{jt}}} = h_2 \left(\text{vec}(\widehat{A^*}), \text{vech}(\widehat{\Sigma}); \beta \right)$, $\widehat{[\psi_j^{ortog}]_U} = h_3 \left(\text{vec}(\widehat{A^*}), \text{vech}(\widehat{\Sigma}); \beta \right)$ e $\widehat{GIR_U} = h_4 \left(\text{vec}(\widehat{A^*}), \text{vech}(\widehat{\Sigma}); \beta \right)$, em que h_1, h_2, h_3 e h_4 são funções contínuas em $\text{vec}(\widehat{A^*})$ e $\text{vech}(\widehat{\Sigma})$.

Seja $\hat{\theta}_T = \left(\text{vec}(\widehat{A^*})', \text{vech}(\widehat{\Sigma})' \right)'$ e $F = \begin{bmatrix} \text{var}(A^*) & 0 \\ 0 & \text{var}(\text{vech}(\Sigma)) \end{bmatrix}$.²⁵ O

Método Delta define que se $\hat{\theta}_T$ é uma sequência de vetores aleatórios tal que $\sqrt{T} \left[\hat{\theta}_T - \theta_o \right] \xrightarrow{d} N(\mathbf{0}, F)$ e se $g(\hat{\theta})$ é uma função contínua de $\hat{\theta}_T$, então $\sqrt{T} \left[g(\hat{\theta}_T) - g(\theta_o) \right] \xrightarrow{d} N(\mathbf{0}, c(\theta) F c'(\theta))$, em que $c(\theta)$ é o vetor linha de derivadas $\partial g(\theta_o) / \partial \theta_o'$.

²⁵ Ver Hamilton (1994) pp. 300-302.

Capítulo 4

Are Business Cycles All Alike in Europe?*

Márcio Antônio Salvato[†]

Abstract

We investigate whether business cycles are all alike computing the welfare costs of business cycles for European-Union (EU) as the solution of the problem proposed by Lucas (1987). Because these countries have a long tradition of integration and trade, it is a “natural experiment” to investigate how similar their welfare costs of business cycles are. Using standard assumptions on preferences and a reasonable reduced form for consumption, we computed welfare costs using three alternative trend-cycle decomposition methods, but focusing on the multivariate Beveridge-Nelson decomposition. Our results show that welfare costs are very different across EU countries and between US and EU countries, and thus it is a strong evidence that business cycles are not alike in Europe.

JEL Codes: E32; C32; C53.

Keywords: Business cycles, welfare costs, cointegration, Beveridge-Nelson decomposition.

*João Victor Issler (EPGE/FGV) and Angelo Mont'alverne Duarte (EPGE/FGV) are co-authors in this article. We gratefully acknowledge the suggestions of Osmani T. Guillén and Afonso Arinos de Mello Franco. All errors are ours. We thank CNPq-Brazil, PUC.Minas and PRONEX for financial support.

[†]EPGE/FGV, PUC.Minas and CEPE, salvato@fgvmail.br.

4.1 Introduction

A discussion with a long tradition in macroeconomics is on what generates business cycles. According to one view, which we label the *institutional* view, business cycles are generated by large and infrequent shocks that hit macroeconomic variables, leading them to fluctuate about their trend. Because institutional settings vary from country to country, these shocks are different across countries and business cycles are not all alike. According to a different view, which we label the *dynamic-stochastic-general-equilibrium* – *DSGE* – view, business cycles are generated by small and frequent white-noise shocks that hit macroeconomic variables, which have a dynamic path qualitatively well approximated by a dynamic stochastic general equilibrium model. Because these shocks are small (low variance), and there is no reason to believe that propagation mechanisms vary from country to country, business cycles are all alike. Indeed, Lucas (1977), in his opening statement of this classic paper, asks: “Why is it that, in capitalist economies, aggregate variables undergo repeated fluctuations about trend, all of essentially the same character?”

Of course, it is not trivial to investigate if business cycles are all alike. First, one has to define in which sense they should be alike and different ways to measure similarities. A first approach, followed by Blanchard and Watson (1986), is to look directly into shocks themselves, investigating whether they are small or large, as well as their nature. Usually this is done using a structural econometric model. Since there is no consensus on how shock identification should be performed, and several shock-identification techniques have been criticized on different grounds, it is hard to come out with a satisfactory answer once this direct approach is followed.

The shortcomings of the direct approach can be overcome if instead of focusing directly on shocks, one uses an indirect approach, focusing on a fundamental difference in the nature of business cycles entailed by these two types of shocks. A concept that has received some attention recently, and that can be used to investigate whether business cycles are alike is the welfare cost of business cycles. The idea is straightforward: Lucas (1987) calculates the proportion of extra consumption, in all dates and states of nature, a rational consumer would require in order to be indifferent between an infinite sequence of consumption under uncertainty and a certain sequence which is cycle free. This proportion is labelled the *welfare cost of business cycles*, and can be directly computed using consumption data and a parametric version of the utility function; see the variants in Imrohoroglu (1989), Obstfeld

(1994), Van Wincoop (1994), Atkeson and Phelan (1995), Pemberton (1996), Dolmas (1998), Tallarini (2000), Otrok (2001), and Franco, Guillen and Issler (2003).

If shocks are frequent and similar across countries, in which they have a low variance, and if the propagation mechanism is similar in nature to that in dynamic stochastic general equilibrium models, one should find that the welfare costs of business cycles across economies are all similar. However, if institutional factors are important, shocks will be different in nature and the welfare costs of business cycles will be different across economies. Of course, one can always find a set of countries that have similar institutional settings. For them, finding similar welfare costs of business cycles may just be a consequence of similar institutions. However, if the opposite is true for this set of countries, then it is hard to argue for the DSGE view.

In this paper, we investigate whether business cycles are all alike computing the welfare costs of business cycles for an important subset of European countries – European-Union (EU) countries: Austria, Belgium, Denmark, Finland, France, Germany, Great Britain, Greece, Ireland, Italy, Luxembourg, Portugal, Spain, Sweden and The Netherlands. As is well known, these countries have a long tradition of integration and trade dating well before the common-currency Euro area was implemented. Because of this feature, it is a “natural experiment” to investigate how similar their welfare costs of business cycles are, in the sense that we will be surprised to find major differences between them.

In computing the welfare costs of business cycles for EU countries we use the techniques in Beveridge and Nelson (1981) to decompose (the log of) consumption in a trend and a cyclical component. In this case, the trend will be stochastic and modeled as a random walk¹. This choice relies on a sizable amount of econometric evidence available on consumption, or, alternatively, on the amount of authors that have used the unit-root specification, e.g., Hall (1978), Nelson and Plosser (1982), Campbell (1987), King et al. (1991), Cochrane (1994), *inter alia*. Moreover, to make our results comparable to previous work, we also modelled the trend as either a deterministic linear process (with and without a break) or following a slowly evolving secular process captured by the Hodrick and Prescott (1997) filter.

Our results show that the welfare costs of business cycles are very different across EU countries. Using the Beveridge and Nelson decomposition, and plausible values for the risk aversion coefficient and the discount rate of future utility, we find that the welfare cost of

¹Lucas (1987, pp. 22-23, footnote 1) explicitly considers the possibility that the trend in consumption is stochastic as in Nelson and Plosser (1982).

Spain (4.1% of consumption) is almost ten times that of the UK (0.45% of consumption) – median of 2.85%. Major differences in welfare costs are also found when alternative trend-cycle decomposition methods are employed, although they are not as pronounced as the ones obtained using the Beveridge and Nelson decomposition.

The paper is divided as follows. Section 2 provides a theoretical and statistical framework to evaluate the welfare costs of business cycles. Section 3 provides the estimates that are used in calculating them. Section 4 provides the calculations results, and Section 5 concludes. There is also an Appendix providing the econometric background necessary to implement the calculations carried out in the paper.

4.2 The Problem

Lucas (1987) proposed a way to evaluate the welfare gains of cycle smoothing. Suppose an agent that choose a consumption sequence $\{c_t\}_{t=0}^{\infty}$ that maximizes intertemporal utility, U , subject to a budget constraint:

$$U = E_0 \sum_{t=0}^{\infty} \beta^t u(c_t), \quad (1)$$

where $E_t(\cdot) = E(\cdot | \Omega_t)$ is the conditional expectation operator of a random variable, using Ω_t as the information set, and $\beta \in (0, 1)$ is a constant discount factor. He worked with a class of consumption streams with trend and cycle components such as:

$$c_t = \alpha_0 (1 + \alpha_1)^t \exp\left(-\frac{1}{2}\sigma_z^2\right) z_t, \quad t = 0, 1, \dots, \quad (2)$$

where $\{z_t\}$ is a stationary stochastic process with a stationary distribution given by $\ln(z_t) \sim N(0, \sigma_z^2)$. Cycle-free consumption will be the sequence $\{c_t^*\}_{t=0}^{\infty}$, where $c_t^* = E(c_t) = \alpha_0 (1 + \alpha_1)^t$ since $E\left[\exp\left(-\frac{1}{2}\sigma_z^2\right) z_t\right] = 1$. Notice that $\{c_t^*\}_{t=0}^{\infty}$ is the resulting sequence when we replace the random variable c_t with its unconditional mean. Hence, for any particular time period, c_t represents a mean-preserving spread of c_t^* .

