

UNIVERSIDADE FEDERAL FLUMINENSE
DEPARTAMENTO DE ECONOMIA
DOUTORADO EM ECONOMIA

ENSAIOS SOBRE SEGREGAÇÃO NO MERCADO DE TRABALHO,
AVALIAÇÃO DE PROGRAMAS SOCIAIS E IDENTIFICAÇÃO
NÃO-PARAMÉTRICA

Miguel Nathan Foguel

Orientadora: Rosane Silva Pinto de Mendonça

Niterói
Novembro de 2007

UNIVERSIDADE FEDERAL FLUMINENSE
DEPARTAMENTO DE ECONOMIA
DOUTORADO EM ECONOMIA

ENSAIOS SOBRE SEGREGAÇÃO NO MERCADO DE TRABALHO,
AVALIAÇÃO DE PROGRAMAS SOCIAIS E IDENTIFICAÇÃO
NÃO-PARAMÉTRICA

Miguel Nathan Foguel

Tese apresentada para o programa de doutorado em
economia da Universidade Federal Fluminense como
parte dos requisitos necessários para a obtenção do
título de Doutor em Economia

Orientadora: Rosane Silva Pinto de Mendonça

Niterói
Novembro de 2007

MIGUEL NATHAN FOGUEL

ENSAIOS SOBRE SEGREGAÇÃO NO MERCADO DE TRABALHO,
AVALIAÇÃO DE PROGRAMAS SOCIAIS E IDENTIFICAÇÃO
NÃO-PARAMÉTRICA

Tese apresentada para o programa de doutorado em
economia da Universidade Federal Fluminense como
parte dos requisitos necessários para a obtenção do
título de Doutor em Economia

BANCA EXAMINADORA

Professora Doutora Rosane Silva Pinto de Mendonça (orientadora) - UFF

Professor Doutor Ricardo Paes de Barros - IPEA

Professora Doutora Simone Wajnman - UFMG

Professora Doutora Danielle Carusi Machado - UFF

Professor Doutor Sergio Pinheiro Firpo - PUC/RJ

Niterói
Departamento de Economia - UFF
Novembro de 2007

Agradecimentos

Gostaria de agradecer a Rosane Mendonça por me ajudar a elaborar a presente tese. Os comentários, a energia e os incentivos que me deu foram fundamentais para que eu pudesse aprimorar e finalizar este trabalho. Sempre disponível, atenciosa e generosa, para com ela, tenho uma enorme dívida de gratidão.

Gostaria também de agradecer a Ricardo Paes de Barros, com quem tive o privilégio de discutir em detalhes diversas partes desta tese. Sua clareza intelectual e profundo conhecimento de economia e econometria foram cruciais para o desenvolvimento deste trabalho. Ele tem sido e continuará sendo uma importante fonte de inspiração para minha carreira de pesquisador. Para além do suporte estritamente acadêmico, quero também agradecer-lhe pela amizade e pela generosidade com que me ajudou a contornar os momentos mais difíceis da árdua travessia do doutorado. Minha gratidão para com ele é incalculável.

Diversos outros pesquisadores contribuíram direta ou indiretamente para a elaboração desta tese. Inicialmente, gostaria de agradecer aos meus colegas do Instituto de Pesquisa Econômica Aplicada (IPEA), em especial a Carlos Henrique Corseuil, que generosamente me concedeu inúmeras horas do seu tempo para me ajudar nos problemas do doutorado e discutir o conteúdo dos capítulos que compõem o corpo principal da tese. Quero também agradecer em particular a João Pedro Azevedo, Claudio Ferraz, Gabriel Ulyssea, Dimitri Szerman, Ajax Moreira, Waldir Lobão, Solange Kanso, Samuel Franco e Mirela de Carvalho. Todos me ajudaram muito ao longo dos últimos anos, tanto no plano pessoal quanto no acadêmico.

Agradeço a Elena Martínez-Sanchís, Pierre Hoonhout, Mario Fiorini, Yong-Woo Lee, Giovanni Gallipoli e Marco Cozzi, que foram meus mais próximos companheiros nos primeiros anos de doutorado na *University College London (UCL)*. Além de terem me animado em diversos momentos difíceis, todos eles me deram comentários relevantes sobre a tese. Quero também expressar minha gratidão a Fábio Veras e Sara Lemos pela ajuda pessoal no desenvolvimento dessa pesquisa.

Gostaria de agradecer a Sérgio Firpo, Alexandre Ywata Carvalho, Luiz Eduardo Mirando Cruz e Marcelo Jovita Moreira por terem me ajudado a compreender melhor

diversos tópicos intrincados dos capítulos da tese. Agradeço também a Hidehiko Ichimura, Andrew Chesher, Richard Blundell e Costas Meghir, que foram meus professores na *UCL* e me ajudaram a iniciar o desenvolvimento de alguns capítulos desta tese. Por sua generosidade e discussões sobre diversas dimensões da vida em geral, expresso aqui minha gratidão à Professora Victoria Chick. Quero agradecer também a Gustavo Gonzaga pela atenção e força que me deu para superar alguns dos momentos mais complicados do doutorado.

Sou grato também aos inúmeros participantes dos seminários que apresentei sobre os capítulos da tese na Universidade Federal Fluminense (UFF), Pontifícia Universidade Católica do Rio de Janeiro (PUC/RJ), Escola de Pós-Graduação em Economia (EPGE-FGV/RJ), *UCL*, *Universidad de las Baleares*, *Latin American and Caribbean Economic Association (LACEA)*, *Latin American Meeting of the Econometric Society (LAMES)*, Sociedade Brasileira de Econometria (SBE) e IPEA.

Quero deixar um especial agradecimento aos Professores do Departamento de Economia da UFF, em particular à Professora Carmen Feijó. Sem a carinhosa acolhida que me deram, certamente eu não teria conseguido finalizar o trabalho que agora apresento. Eterna será a minha gratidão para com eles.

Quero agradecer também aos meus amigos pessoais, que me ouviram pacientemente e sempre me incentivaram a seguir em frente. Em especial, agradeço a Aloysio d’Abreu por sua disponibilidade e generosidade ao longo dos últimos anos.

Diego, meu querido menino, é o que de melhor aconteceu em minha vida. Seus sorrisos, amor e *joie de vivre* são a base de todas as forças que me fizeram progredir. Cristine, querida amiga e mãe de Diego, foi a pessoa que mais me deu suporte ao longo de todo o doutorado. Também, sem o afeto e o amparo que recebi nesses últimos anos de minha querida Juliana, decerto eu não teria chegado até aqui. Eu nunca serei capaz de agradecer-lhes o suficiente por todo o carinho e apoio que me deram nessa trajetória. Eu dedico esta tese a eles, e também a minha mãe, irmã e cunhado por sua infindável amizade, afeição e ajuda emocional. Na realidade, de várias maneiras, todos eles são importantes “co-autores” desse trabalho.

A elaboração da presente tese não teria sido possível sem o financiamento concedido pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). O apoio institucional que recebi do IPEA foi também fundamental para realização e conclusão deste trabalho.

Resumo

O corpo principal desta tese possui três capítulos. No capítulo 1, estudamos o efeito da segregação por sexo no nível do estabelecimento sobre os salários de homens e mulheres. Enquanto a literatura disponível sobre o tema tem utilizado dados de *cross-section*, nesse capítulo fazemos uso de um extenso painel de informações combinadas de empregadores e empregados para o Brasil (Relação Anual de Informações Sociais - RAIS). As estimativas obtidas com base em métodos longitudinais mostram que o sinal do efeito de interesse parece ser negativo. No entanto, embora esse resultado esteja em linha com as estimativas de *cross-section*, a análise longitudinal revela que a magnitude do efeito é menor (isto é, menos negativa). Essa diferença de magnitude indica que os efeitos específicos não-observáveis de trabalhadores e estabelecimentos são (negativamente) correlacionados com a proporção de mulheres nos estabelecimentos, o que implica que as estimativas tradicionais de *cross-section* são enviesadas.

Os capítulos 2 e 3 da tese investigam a identificação não-paramétrica de certos efeitos de interesse quando um sub-conjunto das covariadas do modelo são funcionalmente dependentes das demais. Num contexto não-paramétrico, essa dependência funcional entre covariadas é equivalente à existência de perfeita multi-colinearidade num modelo de regressão linear. Propomos um conjunto de condições suficientes que são capazes de gerar a identificação não-paramétrica de alguns aspectos dos efeitos das covariadas funcionalmente dependentes. No capítulo 2, estudamos a identificação de dois efeitos relevantes para a área de avaliação de programas sociais: (i) o efeito temporal do programa em si; e, (ii) o efeito associado com a influência das condições econômicas externas ao programa. O capítulo 3 apresenta uma generalização da análise de identificação do capítulo 2 para outras áreas de aplicação.

Palavras chaves: Diferenciais de Salários, Discriminação, Identificação, Modelos Não-Paramétricos, Avaliação de Programas.

Abstract

The main body of this thesis contains three chapters. In chapter 1, we study the effect of gender segregation at the establishment level on the wages of males and females. While the available literature on this issue has used cross-section data, in that chapter we make use of a large panel of matched employer-employee data for Brazil (“Annual Social Information Report” - *Relação Anual de Informações Sociais* - RAIS). Estimates obtained with longitudinal methods show that the sign of the effect of interest seems to be negative. However, though this result is in line with the cross-section estimates, the longitudinal analysis reveals that the magnitude of the effect is smaller (that is, less negative). This difference in magnitude indicates that unobserved worker- and establishment-specific effects are (negatively) correlated with the proportion of females in establishments, implying that the traditional cross-section estimates are biased.

Chapters 2 and 3 of the thesis investigate the non-parametric identification of certain effects of interest when a subset of the covariates are functionally dependent of the other covariates in the model. In a non-parametric setting, this functional dependence is equivalent to the existence of perfect multi-collinearity in a linear regression model. We propose a set of sufficient conditions upon which some aspects of the effects of the functionally dependent covariates are identified in a non-parametric fashion. In chapter 2, we study the identification of two effects that are relevant to the area of programme evaluation: (i) the time effect of the programme itself; and (ii) the effect associated with the influence of economic conditions that are external to the programme. Chapter 3 presents a generalisation of the identification analysis of chapter 2 to other areas of application.

Keywords: Wage Differentials, Discrimination, Identification, Non-Parametric Models, Programme Evaluation.

Sumário

Introdução	1
1 Os Efeitos da Segregação por Gênero sobre Salários no Nível do Estabelecimento: Uma Análise Empírica Usando um Painel de Dados Combinados de Empregadores e Empregados	6
1.1 Introdução	6
1.2 Literatura Empírica	10
1.3 Literatura Teórica	13
1.3.1 Teorias Gerais	13
1.3.2 Discriminação pelo Empregador num Mercado de Trabalho Monopsonístico	24
1.3.3 Discriminação Estatística e <i>Quality Sorting</i>	28
1.4 Base de Dados	35
1.4.1 Descrição	35
1.4.2 Variável de Salário	36
1.4.3 Construção das Covariadas	36
1.4.4 Seleção da Amostra	37
1.4.5 Estatísticas Descritivas	38
1.5 Medindo Segregação Feminina entre Estabelecimentos	40
1.6 Metodologia Empírica	44
1.7 Resultados	48
1.7.1 Por Sexo	48
1.7.2 Por Sexo e Grupo Ocupacional	50
1.8 Robustez dos Resultados	53
1.8.1 Trabalhadores Permanentes	55
1.8.2 Todos os Trabalhadores	63
1.9 Discussão	68
1.9.1 Sumário dos Resultados Empíricos	68
1.9.2 Comparação com Resultados de <i>Cross-Section</i>	70
1.9.3 Efeito sobre o Hiato Salarial por Gênero	73

1.9.4	Considerações de Política	77
1.10	Conclusões	80
2	Efeitos do Tempo, Efeitos Econômicos e Impactos de Programas Sociais	83
2.1	Introdução	83
2.2	Múltiplas Regiões e Coorte Única	86
2.2.1	Identificação do EIT: Múltiplas Regiões	87
2.2.2	Identificação do EEE: Múltiplas Regiões	90
2.3	Múltiplas Coortes e Região Única	94
2.3.1	Identificação do EIT: Múltiplas Coortes	95
2.3.2	Identificação do EEE: Múltiplas Coortes	97
2.4	Exercício Empírico	102
2.4.1	Descrição do Programa	102
2.4.2	Base de Dados	103
2.4.3	Metodologia	103
2.4.4	Resultados	109
2.5	Conclusões	117
3	Identificação de Diferenças Parciais com Dependência Funcional entre Covariadas	120
3.1	Introdução	120
3.2	Literatura Relacionada	123
3.3	Modelo e Resultados de Identificação	124
3.3.1	Modelo	124
3.3.2	Resultados de Identificação	126
3.4	Um Exemplo Analítico	134
3.5	Conclusões	138
	Conclusões e Pesquisas Futuras	141
	Referências Bibliográficas	144

Lista de Tabelas

1.1	Características e Resultados dos Estudos sobre a Relação entre Salários e Segregação por Gênero no Estabelecimento	11
1.2	Os sinais da relação entre os salários de homens e mulheres e o grau de feminização das firmas ($\frac{dw_{H^*}}{d(\frac{M^*}{H^*})}$ e $\frac{dw_{M^*}}{d(\frac{M^*}{H^*})}$)	29
1.3	Estatísticas Descritivas: Município de Belo Horizonte	39
1.4	Medidas de Segregação por Gênero entre Estabelecimentos: Município de Belo Horizonte	43
1.5	Estimativas do Modelo Dinâmico por Sexo	49
1.6	Estimativas do Modelo Dinâmico por Sexo: Ocupações Ligadas à Produção	51
1.7	Estimativas do Modelo Dinâmico por Sexo: Ocupações Administrativas	52
1.8	Estimativas do Modelo Dinâmico por Sexo: Ocupações Profissionais	53
1.9	Estimativas do Modelo Dinâmico por Sexo: Ocupações Gerenciais	54
1.10	Estimativas dos Modelos de Intra-Grupos e de Primeiras-Diferenças para o Coeficiente da Variável Proporção de Mulheres no Estabelecimento por Sexo e Grupo Ocupacional: Trabalhadores Permanentes	56
1.11	Estimativas do Modelo Multi-Nível por Sexo: Trabalhadores Permanentes	59
1.12	Estimativas do Modelo Multi-Nível para Mulheres por Grupo Ocupacional: Trabalhadores Permanentes	61
1.13	Estimativas do Modelo Multi-Nível para Homens por Grupo Ocupacional: Trabalhadores Permanentes	62
1.14	Estimativas do Modelo Multi-Nível por Sexo: Todos os Trabalhadores	64
1.15	Estimativas do Modelo Multi-Nível para Mulheres por Grupo Ocupacional: Todos os Trabalhadores	65
1.16	Estimativas do Modelo Multi-Nível para Homens por Grupo Ocupacional: Todos os Trabalhadores	66
1.17	Estimativas do Modelo de Intra-Grupos para o Coeficiente da Variável Proporção de Mulheres no Estabelecimento por Sexo e Grupo Ocupacional: Todos os Trabalhadores	68
1.18	Média das Estimativas do Coeficiente da Variável Proporção de Mulheres no Estabelecimento	70
1.19	Estimativas por Mínimos Quadrados Ordinários: Dados de Cross-Section	72

1.20	Decomposições de Oaxaca-Blinder para o Hiato Salarial por Gênero	76
1.21	Decomposição do Impacto de um Programa que Fixa a Proporção de Mulheres nos Estabelecimentos	79
2.1	Tamanho da Amostra por Região, Sexo e Status de Tratamento	103
2.2	Estatísticas Descritivas da Amostra por Região e Status de Tratamento . .	104
2.3	O Efeito Interno Temporal por Sexo e Região Metropolitana: Estimativas Locais	110
2.4	O Efeito Econômico Externo por Sexo: Estimativas Locais	112

Lista de Figuras

1.1	Curvas de Salários para Homens (H) e Mulheres (M)	22
1.2	Curvas de Custo para Diferentes Níveis de Habilidade Inata	33
2.1	Impactos Hipotéticos Médios de um Programa	85
2.2	Superfícies da Função Hipotética m para Duas Regiões e o Mesmo Tipo de Indivíduos	89
2.3	Superfícies da Função Hipotética m para Duas Regiões e o Mesmo Tipo de Indivíduos	92
2.4	Superfícies da Função Hipotética h para Duas Coortes e o Mesmo Tipo de Indivíduos	97
2.5	Superfícies da Função Hipotética h para Duas Coortes e o Mesmo Tipo de Indivíduos	100
2.6	Impacto Médio sobre o Rendimento do Trabalho por Região Metropolitana: Homens	105
2.7	Impacto Médio sobre o Rendimento do Trabalho por Região Metropolitana: Mulheres	106
2.8	Contrações Líquidas por Região Metropolitana: Homens	108
2.9	Contrações Líquidas por Região Metropolitana: Mulheres	108
2.10	Estimativa da Função m_1 para Homens	113
2.11	Estimativa da Função m_1 para Mulheres	114
2.12	Efeito Interno Temporal para Homens: Estimativas Semi-Paramétricas . . .	115
2.13	Efeito Econômico Externo para Homens: Estimativas Semi-Paramétricas . .	116
2.14	Efeito Interno Temporal para Mulheres: Estimativas Semi-Paramétricas . .	116
2.15	Efeito Econômico Externo para Mulheres: Estimativas Semi-Paramétricas .	117
3.1	Funções Hipotéticas m e f : Identificação de Δ_{Z_l}	128
3.2	Funções Hipotéticas m e f : Identificação de Δ_{X_k}	132

Introdução

O corpo principal desta tese está dividido em três capítulos, cada qual podendo ser considerado um artigo individual sobre o tópico que procura investigar. O segundo e o terceiro capítulos estão mais diretamente conectados, uma vez que tratam de uma questão semelhante: a identificação não-paramétrica de certos efeitos de interesse. O primeiro capítulo lida com modelos paramétricos, os quais são utilizados para estimar o efeito sobre os salários da segregação de mulheres no nível do estabelecimento. No que se segue, procuramos apresentar as principais motivações por trás de cada um desses capítulos, os quais também são brevemente descritos ao longo desta introdução.

A literatura empírica sobre a relação entre salários e o grau de segregação feminina nos estabelecimentos mostra que tanto o salário das mulheres quanto o dos homens são negativamente relacionados com a proporção de mulheres nos estabelecimentos. Esses resultados foram obtidos para diferentes países, períodos de tempo, ocupações e setores de atividade, indicando a existência de uma regularidade (estatística) entre trabalhar em estabelecimentos predominantemente femininos e os salários dos trabalhadores.

Embora esses resultados sejam consistentes entre os estudos dessa literatura, eles podem sofrer de uma importante limitação. Especificamente, todos os resultados disponíveis foram obtidos com dados de *cross-section*, o que implica que as estimativas da relação entre salários e segregação feminina nos estabelecimentos não controlaram para as características não-observáveis dos trabalhadores (por exemplo, habilidade inata e preferências) e dos estabelecimentos (por exemplo, eficiência gerencial e grau de discriminação). Assim, como é possível que os salários dos trabalhadores e a proporção de mulheres nos estabelecimentos sejam correlacionados com essas características não-observáveis, as estimativas disponíveis na literatura podem ser enviesadas.

O principal objetivo do primeiro capítulo da presente tese é checar se os resultados de *cross-section* são mantidos quando dados de painel são utilizados para estimar o efeito da segregação feminina no nível do estabelecimento sobre os salários dos trabalhadores. Para tanto, empregamos vários modelos longitudinais, os quais são capazes de incorporar a presença de efeitos específicos não-observáveis de trabalhadores e estabelecimentos. Esses modelos são estimados com o uso de uma extensa base de dados que combina informações de trabalhadores e estabelecimentos no Brasil. Especificamente, utilizamos os

micro-dados da Relação Anual de Informações Sociais (RAIS), a qual contém dados administrativos reunidos pelo governo federal a partir de informações anuais fornecidas pelos estabelecimentos registrados no país. Essa base de dados pode ser considerada rica, pois possui informações sobre diversas características tanto dos trabalhadores (por exemplo, sexo, idade, escolaridade e ocupação) quanto dos estabelecimentos (por exemplo, localização e setor de atividade). A presença de identificadores únicos para os trabalhadores e os estabelecimentos permitiu-nos construir um painel de dados entre os anos de 2000 e 2002.

Em linhas gerais, nossas estimativas longitudinais mostram que o efeito da segregação feminina no nível do estabelecimento é negativo sobre os salários de ambos os sexos. Esse resultado está em linha com o que foi previamente encontrado na literatura de *cross-section*.¹ Entretanto, nossas estimativas longitudinais são menos negativas (isto é, menores em termos absolutos) do que as que obtivemos com base em dados de *cross-section*. Essa diferença de magnitude é um resultado significativo, pois demonstra que controlar para fatores específicos não-observáveis dos trabalhadores e estabelecimentos realmente importa para a estimação do efeito de interesse. Ademais, o fato de que as estimativas longitudinais são menos negativas que as de *cross-section* indica que os efeitos específicos não-observáveis são negativamente correlacionados com o grau de segregação feminina dos estabelecimentos. Uma possível explicação para isso é que os empregadores com maior produtividade não-observada e/ou mais elevado grau de discriminação contra as mulheres tendem a contratar relativamente mais homens. Uma explicação alternativa é que os trabalhadores de ambos os sexos com menores níveis de atributos produtivos não-observáveis (*e.g.* habilidade inata) acabam alocados em estabelecimentos predominantemente femininos.

Uma das maiores contribuições da área de econometria para o conhecimento científico em geral é o que se costuma denominar *problema de identificação* (Heckman 2000). Um problema de identificação nasce porque existem distintos modelos econométricos que são compatíveis com os mesmos dados observados. O fato de que diferentes modelos econométricos são capazes de gerar os mesmos dados leva a que os objetos de interesse (*e.g.* o efeito de uma covariada sobre a variável de resposta) não possam ser identificados sem a imposição de restrições sobre a classe de modelos em consideração.²

Diversos tipos de restrições têm sido empregadas no estudo de identificação. Exemplos típicos são restrições sobre a distribuição dos termos de erro não-observáveis (*e.g.* normalidade e independência condicional), restrições de exclusão sobre as covariadas (*e.g.* a

¹No primeiro capítulo, desenvolvemos dois modelos teóricos simples que procuram explicar porque esse efeito é negativo sobre os salários tanto de homens quanto de mulheres.

²Os fundamentos da área de identificação em econometria foram propostos pelos pesquisadores que formaram a chamada *Cowles Commission*. Um dos maiores legados deixados por esses pesquisadores é a clarificação de que a identificação dos objetos de interesse é contingente ao conjunto de restrições utilizadas. Ver Christ (1994) para um relato da história das principais contribuições dessa Comissão para o campo de econometria.

existência de variáveis instrumentais) e a imposição de formas funcionais específicas sobre as equações que compõem o modelo (*e.g.* parametrizações).

Este último tipo de restrição é certamente uma das mais adotadas nos modelos econométricos. No entanto, como o seu uso não é normalmente baseado em conhecimento prévio sobre as formas funcionais do modelo, a imposição desse tipo de restrição pode levar a conclusões enganosas sobre a identificação dos efeitos de interesse. Por exemplo, se o modelo gerador dos dados é não-paramétrico, a imposição de parametrizações tenderá a produzir julgamentos equivocados sobre a identificação dos verdadeiros efeitos de interesse. Esse problema torna-se mais grave na medida em que a teoria econômica tipicamente não fornece informações sobre as formas funcionais exatas do processo gerador dos dados. Assim, como enfatizado por diversos autores (*e.g.* Roehrig (1988) e Chesher (2007)), a menos que haja um conhecimento prévio sobre as formas funcionais das equações do modelo, a análise de identificação deveria ser implementada com base em arcabouços flexíveis, entre os quais o não-paramétrico é sem dúvida o mais geral.³

O segundo e o terceiro capítulos da tese utilizam um arcabouço não-paramétrico para estudar a identificação de certos efeitos de interesse. O segundo capítulo é dedicado a estudar a identificação de dois efeitos que se relacionam com a área de avaliação de programas sociais. O terceiro capítulo procura generalizar os resultados de identificação do segundo capítulo para outras áreas de aplicação. A seguir, descrevemos brevemente o conteúdo desses dois capítulos.

A vasta literatura empírica da área de avaliação de programas tem mostrado que os impactos médios de programas sociais variam ao longo do tempo. Muitos analistas tendem a considerar essas variações como os verdadeiros efeitos temporais dos programas. Entretanto, essa constatação não é necessariamente correta, uma vez que podem existir outros fatores que influenciam as trajetórias temporais dos impactos médios observados. Em particular, um fator importante é o ambiente econômico no qual estão inseridos os tratados (e não-tratados) pela intervenção. Por exemplo, tomando o caso de um programa de treinamento de mão-de-obra, é de se esperar que as mudanças na demanda (relativa) por trabalhadores treinados afetem a evolução temporal do impacto médio do programa sobre os salários (relativos) dos tratados. Portanto, em princípio, os impactos médios que observamos podem conter tanto o efeito do programa em si, quanto o efeito das condições econômicas externas ao programa. A fim de distinguir as variações temporais no verdadeiro impacto do programa das influências das condições econômicas, definimos dois efeitos que denominamos respectivamente de *efeito interno temporal* e *efeito econômico externo*.

O principal objetivo do segundo capítulo é estudar que aspectos desses dois efeitos podem ser identificados não-parametricamente. Como discutido acima, inserimos essa análise

³Observe-se, no entanto, que o uso da abordagem não-paramétrica para o estudo de identificação não significa que formas funcionais específicas não podem ser empregadas para propósitos de estimação. De fato, tendo assegurado a identificação não-paramétrica dos objetos de interesse, o analista pode optar por utilizar formas funcionais menos flexíveis na fase de estimação.

num arcabouço não-paramétrico, pois não dispomos de conhecimento prévio sobre como os impactos médios dos programas sociais são determinados. Nossos resultados mostram que somente o efeito interno temporal é identificado de forma totalmente não-paramétrica. A razão que impede o efeito econômico externo de ser integralmente identificado de forma não-paramétrica é que as condições econômicas não podem ser variadas independentemente dos demais argumentos que determinam o impacto médio do programa.⁴

A análise de identificação dos dois efeitos de interesse é conduzida para dois tipos de dados que um analista costuma dispor. No primeiro, existem informações para diferentes regiões de um país sobre uma coorte de participantes do programa. No segundo, há disponibilidade de dados para múltiplas coortes de participantes em uma única região do país (ou país inteiro). Em ambos os casos, assumimos que o analista conhece (ou foi capaz de estimar consistentemente) o impacto médio do programa para vários períodos de tempo. Assumimos também que a escala do programa é pequena relativamente ao tamanho da economia. Esta última hipótese permite ignorar os efeitos de *feedback* do próprio programa sobre a economia.

O segundo capítulo também contém um exemplo empírico que procura ilustrar os resultados de identificação nele obtidos. Especificamente, utilizamos dados experimentais que foram coletados para avaliar a efetividade de um programa público de treinamento no Brasil (Plano Nacional de Qualificação do Trabalhador - PLANFOR). Como os dados se referem a uma coorte de participantes em duas regiões metropolitanas do país (Rio de Janeiro e Fortaleza), o exercício exemplifica os resultados de identificação obtidos com o primeiro tipo de dados mencionado anteriormente. Utilizamos também informações de uma outra fonte (Cadastro Geral de Empregados e Desempregados - CAGED) para calcular uma medida das condições econômicas nessas duas regiões. As estimativas dos dois efeitos de interesse indicam que o efeito interno temporal interage com essa medida das condições econômicas, e que o efeito econômico externo varia para diferentes períodos de tempo.

Sabe-se que os coeficientes de uma regressão linear, paramétrica não são identificados se houver perfeita multi-colinearidade entre as covariadas. Num contexto não-paramétrico, esse tipo de problema aparece quando há restrições funcionais entre as covariadas, ou seja, restrições (não necessariamente lineares) que impedem que o valor de algumas covariadas seja variado de forma independente do das demais. Em outras palavras, enquanto no primeiro caso a presença de dependência linear obstrui a identificação do efeito de variações isoladas nas covariadas do modelo, no segundo é a existência de *dependência funcional* entre as covariadas que impossibilita identificar esse efeito. A impossibilidade de variar as medidas de condições econômicas das demais covariadas que determinam o impacto médio dos programas é um exemplo desse tipo de dependência funcional.

No terceiro capítulo da tese, procuramos estudar a identificação do efeito de variar

⁴As implicações desse tipo de problema num contexto não-paramétrico geral são o objeto do terceiro capítulo da tese.

o valor de covariadas que são funcionalmente dependentes num modelo não-paramétrico geral. Estudamos também a identificação do efeito de variações nas covariadas que não são funcionalmente dependentes no modelo. Como no segundo capítulo, o principal objetivo é investigar quais aspectos desses efeitos podem ser identificados de forma não-paramétrica. As condições que propomos são suficientes para identificar o segundo tipo de efeito de forma totalmente não-paramétrica. Todavia, somente uma parte do primeiro efeito é identificado não-parametricamente. O motivo principal para essa identificação parcialmente não-paramétrica é a imposição de uma restrição de separabilidade aditiva. No entanto, essa restrição requer que apenas *uma* covariada apareça de forma aditiva no modelo. Essa característica pode ser vista como uma vantagem, já que um modelo tipicamente aditivo (*e.g.* totalmente parametrizado) não permite explorar as interações entre as demais covariadas de maneira tão flexível.

Após o terceiro capítulo, apresentamos as principais conclusões da tese e algumas possibilidades de pesquisa futura.

Capítulo 1

Os Efeitos da Segregação por Gênero sobre Salários no Nível do Estabelecimento: Uma Análise Empírica Usando um Painel de Dados Combinados de Empregadores e Empregados

1.1 Introdução

A existência de diferenciais salariais por gênero é um fenômeno que tem sido documentado para diferentes países e períodos de tempo. Uma das principais regularidades empíricas encontradas nessa literatura é a existência de uma relação entre os salários de homens e mulheres e a composição por gênero de ocupações, setores e firmas. A investigação sobre essa relação tem sido importante não só para aperfeiçoar o nosso entendimento sobre os canais por meio dos quais os diferenciais de salários por gênero são gerados, mas também para analisar a efetividade de diversas políticas públicas que visam a reduzir o hiato salarial entre os sexos.

A maior parte dessa literatura tem concentrado atenção na relação entre salários e composição por gênero nos níveis ocupacional e setorial.¹ Um dos resultados encontrados nesses estudos é que os salários dos trabalhadores variam sistematicamente com a composição por gênero das ocupações e dos setores. Mais especificamente, a evidência empírica mostra que a proporção de mulheres nas ocupações e nos setores é negativamente relacionada aos salários de homens e mulheres.

¹Entre os vários estudos internacionais, ver, por exemplo: Bergmann (1974), Polachek (1979), Treiman & Hartmann (1981), Blau (1984), Johnson & Solon (1986), Sorensen (1989), Filer (1989), Killingsworth (1990), Blau & Ferber (1992), England (1992), Macpherson & Hirsch (1995), Fields & Wolff (1995). Para estudos brasileiros, ver: Barros et al. (1997), Ometto et al. (1997), Ometto et al. (1999) e Oliveira (2001).