Risk averse consumers prefer $\{c_t^*\}_{t=0}^{\infty}$ to $\{c_t\}_{t=0}^{\infty}$, so the costs of the economic instability can be measured by calculating λ which solves the following equation²:

²Notice that Lucas (1987) uses the unconditional mean operator instead of the conditional mean operator in (3). The same problem can be proposed using the conditional expectation instead. This is exactly how

$$E \left(E_0 \sum_{t=0}^{\infty} \beta^t u((1+\lambda)c_t) \right) = \sum_{t=0}^{\infty} \beta^t u(c_t^*). \quad (3)$$

Then λ is the compensation required by consumers that makes them indifferent between the uncertain stream $\{c_t\}_{t=0}^{\infty}$ and the stream $\{c_t^*\}_{t=0}^{\infty}$. Notice that uncertainty here comes in the form of stochastic business cycles, since the trend in consumption is deterministic.

Lucas (1987) assumed that the utility function is in CES class:

$$u(c_t) = \frac{c_t^{1-\phi} - 1}{1-\phi}, \quad (4)$$

where $\phi > 0$ is the constant coefficient of relative risk-aversion and $u(c_t)$ converges to $\ln(c_t)$ as $\phi \rightarrow 1$. It calculated λ that satisfies (3) for some values of β and ϕ using US data for post-war period.

Obviously there are others forms of c_t besides (2). If we suppose c_t is difference stationary then it can be decomposed as the sum of a deterministic trend, a random walk trend and a stationary cycle (ARMA process), as shown in Beveridge and Nelson (1981),

$$\begin{aligned} \ln(c_t) &= \ln(\alpha_0) + \ln(1 + \alpha_1) \cdot t - \frac{\omega_t^2}{2} + \sum_{i=1}^t \epsilon_i + \sum_{j=0}^{t-1} \psi_j \mu_{t-j} \\ &= \ln(\alpha_0 (1 + \alpha_1)^t) - \frac{\omega_t^2}{2} + \ln(X_t) + \ln(Y_t), \end{aligned} \quad (5)$$

where $\ln[\alpha_0 (1 + \alpha_1)^t \exp(-\omega_t^2/2)]$ is the deterministic term, $\ln(X_t) = \sum_{i=1}^t \epsilon_i$ is the random walk component, $\ln(Y_t) = \sum_{j=0}^{t-1} \psi_j \mu_{t-j}$ is the $MA(\infty)$ representation of the stationary part (cycle), and $\omega_t^2 = \sigma_{11}t + 2\sigma_{12} \sum_{j=0}^{t-1} \psi_j + \sigma_{22} \sum_{j=0}^{t-1} \psi_j^2$ is the conditional variance of $\ln(c_t)$. The permanent shock ϵ_t and the transitory shock μ_t are assumed to have a bi-variate normal distribution as follows,

$$\begin{pmatrix} \epsilon_t \\ \mu_t \end{pmatrix} \sim IIDN \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \right), \quad (6)$$

i.e., shocks are independent, thus serially uncorrelated, but contemporaneously correlated if $\sigma_{12} \neq 0$.

we proceed in this paper.

Calculating the welfare cost of business cycles for the difference-stationary case requires first a discussion on how to deal with the fact that now uncertainty comes both in the trend and the cyclical component of $\ln(c_t)$. Moreover, since the trend component has a unit root, its unconditional mean and variance are not defined. Notice that, in the exercise proposed by Lucas, all the cyclical variation in $\ln(c_t)$ is eliminated, which is equivalent to eliminating *all* its variability, since the trend is deterministic. Here, this equivalence is lost, because the trend is stochastic as well.

To deal with this issue, we follow Obstfeld (1994) in considering the conditional expectation operator $E_0(\cdot)$ in (3), in spite of the unconditional expectation operator $E(\cdot)$. In this case, c_t^* is now redefined as $c_t^* = E_0(c_t)$. Therefore, we are assuming that it is possible to offer the consumer an certain consumption stream c_t^* (with no trend and cyclical variation) based on information available at the outset of the problem. Of course, the alternative for the consumer is to face c_t , which has a conditional variance that depends on ω_t^2 . Consumption has now a unit root and so $\omega_t^2 \rightarrow \infty$, as $t \rightarrow \infty$ (although $\omega_t^2 < \infty$ for all t finite). Hence, uncertainty can get relatively large as the horizon increases, which may be balanced by the fact that there is discounting in the welfare function.

As in Obstfeld (1994), the problem we propose solving here is

$$E_0 \sum_{t=0}^{\infty} \beta^t u((1+\lambda)c_t) = \sum_{t=0}^{\infty} \beta^t u(E_0(c_t)). \quad (7)$$

Under (4), (5) and (6), and using the properties of the moments of log-normal distributions, we can calculate (7). Apart from an irrelevant constant term, its left-hand side is given by

$$E_0 \sum_{t=0}^{\infty} \beta^t u((1+\lambda)c_t) = \frac{[\alpha_0(1+\lambda)]^{1-\phi}}{1-\phi} \sum_{t=0}^{\infty} \left[\beta(1+\alpha_1)^{1-\phi} \right]^t \exp \left[-\frac{(1-\phi)\phi\omega_t^2}{2} \right]. \quad (8)$$

Notice that, (8) converges if $\beta(1+\alpha_1)^{1-\phi} \exp \left[-\frac{(1-\phi)\phi\sigma_{11}}{2} \right] < 1$.

Calculating the conditional mean of c_t yields $c_t^* = E_0(c_t) = \alpha_0(1+\alpha_1)^t \exp \left(-\frac{\omega_t^2}{2} \right) E_0(X_t Y_t) = \alpha_0(1+\alpha_1)^t$. Hence, apart from an irrelevant constant term, the right-hand side of (7) is

$$\sum_{i=0}^{\infty} \beta^t u(c_t^*) = \frac{\alpha_0^{1-\phi}}{1-\phi} \sum_{t=0}^{\infty} \left[\beta(1+\alpha_1)^{1-\phi} \right]^t, \quad (9)$$

which converges if $\beta(1+\alpha_1)^{1-\phi} < 1$.

Given the parameters defining the processes $\{c_t^*\}_{t=0}^{\infty}$ and $\{c_t\}_{t=0}^{\infty}$, $\lambda(\phi, \beta)$ is

$$\lambda(\phi, \beta) = \left\{ \frac{\sum_{t=0}^{\infty} [\beta(1 + \alpha_1)^{1-\phi}]^t}{\sum_{t=0}^{\infty} [\beta(1 + \alpha_1)^{1-\phi}]^t \exp\left[-\frac{(1-\phi)\phi\omega_t^2}{2}\right]} \right\}^{1/(1-\phi)} - 1. \quad (10)$$

In the definition of ω_t^2 in (10), we replace $\sigma_{12} \sum_{j=0}^{t-1} \psi_j$ and $\sigma_{22} \sum_{j=0}^{t-1} \psi_j^2$ by their respective unconditional counterparts, $\tilde{\sigma}_{12} = \sigma_{12} \sum_{j=0}^{\infty} \psi_j$ and $\tilde{\sigma}_{22} = \sigma_{22} \sum_{j=0}^{\infty} \psi_j^2$ (which may be a reasonable approximation even for relatively small t , and a very good approximation for intermediate and large t), making $\omega_t^2 = \sigma_{11} \cdot t + 2\tilde{\sigma}_{12} + \tilde{\sigma}_{22}$. Assuming that the conditions for (8) and (9) holds, (10) converges to³

$$\lambda(\phi, \beta) = \begin{cases} \exp\left[\frac{\phi(2\tilde{\sigma}_{12} + \tilde{\sigma}_{22})}{2}\right] \left\{ \frac{(1-\beta(1+\alpha_1)^{1-\phi} \exp[-\frac{(1-\phi)\phi\sigma_{11}}{2}])}{(1-\beta(1+\alpha_1)^{1-\phi})} \right\}^{1/(1-\phi)} - 1, & \text{if } \phi \neq 1 \\ \exp\left[\frac{1}{2} \left(\frac{\beta\sigma_{11}}{(1-\beta)} + 2\tilde{\sigma}_{12} + \tilde{\sigma}_{22} \right)\right] - 1, & \text{if } \phi = 1 \end{cases} \quad (11)$$

which shows the way we chose to estimate $\lambda(\phi, \beta)$ in this paper⁴. In subsection 2.1 we discuss a methodology for calculating $\lambda(\phi, \beta)$ estimates standard errors. It's straightforward to see that $\lambda(\phi, \beta)$ is increasing in β ⁵, thus welfare cost of fluctuations is as large as agents are patient.