Menos atenção tem sido dada à conexão entre salários e segregação de mulheres no nível da firma.² Os primeiros trabalhos que utilizaram informações sobre empregadores e empregados (McNulty 1967, Buckley 1971, Blau 1977) encontraram que as mulheres são segregadas em firmas que pagam menores salários, um resultado que tem sido constantemente obtido em estudos mais recentes (Groschen 1991, Carrington & Troske 1995, 1998, Bayard et al. 1999, Reilly & Wirjanto 1999a, Vieira et al. 2003). Tal como no caso de ocupações e setores, as evidências desse conjunto de estudos mostram que a relação entre os salários de homens e mulheres e a proporção de mulheres na firma é negativa.³

Apesar desse resultado ter sido repetidamente encontrado pelos estudos dessa literatura, ele pode sofrer de uma importante limitação. Especificamente, todos os estudos disponíveis foram baseados em dados de *cross-section*, o que implica que as estimativas da relação entre salários e segregação feminina nas firmas não controlaram para as características não-observáveis dos trabalhadores (*e.g.* habilidades produtivas e preferências) e dos empregadores (*e.g.* grau de discriminação e eficiência gerencial). Isso pode representar um problema, já que, se os salários dos trabalhadores e composição por gênero das firmas forem correlacionados com essas características não-observadas, as estimativas até agora obtidas tenderão a ser enviesadas. Em outras palavras, sem a inclusão desses fatores não-observáveis no modelo, não é possível rejeitar a hipótese de que a proporção de mulheres nas firmas funcione como uma *proxy* para diferenças de características não-observadas entre trabalhadores e empregadores que sejam correlacionados com os salários.

O principal objetivo deste capítulo é estimar a relação de interesse utilizando dados de painel e métodos longitudinais. Distintamente das bases de *cross-section*, os dados de painel permitem a inclusão de fatores não-observáveis (invariantes no tempo) de trabalhadores e empregadores, o que torna possível estimar o efeito da segregação feminina no nível do estabelecimento sobre salários, controlando para os efeitos desses fatores específicos não-mensurados. Adicionalmente, o uso de métodos longitudinais permite estimar modelos econométricos mais flexíveis, que admitem estruturas mais gerais para o termo de erro da regressão (*e.g.* modelos auto-regressivos). Tanto quanto sabemos, este é o primeiro estudo que investiga a conexão entre salários e segregação de mulheres no nível do estabelecimento utilizando dados de painel com informações combinadas de empregadores e empregados.⁴

²O principal motivo para essa menor atenção é a relativa carência de micro-dados públicos com informações combinadas sobre empregadores e empregados. De fato, somente com a recente disponibilização de bases de dados desse tipo é que começaram a aparecer mais estudos utilizando informações no nível da firma, incluindo os que focam na questão dos diferenciais de salários por gênero. Para uma resenha sobre a disponibilidade e o uso econométrico dessas bases de dados, ver, por exemplo, Abowd & Kramarz (1999).

³Não existem estudos que investigaram essa relação ao nível do empregador para o Brasil.

⁴Abowd et al. (1999) usam dados de painel com informações combinadas de empregadores e empregados para estimar os efeitos de variáveis de capital humano (*e.g.* escolaridade, experiência geral e específica dos trabalhadores) sobre os salários de homens e mulheres. No entanto, esses autores não incluem medidas de segregação de mulheres no nível do empregador, que é o nosso foco aqui.

O painel de dados que utilizamos é baseado em arquivos administrativos mantidos pelo governo federal brasileiro, a saber: a Relação Anual de Informações Sociais - RAIS. Trata-se de uma base de dados bastante rica, uma vez que contém uma série de informações relevantes sobre empregadores (*e.g.* localização e setor de atividade) e trabalhadores (*e.g.* salário, sexo, escolaridade e experiência no estabelecimento). Ademais, o número de observações disponíveis é bastante elevado, o que torna possível obter estimativas mais precisas dos parâmetros de interesse. Metodologicamente, nossa abordagem mais geral é a que se costuma denominar de modelo de painel dinâmico. Implementamos também vários outros métodos longitudinais para checar a robustez do modelo dinâmico. Antecipando os resultados, nossas estimativas longitudinais indicam que o efeito da segregação feminina nos estabelecimentos sobre os salários de homens e mulheres é negativo. No entanto, a comparação com nossos resultados de *cross-section* mostra uma redução desse efeito (isto é, ele se torna menos negativo) para ambos os grupos de gênero, o que indica que as estimativas de *cross-section* são enviesadas para baixo.

Existem diversas teorias que procurar explicar a relação entre segregação por sexo no nível da firma e os salários de homens e mulheres. Por exemplo, [Becker \(1971\)](#) e [Arrow \(1973a,b\)](#) predizem que a existência de discriminação por parte do empregador, co-trabalhador ou cliente pode levar a segregação por gênero entre firmas e a aparição de um hiato salarial entre homens e mulheres. Uma outra teoria é baseada nos modelos de discriminação estatística, que assumem que os empregadores têm informação imperfeita acerca da verdadeira produtividade dos indivíduos, e acreditam que, em média, as mulheres são menos produtivas que os homens (*e.g.* [Arrow \(1973b\)](#)). Existe também uma outra teoria – usualmente denominada de *quality sorting* – que argumenta que algum mecanismo de natureza histórica, social ou econômica levou a que a composição por gênero das firmas funcione como um indicador de qualidade para a contratação de trabalhadores ([Macpherson & Hirsch \(1995\)](#)).⁵ [Reilly & Wirjanto \(1999b\)](#) propõem um outro argumento baseado em diferenciais compensatórios de salários – denominado pelos autores de “coincidência de necessidades” entre empregadores e empregados –, que seria capaz de gerar a existência de segregação por sexo e hiatos salariais entre homens e mulheres.

Embora não busquemos evidências para testar qualquer dessas teorias, desenvolvemos dois modelos simples cujas predições são compatíveis com nossos resultados empíricos e as evidências disponíveis sobre a relação entre segregação feminina no nível do estabelecimento e os salários de homens e mulheres. No primeiro modelo, inserimos o modelo de Becker/Arrow de discriminação pelo empregador num arcabouço monopsonístico, a fim de permitir o aparecimento de salários heterogêneos entre as firmas.⁶ No segundo modelo,

⁵Por exemplo, se as barreiras discriminatórias (passadas ou presentes) produzem a aglomeração de mulheres em firmas com menores salários, então a composição por gênero pode passar a ser vista como um sinalizador da produtividade da força de trabalho das firmas.

⁶O modelo de Becker/Arrow é competitivo, levando a que todas as firmas paguem o mesmo salário em equilíbrio.

combinamos algumas idéias das teorias de discriminação estatística e de *quality sorting* para construir um arcabouço que, acreditamos, é a primeira tentativa de colocar formalmente essas duas teorias juntas.

A análise deste capítulo também é motivada por questões de política pública. Várias iniciativas que objetivam reduzir o hiato salarial entre homens e mulheres são baseadas no argumento de que os salários pagos a trabalhadores em postos de trabalho de valor similar (*e.g.* mesmos requisitos de qualificação, responsabilidade e condições de trabalho) deveriam ser iguais. Por exemplo, o principal objetivo da política de “valor comparável” (*comparable worth*) é equalizar os salários de ocupações tipicamente femininas aos de ocupações tipicamente masculinas consideradas de valor similar. A implementação desse tipo de política é normalmente feita no âmbito de cada firma, implicando que as disparidades salariais *intra* firma são afetadas pela política, ao passo que as discrepâncias existentes *entre* firmas não o são.⁷ Um outro tipo de política busca garantir a igualdade de oportunidades de emprego às mulheres no mercado de trabalho. Essas iniciativas procuram lidar diretamente com a existência de segregação de mulheres no nível da firma e, nesse sentido, difere bastante da política de “valor comparável”.

Esse capítulo está estruturado da seguinte maneira. Na seção 2, apresentamos uma breve revisão da literatura empírica que tem investigado a relação entre salários e segregação feminina no nível do estabelecimento. A seção 3 é dedicada a rever as teorias que procuram explicar essa relação. Nessa seção, apresentamos também os dois modelos estilizados a que fizemos referência acima. A seção 4 descreve a base de dados que utilizamos na análise empírica e apresenta estatísticas descritivas sobre as principais variáveis dessa base. Na seção 5, medimos o grau de segregação por gênero entre estabelecimentos. Como parte dessa segregação pode ser atribuída a uma alocação aleatória de indivíduos entre estabelecimentos, utilizamos a metodologia proposta por Carrington & Troske (1997) para separar o que esses autores denominam de segregação sistemática e aleatória. A seção 6 descreve a metodologia empírica empregada para estimar o efeito de interesse, com os resultados obtidos sendo apresentados na seção 7. A seção 8 contém as metodologias e os resultados dos diversos testes de robustez que implementamos. A seção 9 é uma seção de discussão na qual sumariamos nossos resultados, comparamo-os com as estimativas de cross-section e implementamos decomposições do diferencial de salários por gênero, as quais são utilizadas para discutir algumas implicações de política. As conclusões são apresentadas na seção 10.

⁷Nesse sentido, a efetividade dessa política para combater o hiato salarial será reduzida se as mulheres estiverem concentradas em firmas que pagam menores salários. Esse ponto é enfatizado por Johnson & Solon (1986).

1.2 Literatura Empírica

Nesta seção, resumiremos os resultados dos estudos empíricos que investigaram a relação entre salários e segregação feminina no nível do estabelecimento. A Tabela 1.1 apresenta algumas das principais características desses estudos, assim como as estimativas dos parâmetros que medem a relação de interesse. Nessa Tabela constam somente os trabalhos que incluem informações no nível do estabelecimento/firma.

McNulty (1967) e Buckley (1971) foram os primeiros a indicar que a segregação feminina entre firmas representava um fator importante relacionado ao hiato salarial por gênero. De fato, utilizando dados não-publicados, esses autores mostram que as mulheres tendem a trabalhar em firmas que pagam menores salários, com o oposto ocorrendo com os homens. Por exemplo, McNulty (1967) encontrou que as firmas que empregavam somente guardas do sexo masculino pagavam salários 37% maiores do que aquelas que só contratavam guardas do sexo feminino. Distintamente, os salários dos homens eram somente 18% superiores aos das mulheres em firmas que empregavam guardas de ambos os sexos. Similarmente, Buckley (1971) encontrou que a média (não-ponderada) das razões de salário homem-mulher entre diferentes ocupações era de 1,22 para firmas totalmente segregadas, sendo igual a 1,11 para firmas integradas. Blau (1977) generalizou esses resultados computando a correlação entre os *rankings* de salários (médios) e de proporções de mulheres das firmas. Esses *rankings* foram calculados com base nos coeficientes estimados das *dummies* de firma obtidos de duas regressões separadas: uma na qual a variável dependente era o salário dos trabalhadores e outra em que a variável dependente era a proporção de mulheres na firma. A correlação estimada foi negativa para as três áreas metropolitanas usadas no estudo, evidenciando que quanto maior a proporção de mulheres nos quadros das firmas, menor o salário (médio) recebido pelos trabalhadores.

Groshen (1991) estudou cinco setores específicos e encontrou uma relação negativa entre os salários e a segregação feminina no nível do estabelecimento para todos os setores. Para alguns setores estudados, essa autora também encontrou que a segregação de mulheres entre estabelecimentos representava um dos componentes mais importantes do diferencial de salários por gênero. Para outros setores, entretanto, a segregação feminina entre ocupações desempenhou o papel mais importante. A base de dados utilizada nesse estudo não dispõe de informações sobre características produtivas dos trabalhadores e estabelecimentos, o que representa uma limitação, uma vez que esse tipo de controle é considerado importante em regressões salariais que envolvem gênero.

Tabela 1.1: Características e Resultados dos Estudos sobre a Relação entre Salários e Segregação por Gênero no Estabelecimento

Estudo (1)	Base de dados/País/Período (2)	Método de Estimação (3)	Variável Dependente (4)	Covariadas (5)	Coefficiente da Proporção de Mulheres no Estabelecimento (6)
1.McNulty (1967)	Occupational Earnings Survey/EUA/1966	Comparação das razões entre médias salariais por gênero entre estabelecimentos integrados e completamente segregados por ocupação	Salário horário	NA	Guardas em plantas integradas: Razão: 1.18 Guardas em plantas segregadas: Razão: 1.37 Comparação: 1.37/1.18=1.16
2.Buckley (1971)	Occupational Earnings Survey/EUA/1970	Comparação da média das razões salariais homem-mulher entre ocupações para estabelecimentos integrados e completamente segregados	Salário horário	NA	Média não-ponderada em plantas integradas: Razão: 1.11 Média não-ponderada em plantas segregadas: Razão: 1.22 Comparação: 1.22/1.11=1.10
3.Blau(1977, capítulo 5)	Area Wage Surveys/EUA/1970	Correlação entre coeficientes das dummies de firma de regressões de salário e proporção de mulheres	NA	Regressão de salário: dummy para mulher, ocupação e dummies de firma Regressão de proporção de mulheres: ocupação e dummies de firma	Boston: corr = -0.35 New York: corr = -0.27 Philadelphia: corr = -0.26
4.Groshen (1991)	Industry Wage Surveys/EUA: Plásticos (1974), Máquinas (1983), Seguros (1980), Bancário(1980), Computação (1983)	Regressão MQO empilhada	Log salário horário	Dummy para mulher, proporção de mulheres na ocupação, estabelecimento e estabelecimento-ocupação (job-cell)	Plásticos: -0.099 Máquinas: -0.330 Seguros: -0.256 Bancário: -0.385 Computação: -0.339
5.Carrington & Troske (1995)	Characteristics of Business Owners Survey/EUA/1982	Regressão MQO empilhada no nível da firma	Log média salarial dos empregados	Tamanho da firma, escolaridade média e sexo proprietário, dummies de setor, idade da firma e proporção de mulheres na firma (variável categórica)	Prop. mulheres firma: 0%:0.19, 1-9%:0.14, 10-24%:0.36, 25-49%:0.26, 50-74%:0.09, 75-100%: categoria excluída
6.Carrington & Troske (1998)	Worker-Establishment Characteristics Database/EUA/1990	Regressão MQO empilhada	Log salário horário	Dummy para mulher, experiência, escolaridade, raça, dummies de ocupação, tamanho do estabelecimento, proporção de mulheres no estabelecimento, interação dummy p/ mulher e proporção de mulheres no estabelecimento	Homens: -0.214 Mulheres: -0.305
7.Bayard, Hellerstein, Neumark, & Troske (1999)	New Worker-Establishment Characteristics Database/EUA/1990	Regressão MQO empilhada Regressão MQO por sexo	Log salário horário	Dummy para mulher, proporção de mulheres na ocupação, estabelecimento, estabelecimento-ocupação (job cell), variáveis de capital humano	Empilhada: -0.162 Homens: -0.127 Mulheres: -0.157
8.Reilly & Wirjanto (1999)	General Segmentation Survey/Canadá/1979	Regressão MLG empilhada Regressão MLG por sexo	Log salário horário	Escolaridade, experiência, tempo na firma, densidade sindical, dummies de ocupação setor, região e estado civil, proporção de mulheres no estabelecimento	Empilhada: -0.439 Homens: -0.223 Mulheres: -0.293
9.Vieira, Cardoso, & Portela (2003)	Base do Ministério da Segurança Social e Trabalho/Portugal/1985-1995	Regressão OLS por sexo	Log salário horário	Escolaridade, experiência, tempo na firma dummies de ocupação, setor e região, tamanho do estabelecimento, proporção de mulheres no estabelecimento	Homens (1985): 0.0296 Homens (1995): 0.0344 Mulheres (1985): -0.0393 Mulheres (1995): -0.0283

Notas: Todos os resultados são baseados em dados de cross-section. NA = Não Aplicável; MQO = Mínimos Quadrados Ordinários; MLG = Modelo Linear Generalizado.