We now turn to other possible ways of modelling the trend component. If the trend is modelled as a deterministic function of time, as in (2), then the analysis is done as originally proposed by Lucas (1987). In spite of the fact that Lucas has proposed the analysis as in (3) above, he actually implemented it in a different way (see Lucas, 1987, footnote 2, p. 23), removing the trend in consumption using the filtering procedure proposed in Hodrick and Prescott (1997). The filter is two sided, i.e., uses past and future consumption values to get the slowly-moving trend. In principle, the trend removed using such a procedure should be treated as a random variable. However, for simplicity, Lucas treated the trend as deterministic, which we also do here. Hence, when using the Hodrick and Prescott trend,

³Equation (11) for $\phi = 1$ is derived on appendix A.

⁴In our results we have observed that, for all values of (ϕ, β) we considered here, $\beta(1 + \alpha_1)^{1-\phi} < 1$. However, it was not always the case that $\beta(1 + \alpha_1)^{1-\phi} \exp\left[-\frac{(1-\phi)\phi\sigma_{11}}{2}\right] < 1$, since the term $\exp\left[-\frac{(1-\phi)\phi\sigma_{11}}{2}\right]$ was always greater than unity, and sometimes large enough as to prevent the convergence condition to hold.

⁵The term $\exp\left[-\frac{(1-\phi)\phi\sigma_{11}}{2}\right]$ is always greater than unity.

our results should be viewed as a lower-bound for the welfare cost of business cycles. To implement the calculations in this case, we computed the deterministic growth rate present in the Hodrick and Prescott trend, treating the cyclical component as in (7) above. Hence, $c_t = \alpha'_0 (1 + \alpha'_1)^t \exp(-\sigma_{z'}^2/2) z'_t$, $\ln(z'_t) \sim N(0, \sigma_{z'}^2)$, and $c_t^* = \alpha'_0 (1 + \alpha'_1)^t$, where α'_0 and α'_1 are now the deterministic components associated with the Hodrick-Prescott trend, and z'_t is the residual cyclical component associated with it. We may observe that for linear and Hodrick-Prescott trend, $\sigma_{11} = \tilde{\sigma}_{12} = 0$, and so λ in equation (11) does not depend of β and α_1 and is monotonically increasing in ϕ .

4.2.1 Standard Errors of $\lambda(\phi, \beta)$ Estimates

Let Ω the variance-covariance matrix of the permanent shock ϵ_t and the transitory shock μ_t of the log of consumption, as presented in equation (6), and $\hat{\Omega}$ the maximum likelihood estimator of Ω ,⁶ thus,

$$\sqrt{T} \begin{bmatrix} \hat{\sigma}_{11,T} - \sigma_{11} \\ \hat{\sigma}_{21,T} - \sigma_{21} \\ \hat{\sigma}_{22,T} - \sigma_{22} \end{bmatrix} \xrightarrow{d} N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2\sigma_{11}^2 & 2\sigma_{11}\sigma_{21} & 2\sigma_{21}^2 \\ 2\sigma_{11}\sigma_{21} & \sigma_{11}\sigma_{22} + \sigma_{21}^2 & 2\sigma_{21}\sigma_{22} \\ 2\sigma_{21}^2 & 2\sigma_{21}\sigma_{22} & 2\sigma_{22}^2 \end{bmatrix} \right). \quad (12)$$

Let $\hat{\alpha}_1$ a consistent and asymptotically normally-distributed estimator of α_1 , i.e.,

$$\sqrt{T}(\hat{\alpha}_1 - \alpha_1) \xrightarrow{d} N(0, \sigma_\alpha^2). \quad (13)$$

Let $\theta_0 = (\alpha_1, \sigma_{11}, \sigma_{21}, \sigma_{22})'$ and $\hat{\theta}_T = (\hat{\alpha}_{1,T}, \hat{\sigma}_{11,T}, \hat{\sigma}_{21,T}, \hat{\sigma}_{22,T})'$. From (12) and (13) and applying Delta Method, we have,

$$\sqrt{T} [\lambda(\hat{\theta}_T) - \lambda(\theta_0)] \xrightarrow{d} N \left[\mathbf{0}, \begin{bmatrix} \mathbf{C}(\theta_0) & \sigma_\alpha^2 & \mathbf{0} \\ \mathbf{0} & \Omega & \mathbf{C}(\theta_0)' \end{bmatrix} \right],$$

where $\mathbf{C}(\theta_0)$ is the vector of partial derivatives of λ with respect to θ'_0 .

⁶See Hamilton (1994), pages 300-301.

4.3 Reduced Form and Long-Run Constraints

Let $y_t = (\ln(c_t), \ln(I_t))'$ is a 2×1 vector containing the logarithms of consumption and disposable income.⁷ Assume that both series individually contain a unit-root, and are generated by a p -th order vector autoregression (VAR),

$$y_t = \pi_1 y_{t-1} + \pi_2 y_{t-2} + \cdots + \pi_p y_{t-p} + \varepsilon_t, \text{ or,} \\ \Pi(L) y_t = \varepsilon_t,$$

where $\Pi(L) = I_n - \pi_1 L - \pi_2 L^2 - \cdots - \pi_p L^p$. Decomposing $\Pi(L)$ as

$$\Pi(L) = -\Pi(1) L^p + (1 - L) \Gamma(L),$$

leading to the vector error-correction model (VECM)

$$\Delta y_t = \Gamma_1 \Delta y_{t-1} + \Gamma_2 \Delta y_{t-2} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} - \Pi(1) y_{t-p} + \varepsilon_t, \quad (14)$$

where $\Pi(1) = \gamma \alpha'$, $\Gamma_j = -I_n + \sum_{i=1}^j \pi_i$, $j = 1, 2, \dots, p-1$, α is the cointegration vector and γ is a 2×1 constant vector.

Cointegration between the logarithms of consumption and income may be explained using the theory of permanent-income. In this theory, consumption can be viewed as proportional to the expected present discounted value of all income stream. Hence, the expected present value of consumption and income are equal, and both series are proportional in the long run,⁸ moreover, the cointegrating vector will be $\alpha = (-1, 1)'$.

We turn now to the discussion of how to extract trends and cycles from (14). First, put the system (14) in state-space form, as discussed in Proietti (1997),

$$\Delta y_t = Z f_t \\ f_t = T f_{t-1} + Z' \varepsilon_t, \quad (15)$$

⁷A full discussion of the econometric models employed here can be found in Beveridge and Nelson (1981), Stock and Watson (1988), Engle and Granger (1987), Campbell (1987), Campbell and Deaton (1989), and Proietti (1997).

⁸See Campbell (1987) and Campbell and Deaton (1989).

where,

$$f_t = \begin{bmatrix} \Delta y_t \\ \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-p+1} \\ \alpha' y_{t-p} \end{bmatrix}, \quad T = \begin{bmatrix} \Gamma_1 & \cdots & \Gamma_{p-1} & -\gamma\alpha' & -\gamma \\ & I_{2(p-1)} & \mathbf{0} & \mathbf{0} & \\ & \mathbf{0}_{1 \times 2(p-1)} & \alpha' & 1 & \end{bmatrix}, \quad Z = [I_2 \quad \mathbf{0}_{2 \times (2p-1)}]$$

From the work of Beveridge and Nelson (1981), and Stock and Watson (1988), ignoring initial conditions and deterministic components, the series in y_t can be decomposed into a trend (τ_t) and a cyclical component (ψ_t), as follows,

$$y_t = \tau_t + \psi_t,$$

where,

$$\tau_t = y_t + \lim_{k \rightarrow \infty} \sum_{i=1}^k E_t [\Delta y_{t+i}] \quad \text{and} \quad \psi_t = - \lim_{k \rightarrow \infty} \sum_{i=1}^k E_t [\Delta y_{t+i}]. \quad (16)$$

It is straightforward to show that τ_t is a multivariate random-walk. Using the state-space representation (15), we can compute the limits above. The cyclical and trend components will be, respectively,⁹

$$\begin{aligned} \psi_t &= -Z [I_m - T]^{-1} T f_t, \\ \tau_t &= y_t - \psi_t. \end{aligned} \quad (17)$$

where $m = 2p + 1$, or, using formulas (6) and (7) in Proietti (1997),

$$\psi_t = -K\Gamma^*(L) \Delta y_t + P y_t, \quad \text{and} \quad (18)$$

$$\tau_t = K \sum_{i=1}^t \varepsilon_i, \quad (19)$$

where $K = (I_N - P)(\Gamma(1) + \gamma\alpha')^{-1}$ and $P = (\Gamma(1) + \gamma\alpha')^{-1} \gamma [\alpha'(\Gamma(1) + \gamma\alpha')^{-1} \gamma] \alpha'$ are projection matrices.

We can also use (15) to forecast trend and cyclical components at any horizon into the future. The forecast of ψ_{t+s} , given information up to t , is:

$$\widehat{\psi}_{t+st} = E_t [\psi_{t+s}] = -K\Gamma^*(L) Z T f_{t+s-1} + P y_t + P Z \left(\sum_{i=1}^s T^i \right) f_t,$$

⁹See appendix B for cycle and trend equations derivation.

and the forecast of τ_{t+s} , given information up to t , is:

$$\widehat{\tau}_{t+s} = \tau_t,$$

since the best forecast of a random walk $t + s$ periods ahead is simply its value today.

To fully characterize the elements in (11), we need to compute the variance and the covariance of forecasts of trend and cyclical components. Recall that the conditional expectation of a log-normal random variable is just a function of the mean and variance of the normal distribution associated with it. Hence, to compute the variances of these forecasts, we have just to apply standard results of state-space representations. It is straightforward to show that:

$$E_t \left[(\tau_{t+s} - \widehat{\tau}_{t+s|t}) (\tau_{t+s} - \widehat{\tau}_{t+s|t})' \right] = s \cdot K Q K',$$

where $E_t [\varepsilon_{t+i} \varepsilon_{t+i}'] = Q$, and that,

$$E_t \left[(\psi_{t+s} - \widehat{\psi}_{t+s|t}) (\psi_{t+s} - \widehat{\psi}_{t+s|t})' \right] = V Q V' + P \left(\sum_{i=1}^{s-1} W(i) Q W(i)' \right) P'$$

and

$$E_t \left[(\tau_{t+s} - \widehat{\tau}_{t+s|t}) (\psi_{t+s} - \widehat{\psi}_{t+s|t})' \right] = K Q V' + K \left[\sum_{i=1}^{s-1} Q W(i)' \right] P',$$

where $V = [P - K \Gamma^* (1)]$, as computed in the appendix C.

Based on these last three covariance matrices, the correlations between trend and cyclical components of the data can be fully characterized. Hence, to get the corresponding element of means, variances, and covariances associated with $\ln(c_t)$, one has simply to choose the appropriate elements of these vectors and matrices.

4.4 Data

European Union (EU-15) countries¹⁰ annual data for real income and population were obtained from Penn World Table (Summers & Heston) from 1950 to 2000. Annual data for

¹⁰ At presente European Union is composed by 15 countries: Austria, Belgium, Denmark, Finland, France, Germany, Great Britain, Greece, Ireland, Italy, Luxembourg, Portugal, Spain, Sweden and The Netherlands.

household consumption were extract from EUROSTAT, Statistics Sweden and Penn World Table from 1950 to 2000.¹¹

We tested cointegration between series of logs of per capita consumption and income of each country and EU-15. Table 1 presents results of the Johansen (1988, 1991) cointegration test.

Table 1 - Johansen Cointegration Test

Country	VAR Lag	Ho: n° of Coint Eq = None		Ho: n° of Coint Eq = At most 1	
		Trace Statistic	Max-Eigen Statistic	Trace Statistic	Max-Eigen Statistic
EU-15	6	49.71 **	44.01 **	5.70	5.70
Austria	2	27.98 **	23.47 **	4.51	4.51
Belgium	2	35.21 **	26.34 **	8.86	8.86
Denmark	1	4.78	4.21	0.57	0.57
Finland	3	29.61 **	23.31 **	6.29	6.29
France	3	13.10	12.00	1.10	1.10
Germany	1	6.46	5.20	1.26	1.26
Greece	1	9.81	8.12	1.69	1.69
Ireland	2	32.30 **	27.27 **	5.04	5.04
Italy	3	11.43	11.18	0.25	0.25
Luxembourg	1	29.03 **	23.87 **	5.17	5.17
Portugal	6	28.40 **	20.05 *	8.35	8.35
Spain	2	32.14 **	24.83 **	7.31	7.31
Sweden	2	22.03 *	16.37 *	5.66	5.66
The Netherlands	1	14.77	11.99	2.78	2.78
United Kingdom	3	27.35 **	23.02 **	4.32	4.32
Critical Values:	5%	15.41	14.07	9.24	9.24
	1%	24.60	20.20	12.97	12.97

*(**) indicates null hypothesis was rejected at 5%(1%) of significance.

VAR Lag indicates VAR order used in tests

The hypothesis of no cointegration equation was reject and the hypothesis of at most one cointegration equation was not rejected at 5% significance, except to Denmark, France, Germany, Greece, Italy and The Netherlands. Conditioning on the existence of one cointegrating vector, we tested the restriction that it was equal to $(-1, 1)'$ using the likelihood-ratio test in Johansen (1991). This hypothesis was not rejected for Austria, Ireland, Luxembourg, Sweden and United Kingdom (UK). Results are reported on table 2.

¹¹Data for Greece was available from 1951 to 2000 and for Germany from 1970 to 2000.

Table 2 - Testing $H_0 : \alpha' = (1, -1)$

Country	LR test for binding restrictions (rank = 1)		
	Cointegration Vector	Statistic Test $[\chi(1)]$	Probability
EU-15	(1 -1.3036)	22.992	0.0000
Austria	(1 -1.1301)	0.011	0.9179
Belgium	(1 -0.9169)	9.034	0.0027
Finland	(1 -0.9216)	7.097	0.0077
Ireland	(1 -0.566)	2.832	0.0924
Luxembourg	(1 -0.9523)	0.307	0.5794
Portugal	(1 -0.8471)	11.699	0.0006
Spain	(1 -0.9142)	9.461	0.0021
Sweden	(1 -0.605)	0.551	0.4578
United Kingdom	(1 -1.0709)	1.109	0.2923

The presence of unit root was investigated in consumption and income series for those countries which series do not cointegrate. At 5%, the unit root hypothesis was not rejected in all cases using the ADF test; see the same results obtained using the KPSS test.

Table 3 - Unit Root Tests

Country	Level / 1 st difference	Log per capita Consumption		Log per capita GDP	
		KPSS Statistic	ADF Statistic	KPSS Statistic	ADF Statistic
Denmark	level	0.09667	-0.11656	0.19865 **	-1.20567
	1 st difference	0.07542	-6.24076 ***	0.20865	-3.78836 ***
France	level	0.24246 ***	-0.33044	0.23844 ***	-0.46426
	1 st difference	0.06699	-4.53792 ***	0.07984	-4.21962 ***
Germany	level	0.70563 **	-0.52329	0.72877 **	-0.98216
	1 st difference	0.07968	-3.28340 ***	0.11754	-2.54650 **
Greece	level	0.21436 ***	-1.85539	0.86037 ***	-2.86528 *
	1 st difference	0.33663	-1.88616 *	0.11701	-1.34423 *
Italy	level	0.93510 ***	-2.30782	0.23836 ***	-0.48263
	1 st difference	0.07038	-2.03137 **	0.04587	-3.89228 **
The Netherlands	level	0.15361 **	-0.82406	0.20894 **	-1.99065
	1 st difference	0.09047	-2.85621 ***	0.30862	-2.12404 **

*(**)(***) indicates null hypothesis was rejected at 10%(5%)(1%) of significance.

Ho ADF test: serie has unit root; Ho KPSS test: serie is stationary

4.5 Empirical Results

A p th-order vector error-correction model (VECM) with an unrestricted constant term for the logs of consumption and income was fitted using data for each country where we found cointegration. Otherwise, a vector autoregression model for the first differences of those series was estimated. We selected lag length by the use of information criteria, coupled with diagnostic test results. Based on VECM estimates we implemented the multivariate Beveridge and Nelson decomposition as suggested in Proietti (1997). We compute trend and cycle components of consumption using either equations (18) and (19) or equation (16).

Welfare costs of business cycles (λ) for EU-15 and EU countries was computed using equation (11) considering Beveridge-Nelson decomposition, linear time trend and Hodrick-Prescott trend. As a benchmark, we also computed the welfare cost of business cycles for the USA using aggregated consumption data from 1950 to 2000.