Utilizando uma amostra de pequenas empresas, Carrington & Troske (1995) estimam uma regressão no nível da firma, na qual a folha salarial por empregado é regredida contra um conjunto de controles para firmas e uma variável categórica medindo a proporção de mulheres nas firmas. Os resultados por eles encontrados mostram que as firmas que possuem uma força de trabalho predominantemente masculina tendem a pagar maiores salários médios do que as firmas predominantemente femininas. Carrington & Troske (1998) utilizam uma amostra de estabelecimentos do setor industrial, que é combinada com informações do censo demográfico americano de 2000 para construir uma base de dados de empregadores e trabalhadores. Incluindo variáveis para características dos estabelecimentos e indivíduos, esses autores regredem os salários dos trabalhadores contra uma *dummy* para mulheres, uma variável para a proporção de mulheres no estabelecimento e uma interação entre essas duas variáveis. Esse termo de interação permite medir separadamente o efeito da segregação feminina no nível do estabelecimento sobre os salários dos homens e das mulheres. Os resultados encontrados mostram que esse efeito é negativo para ambos os sexos, sendo mais intenso (isto é, mais negativo) para as mulheres.

Bayard et al. (1999) usam uma base de dados similar a de Carrington & Troske (1998). Diferentemente de Groshen (1991), os autores incluem uma série de controles para características produtivas dos trabalhadores e estabelecimentos, porém também encontram que a segregação de mulheres entre estabelecimentos é um dos fatores mais importantes para explicar o diferencial de salários por gênero. Das regressões estimadas por sexo, o coeficiente da variável proporção de mulheres no estabelecimento é negativo para ambos os sexos, sendo mais negativo para o caso das mulheres.

Reilly & Wirjanto (1999a) utilizam uma pequena amostra com informações combinadas de empregadores e empregados do Canadá para estimar uma equação de salários com regressores que incluem um conjunto de características desses grupos e a variável proporção de mulheres nos estabelecimentos. Os resultados confirmam para o Canadá o que foi encontrado para o caso americano: o efeito da segregação feminina entre estabelecimentos é negativo sobre os salários de homens e de mulheres, sendo mais intenso para este último grupo.

Finalmente, Vieira et al. (2003) fazem uso de uma base de dados de empregadores e empregados para Portugal, a qual é semelhante em muitos aspectos à base que utilizamos neste capítulo. Entretanto, embora os autores disponham de informações para vários anos, a base de dados por eles utilizada não é longitudinal, implicando que a análise empírica implementada no estudo é baseada em regressões de *cross-section*. Incluindo controles para características dos trabalhadores e estabelecimentos nas regressões, os autores encontram que o efeito da segregação feminina no nível do estabelecimento é negativo sobre os salários das mulheres porém positivo no caso dos homens. Esse é o único estudo no qual se encontrou um impacto positivo sobre salários da segregação feminina entre estabelecimentos.

Em suma, a análise da literatura empírica existente sobre o tema sugere que o efeito de

trabalhar em estabelecimentos predominantemente femininos é negativo sobre os salários de homens e de mulheres, sendo esse efeito mais negativo para este último grupo.

É importante enfatizar que todos os estudos disponíveis utilizaram dados de *cross-section*. Uma principais motivações deste capítulo é checar se o efeito negativo da segregação feminina sobre salários é confirmado quando se empregam dados de painel e métodos longitudinais. Na seção 1.9.2, comparamos nossos resultados baseados nas abordagens de *cross-section* e longitudinal.

1.3 Literatura Teórica

Nesta seção, apresentamos uma breve revisão da literatura teórica que trata dos diferenciais de salário por gênero. Naturalmente, dado o nosso objeto de estudo, procuramos dar maior ênfase àquelas teorias que, de alguma forma, estão centradas no nível da firma. Também incluímos nesta seção, os dois modelos teóricos que desenvolvemos para tentar explicar a relação entre salários e a composição por gênero das firmas. A sub-seção 1.3.1 está dedicada à revisão das teorias gerais, e as sub-seções 1.3.2 e 1.3.3 contêm os nossos dois modelos.

1.3.1 Teorias Gerais

Em linhas gerais, existem dois grupos de teorias que procuram explicar a existência de diferenças de resultados entre homens e mulheres no mercado de trabalho: a teoria do capital humano e as teorias de discriminação. A primeira é essencialmente baseada no lado oferta, enfatizando as escolhas (voluntárias) dos indivíduos em relação tanto à quantidade quanto ao tipo de investimento por eles realizados em capital humano. Essa teoria provê uma explicação consistente de o porquê homens e mulheres acabam tendo diferentes características produtivas e ocupações. Já as teorias de discriminação são mais baseadas no lado da demanda, pois enfatizam os efeitos que a discriminação pelos empregadores, co-trabalhadores e clientes podem ter sobre a geração de diferenças salariais e de emprego entre indivíduos igualmente produtivos.⁸

O modelo de capital humano (Mincer & Polachek 1974) parte da hipótese de que existe uma divisão tradicional de trabalho entre os sexos dentro da família. Essa divisão faz com que as mulheres tenham uma participação mais curta e intermitente no mercado de trabalho do que os homens, o que leva a que, *ceteris paribus*, os investimentos em educação e em treinamento geral e específico sejam relativamente mais baixos para o primeiro grupo. Assim, dados os seus vínculos mais curtos e descontínuos com o mercado de trabalho, a resposta ótima esperada das mulheres é adquirir menos capital humano educacional e em experiência no mercado de trabalho. A implicação disso é que parte do hiato salarial

⁸Para uma resenha detalhada das teorias de discriminação no mercado de trabalho, ver Cain (1986). Para uma resenha geral das teorias sobre diferenças de resultados por gênero no mercado de trabalho, ver Altonji & Blank (1999).

observado entre homens e mulheres se deve às diferenças de atributos produtivos entre os sexos relativas ao acúmulo de capital humano.

A teoria do capital humano também oferece uma explicação para as escolhas ocupacionais (Polachek 1979, 1981). De acordo com esse modelo, as mulheres tendem a escolher ocupações que requerem menos investimento em treinamento no posto de trabalho (*on-the-job training*) e menores taxas de depreciação durante o período em que elas saem da força de trabalho. Nesse sentido, dever-se-ia esperar que as ocupações predominantemente femininas possuam curvas de salário-idade mais horizontais.

Proximamente relacionados à teoria de capital humano estão os argumentos baseados na existência de diferenças de preferências e vantagens comparativas entre homens e mulheres. As diferenças de preferências entre os indivíduos (no que tange ao mercado de trabalho) podem se relacionar a uma diversidade de aspectos: trabalho *versus* lazer, trabalhar no mercado *versus* trabalho doméstico, preferências por certas características do posto de trabalho (*e.g.* horário flexível *versus* rígido) etc.⁹ Argumenta-se também que as diferenças biológicas entre os sexos (*e.g.* força física) criam vantagens comparativas que induzem as mulheres a escolher certos tipos de ocupações e postos de trabalho. Segundo essa argumentação, a distribuição de vantagens comparativas e preferências por certas atividades e o valor para os empregadores de oferecer postos de trabalho com características distintas influenciarão a distribuição final de emprego e salário das ocupações, setores e firmas para homens e mulheres. Esses argumentos encontram-se no interior da teoria dos diferenciais compensatórios de salários, segundo a qual empregadores e trabalhadores estabelecem trocas entre si sobre os aspectos pecuniários e não-pecuniários dos postos de trabalho.

Os argumentos baseados em diferenciais salariais compensatórios foram utilizados por Reilly & Wirjanto (1999a,b) para explicar a correlação negativa observada entre os salários e o grau de feminização das firmas. Especificamente, os autores sustentam que pode existir uma “coincidência de necessidades” entre empregadores e empregados que é vantajosa para ambas as partes: os primeiros utilizam tecnologias diferentes e oferecem distintos contratos de trabalho aos segundos, nos quais salários mais baixos são trocados por características tipicamente valorizadas por mulheres (condições de trabalho, horários mais flexíveis, cre-

⁹Uma das fontes mais citadas para o aparecimento de diferenças de preferências entre os sexos é a existência de discriminação social, isto é, um tipo de discriminação que envolve um conjunto de fatores presentes no funcionamento das sociedades que acabam por determinar os papéis que as mulheres devem desempenhar. Exemplos desses papéis são o cuidado com os filhos e a manutenção dos lares. Entre os fatores sociais que ajudam a moldar as preferências de homens e mulheres no mercado de trabalho está a influência dos pais em direcionar as escolhas das crianças para certas ocupações tradicionalmente vistas como femininas ou masculinas. Um outro fator são os estereótipos sociais. Por exemplo, acredita-se que as mulheres são melhores em tarefas que envolvem cuidados especiais (enfermeira, assistente social), experiência com atividades domiciliares (arrumadeira, faxineira), destreza manual (costureira, datilógrafa) e aparência física (receptionista, vendedora). Ver Anker (1997, Tabela 1) para uma descrição detalhada da formação de estereótipos entre os sexos que se relacionam com o mercado de trabalho.

ches, entre outras). De acordo com essa visão, a alocação das mulheres (e homens) entre as firmas e o diferencial (compensatório) de salário por gênero a ela associada refletiriam uma troca mutuamente vantajosa para empregadores e trabalhadores que ocorre em equilíbrio. Nesse sentido, políticas públicas que tentam equalizar os salários entre os sexos tenderiam a reduzir o nível de bem-estar de empregadores e trabalhadores.

As teorias que enfatizam as diferenças entre os sexos em termos de preferências, vantagens comparativas e acumulação de capital humano implicitamente interagem com as teorias de discriminação que discutiremos a seguir. Por exemplo, se as mulheres percebem que elas são discriminadas no mercado de trabalho, então, é provável que isso influencie suas decisões de investimento em capital humano. Esse ponto tem implicações para as análises empíricas sobre o hiato salarial por gênero, pois torna mais difícil separar os fatores (exógenos) que produzem diferenças de características produtivas entre os sexos dos efeitos da discriminação no mercado de trabalho. Nesse sentido, as análises que tomam variáveis tais como educação e experiência como verdadeiramente exógenas tendem a sub-estimar os impactos da discriminação contra a mulher no mercado de trabalho.

Existem basicamente duas classes de modelos de discriminação no mercado de trabalho. A primeira é baseada na idéia de que os indivíduos do grupo majoritário (homens, brancos) possuem algum grau de insatisfação em interagir economicamente com indivíduos do grupo minoritário (mulheres, negros).¹⁰ A segunda, denominada de discriminação estatística, toma como base a existência de informação imperfeita sobre as habilidades produtivas dos indivíduos e o uso do gênero por parte dos empregadores como instrumento para discriminar os trabalhadores considerados de menor ou maior qualidade.

As fundações da teoria de discriminação no mercado de trabalho foram propostas por Becker (1971). Existem três classes de modelos nessa teoria, as quais se diferenciam entre si dependendo de se as preferências discriminatórias se originam dos empregadores, co-trabalhadores ou clientes. Em todos os casos, assume-se que homens e mulheres são igualmente produtivos (i.e. eles são perfeitamente substitutos na produção), o que implica que seus salários deveriam ser iguais na ausência de discriminação. Ademais, em todos os modelos supõe-se que as firmas operam em mercados perfeitamente competitivos de insumos e produtos.

Na sua forma mais simples, o modelo de Becker de discriminação por parte do empregador assume que firmas homogêneas maximizam uma função utilidade composta pela soma de duas partes: o lucro e o valor (monetário) de contratar mulheres. Mais especificamente, a firma representativa maximiza

$$u = f(H + M) - w_H H - w_M M - dM, \quad (1.1)$$

onde $f(\cdot)$ representa a função de produção, H e M denotam respectivamente o nível de

¹⁰Embora nossa apresentação dos modelos de discriminação seja focada na discriminação por gênero, ela também se aplica, *mutatis mutandis*, ao caso da discriminação por raça.

emprego de homens e mulheres, w_g , $g = H, M$, é o salário pago ao gênero g e $d > 0$ representa o parâmetro de discriminação contra as mulheres (denominado “coeficiente de discriminação” por Becker).¹¹ Das condições de primeira ordem (CPO), tem-se que

$$f_H = w_H \quad (1.2)$$

e

$$f_M = w_M + d, \quad (1.3)$$

onde f_g , $g = H, M$, representa o produto marginal do gênero g . Note que a equação (1.3) implica que as mulheres recebem um salário menor que o seu produto marginal.

Dado que homens e mulheres são substitutos perfeitos na produção: $f_H = f_M = f_L$, onde $L = H + M$. Assim, de (1.2) e (1.3), obtém-se que

$$w_H - w_M = d > 0. \quad (1.4)$$

Portanto, como esperado, o equilíbrio requer que o salário dos homens exceda o das mulheres. Uma vez que todas as firmas são idênticas, cada uma contrata o mesmo número de homens e mulheres. As implicações disso são que a alocação de trabalho na economia é eficiente e o efeito da discriminação é puramente distributivo.

Arrow (1973b) generalizou o modelo de Becker de discriminação pelo empregador. Arrow utilizou uma forma mais geral para a função utilidade das firmas e permitiu que elas tivessem preferências heterogêneas em relação à discriminação contra as mulheres e a favor dos homens. No modelo de Arrow, as firmas maximizam uma função de utilidade geral $U(\pi, M, H)$, onde π denota o lucro. As taxas marginais de substituição dos lucros por mulheres e por homens são dadas respectivamente por $d_M = -U_\pi/U_M$ e $d_H = -U_\pi/U_H$. Assume-se que $U_M < 0$ e $U_H > 0$, ou seja, que os empregadores derivam utilidade marginal negativa (positiva) ao contratar mulheres (homens). Das CPO, tem-se agora que $f_H = w_H + d_H$ e $f_M = w_M + d_M$, implicando que

$$w_H - w_M = d_M - d_H > 0. \quad (1.5)$$

Note, mais uma vez, que aparece um hiato salarial entre homens e mulheres em equilíbrio. O efeito da discriminação sobre os lucros pode ser calculado comparando-se os níveis de lucro nas situações de discriminação (D) e não-discriminação (ND). No primeiro caso, pode-se escrever $\pi_D = f(L) - f_L L + d_H H + d_M M$; no segundo, tem-se $\pi_{ND} = f(L) - f_L L$.

¹¹No modelo de Becker existe um segundo parâmetro que reflete a discriminação a favor dos homens (“coeficiente de nepotismo”). Mais adiante, apresentamos o modelo de Arrow, no qual esse coeficiente aparece explicitamente.

A diferença de lucro nas duas situações é

$$\pi_D - \pi_{ND} = d_H H + d_M M. \quad (1.6)$$

A equação (1.6) pode ser interpretada como o nível de lucros requerido para compensar o empregador por um aumento marginal na força de trabalho que preserve a composição dos sexos. Uma hipótese considerada plausível por Arrow é que a função utilidade dos empregadores depende da razão entre mulheres e homens na firma.¹² Uma das implicações dessa hipótese é que a firma não aumenta nem reduz os lucros discriminando contra as mulheres ou a favor dos homens, isto é:

$$d_H H + d_M M = 0. \quad (1.7)$$

As expressões (1.5) e (1.7) constituem um sistema de equações lineares em d_H e d_M que, quando resolvido, fornece as seguintes relações para a proporção de homens e mulheres na firma:

$$H/L = d_M/(w_H - w_M) \quad \text{e} \quad M/L = -d_H/(w_H - w_M). \quad (1.8)$$

Esta última expressão mostra que as diferenças de proporção de mulheres entre as firmas variam de acordo com o “grau” de discriminação destas: quanto menos (mais) discriminadora for a firma, maior (menor) a proporção de mulheres na sua mão-de-obra. Cabe observar que, em contraste com o modelo de Becker, o produto marginal do trabalho não é mais o mesmo entre as firmas, o que implica que a alocação de trabalho deixa de ser eficiente.¹³

Um dos problemas do modelo de discriminação por parte do empregador é que suas implicações não são necessariamente persistentes no longo prazo. De fato, as firmas menos discriminadoras nesse modelo têm custos menores (elas podem contratar relativamente mais mulheres, que têm salários mais baixos) e, portanto, podem expandir sua produção, retirando as firmas mais discriminadoras do mercado no longo prazo.¹⁴ Se a oferta de trabalho feminina for pequena em relação à demanda por trabalho das firmas menos discriminadoras, então, o salário das mulheres tenderá a subir, fazendo com que o hiato salarial por gênero desapareça. Implícito nessa discussão está o conflito entre o comportamento maximizador de utilidade *versus* o lucro das firmas que operam em mercados competitivos. Em outras palavras, as pressões competitivas tenderão a expulsar aqueles

¹²Essa hipótese é equivalente à suposição de que a função utilidade é homogênea de grau um.