Results for reasonable preference parameter and discount values ($\beta = 0.971, \phi = 2$) are presentend in Table 4. Standard errors were calculated using Delta Method as discussed above and, as we may observe, they are negligible if compared to λ . Thus, walfare cost estimates are statistically different from zero at 1% of significance. Results for $\beta = \{0.950; 0.971; 0.985\}$ and $\phi = \{1; 5; 10; 20\}$ are presented in Appendix D.

Table 4 - Welfare Cost of Business Cycles

Country	$\lambda(\%)$ for $\beta = 0.971$ and $\phi=2$		
	Beveridge-Nelson Decomposition	Hodrick-Prescott for Trend	Linear Time Trend
USA [Franco, Guillen and Issler (2003)]	0.25	0.04	0.40
USA	0.75 (0.0227)	0.04 (0.0011)	0.10 (0.0029)
European Union Countries	EU-15	0.18 (0.0041)	0.21 (0.0059)
	The Netherlands	2.33 (0.0720)	0.43 (0.0121)
	Italy	2.85 (0.0876)	0.80 (0.0228)
	United Kingdom	0.45 (0.0134)	0.09 (0.0024)
	France	2.57 (0.0787)	0.89 (0.0254)
	Austria	1.33 (0.0375)	0.23 (0.0066)
	Spain	4.10 (0.1171)	1.49 (0.0426)
	Portugal	2.82 (0.0631)	1.28 (0.0365)
	Belgium	2.91 (0.0698)	0.35 (0.0099)
	Finland	3.72 (0.1051)	0.67 (0.0190)
	Denmark	2.96 (0.0919)	0.57 (0.0162)
	Ireland	2.42 (0.0737)	0.57 (0.0161)
	Greece	3.26 (0.1005)	0.92 (0.0262)
	Sweden	0.80 (0.0243)	0.43 (0.0122)
	Luxembourg	1.23 (0.0297)	0.62 (0.0175)
	Germany	3.91 (0.1233)	0.36 (0.0102)

Note: Standard error in parenthesis

On the one hand, for Beveridge-Nelson decomposition welfare costs for most EU countries is much greater than that for EU-15 as a whole and for the USA. Numbers for UK (0.45%) and Sweden (0.80%) are of the same order magnitude as for USA (0.75%). However, the result for the EU-15 as a whole is even smaller. On the other hand, there is a group of countries whose welfare costs are more than 2.5%: Spain (4.10%), Finland (3.72%), Germany (3.91%), Greece (3.26%), Belgium (2.91%), Italy (2.85%) and Portugal (2.82%). Comparing with Franco, Guillen and Issler (2003) results for USA for post-WWII period¹², our result is

¹²They use non-durables and services annually data from 1947-2000.

three times greater.

Using Hodrick-Prescott Filtering we were able to reproduce Lucas (1987) and Franco, Guillen and Issler (2003) results for USA, i.e. $\lambda_{USA} = 0.04\%$. Welfare cost for EU-15 as a whole (0.02%) is lower than that for USA. Results for France (0.03%), UK (0.04%), Belgium (0.05%) and Italy (0.06%) are similar to that of the USA. For the remaining EU countries, particularly Portugal (0.31%) and Luxembourg (0.27%), λ is between 4 and 8 times that of the USA.

Summarily, welfare costs are very different across EU countries and between US and EU countries, and thus it is a strong evidence that business cycles are not alike in Europe. Differences in institutional settings from country to country, and consequently the effects of shocks in the economies, are good explanation for variations in business cycles. Thus, our result is a contrary evidence of the *dynamic-stochastic-general-equilibrium* view.

4.6 Conclusions

In this paper, we investigate whether business cycles are all alike computing the welfare costs of business cycles for an important subset of European countries – European-Union (EU) countries: Austria, Belgium, Denmark, Finland, France, Germany, Great Britain, Greece, Ireland, Italy, Luxembourg, Portugal, Spain, Sweden and The Netherlands. As is well known, these countries have a long tradition of integration and trade dating well before the common-currency Euro area was implemented. Because of this feature, it is a “natural experiment” to investigate how similar their welfare costs of business cycles are, in the sense that we will be surprised to find major differences between them.

In computing the welfare costs of business cycles for EU countries we use the techniques in Beveridge and Nelson (1981) to decompose (the log of) consumption in a trend and a cyclical component. In this case, the trend will be stochastic and modeled as a random walk. Moreover, to make our results comparable to previous work, we also modelled the trend as either a deterministic linear process (with and without a break) or following a slowly evolving secular process captured by the Hodrick and Prescott (1997) filter.

Our results show that the welfare costs of business cycles are very different across EU countries. Using the Beveridge and Nelson decomposition, and plausible values for the risk aversion coefficient and the discount rate of future utility, we find that the welfare cost of

Spain (4.1% of consumption) is almost ten times that of the UK (0.45% of consumption) – median of 2.85%. Major differences in welfare costs are also found when alternative trend-cycle decomposition methods are employed, although they are not as pronounced as the ones obtained using the Beveridge and Nelson decomposition.

References

- [1] Atkeson, A. and Phelan, C., 1995, "Reconsidering the Cost of Business Cycles with Incomplete Markets", *NBER Macroeconomics Annual*, 187-207, with discussions.
- [2] Beveridge, S. and Nelson, C.R., 1981, "A New Approach to Decomposition of Economic Time Series into a Permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle'," *Journal of Monetary Economics*, 7, 151-174.
- [3] Blanchard, O., and Watson, M., 1986, "Are Business Cycles all Alike?", *NBER Working Paper Series*, 1392.
- [4] Campbell, J. 1987, "Does Saving Anticipate Declining Labor Income? An Alternative Test of the Permanent Income Hypothesis," *Econometrica*, vol. 55(6), pp. 1249-73.
- [5] Campbell, John Y. and Deaton, Angus 1989, "Why is Consumption So Smooth?", *The Review of Economic Studies* 56:357-374.
- [6] Cochrane, J.H., 1994, "Permanent and Transitory Components of GNP and Stock Prices," *Quarterly Journal of Economics*, 30, 241-265.
- [7] Dolmas, J., 1998, "Risk Preferences and the Welfare Cost of Business Cycles", *Review of Economic Dynamics*, 1, 646-676.
- [8] Engle, R.F. and Granger C.W.J., 1987, "Cointegration and Error Correction: Representation, Estimation and Testing", *Econometrica*, 55, 251-276.
- [9] Hall, R.E., 1978, "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy*, 86, 971-987.
- [10] Hamilton, J., 1994, *Time Series Analysis*, Princeton University Press.

- [11] Hodrick, R.J. and Prescott, E.C., 1997, "Postwar U.S. Business Cycles: An Empirical Investigation," *Journal of Money, Credit, and Banking*, 29, 116.
- [12] Imrohoroglu, Ayse, 1989, "Cost of Business Cycles With Indivisibilities and Liquidity Constraints", *Journal of Political Economy*, 97 (6) , 1364-1383.
- [13] Issler, J.V. and Vahid, F., 2001, "Common Cycles and the Importance of Transitory Shocks to Macroeconomic Aggregates," *Journal of Monetary Economics*, 47, 449-475.
- [14] Franco, A. R., Guillen, O., and Issler, J.V., 2003, "On the Welfare Costs of Business Cycles in the 20th Century", *Ensaios Econômicos da EPGE*, 481.
- [15] Johansen, S., 1988, "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, pp. 231-254.
- [16] Johansen, S., 1991, "Estimation and Hypothesis Testing of Cointegrated Vectors in Gaussian Vector Autoregressions", *Econometrica*, vol. 59-6, pp. 1551-1580.
- [17] King, R.G., Plosser, C.I., Stock, J.H. and Watson, M.W., 1991, "Stochastic Trends and Economic Fluctuations", *American Economics Review*, 81, 819-840.
- [18] Kuznets S., 1961, "Capital in the American Economy: Its formation and financing," Princeton University Press.
- [19] Lucas, R., 1977, "Understanding Business Cycles," *Carnegie-Rochester Series in Public Policy*, vol. 5, eds. Karl Brunner and Allan Meltzer, pp. 7-29.
- [20] Lucas, R., 1987, "*Models of Business Cycles*", Oxford: Blackwell.
- [21] Nelson, C.R. and Plosser, C., 1982, "Trends and Random Walks in Macroeconomics Time Series," *Journal of Monetary Economics*, 10, 1045-1066.
- [22] Obstfeld, M., 1994, "Evaluating Risky Consumption Paths: The Role of Intertemporal Substitutability," *European Economic Review*, 38, 1471-1486.
- [23] Otrok, C., 2001, "On Measuring the Welfare Cost of Business Cycles," *Journal of Monetary Economics*, 47, 61-92.

- [24] Pemberton, J., 1996, "Growth Trends, Cyclical Fluctuations, and Welfare with Non-Expected Utility Preferences," *Economic Letters*, 50, 387-392.
- [25] Proietti, T., 1997, "Short-run Dynamics in Cointegrated Systems", *Oxford Bulletin of Economics and Statistics*, 59 (3), 405-422.
- [26] Stock, J.H. and Watson, M.W., 1988, "Testing for Common Trends," *Journal of the American Statistical Association*, 83, 1097-1107.
- [27] Tallarini Jr., T.D., 2000, "Risk-sensitive Real Business Cycles", *Journal of Monetary Economics*, 45, 507-532.
- [28] Vahid, F. and Engle, R.F., 1993, "Common Trends and common Cycles", *Journal of Applied Econometrics*, 8, 341-360.
- [29] Watson, M.W., 1986, "Univariate Detrending Methods with Stochastic Trends", *Journal of Monetary Economics*, 18, 49-75.
- [30] Van Wincoop, E., 1994, "Welfare Gains From International Risksharing", *Journal of Monetary Economics*, 34, 175-200.

4.A Convergence of $\lambda(\phi, \beta)$ for $\phi \rightarrow 1$

Let $A(\phi) \equiv \exp\left[\frac{\phi(2\tilde{\sigma}_{12} + \tilde{\sigma}_{22})}{2}\right]$ and $B(\phi, \beta) = \left\{ \frac{1 - \beta(1 + \alpha_1)^{1-\phi} \exp\left[-\frac{(1-\phi)\phi\sigma_{11}}{2}\right]}{1 - \beta(1 + \alpha_1)^{1-\phi}} \right\}^{1/(1-\phi)}$, then $\lambda(\phi, \beta) = A(\phi) \cdot B(\phi, \beta) - 1$.

Rewriting $B(\phi, \beta)$,

$$\begin{aligned} B(\phi, \beta) &= \left\{ \frac{1 - \beta(1 + \alpha_1)^{1-\phi} \exp\left[-\frac{(1-\phi)\phi\sigma_{11}}{2}\right]}{1 - \beta(1 + \alpha_1)^{1-\phi}} \right\}^{1/(1-\phi)} \\ &= \left\{ 1 + \frac{\beta(1 + \alpha_1)^{1-\phi} \left(1 - \exp\left[-\frac{(1-\phi)\phi\sigma_{11}}{2}\right]\right)}{1 - \beta(1 + \alpha_1)^{1-\phi}} \right\}^{1/(1-\phi)} \\ &\equiv \left\{ [1 + B_1(\phi, \beta)]^{\frac{1}{B_1(\phi, \beta)}} \right\}^{B_2(\phi, \beta)}. \end{aligned} \quad (20)$$

where $B_1(\phi, \beta) \equiv \frac{\beta(1 + \alpha_1)^{1-\phi} \left(1 - \exp\left[-\frac{(1-\phi)\phi\sigma_{11}}{2}\right]\right)}{1 - \beta(1 + \alpha_1)^{1-\phi}}$ and $B_2(\phi, \beta) \equiv \frac{\beta(1 + \alpha_1)^{1-\phi} \left(1 - \exp\left[-\frac{(1-\phi)\phi\sigma_{11}}{2}\right]\right)}{(1 - \beta(1 + \alpha_1)^{1-\phi})(1 - \phi)}$.

Rewriting $B_2(\phi, \beta)$,

$$B_2(\phi, \beta) = \frac{\beta(1 + \alpha_1)^{1-\phi}}{(1 - \beta(1 + \alpha_1)^{1-\phi})} \frac{\left(1 - \exp\left[-\frac{(1-\phi)\phi\sigma_{11}}{2}\right]\right)}{(1 - \phi)} \equiv B_{21}(\phi, \beta) \cdot B_{22}(\phi). \quad (21)$$

Thus,

$$\lim_{\phi \rightarrow 1} B_{21}(\phi, \beta) = \frac{\beta}{1 - \beta}. \quad (22)$$

Applying L'Hospital's Rule in $B_{22}(\phi)$,

$$\lim_{\phi \rightarrow 1} B_{22}(\phi) = \frac{\lim_{\phi \rightarrow 1} \left(\frac{\sigma_{11}}{2} (1 - 2\phi) \exp\left[-\frac{(1-\phi)\phi\sigma_{11}}{2}\right] \right)}{(-1)} = \frac{\sigma_{11}}{2}. \quad (23)$$

From (21), (22) and (23) we have,

$$\lim_{\phi \rightarrow 1} B_2(\phi, \beta) = \frac{\beta\sigma_{11}}{2(1 - \beta)} \quad (24)$$

Since that $\lim_{\phi \rightarrow 1} B_1(\phi, \beta) = 0$, applying the definition of e (base of the natural logarithm) and using (24) in (20) we have,

$$\lim_{\phi \rightarrow 1} B(\phi, \beta) = \exp \left[\frac{\beta \sigma_{11}}{2(1 - \beta)} \right]. \quad (25)$$

Thus, we have,

$$\lim_{\phi \rightarrow 1} \lambda(\phi, \beta) = \exp \left[\frac{1}{2} \left(\frac{\beta \sigma_{11}}{(1 - \beta)} + 2\tilde{\sigma}_{12} + \tilde{\sigma}_{22} \right) \right] - 1.$$

4.B Deriving Trend and Cycle Formulae from Space-State Form

Space-state form:

$$\Delta y_t = Z f_t, \quad (26)$$

and

$$f_t = T f_{t-1} + Z' \varepsilon_t. \quad (27)$$

Trend-cycle representation from Beveridge and Nelson (1981):

$$y_t = \tau_t + \psi_t,$$

where

$$\tau_t = y_t + \lim_{k \rightarrow \infty} \sum_{i=1}^k E_t [\Delta y_{t+i}] \quad \text{and} \quad \psi_t = - \lim_{k \rightarrow \infty} \sum_{i=1}^k E_t [\Delta y_{t+i}]. \quad (28)$$

Applying (27) on (26) recursively we get

$$\Delta y_{t+i} = Z T^i f_t + \sum_{j=1}^i Z T^{i-j} Z' \varepsilon_{t+j}. \quad (29)$$

Applying E_t on both sides of (29) and summing up from $i = 1$ to ∞ we have:

$$\begin{aligned} \sum_{i=1}^{\infty} E_t [\Delta y_{t+i}] &= Z \left(\sum_{i=1}^{\infty} T^i \right) f_t = Z \left(\sum_{i=0}^{\infty} T^i \right) T f_t \\ &= Z [I_m - T]^{-1} T f_t. \end{aligned} \quad (30)$$

where $m = 2p + 1$.

From (28) and (30) the cyclical and trend components will be, respectively,

$$\psi_t = -Z [I_m - T]^{-1} T f_t, \quad (31)$$

and,

$$\tau_t = y_t - \psi_t. \quad (32)$$

It is straightforward to see that $\Delta \tau_t = Z [I_m - T]^{-1} Z' \varepsilon_t$, i.e. τ_t is a multivariate random-walk.

4.C Computing Conditional Covariances

From Proposition 2 in Proietti (1997),

$$\psi_{t+1} = -(I_N - P)(\Gamma(1) + \gamma\alpha')^{-1}\Gamma^*(L) \Delta y_{t+1} + P y_{t+1}, \quad (33)$$

and,

$$\tau_{t+1} = (I_N - P)(\Gamma(1) + \gamma\alpha')^{-1}\Gamma(L)y_{t+1}$$

or,

$$\Delta \tau_{t+1} = (I_N - P)(\Gamma(1) + \gamma\alpha')^{-1}\varepsilon_{t+1}, \quad (34)$$

where $P = (\Gamma(1) + \gamma\alpha')^{-1}\gamma[\alpha'(\Gamma(1) + \gamma\alpha')^{-1}\gamma]\alpha'$, and $\Gamma(L) = I_2 - \Gamma_1 L$, which is decomposed as:

$$\begin{aligned} \Gamma(L) &= \Gamma(1) + (1 - L)\Gamma^*(L), \text{ where,} \\ \Gamma^*(L) &= \Gamma_1, \end{aligned}$$

in the present context.