¹³Vale enfatizar que o modelo de Becker/Arrow é baseado na hipótese de mercados perfeitamente competitivos. Assim, uma vez que nesse modelo existe somente um par de salários de equilíbrio para homens e mulheres, a única margem de ajuste para a firma individual é escolher os níveis ótimos de emprego para cada sexo. Mais adiante, inserimos o modelo de Becker/Arrow num arcabouço monopsonístico, que permite às firmas pagarem salários distintos e escolherem diferentes composições por sexo das suas forças de trabalho.

¹⁴Para isso, bastaria assumir que existe entrada livre no mercado ou que a função de produção possui retornos constantes de escala.

empregadores que se “dão ao luxo” de sacrificar seus lucros para satisfazer suas preferências discriminatórias.

Algumas alternativas têm sido propostas na literatura para resolver os problemas de longo prazo embutidos no modelo de discriminação pelo empregador. Algumas delas serão discutidas mais adiante na apresentação dos modelos de discriminação por parte dos co-trabalhadores e dos clientes. Outras possibilidades são baseadas em argumentos de mercados não-competitivos. Primeiramente, uma vez que firmas monopolistas auferem rendas não-competitivas positivas, elas podem trocar parte dos seus lucros pela “satisfação” de discriminar. Nesse sentido, dever-se-ia esperar que a discriminação seja, em média, mais severa em setores monopolistas do que naqueles ramos mais competitivos.¹⁵ Em segundo lugar, argumenta-se também que o poder de monopsonio dos empregadores pode sustentar a existência no longo-prazo do hiato salarial por gênero (Madden 1973). Como se sabe, no modelo de monopsonio, a firma se depara com uma curva de oferta de trabalho positivamente inclinada. No ponto de ótimo, a firma monopsonista equaliza o produto da receita marginal ao custo marginal para obter a demanda ótima por trabalho. O salário de equilíbrio, que é determinado pela curva de oferta de trabalho, é menor que o produto da receita marginal, levando a que as firmas obtenham lucros sobre-normais. Assim, se se assume que a oferta de trabalho das mulheres é menos elástica que a dos homens, então, deve aparecer um diferencial salarial por gênero.

Um problema da explicação monopsonística é que ela se baseia na existência de diferenças nas ofertas de trabalho entre os sexos. Alguns argumentam que a oferta de trabalho feminino tende a responder menos a variações nos salários devido a fatores tais como segregação ocupacional, menor esforço na procura por trabalho e pela ênfase na busca de empregos próximos de casa por parte das mulheres (Madden 1973). Entretanto, outros argumentos têm sido propostos que poderiam fazer com a elasticidade da oferta de trabalho das mulheres fosse maior. Por exemplo, pode ser que a disponibilidade de trabalhos domésticos proveja uma atividade alternativa para as mulheres (pelo menos para aquelas que estão na margem da participação no mercado de trabalho). Adicionalmente, se as mulheres têm maior mobilidade no emprego – por exemplo, porque possuem menor treinamento específico na firma –, então esse fator deveria contribuir para aumentar a elasticidade da oferta de trabalho feminina (Blau & Jusenius 1976). Como existem argumentos nas duas direções, a discussão sobre as possíveis diferenças entre as ofertas de trabalho de mulheres e homens passa a ser uma questão empírica. Na realidade, embora existam evidências de que a elasticidade da oferta agregada de trabalho feminina é mais alta que a masculina, pouco ainda se sabe sobre essas diferenças no que tange a firmas individuais.¹⁶ Ademais, procurar explicar os diferenciais de salário por gênero por meio

¹⁵Existe algum suporte empírico para essa predição. Ver Ashenfelter & Hannan (1986).

¹⁶Utilizando um modelo de busca por trabalho com concorrência monopsonística, Manning (2003, cap. 7) mostra que a elasticidade da oferta de trabalho de mulheres para empregadores individuais pode ser maior ou menor que a dos homens.

de discrepâncias nos comportamentos da oferta de trabalho feminino e masculino introduz um elemento desconectado do argumento principal, que é o de discriminação por parte do empregador. Em outras palavras, a existência de diferenciais de salário por gênero pode ser explicada por diferenças na oferta de trabalho de homens e mulheres mesmo que os empregadores não sejam discriminadores.

Tanto Becker como Arrow também consideram o caso de discriminação por parte do co-trabalhador. Nesse modelo, os empregados do sexo masculino possuem algum grau de insatisfação em trabalhar com mulheres e requerem um prêmio salarial que seja suficiente para induzi-los a tal. Existem duas variantes principais desse modelo. Na primeira, há um terceiro insumo de produção (o supervisor) que é complementar aos fatores-trabalho masculino e feminino. Assume-se que esses supervisores são discriminadores em relação às mulheres. Ademais, supõe-se que cada supervisor escolhe oportunidades de emprego levando em consideração uma relação de preferência inversa entre salários e a razão entre homens e mulheres observada nas firmas. Com base nessas hipóteses, pode-se provar (ver [Arrow \(1973a\)](#)) que os empregadores pagarão às mulheres um salário inferior ao dos homens, gerando, portanto, um hiato salarial por gênero. Essa variante do modelo de discriminação do co-trabalhador não tem uma predição clara sobre a magnitude da segregação feminina: em equilíbrio, todas as firmas podem ter o mesmo nível de emprego para os três tipos de trabalhadores, mas é também possível que apareçam diferentes “nichos” de firmas caracterizados por diferentes razões entre trabalhadores homens e mulheres ([Arrow 1973a](#), p. 116).

Na segunda variante do modelo de discriminação do co-trabalhador, não existe um terceiro fator de produção, mas assume-se que os próprios empregados do sexo masculino possuem preferências discriminatórias contra trabalhar com mulheres. Nesta variante, são os empregados homens (e não os supervisores) que levam em consideração uma relação de preferências inversa entre salários e a proporção de homens na firma. Nesse caso, o modelo prediz que não aparecerá um diferencial de salário por gênero e todas as firmas terminarão completamente segregadas em equilíbrio. Os argumentos que provam essas predições podem ser sumariados da seguinte forma. Primeiramente, o salário requerido pelos empregados do sexo masculino para trabalhar numa firma que contrata somente homens será sempre menor que o salário requerido para trabalhar numa firma integrada. Dado isso, tomemos dois casos polares. Se o salário requerido pelos homens para trabalhar numa firma totalmente masculina for maior que o salário das mulheres, então, a escolha mais barata para a firma é contratar somente mulheres; se o caso oposto ocorrer, então, a escolha ótima da firma é contratar somente homens. Dado que esse é um modelo de pleno emprego, o equilíbrio requer que os salários de homens e mulheres sejam iguais. Portanto, uma vez que firmas integradas teriam que pagar salários mais elevados para os seus empregados homens, em equilíbrio todas as firmas se tornam completamente segregadas, não existindo um diferencial de salários entre os sexos.

O terceiro modelo de discriminação considerado por Becker e Arrow é baseado na hipótese de que os clientes da firma são discriminadores, isto é, de que eles têm algum grau de insatisfação em adquirir produtos vendidos por mulheres. Como resultado, os clientes somente comprarão produtos de mulheres se o preço for mais baixo. Isso implica que os empregadores reduzirão os salários femininos de forma a compensá-los pela menor receita obtida pela venda dos produtos por mulheres. Claramente, esse modelo é mais relevante para aqueles setores onde existe bastante contato de clientes com vendedores, como é o caso do comércio.

Como mencionado anteriormente, um aspecto importante dos modelos baseados na discriminação dos co-trabalhadores e dos clientes é que eles são compatíveis com equilíbrios competitivos no longo prazo. De fato, como nesses dois tipos de modelos o comportamento discriminador não se origina da estrutura de preferências dos empregadores, a maximização de lucros da firma não é afetada. Portanto, as pressões competitivas podem co-existir com a presença de discriminação e de um diferencial de salários por gênero.

Como referido na introdução deste capítulo, a maior parte da literatura empírica focaliza a questão da segregação feminina e dos diferenciais de salário por gênero nas dimensões ocupacional e setorial. Bergmann (1974) desenvolveu uma teoria – o usualmente denominado modelo de “*crowding*” –, na qual a discriminação contra as mulheres leva à segregação dos sexos em dois conjuntos de ocupações (ou setores).¹⁷ Nessa teoria, assume-se que não há mobilidade perfeita de trabalhadores entre esses conjuntos de ocupações. Assim, se as oportunidades de emprego nas ocupações femininas forem reduzidas relativamente à oferta de trabalho das mulheres, deverá aparecer uma “super-concentração” (*crowding*) de trabalhadoras para essas ocupações, o que tenderá a deprimir os salários das mulheres. Isso pode levar, portanto, ao aparecimento de um hiato salarial entre os sexos.

Todos os modelos de discriminação anteriormente discutidos foram baseados na hipótese de que os membros do grupo majoritário possuem preferências discriminatórias contra o grupo minoritário. A segunda classe de modelos de discriminação não parte dessa suposição. Distintamente, a existência de discriminação nesses modelos se baseia na idéia de que os empregadores possuem informação imperfeita sobre a produtividade dos trabalhadores e utilizam características facilmente observadas (por exemplo, sexo ou raça) para inferir a produtividade dos indivíduos. Existem dois tipos de modelos nessa literatura de discriminação estatística. O primeiro se baseia na suposição de que os empregadores discriminam as mulheres porque acreditam que a produtividade delas é menos confiavelmente predita que a dos homens (Phelps 1972, Aigner & Cain 1977, Lundberg & Startz 1983). O segundo tipo de modelos é baseado na idéia de que os empregadores são discriminadores porque eles crêem que a produtividade média das mulheres é menor que a dos homens (Arrow 1973b, Coate & Loury 1993). Uma vez que são calcados na hipótese de informação

¹⁷ Isso poderia ocorrer porque empregadores, co-trabalhadores ou clientes discriminam mulheres somente em algumas ocupações.

imperfeita, esses modelos oferecem uma outra rota pela qual discriminação pode co-existir com o comportamento maximizador das firmas.

Um modelo simples do primeiro tipo foi postulado por [Lundberg & Startz \(1983\)](#). Nele, a verdadeira produtividade do trabalhador, p , é dada pela soma de dois componentes aleatórios: $p = a + be$, onde a e e representam respectivamente a habilidade inata e adquirida pelo indivíduo, e b é um escalar positivo. Os trabalhadores incorrem em custos para adquirir habilidades. Assume-se o custo marginal de obter e é linear, isto é: $c'(e) = ce$, com c sendo um escalar positivo comum a todos os trabalhadores. Ao escolherem o nível ótimo de habilidade adquirida, os trabalhadores equalizam o incremento marginal nos salários ao custo marginal de e . Os empregadores baseiam as suas decisões de contratação em um indicador de produtividade do trabalho, t , que mede a verdadeira produtividade dos trabalhadores com um certo erro aleatório: $t = p + u$, onde u denota esse erro. Na determinação do equilíbrio desse modelo, os trabalhadores decidem a quantidade ótima de investimento em e considerando a curva ótima de salário que as firmas derivam a partir da distribuição conjunta de p e t .¹⁸ Supõe-se que os termos não-observáveis a e u são independentes e normalmente distribuídos. Os empregadores pagam salários iguais à produtividade esperada dos trabalhadores, isto é: $w = E[p|t] = t - E[u|t]$. Pode-se então provar que a curva de salários percebida pelos trabalhadores é dada por $w = \bar{p} + \beta(t - \bar{t})$, onde β é um parâmetro fixo que depende de outros parâmetros do modelo.¹⁹ Uma vez que incrementos marginais em e elevam p e t pelo montante dado pelo escalar b , o retorno ao investimento em e é igual a $b\beta$, levando a que o nível de equilíbrio para a habilidade adquirida seja $e = \beta b/c$. Isso implica que e é não-estocástico em equilíbrio, o que, como mostram os autores, leva a que $\beta = \sigma_a^2/\sigma_t^2 = \sigma_a^2/(\sigma_a^2 + \sigma_u^2) < 1$.

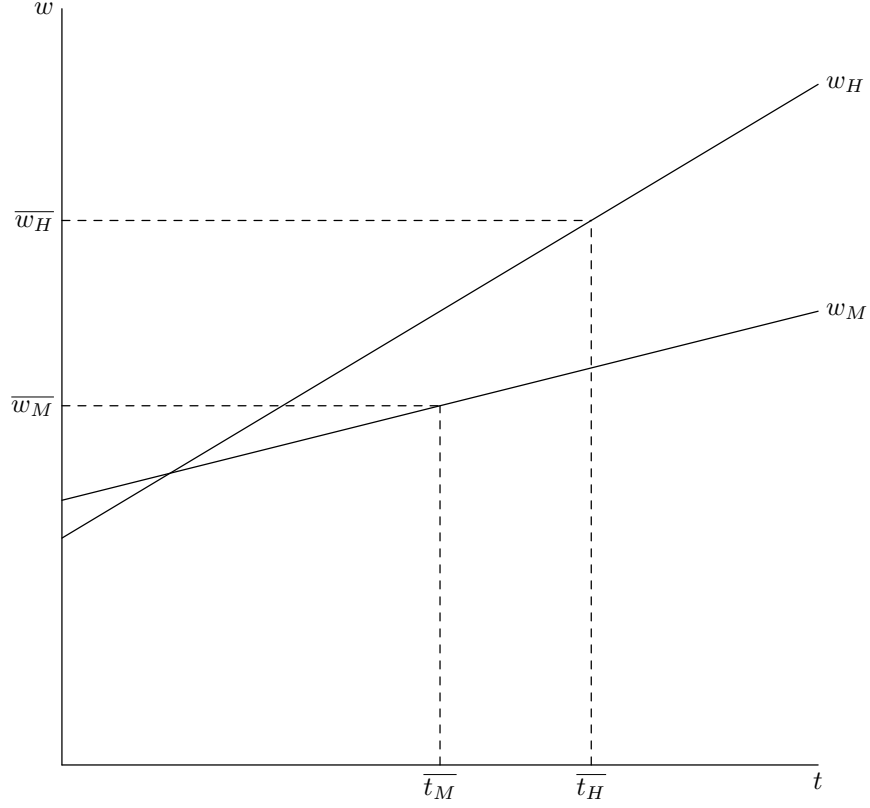
Assume-se que homens e mulheres têm médias idênticas tanto de habilidade inata (\bar{a}), quanto do erro de medida da produtividade (\bar{u}). Supõe-se também que os dois grupos diferem no sentido de que $\sigma_{aH}^2 > \sigma_{aM}^2$ ou $\sigma_{uH}^2 < \sigma_{uM}^2$, ou seja, supõe-se que existe relativamente menos heterogeneidade na habilidade inata das mulheres (M) ou relativamente mais homogeneidade na confiabilidade com que a produtividade dos homens (H) é medida. Em qualquer dos dois casos, as curvas ótimas de salários para homens e mulheres são diferentes, com $\beta_H > \beta_M$, implicando que os empregadores racionalmente discriminam entre os dois grupos. Dado que o retorno marginal líquido para obter e é maior para os homens do que para as mulheres (i.e. $\beta_H b/c > \beta_M b/c$), em equilíbrio o primeiro grupo obtém mais habilidade adquirida que o segundo (i.e. $\bar{e}_H > \bar{e}_M$). Como consequência, a produtividade média masculina é maior que a feminina, levando a que os salários dos homens sejam superiores em média aos das mulheres.

A Figura 1.1 (que é baseada na Fig. 2 de [Lundberg & Startz \(1983\)](#)) mostra as curvas

¹⁸ Assume-se que os parâmetros que caracterizam as distribuições (conjuntas) das variáveis aleatórias nesse modelo são conhecidos por todos. O mesmo vale para os parâmetros b e c .

¹⁹ Ver [Lundberg & Startz \(1983, p. 343\)](#) para a derivação desse resultado.

Figura 1.1: Curvas de Salários para Homens (H) e Mulheres (M)



de salários dos dois grupos. Dado que $\beta_H > \beta_M$, a curva dos homens é mais inclinada que a das mulheres, o que significa que os primeiros recebem aumentos salariais maiores para uma mesma variação no valor do indicador de produtividade. Além disso, uma vez que $\bar{t}_H > \bar{t}_M$, os homens têm um salário médio mais alto, o que resulta no aparecimento de um diferencial de salários por gênero em equilíbrio.²⁰ É interessante notar que, como os trabalhadores são pagos de acordo com suas produtividades marginais esperadas e o salário médio de cada grupo é igual a produtividade marginal média do grupo, um economista poderia considerar essa situação como um equilíbrio não-discriminatório. Entretanto, se se define discriminação econômica para aquelas situações nas quais os grupos com idêntica habilidade inata média não recebem salários iguais (Lundberg & Startz 1983, p. 342), então, o equilíbrio que aparece nesse modelo é discriminatório.²¹

O segundo tipo de modelos de discriminação estatística pode ser exemplificado utilizando-se Arrow (1973b). Nesse modelo, existem dois tipos de postos de trabalho complementares na firma: trabalho qualificado e não-qualificado. Todos os trabalhadores estão capacitados para ocupar os postos de trabalho não-qualificados, mas somente alguns podem

²⁰Dado que nesse modelo homens e mulheres têm idênticos \bar{a} e \bar{u} , o fato de que $\bar{e}_H > \bar{e}_M$ leva a que $\bar{p}_H > \bar{p}_M$ e, portanto, a que $\bar{t}_H > \bar{t}_M$.

²¹Ver Aigner & Cain (1977) para uma discussão sobre diferentes definições de discriminação no mercado de trabalho.

desempenhar as tarefas dos postos qualificados. Assume-se que os empregadores não conhecem *a priori* a produtividade dos indivíduos e que têm que investir para descobrir se o trabalhador está capacitado para ocupar um posto qualificado (*e.g.* período de experiência). Supõe-se também que os empregadores têm uma pré-concepção sobre a distribuição de produtividade de homens e mulheres. Especificamente, os empregadores acreditam que um indivíduo aleatório do sexo masculino é capacitado para entrar num posto qualificado com probabilidade p_H , e um indivíduo aleatório do sexo feminino com probabilidade p_M . Seja mp_Q a produtividade marginal dos trabalhadores qualificados e r o retorno esperado pelo investimento realizado para determinar a verdadeira produtividade do trabalhador. Supõe-se que o ganho líquido para a firma em contratar um indivíduo que acaba sendo não-qualificado é zero. Assumindo-se empregadores risco-neutros, tem-se que

$$r = (mp_Q - w_H)p_H = (mp_Q - w_M)p_M, \quad (1.9)$$

onde w_g é o salário pago ao trabalhador qualificado do gênero $g = H, M$. Seja $q = p_M/p_H$. Tem-se, então, que

$$w_H = qw_M + (1 - q)mp_Q. \quad (1.10)$$

Se os empregadores têm crenças de que $p_H > p_M$, então, a equação (1.10) mostra que w_H é uma combinação convexa de w_M e mp_Q . Notando-se que $w_M < mp_Q$ – de forma a que custo do investimento em determinar a produtividade dos trabalhadores seja recuperado, ver equação (1.9) –, segue-se que $w_H > w_M$, ou seja, aparece um diferencial de salários entre homens e mulheres.

Uma questão importante relativa a esse modelo é o que faz com que os empregadores mantenham diferentes crenças sobre as probabilidades de que indivíduos dos dois sexos sejam qualificados. Arrow sugere uma possível explicação baseada na teoria da ressonância cognitiva, que pode ser sintetizada como um processo no qual crenças e ações entram em equilíbrio, isto é, onde atos discriminatórios dão origem a crenças discriminatórias, as quais, por sua vez, justificam os atos discriminatórios iniciais.²² Esses efeitos retroativos podem levar, então, a um tipo de equilíbrio socioeconômico em que há uma perpetuação da discriminação no longo prazo. Arrow também propõe uma explicação mais baseada em argumentos puramente econômicos, na qual as decisões dos indivíduos de se qualificar são endógenas, ou seja, na qual os trabalhadores passam a poder decidir em investir em si mesmos considerando os ganhos de se qualificarem.²³ Arrow (1973a,b) mostra que, embora um equilíbrio simétrico e não-discriminatório seja possível, ele não é estável. O argumento para essa instabilidade é simples. Partindo de um equilíbrio não-discriminatório, suponha que a

²²Arrow cita (Arrow 1973b, p. 159) a teoria desenvolvida por Festinger, Leon (1957): *A Theory of Cognitive Dissonance*, Row Peterson, Evanston, Illinois. Ver Akerlof & Dickens (1982) para uma discussão sobre como a ressonância cognitiva pode produzir efeitos de natureza econômica.

²³Como esse investimento não é perfeitamente observado pelos empregadores, ele envolve a aquisição de habilidades tais como hábitos de trabalho (iniciativa, seriedade etc.).

proporção de homens qualificados torne-se um pouco superior a das mulheres. Nesse caso, deverá aparecer um hiato salarial entre os dois grupos, que dará aos homens (mulheres) um maior (menor) incentivo para investir em qualificação. Isso, por sua vez, amplificará o hiato salarial existente, levando a um tipo de equilíbrio auto-sustentado no qual a percepção dos empregadores sobre as diferenças de qualificação entre os sexos é corroborada pela evidência concreta de produtividades diferentes entre os dois grupos. Naturalmente, portanto, tenderá a aparecer um hiato salarial entre homens e mulheres que se sustenta no longo prazo.²⁴

Na análise empírica que apresentaremos mais adiante, o interesse se centra na relação entre segregação feminina no nível do estabelecimento e os salários de homens e mulheres. À exceção do modelo de Becker/Arrow de discriminação pelo empregador, nenhum dos demais modelos apresentados anteriormente trata diretamente dessa relação. No que se segue, desenvolvemos dois modelos distintos que lidam com essa relação explicitamente. Todavia, antes de apresentarmos esses modelos, vale a pena mencionar uma teoria frequentemente citada na literatura mas nunca formalizada: a usualmente denominada hipótese de *quality sorting* (Macpherson & Hirsch 1995, Carrington & Troske 1998, Reilly & Wirjanto 1999a). Assumindo que as firmas são heterogêneas em termos das suas demandas relativas por qualificação, essa teoria argumenta que existe algum mecanismo de natureza histórica, social ou econômica que faz com que as firmas que requerem relativamente mais (menos) trabalho não-qualificado tenham uma maior (menor) densidade de mulheres em seus quadros.²⁵ Como essas firmas pagam menores (maiores) salários, essa teoria prediz que a proporção de mulheres na firma será negativamente correlacionada com os salários dos trabalhadores de ambos os sexos. Em um dos modelos que apresentamos a seguir, formalizamos algumas das idéias contidas nessa teoria utilizando um mecanismo baseado em discriminação estatística. Entretanto, em nosso primeiro modelo, tomamos um caminho diferente: inserimos o modelo de Becker/Arrow de discriminação pelo empregador num arcabouço de concorrência monopsonística.

1.3.2 Discriminação pelo Empregador num Mercado de Trabalho Monopsonístico

Como no modelo de Becker/Arrow, os empregadores possuem preferências discriminatórias contra as mulheres. As firmas operam em um mercado de trabalho monopsonístico, o que significa que a curva de oferta de trabalho para a firma individual é positivamente inclinada.

²⁴Coate & Loury (1993) também apresentam um modelo de discriminação estatística no qual as crenças *ex-ante* dos empregadores sobre a qualificação de homens e mulheres são confirmadas num equilíbrio cujos níveis *ex-post* de qualificação dos dois grupos são de fato distintos.

²⁵Uma via pela qual isso pode ocorrer é que a discriminação cria barreiras à entrada de mulheres aos melhores postos de trabalho. Uma outra via seria que as mulheres acumulam relativamente menos capital humano (*e.g.* experiência no mercado de trabalho) do que os homens.

Assume-se que as ofertas de trabalho de homens e mulheres para a firma são idênticas.²⁶ Supõe-se também que as firmas possuem tecnologias heterogêneas, cujo único insumo é o fator trabalho. Homens e mulheres são perfeitamente substitutos na produção.

Dadas essas suposições, a firma $j = 1, \dots, J$ maximiza uma função utilidade escolhendo o total de homens, H , e mulheres, M , de acordo com

$$\max_{H_j, M_j} u_j = f(H_j + M_j, \theta_j) - w_{jH}(H_j)H_j - w_{jM}(M_j)M_j - d_j M_j, \quad (1.11)$$

onde $f(\cdot)$ é uma função de produção bem comportada, e $w_{jH}(\cdot)$ e $w_{jM}(\cdot)$ denotam a oferta de trabalho para a firma de homens e mulheres, respectivamente. A heterogeneidade na produção entre as firmas é representada pelo parâmetro $\theta_j > 0$, enquanto a heterogeneidade na discriminação contra as mulheres é capturada pelo parâmetro $d_j > 0$.

Defina $L_j \equiv H_j + M_j$. Abandonando o sub-escrito j por simplicidade notacional, as CPO são

$$\frac{\partial u}{\partial H} = f_L(L^*(\theta, d), \theta) - w'_H(H^*(\theta, d))H^*(\theta, d) - w_H(H^*(\theta, d)) \equiv 0 \quad (1.12)$$

e

$$\frac{\partial u}{\partial M} = f_L(L^*(\theta, d), \theta) - w'_M(M^*(\theta, d))M^*(\theta, d) - w_M(M^*(\theta, d)) - d \equiv 0, \quad (1.13)$$

onde $f_L \equiv \frac{\partial f}{\partial L}$ e $w'_g \equiv \frac{\partial w_g}{\partial g}$, $g = H, M$.

Das CPO, podemos analisar como as demandas por trabalho de homens e mulheres respondem a mudanças nos parâmetros tecnológico e de discriminação:

$$\frac{\partial H^*}{\partial \theta} = \frac{f_{L\theta}[w''_M M^* + 2w'_M]}{\Delta_2} \quad (1.14)$$

$$\frac{\partial M^*}{\partial \theta} = \frac{f_{L\theta}[w''_H H^* + 2w'_H]}{\Delta_2} \quad (1.15)$$

$$\frac{\partial H^*}{\partial d} = \frac{-f_{LL}}{\Delta_2} \quad (1.16)$$

$$\frac{\partial M^*}{\partial d} = \frac{\Delta_1}{\Delta_2}, \quad (1.17)$$

onde $\Delta_1 = f_{LL} - w''_H H^* - 2w'_H < 0$ e $\Delta_2 = (f_{LL} - w''_H H^* - 2w'_H)(f_{LL} - w''_M M^* - 2w'_M) - (f_{LL})^2 > 0$.²⁷ Assim, uma vez que as curvas de oferta de trabalho para a firma são positivamente inclinadas (i.e. $w'_g > 0$, $g = H, M$), as demandas por trabalho de homens e mulheres aumentam com θ se duas condições (suficientes) forem atendidas: $f_{L\theta} > 0$ e

²⁶Como discutido anteriormente, essa suposição tem a finalidade de não confundir os efeitos da discriminação com os gerados por possíveis diferenças entre as ofertas de trabalho dos dois sexos.

²⁷Os sinais de Δ_1 e Δ_2 são dados pelas usuais condições de maximização da utilidade.

$w_g'' \geq 0$, $g = H, M$. A primeira condição parece uma suposição razoável, já que é válida para diversas especificações usuais da função de produção. A segunda condição, a qual requer que as funções de oferta de trabalho sejam convexas, é atendida no caso linear. Ao custo de perder alguma generalidade, por simplicidade assumiremos abaixo que essas funções de oferta são lineares. Como esperado, quanto mais discriminador for o empregador (i.e. quanto mais elevado for o valor de d), mais alta (baixa) a demanda por trabalho dos homens (mulheres).

A esse nível de generalidade, torna-se difícil obter expressões de “forma fechada” (*closed form*) que permitam uma análise direta da relação entre segregação feminina nas firmas e os salários pagos a homens e mulheres. Assim, simplificaremos o problema fazendo algumas suposições sobre certas formas funcionais. Especificamente, supomos que

$$f(L) = C + \theta L - L^2/2, \quad C > 0, \theta > 0 \quad (1.18)$$

e

$$w_g = \alpha + \frac{\beta}{2}g, \quad g = H, M, \alpha > 0, \beta > 0. \quad (1.19)$$

Das CPO (1.12) e (1.13), podemos então obter explicitamente a demanda ótima por trabalho de homens e mulheres, que são dadas por:

$$H^*(\theta, d) = \frac{(\theta - \alpha)\beta + d}{\beta(2 + \beta)} \quad (1.20)$$

e

$$M^*(\theta, d) = \frac{(\theta - \alpha)\beta - (1 + \beta)d}{\beta(2 + \beta)}. \quad (1.21)$$

Para referência futura, assinala-se que, se a firma contrata pelo menos uma mulher (i.e. se $M^*(\theta, d) > 0$), então:

$$\theta - \alpha > \frac{1 + \beta}{\beta}d > 0. \quad (1.22)$$

Inserindo as expressões (1.20) e (1.21) na equação (1.19), os salários pagos a homens e mulheres por qualquer firma específica são dados respectivamente por:²⁸

$$w(H^*(\theta, d)) = \frac{2\alpha(2 + \beta) + (\theta - \alpha)\beta + d}{2(2 + \beta)} \quad (1.23)$$

e

$$w(M^*(\theta, d)) = \frac{2\alpha(2 + \beta) + (\theta - \alpha)\beta - (1 + \beta)d}{2(2 + \beta)}. \quad (1.24)$$

A fim de analisar a conexão entre a segregação feminina e os salários de homens e

²⁸Há pelo menos duas teorias monopsonísticas que predizem a existência de dispersão salarial entre as firmas em equilíbrio: a primeira é o modelo de procura por trabalho de [Burdett & Mortensen \(1998\)](#) e a segunda é o modelo de localização das firmas desenvolvido por [Bhaskar & To \(2003\)](#). [Manning \(2003\)](#) e [Bhaskar et al. \(2002\)](#) apresentam resenhas desses tipos de modelos.

mulheres, propomos a existência de uma relação entre o comportamento discriminador e a produtividade das firmas. Especificamente, supomos que

$$d = m(\theta - \alpha) = m(\theta^*), \quad (1.25)$$

onde $\theta^* = \theta - \alpha$ e $m(\cdot)$ é uma função desconhecida, porém continuamente diferenciável. Das equações (1.23) e (1.24), pode-se escrever que os salários pagos a homens e mulheres são dados respectivamente por:

$$w_{H^*} = \frac{2\alpha(2 + \beta) + (\theta - \alpha)\beta + d}{2(2 + \beta)} = \gamma_1 + \gamma_2\theta^* + \gamma_3m(\theta^*) \quad (1.26)$$

e

$$w_{M^*} = \frac{2\alpha(2 + \beta) + (\theta - \alpha)\beta - (1 + \beta)d}{2(2 + \beta)} = \gamma_1 + \gamma_2\theta^* - \gamma_4m(\theta^*), \quad (1.27)$$

onde $\gamma_1 = \alpha$, $\gamma_2 = \frac{\beta}{2(2+\beta)} > 0$, $\gamma_3 = \frac{1}{2(2+\beta)} > 0$, e $\gamma_4 = \frac{(1+\beta)}{2(2+\beta)} > 0$.