From (34) we have,

$$\tau_{t+s} = \tau_t + (I_N - P)(\Gamma(1) + \gamma\alpha')^{-1} \sum_{i=1}^s \varepsilon_{t+i},$$

which implies that $\hat{\tau}_{t+s|t} = E_t[\tau_{t+s}] = \tau_t$. Denoting $K = (I_N - P)(\Gamma(1) + \gamma\alpha')^{-1}$, and $(\tau_{t+s} - \hat{\tau}_{t+s|t}) = K \sum_{i=1}^s \varepsilon_{t+i}$ we have,

$$\begin{aligned} E \left[(\tau_{t+s} - \hat{\tau}_{t+s|t}) (\tau_{t+s} - \hat{\tau}_{t+s|t})' \right] &= K \left(\sum_{i=1}^s Q_i \right) K' \\ &= s \cdot K Q K', \end{aligned} \quad (35)$$

where $E[\varepsilon_{t+i} \varepsilon_{t+i}'] = Q_i = Q$. On the other hand, from (33),

$$\psi_{t+s} = -K\Gamma^*(L) \Delta y_{t+s} + P y_{t+s},$$

but, $\Delta y_{t+1} = ZTf_t + \varepsilon_{t+1}$, which implies that $\Delta y_{t+s} = ZTf_{t+s-1} + \varepsilon_{t+s}$. However, $y_{t+1} = y_t + ZTf_t + \varepsilon_{t+1}$, which implies that $y_{t+s} = y_t + Z \left(\sum_{i=1}^s T^i \right) f_t + \sum_{i=1}^{s-1} \left(I_N + Z \left\{ \sum_{j=1}^{s-i} T^j \right\} Z' \right) \varepsilon_{t+i} + \varepsilon_{t+s}$. Hence, $\hat{\psi}_{t+s|t} = E_t[\psi_{t+s}] = -K\Gamma^*(1) ZTf_{t+s-1} + P y_t + P Z \left(\sum_{i=1}^s T^i \right) f_t$, which implies that $(\psi_{t+s} - \hat{\psi}_{t+s|t}) = [P - K\Gamma^*(1)] \varepsilon_{t+s} + P \sum_{i=1}^{s-1} \left(I_N + Z \left\{ \sum_{j=1}^{s-i} T^j \right\} Z' \right) \varepsilon_{t+i}$. Denoting $V = [P - K\Gamma^*(1)]$ and $W(i) = \left(I_N + Z \left\{ \sum_{j=1}^{s-i} T^j \right\} Z' \right)$. Thus,

$$E \left[(\psi_{t+s} - \hat{\psi}_{t+s|t}) (\psi_{t+s} - \hat{\psi}_{t+s|t})' \right] = V Q V' + P \left(\sum_{i=1}^{s-1} W(i) Q W(i)' \right) P' \quad (36)$$

and

$$E \left[(\tau_{t+s} - \hat{\tau}_{t+s|t}) (\psi_{t+s} - \hat{\psi}_{t+s|t})' \right] = K Q V' + K \left[\sum_{i=1}^{s-1} Q W(i)' \right] P'. \quad (37)$$

It is straightforward to extract the estimates of σ_{11} , σ_{22} , and σ_{12} , in (6) from equations (35), (36) and (37). One has simply to evaluate them at sample estimates selecting their first diagonal element, which will correspond to consumption moments.

4.D Tables

Tabela D.1 - Welfare Cost of Business Cycles ($\lambda\%$)

(continues)

Lucas(1987) Benchmark Values (USA Data)		$\beta = 0.950, 0.971, 0.985$	$\phi=1$	$\phi=5$	$\phi=10$	$\phi=20$
			0.008	0.042	0.08	0.17
Country	Model	β	$\phi=1$	$\phi=5$	$\phi=10$	$\phi=20$
USA (Franco et alii, 2003)	Beveridge-Nelson Decomposition	0.950	0.14	0.25	0.27	0.30
		0.971	0.24	0.29	0.30	0.31
		0.985	0.47	0.33	0.32	0.32
	Hodrick-Prescott for Trend	0.950, 0.971, 0.985	0.01	0.04	0.08	0.17
	Linear Time Trend	0.950, 0.971, 0.985	0.14	0.70	1.41	2.84
USA (Total Consumption)	Beveridge-Nelson Decomposition	0.950	0.38 (0.0115)	0.66 (0.0200)	0.70 (0.0217)	0.68 (0.0226)
		0.971	0.67 (0.0203)	0.79 (0.0240)	0.78 (0.0242)	0.73 (0.0242)
		0.985	1.33 (0.0402)	0.90 (0.0277)	0.83 (0.0262)	0.76 (0.0254)
	Hodrick-Prescott for Trend	0.950, 0.971, 0.985	0.02 (0.0005)	0.09 (0.0027)	0.19 (0.0054)	0.38 (0.0107)
	Linear Time Trend	0.950, 0.971, 0.985	0.05 (0.0015)	0.26 (0.0073)	0.51 (0.0146)	1.03 (0.0294)
European Union (EU-15)	Beveridge-Nelson Decomposition	0.950	0.10 (0.0021)	0.27 (0.0062)	0.44 (0.0115)	0.77 (0.0225)
		0.971	0.14 (0.0034)	0.28 (0.0065)	0.45 (0.0116)	0.78 (0.0226)
		0.985	0.25 (0.0065)	0.30 (0.0067)	0.46 (0.0117)	0.78 (0.0226)
	Hodrick-Prescott for Trend	0.950, 0.971, 0.985	0.01 (0.0003)	0.06 (0.0017)	0.12 (0.0035)	0.25 (0.0069)
	Linear Time Trend	0.950, 0.971, 0.985	0.10 (0.0029)	0.52 (0.0148)	1.05 (0.0298)	2.11 (0.0602)
The Netherlands	Beveridge-Nelson Decomposition	0.950	1.33 (0.0403)	1.98 (0.0630)	2.11 (0.0730)	2.32 (0.1012)
		0.971	2.36 (0.0717)	2.31 (0.0744)	2.31 (0.0810)	2.47 (0.1094)
		0.985	4.68 (0.1439)	2.60 (0.0842)	2.47 (0.0870)	2.57 (0.1153)
	Hodrick-Prescott for Trend	0.950, 0.971, 0.985	0.07 (0.0019)	0.34 (0.0096)	0.68 (0.0192)	1.36 (0.0387)
	Linear Time Trend	0.950, 0.971, 0.985	0.21 (0.0060)	1.07 (0.0304)	2.15 (0.0613)	4.34 (0.1253)
Italy	Beveridge-Nelson Decomposition	0.950	1.74 (0.0522)	2.16 (0.0682)	2.00 (0.0711)	1.78 (0.1018)
		0.971	3.13 (0.0952)	2.54 (0.0817)	2.24 (0.0808)	1.97 (0.1141)
		0.985	6.29 (0.1947)	2.86 (0.0931)	2.41 (0.0882)	2.10 (0.1232)
	Hodrick-Prescott for Trend	0.950, 0.971, 0.985	0.03 (0.0008)	0.14 (0.0039)	0.28 (0.0079)	0.56 (0.0158)
	Linear Time Trend	0.950, 0.971, 0.985	0.40 (0.0114)	2.02 (0.0577)	4.08 (0.1178)	8.33 (0.2453)

Tabela D.1 - Welfare Cost of Business Cycles ($\lambda\%$)

(continues)