Definindo o grau de feminização de uma firma como a razão entre a demanda por mulheres e a demanda por homens, pode-se escrevê-lo como:

$$\begin{aligned} \frac{M^*}{H^*} &= \frac{(\theta - \alpha)\beta - (1 + \beta)d}{(\theta - \alpha)\beta + d} \\ &= \frac{\frac{\theta^*}{d} - \frac{1 + \beta}{\beta}}{\frac{\theta^*}{d} + \frac{1}{\beta}} = 1 - \frac{\frac{2 + \beta}{\beta}}{\frac{\theta^*}{d} + \frac{1}{\beta}} \\ &= r \left(\frac{\theta^*}{d} \right) = r \left(\frac{\theta^*}{m(\theta^*)} \right). \end{aligned} \quad (1.28)$$

Podemos, então, analisar como os salários dos dois sexos variam com o grau de feminização de uma firma qualquer. Tem-se que

$$\frac{dw_{H^*}}{d(\frac{M^*}{H^*})} = \frac{\gamma_2 + \gamma_3m'}{r' \cdot \left(\frac{m - \theta^*m'}{m^2} \right)} = \frac{m^2(\gamma_2 + \gamma_3m')}{r' \cdot (m - \theta^*m')} \quad (1.29)$$

e

$$\frac{dw_{M^*}}{d(\frac{M^*}{H^*})} = \frac{\gamma_2 - \gamma_4m'}{r' \cdot \left(\frac{m - \theta^*m'}{m^2} \right)} = \frac{m^2(\gamma_2 - \gamma_4m')}{r' \cdot (m - \theta^*m')}, \quad (1.30)$$

onde $r' = \frac{\partial r(\frac{\theta^*}{d})}{\partial(\frac{\theta^*}{d})} = \frac{\frac{2+\beta}{\beta}}{(\frac{\theta^*}{d} + \frac{1}{\beta})^2} > 0$. Assim, para analisar a direção de $\frac{dw_{H^*}}{d(\frac{M^*}{H^*})}$ e $\frac{dw_{M^*}}{d(\frac{M^*}{H^*})}$, precisamos estudar o sinal de três termos: (i) $[\gamma_2 + \gamma_3m']$, (ii) $[\gamma_2 - \gamma_4m']$, e (iii) $[m - \theta^*m']$. O sinal do primeiro termo é dado por

$$\gamma_2 + \gamma_3m' \geq 0 \Leftrightarrow m' \geq -\frac{\gamma_2}{\gamma_3} = -\beta < 0, \quad (1.31)$$

enquanto o sinal do segundo por

$$\gamma_2 - \gamma_4 m' \geq 0 \Leftrightarrow m' \leq \frac{\gamma_2}{\gamma_4} = \frac{\beta}{1 + \beta} > 0. \quad (1.32)$$

Em relação ao terceiro termo, tem-se que

$$m - \theta^* m' \geq 0 \Leftrightarrow m' \leq \frac{m}{\theta^*} = \frac{d}{\theta^*} = \frac{d}{\theta - \alpha} < \frac{\beta}{1 + \beta}, \quad (1.33)$$

onde utilizamos a definição de θ^* e m (ver equação (1.25)), e a última desigualdade segue do resultado de que, se a firma contrata mulheres, então: $\theta - \alpha > \frac{1+\beta}{\beta} d$ (ver expressão (1.22)).

Esses últimos três resultados mostram que os sinais de $\frac{dw_{H^*}}{d(\frac{M^*}{H^*})}$ e $\frac{dw_{M^*}}{d(\frac{M^*}{H^*})}$ podem ser obtidos pela análise dos valores assumidos por m' . A Tabela 1.2 apresenta os possíveis casos para os sinais em que estamos interessados. Como pode ser visto por essa Tabela, dependendo dos valores dos parâmetros do modelo (α, β, θ, d) e do valor de m' , os salários dos homens e das mulheres podem variar positiva ou negativamente com o grau de feminização das firmas. De fato, podem existir conjuntos de valores assumidos pelos parâmetros nos quais a variação na relação entre discriminação e produtividade (i.e. $\frac{dd}{d\theta^*} = m'$) é tal que: (i) o salário dos homens (mulheres) varia negativamente (positivamente) com o grau de feminização (casos A e D); (ii) ambos os salários de homens e mulheres aumentam com o grau de feminização (caso B); e, (iii) aumentos no grau de feminização reduzem os salários de ambos os grupos (caso C). Vale assinalar que este último caso é compatível com a evidência empírica da literatura de *cross-section* (ver seção 1.2).

Em suma, na medida em que: (i) o mercado de trabalho pode ser razoavelmente descrito pela teoria de concorrência monopsonística; e, (ii) existe uma associação entre o grau de discriminação e a produtividade das firmas, o modelo apresentado acima oferece um arcabouço simples para se entender a conexão entre a segregação feminina entre as firmas e os salários de homens e mulheres. Os resultados do modelo mostram que essa conexão pode ser positiva ou negativa dependendo das características não-observáveis das firmas, (i.e. dos valores dos parâmetros θ e d), da curva de oferta de trabalho (α e β), e da “intensidade” da associação entre as preferências discriminadoras e a produtividade dos empregadores (m'). É interessante notar que, apesar de termos imposto algumas formas funcionais, as previsões do modelo são bastante gerais para gerar vários sinais para a relação de interesse.

1.3.3 Discriminação Estatística e *Quality Sorting*

Nesta seção, apresentamos o modelo que combina algumas idéias das teorias de discriminação estatística e de *quality sorting*. Como mencionado anteriormente, esta última teoria se baseia na existência de mecanismos que produzem uma alocação dos trabalhadores na

Tabela 1.2: Os sinais da relação entre os salários de homens e mulheres e o grau de feminização das firmas ($\frac{dw_{H^*}}{d(\frac{M^*}{H^*})}$ e $\frac{dw_{M^*}}{d(\frac{M^*}{H^*})}$)

Valor de $m'(\theta^*)$	Caso A $(-\infty, -\beta)$	Caso B $[-\beta, \frac{d}{\theta^*})$	Caso C $[\frac{d}{\theta^*}, \frac{1+\beta}{\beta})$	Caso D $[\frac{1+\beta}{\beta}, +\infty)$
$\gamma_2 + \gamma_3 m'(\theta^*)$	—	+	+	+
$\gamma_2 - \gamma_4 m'(\theta^*)$	+	+	+	—
$m(\theta^*) - \theta^* m'(\theta^*)$	+	+	—	—
$\frac{\partial w_{H^*}}{\partial(\frac{M^*}{H^*})}$	—	+	—	—
$\frac{\partial w_{M^*}}{\partial(\frac{M^*}{H^*})}$	+	+	—	+

qual as firmas que requerem relativamente mais trabalho não-qualificado tendem a ser as firmas predominantemente femininas.²⁹ Em nosso modelo, utilizamos um mecanismo inspirado na teoria de discriminação estatística para alocar homens e mulheres entre postos de trabalho qualificados e não-qualificados. Como veremos, esse mecanismo faz com que os homens estejam sobre-representados em postos qualificados, com o oposto ocorrendo com as mulheres. Com a hipótese feita na teoria de *quality sorting* de firmas heterogêneas em termos de suas demandas relativas por qualificação, o modelo leva ao resultado de que quanto maior a proporção de mulheres na firma, menor o salário dos trabalhadores de ambos os sexos. Cabe observar que algumas das idéias aqui formalizadas foram inicialmente propostas por [Reilly & Wirjanto \(1999a\)](#).

Existe um número fixo de firmas (J) no mercado, onde cada firma j possui uma quantidade fixa de capital, k_j . A tecnologia empregada pelas firmas utiliza, além do fator capital,

²⁹[Macpherson & Hirsch \(1995\)](#) propõem que os mecanismos que geram essa alocação estão relacionados à existência de discriminação (passada e presente), que cria barreiras à entrada de mulheres aos postos de trabalho mais qualificados e melhor remunerados. Assim, as firmas relativamente pouco qualificadas possuem uma mão-de-obra mais feminina e pagam menores salários. Nesse sentido, a proporção de mulheres na firma passa a funcionar como um indicador de qualidade para os homens e, em alguma medida, também para as mulheres: ao longo do tempo, as firmas predominantemente femininas e com menores salários tendem a atrair os homens relativamente menos produtivos e a perder as mulheres mais qualificadas. A consequência disso é que se cria uma correlação negativa entre a proporção de mulheres na firma e o salário tanto dos homens quanto das mulheres.

dois tipos de insumo: trabalho qualificado (s) e não-qualificado (u). Cada unidade de capital (*e.g.* uma máquina) requer um número fixo de trabalhadores de cada tipo. Denotando por s_j e u_j esses números fixos de trabalho qualificado e não-qualificado, pode-se expressar a demanda agregada (fixa) de cada tipo de trabalho por:

$$N_s = \sum_{j=1}^J s_j k_j \quad (1.34)$$

e

$$N_u = \sum_{j=1}^J u_j k_j. \quad (1.35)$$

Homens (H) e mulheres (M) ofertam trabalho no mercado. Assume-se que os indivíduos possuem diferentes níveis de habilidade inata e que adquirem habilidades que são relevantes para a produção (*e.g.* educação). Os empregadores não conhecem a verdadeira produtividade dos trabalhadores, mas observam o seu nível de habilidade adquirida, o qual denotamos por e . Supõe-se que os empregadores separam os indivíduos que irão ocupar os postos qualificados e não-qualificados utilizando-se de um nível mínimo para e : acima desse nível os trabalhadores são alocados para as posições qualificadas. Supõe-se também que os empregadores acreditam que o nível mínimo de e para que um homem seja qualificado é *menor* que o nível mínimo para uma mulher, isto é, $\bar{e}_H < \bar{e}_M$, onde a barra superior denota esses cortes fixos de habilidade adquirida.³⁰ Como conseqüência, as mulheres têm que obter relativamente mais habilidades adquiridas do que os homens para sinalizar aos empregadores que elas estão capacitadas a ocupar uma posição qualificada.³¹

Existe um número fixo de indivíduos (T) ofertando trabalho no mercado, os quais podem estar em um de três possíveis estados da força de trabalho: desemprego, empregado num posto não-qualificado ou ocupado numa posição qualificada. Assume-se que todos os trabalhadores têm o mesmo salário de reserva, \bar{w}_r . Ao escolher em que estado eles preferem se encontrar, os trabalhadores observam o salário pago em postos qualificados e não-qualificados, os quais são denotados respectivamente por w_s e w_u . Supõe-se que existe excesso de oferta de trabalho (i.e. $T > N_s + N_u$), o que implica que, em equilíbrio, o salário dos não-qualificados deve ser igual ao salário de reserva, isto é: $w_u^* = \bar{w}_r$. Portanto,

³⁰Podem existir diversos motivos por trás desse comportamento dos empregadores. Por exemplo, se eles vêem as mulheres como mais “falantes” que os homens no ambiente de trabalho, a reação deles poderá ser de exigir maior qualificação das mulheres. Vale assinalar que esse comportamento dos empregadores não se baseia em *preferências* discriminatórias contra às mulheres ou a favor dos homens.

³¹Assume-se que os empregadores observam a produção total da firma, mas não o produto gerado por trabalhador. Nesse sentido, embora as mulheres em postos qualificados tenham mais habilidades adquiridas e, portanto, produzam mais que os homens nesses postos, para os empregadores tudo se passa como se elas produzissem o mesmo que os homens. Isso decorre da crença dos empregadores de que elas tem que compensar com maior e a suposta perda de produtividade por serem mais “dispersas” no ambiente trabalho.

os trabalhadores se tornam indiferentes entre estarem desempregados ou trabalhando em postos não-qualificados.

Assume-se que a distribuição de habilidade inata de homens e mulheres é a mesma, com o primeiro grupo representando α da força de trabalho. Os trabalhadores possuem a mesma função de custo para obter habilidades adquiridas, que é dada por:

$$C = C(a, e), \quad (1.36)$$

onde a representa o nível de habilidade inata. Fazemos as seguintes hipóteses acerca da função de custo $C(., .)$. Primeiramente,

$$C(a, 0) = 0 \quad (1.37)$$

para todos os valores possíveis de a . Em segundo lugar, os indivíduos com habilidade inata mais alta têm menores custos para obter habilidades adquiridas:

$$\frac{\partial C}{\partial a} < 0. \quad (1.38)$$

Em terceiro lugar, quanto maior o nível de habilidade que os trabalhadores decidem adquirir, maior o seu custo:

$$\frac{\partial C}{\partial e} > 0. \quad (1.39)$$

Finalmente, assume-se que, para qualquer nível de habilidade adquirida, o aumento no custo para obter habilidades adicionais é menor para os indivíduos com mais alta habilidade inata, isto é:

$$\frac{\partial^2 C}{\partial a \partial e} < 0. \quad (1.40)$$

Os trabalhadores decidem investir em habilidades adquiridas comparando os custos ao retorno desse investimento, isto é, eles comparam se $C(a, e) \geq w_s - w_u$. Uma vez que o custo de adquirir habilidades é crescente com o nível dessas habilidades (hipótese 1.39), os trabalhadores escolhem ou não obter habilidades adquiridas ou então o nível mínimo exigido pelo empregadores para postos qualificados. Assim, como esse corte mínimo é diferente entre os sexos, os homens e as mulheres que decidem adquirir uma quantidade positiva de habilidade escolhem obter \overline{e}_H e \overline{e}_M , respectivamente.

Defina, respectivamente, por a_H e a_M as menores habilidades inatas do homem e da mulher que decidem obter os níveis mínimos de habilidade adquirida estabelecidos para cada sexo. Formalmente, a_H e a_M podem ser implicitamente definidas por:

$$C(a_H, \overline{e}_H) = w_s - w_u \quad (1.41)$$

$$C(a_M, \overline{e}_M) = w_s - w_u. \quad (1.42)$$

Esses são, portanto, os níveis de habilidade inata dos trabalhadores que estão na margem da indiferença entre obter ou não os mínimos de habilidade para posições qualificadas. Assim, utilizando a hipótese (1.38), tem-se que todos os homens (mulheres) com nível de habilidade inata $a \geq a_H$ ($a \geq a_M$) decidem investir para adquirir o nível mínimo de habilidade.

Denote por $G(a)$ a distribuição acumulada de habilidade inata. Defina $\beta_H = 1 - G(a_H)$ e $\beta_M = 1 - G(a_M)$ como as proporções de homens e mulheres que obtêm os níveis mínimos de habilidade adquirida requeridos para postos qualificados. Seja

$$S_s = T\alpha\beta_H + T(1 - \alpha)\beta_M \quad (1.43)$$

a oferta de trabalho qualificado na economia. Um vez que o salário de equilíbrio para trabalhadores não-qualificados é igual ao salário (fixo) de reserva ($w_u^* = \bar{w}_r$), a igualdade das equações (1.34) e (1.43) (i.e. $N_s = S_s$) define implicitamente o salário de equilíbrio dos trabalhadores qualificados (w_s^*) e, portanto, o diferencial salarial entre trabalho qualificado e não-qualificado: $w_s^* - w_u^*$. Assim, com base nas expressões (1.41) e (1.42), pode-se definir o homem e a mulher que, em equilíbrio, estão na margem de adquirir os níveis mínimos de habilidade por a_H^* e a_M^* . Aqueles trabalhadores que não ocupam posições qualificadas estão em postos não-qualificados ou desempregados. Assume-se que esses trabalhadores são aleatoriamente distribuídos entre o desemprego e os empregos não-qualificados.

Das hipóteses (1.37), (1.38), (1.39), (1.40), e do fato de que $\bar{e}_H < \bar{e}_M$, tem-se que o nível de habilidade inata da mulher na margem de adquirir a quantidade mínima de habilidade deve ser *maior* que o nível de habilidade inata do homem que se encontra nessa margem, isto é:

$$a_M^* > a_H^*. \quad (1.44)$$

Para ver isso, a Figura 1.2 apresenta possíveis curvas de custo variando com a quantidade de habilidade adquirida, para diferentes níveis de habilidade inata.³² Os valores \underline{a} e \bar{a} representam respectivamente o menor e o maior nível possível do suporte da distribuição da variável habilidade inata. O diferencial de salário de equilíbrio $w_s^* - w_u^*$ está marcado no eixo vertical. Dadas as hipóteses (1.38) e (1.40), uma vez que $\bar{e}_H < \bar{e}_M$, a curva de custo associada com a_H^* deve se situar sempre acima da curva relativa a a_M^* . Segue-se, portanto, que $a_H^* < a_M^*$.