Country	Model	β	$\phi=1$	$\phi=5$	$\phi=10$	$\phi=20$
United Kingdom	Beveridge-Nelson Decomposition	0.950	0.23 (0.0067)	0.42 (0.0122)	0.47 (0.0137)	0.52 (0.0154)
		0.971	0.40 (0.0119)	0.50 (0.0146)	0.52 (0.0151)	0.55 (0.0162)
		0.985	0.78 (0.0234)	0.57 (0.0167)	0.56 (0.0162)	0.57 (0.0168)
	Hodrick-Prescott for Trend	0.950, 0.971, 0.985	0.02 (0.0006)	0.10 (0.0028)	0.20 (0.0057)	0.40 (0.0114)
	Linear Time Trend	0.950, 0.971, 0.985	0.04 (0.0012)	0.21 (0.0061)	0.43 (0.0122)	0.86 (0.0244)
France	Beveridge-Nelson Decomposition	0.950	1.35 (0.0402)	2.16 (0.0677)	2.26 (0.0788)	2.44 (0.1240)
		0.971	2.41 (0.0730)	2.61 (0.0837)	2.56 (0.0914)	2.69 (0.1411)
		0.985	4.83 (0.1482)	3.01 (0.0982)	2.79 (0.1014)	2.88 (0.1542)
	Hodrick-Prescott for Trend	0.950, 0.971, 0.985	0.01 (0.0004)	0.07 (0.0020)	0.14 (0.0041)	0.29 (0.0082)
	Linear Time Trend	0.950, 0.971, 0.985	0.45 (0.0126)	2.25 (0.0642)	4.54 (0.1314)	9.29 (0.2747)
Austria	Beveridge-Nelson Decomposition	0.950	0.68 (0.0190)	1.29 (0.0333)	1.56 (0.0394)	1.97 (0.0537)
		0.971	1.17 (0.0339)	1.51 (0.0401)	1.70 (0.0434)	2.05 (0.0561)
		0.985	2.28 (0.0677)	1.70 (0.0463)	1.80 (0.0467)	2.11 (0.0579)
	Hodrick-Prescott for Trend	0.950, 0.971, 0.985	0.05 (0.0013)	0.23 (0.0065)	0.46 (0.0130)	0.92 (0.0261)
	Linear Time Trend	0.950, 0.971, 0.985	0.12 (0.0033)	0.58 (0.0166)	1.17 (0.0333)	2.36 (0.0675)
Spain	Beveridge-Nelson Decomposition	0.950	2.29 (0.0637)	3.88 (0.1023)	4.86 (0.1358)	7.55 (0.3401)
		0.971	4.00 (0.1164)	4.48 (0.1243)	5.27 (0.1541)	8.07 (0.3962)
		0.985	7.90 (0.2401)	5.00 (0.1439)	5.59 (0.1690)	8.47 (0.4432)
	Hodrick-Prescott for Trend	0.950, 0.971, 0.985	0.06 (0.0018)	0.31 (0.0089)	0.63 (0.0178)	1.26 (0.0358)
	Linear Time Trend	0.950, 0.971, 0.985	0.74 (0.0211)	3.78 (0.1089)	7.70 (0.2260)	15.99 (0.4867)
Portugal	Beveridge-Nelson Decomposition	0.950	1.56 (0.0346)	4.54 (0.1152)	7.97 (0.2278)	15.11 (0.4796)
		0.971	2.26 (0.0532)	4.74 (0.1178)	8.09 (0.2288)	15.18 (0.4803)
		0.985	3.83 (0.0997)	4.90 (0.1202)	8.17 (0.2297)	15.23 (0.4807)
	Hodrick-Prescott for Trend	0.950, 0.971, 0.985	0.16 (0.0044)	0.79 (0.0224)	1.58 (0.0451)	3.19 (0.0916)
	Linear Time Trend	0.950, 0.971, 0.985	0.64 (0.0181)	3.24 (0.0931)	6.58 (0.1921)	13.59 (0.4096)

Tabela D.1 - Welfare Cost of Business Cycles ($\lambda\%$)

(continues)

Country	Model	β	$\phi=1$	$\phi=5$	$\phi=10$	$\phi=20$
Belgium	Beveridge-Nelson Decomposition	0.950	1.52 (0.0360)	3.68 (0.0824)	5.65 (0.1397)	9.59 (0.2738)
		0.971	2.43 (0.0628)	4.06 (0.0916)	5.89 (0.1443)	9.78 (0.2774)
		0.985	4.47 (0.1257)	4.38 (0.1005)	6.08 (0.1482)	9.91 (0.2803)
	Hodrick-Prescott for Trend	0.950, 0.971, 0.985	0.03 (0.0008)	0.13 (0.0038)	0.27 (0.0075)	0.53 (0.0151)
	Linear Time Trend	0.950, 0.971, 0.985	0.17 (0.0050)	0.88 (0.0249)	1.76 (0.0503)	3.56 (0.1024)
Finland	Beveridge-Nelson Decomposition	0.950	1.97 (0.0541)	3.63 (0.0942)	4.63 (0.1257)	7.11 (0.2881)
		0.971	3.41 (0.0986)	4.23 (0.1155)	5.05 (0.1435)	7.60 (0.3341)
		0.985	6.70 (0.2023)	4.76 (0.1350)	5.38 (0.1581)	7.98 (0.3727)
	Hodrick-Prescott for Trend	0.950, 0.971, 0.985	0.08 (0.0022)	0.40 (0.0112)	0.79 (0.0225)	1.59 (0.0453)
	Linear Time Trend	0.950, 0.971, 0.985	0.33 (0.0095)	1.68 (0.0479)	3.39 (0.0975)	6.89 (0.2015)
Denmark	Beveridge-Nelson Decomposition	0.950	1.52 (0.0461)	2.73 (0.0887)	3.25 (0.1173)	4.35 (0.2174)
		0.971	2.68 (0.0817)	3.26 (0.1072)	3.60 (0.1327)	4.68 (0.2433)
		0.985	5.30 (0.1636)	3.72 (0.1240)	3.88 (0.1451)	4.94 (0.2637)
	Hodrick-Prescott for Trend	0.950, 0.971, 0.985	0.08 (0.0023)	0.41 (0.0116)	0.82 (0.0233)	1.65 (0.0470)
	Linear Time Trend	0.950, 0.971, 0.985	0.29 (0.0081)	1.43 (0.0408)	2.89 (0.0828)	5.86 (0.1705)
Ireland	Beveridge-Nelson Decomposition	0.950	1.37 (0.0410)	2.03 (0.0629)	2.16 (0.0717)	2.35 (0.1002)
		0.971	2.43 (0.0735)	2.39 (0.0750)	2.37 (0.0802)	2.51 (0.1093)
		0.985	4.83 (0.1484)	2.69 (0.0855)	2.54 (0.0867)	2.63 (0.1159)
	Hodrick-Prescott for Trend	0.950, 0.971, 0.985	0.06 (0.0017)	0.29 (0.0083)	0.59 (0.0167)	1.18 (0.0336)
	Linear Time Trend	0.950, 0.971, 0.985	0.28 (0.0080)	1.43 (0.0407)	2.88 (0.0825)	5.84 (0.1698)
Greece	Beveridge-Nelson Decomposition	0.950	2.00 (0.0602)	2.56 (0.0811)	2.54 (0.0909)	2.72 (0.1542)
		0.971	3.59 (0.1092)	2.99 (0.0965)	2.81 (0.1024)	2.96 (0.1719)
		0.985	7.20 (0.2237)	3.35 (0.1095)	3.01 (0.1112)	3.13 (0.1852)
	Hodrick-Prescott for Trend	0.950, 0.971, 0.985	0.04 (0.0011)	0.20 (0.0057)	0.40 (0.0114)	0.80 (0.0228)
	Linear Time Trend	0.950, 0.971, 0.985	0.46 (0.0130)	2.32 (0.0663)	4.69 (0.1357)	9.59 (0.2840)

Tabela D.1 - Welfare Cost of Business Cycles ($\lambda\%$)

(conclusion)

Country	Model	β	$\phi=1$	$\phi=5$	$\phi=10$	$\phi=20$
Sweden	Beveridge-Nelson Decomposition	0.950	0.37 (0.0110)	0.73 (0.0225)	0.81 (0.0257)	0.81 (0.0281)
		0.971	0.65 (0.0196)	0.92 (0.0283)	0.93 (0.0297)	0.88 (0.0309)
		0.985	1.28 (0.0389)	1.09 (0.0337)	1.02 (0.0329)	0.94 (0.0330)
	Hodrick-Prescott for Trend	0.950, 0.971, 0.985	0.02 (0.0006)	0.10 (0.0029)	0.21 (0.0058)	0.41 (0.0117)
	Linear Time Trend	0.950, 0.971, 0.985	0.09 (0.0025)	0.45 (0.0127)	0.90 (0.0254)	1.80 (0.0514)
Luxembourg	Beveridge-Nelson Decomposition	0.950	0.72 (0.0176)	1.49 (0.0340)	2.22 (0.0569)	3.65 (0.1093)
		0.971	1.16 (0.0306)	1.61 (0.0364)	2.29 (0.0577)	3.69 (0.1096)
		0.985	2.14 (0.0604)	1.71 (0.0386)	2.33 (0.0583)	3.71 (0.1097)
	Hodrick-Prescott for Trend	0.950, 0.971, 0.985	0.13 (0.0038)	0.66 (0.0189)	1.33 (0.0380)	2.68 (0.0769)
	Linear Time Trend	0.950, 0.971, 0.985	0.31 (0.0087)	1.55 (0.0442)	3.13 (0.0898)	6.35 (0.1853)
Germany	Beveridge-Nelson Decomposition	0.950	2.18 (0.0661)	3.27 (0.1106)	3.58 (0.1454)	5.05 (0.3948)
		0.971	3.90 (0.1196)	3.90 (0.1346)	4.02 (0.1673)	5.63 (0.4689)
		0.985	7.83 (0.2448)	4.44 (0.1559)	4.35 (0.1848)	6.09 (0.5316)
	Hodrick-Prescott for Trend	0.950, 0.971, 0.985	0.08 (0.0023)	0.40 (0.0114)	0.81 (0.0229)	1.62 (0.0461)
	Linear Time Trend	0.950, 0.971, 0.985	0.18 (0.0051)	0.90 (0.0255)	1.81 (0.0515)	3.64 (0.1049)