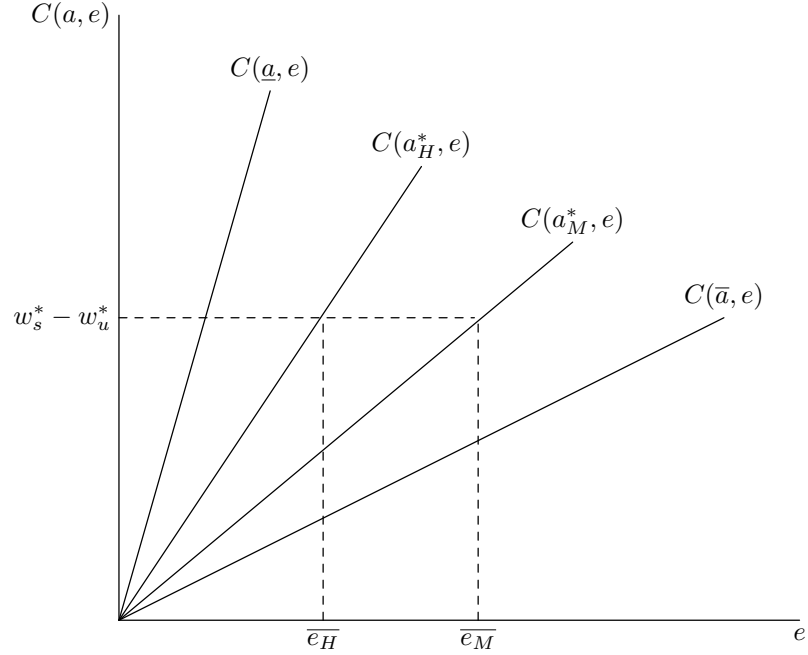
Sejam $\beta_H^* = 1 - G(a_H^*)$ e $\beta_M^* = 1 - G(a_M^*)$ as proporções de homens e mulheres que obtêm empregos qualificados em equilíbrio, respectivamente. Dado que $a_H^* < a_M^*$, tem-se, então, o resultado:

$$\beta_H^* > \beta_M^*, \quad (1.45)$$

ou seja, a proporção de homens em postos qualificados é *maior* do que a proporção de

³²Note-se que as curvas de custo ilustradas na Figura respeitam as hipóteses (1.37), (1.38), (1.39) e (1.40). As curvas estão desenhadas num formato linear apenas por simplicidade gráfica.

Figura 1.2: Curvas de Custo para Diferentes Níveis de Habilidade Inata



mulheres nesse tipo de emprego.³³

Estamos assumindo – como na teoria de *quality sorting* – que as firmas são heterogêneas em termos de suas demandas relativas por trabalho qualificado.³⁴ Seja q_j a parcela da força de trabalho da firma j que é qualificada. Denote por H_j e h_j as proporções de homens da firma j que são contratados respectivamente para posições qualificadas e não-qualificadas. Assume-se que essas proporções são constantes entre as firmas, ou seja: $H_j = H$ e $h_j = h$, $j = 1, \dots, J$.³⁵ Podemos então escrever que:

$$H = \frac{\alpha \beta_H^*}{\alpha \beta_H^* + (1 - \alpha) \beta_M^*} \quad (1.46)$$

e

$$h = \frac{\alpha(1 - \beta_H^*)}{\alpha(1 - \beta_H^*) + (1 - \alpha)(1 - \beta_M^*)}. \quad (1.47)$$

Um resultado importante que se pode obter é que a proporção de homens em posições qualificadas é *maior* que a proporção dos mesmos em postos não-qualificados, isto é: $H >$

³³É interessante notar que os níveis médios de habilidade inata e adquirida das mulheres que estão em posições qualificadas são maiores que os correspondentes níveis dos homens nessas posições.

³⁴Essas demandas relativas são essencialmente determinadas pelos parâmetros tecnológicos s_j e u_j . Ver parágrafo anterior às equações (1.34) e (1.35).

³⁵Essa hipótese pode ser sustentada na medida em que os empregadores são indiferentes entre contratar uma proporção β_H de homens ou β_M de mulheres para postos qualificados (desde que $\bar{e}_M > \bar{e}_H$). Assim, é concebível supor que as firmas tenham proporções iguais de homens (ou mulheres) nesse tipo de posto. Um raciocínio análogo vale para os postos não-qualificados. Note-se que os empregadores em nosso modelo não possuem *preferências* discriminatórias contra as mulheres ou a favor dos homens.

h . Para provar este resultado, seja $k = \frac{H}{\frac{1-H}{h}} = \frac{\frac{\beta_H^*}{\beta_M^*}}{\frac{1-\beta_H^*}{1-\beta_M^*}}$. Da parte anterior do modelo, obtivemos que $\beta_H^* > \beta_M^*$ (ver expressão (1.45)), o que implica que k é claramente maior que a unidade. Segue-se que $\frac{H}{1-H} > \frac{h}{1-h}$ e, portanto, $H > h$.

A proporção de mulheres na firma j é dada por

$$M_j = q_j(1 - H) + (1 - q_j)(1 - h) = (1 - h) - q_j(H - h). \quad (1.48)$$

Uma vez que $H > h$, tem-se, então, o resultado de que: quanto *maior* a fração da força de trabalho da firma j que é qualificada, *menor* a proporção de mulheres na firma, isto é:

$$\frac{\partial M_j}{\partial q_j} = -(H - h) < 0. \quad (1.49)$$

Obviamente, o resultado oposto pode ser obtido para a proporção de homens na firma ($H_j = h + q_j(H - h)$). Portanto, podemos concluir que existe *sorting* de homens e mulheres entre as firmas: aquelas com mais elevadas proporções de postos qualificados (não-qualificados) contratam relativamente mais homens (mulheres) para essas posições.

Os salários médios de homens e mulheres na firma j são dados, respectivamente, por:

$$\begin{aligned} w_j^H &= \frac{q_j H w_s^* + (1 - q_j) h w_u^*}{q_j H + (1 - q_j) h} \\ &= \frac{q_j [H w_s^* - h w_u^*] + h w_u^*}{q_j (H - h) + h} \end{aligned} \quad (1.50)$$

e

$$\begin{aligned} w_j^M &= \frac{q_j (1 - H) w_s^* + (1 - q_j) (1 - h) w_u^*}{q_j (1 - H) + (1 - q_j) (1 - h)} \\ &= \frac{q_j [(1 - H) w_s^* - (1 - h) w_u^*] + (1 - h) w_u^*}{q_j (h - H) + (1 - h)}. \end{aligned} \quad (1.51)$$

Tomando as derivadas das equações (1.50) e (1.51) com respeito a q_j , obtém-se que:³⁶

$$\frac{\partial w_j^H}{\partial q_j} = \frac{H h (w_s^* - w_u^*)}{[q_j (H - h) + h]^2} > 0 \quad (1.52)$$

e

$$\frac{\partial w_j^M}{\partial q_j} = \frac{(1 - H) (1 - h) (w_s^* - w_u^*)}{[q_j (h - H) + (1 - h)]^2} > 0, \quad (1.53)$$

o que implica que os salários médios de homens e mulheres são diretamente relacionados com a fração de postos qualificados na firma j .

Das expressões (1.49), (1.52) e (1.53), pode-se, então, chegar aos resultados de que,

³⁶Note-se que, como os trabalhadores comparam o custo de adquirir habilidades com o diferencial de salários entre postos qualificados e não-qualificados, deve-se ter que $w_s^* - w_u^* > 0$.

quanto maior a fração da força de trabalho da firma j que é qualificada (i.e. quanto mais alto q_j): (1) menor a proporção de mulheres e (2) maior os salários médios de homens e mulheres na firma. Portanto, o modelo prediz a existência de uma correlação *negativa* entre a proporção de mulheres nas firmas e os salários dos trabalhadores dos dois sexos. Essa relação inversa entre o grau de feminização das firmas e os salários de homens e mulheres também foi um dos resultados preditos pelo modelo de concorrência monopsonística apresentado na seção anterior (ver caso C da Tabela 1.2).

1.4 Base de Dados

1.4.1 Descrição

A fonte de informações utilizada na análise empírica deste capítulo são os micro-dados da Relação Anual de Informações Sociais (RAIS). Trata-se de uma base de dados administrativa que contém informações combinadas de empregadores e empregados para todo o Brasil. Os dados são baseados em informações prestadas anualmente ao governo federal por todos os estabelecimentos registrados no país.

A base de dados contém informações sobre cada vínculo de trabalho que o estabelecimento teve com seus empregados num determinado ano. Cada observação presente na base corresponde a uma combinação entre o estabelecimento e o trabalhador, com identificadores únicos para cada um. Para cada observação, há informações sobre o sexo, a idade, o nível de escolaridade, o tempo do trabalhador no estabelecimento, a ocupação, o número de horas contratuais e o salário mensal do trabalhador.³⁷ Além disso, informa-se também o setor de atividade e o município onde o estabelecimento está localizado.

Tivemos acesso aos micro-dados da RAIS de três anos consecutivos: 2000, 2001 e 2002. Utilizando os identificadores dos estabelecimentos e dos trabalhadores, pudemos, então, construir um painel de dados com as informações combinadas de empregadores e empregados. É importante ressaltar que a RAIS reúne informações somente dos estabelecimentos e trabalhadores registrados. Isso implica que, se um trabalhador esteve fora do setor formal num determinado ano (por exemplo, num emprego informal ou trabalhando como conta-própria), as informações sobre ele/ela não estarão disponíveis na base da RAIS daquele ano. De forma semelhante, quando um estabelecimento inteiro passa a ser informal, é improvável que as informações sobre esse estabelecimento e seus empregados estejam presentes na RAIS. A consequência disso é que um conjunto de indivíduos e estabelecimentos deixam de ser continuamente observados durante o período de análise, levando a que a base de dados concatenada tenha um formato de painel não-balanceado.³⁸

³⁷A informação sobre salários é utilizada pelo governo federal na implementação do programa de Abono Salarial, que paga anualmente um benefício igual a um salário mínimo aos trabalhadores que receberam menos de dois salários mínimos em média no ano correspondente.

³⁸Uma outra razão para isso é que um conjunto de estabelecimentos nascem e morrem a cada

A base de dados para o país como um todo têm aproximadamente 60 milhões de observações. Devido a esse elevado tamanho, e ao fato de que a investigação sobre a relação entre salários e a segregação feminina entre estabelecimentos se aplica melhor a mercados de trabalho locais, a análise empírica é baseada nos dados referentes a uma cidade brasileira de tamanho mediano, a saber: Belo Horizonte. Mesmo para esse município isoladamente, a base de dados inicial contém cerca de 3,7 milhões de observações.

1.4.2 Variável de Salário

A informação contida na RAIS sobre salários representa a compensação salarial total dos trabalhadores, isto é, a soma do salário contratual com diversos outros componentes remuneratórios, tais como: pagamento de horas-extras, bônus, comissões, gorjetas etc. Uma vez que o peso desses componentes nos salários dos homens e das mulheres pode ser bastante distinto – em particular o item referente a horas-extras –, torna-se preferível utilizar uma medida de salário que seja mais comparável entre os trabalhadores dos dois sexos. De acordo com a legislação brasileira, todo empregado deve receber o 13º salário, que é igual somente ao salário contratual, e cujo valor é proporcional ao tempo em que o empregado trabalhou no estabelecimento durante o ano.³⁹ Assim, como o salário contratual é uma medida mais comparável da remuneração de homens e mulheres, a variável de salário utilizada na análise empírica é baseada na informação sobre o 13º salário.

Especificamente, a variável de salário foi construída com base nos seguintes passos. Primeiramente, como vários indivíduos trabalharam somente uma parte do ano nos estabelecimentos presentes na base de dados, converteu-se a informação do 13º salário num valor mensal. Para isso, utilizou-se de um fator de conversão igual a $12/e$, onde e é o número de meses de emprego do trabalhador no estabelecimento durante o ano. Em segundo lugar, calculou-se o salário contratual horário para cada trabalhador, que foi obtido por meio da divisão do salário contratual mensal construído por quatro vezes o número de horas contratuais semanais. A variável de salário final é o logaritmo do salário contratual horário real, onde o deflator utilizado foi o Índice Nacional de Preços ao Consumidor (INPC).

1.4.3 Construção das Covariadas

O nível de escolaridade informado é a mais alta série completada pelo trabalhador. Todavia, algumas séries são reportadas em valores categóricos: 1 a 3, 5 a 7, 9 a 10, 12 a 14 e 15 ou mais. Assim, para aqueles indivíduos que estão nessas categorias, utilizou-se o ponto mediano como medida do nível de escolaridade (i.e. 2 para 1 a 3, 6 para 5 a 7, 9,5 para 9 a 10, 13 para 12 a 14). A única exceção foi última categoria, para a qual fixou-se o nível

ano.

³⁹Por exemplo, se esse período for de três meses, então o trabalhador recebe um 13º salário correspondente a $1/4$ do salário contratual.

de escolaridade em 15.

Como medida da experiência geral dos trabalhadores no mercado de trabalho utilizou-se a experiência potencial, a qual foi calculada pela fórmula: $idade - escolaridade - 6$. A RAIS contém informações sobre a experiência do trabalhador no estabelecimento, que é medida em meses.

As ocupações dos trabalhadores foram agrupadas em cinco categorias: ocupações ligadas à produção, ocupações administrativas, profissionais, ocupações de gerência e ocupações diversas. Os setores de atividade dos estabelecimentos também foram divididos em cinco grupos: agricultura, indústria, construção civil, comércio e serviços.

A medida de segregação feminina utilizada foi a proporção de mulheres no estabelecimento. Como medida do tamanho dos estabelecimentos utilizou-se o seu número total de empregados. Ambas as medidas foram calculadas para cada par estabelecimento/ano, o que gera variação nessas duas dimensões.

1.4.4 Seleção da Amostra

Como vimos na seção 1.3, a maior parte das teorias que buscam explicar a existência do diferencial de salários por gênero são baseadas no comportamento maximizador das firmas. Por essa razão, excluíram-se as observações referentes a estabelecimentos que não visam ao lucro, incluindo o setor público. Também foram excluídas as observações correspondentes aos estabelecimentos que reportaram não terem empregados.⁴⁰ Em conjunto, essas exclusões reduziram a amostra para cerca de 2,4 milhões de observações. Foi para essa base de dados que todas as variáveis no nível do estabelecimento foram calculadas, tais como a proporção de mulheres e o tamanho do estabelecimento.

Alguns trabalhadores na RAIS têm mais de um emprego no mesmo mês de um ano. Para não incluir na estimação informações de empregos secundários, procurou-se selecionar as observações que correspondessem ao emprego principal dos indivíduos. Isso foi implementado mantendo-se somente as observações para as quais o salário e o número de horas trabalhadas fossem os mais altos para cada indivíduo com mais de um emprego. Essa operação reduziu a amostra para aproximadamente 2,3 milhões de observações.

Alguns outros filtros foram utilizados na construção da amostra. Em primeiro lugar, excluíram-se as observações dos indivíduos cuja idade na data de suas primeiras aparições na base de dados era menor que 25 anos ou maior que 65 anos. Em segundo lugar, foram mantidos apenas os trabalhadores que permaneceram no mesmo grupo ocupacional ao longo do tempo. Presumivelmente, esses filtros captam os indivíduos que têm um vínculo mais estável com o mercado de trabalho. Em terceiro lugar, como Belo Horizonte pertence a uma área metropolitana, não se incluiu nenhum estabelecimento do setor agrícola. Finalmente, uma vez que a construção civil é um *outlier* em termos da proporção de mulheres que

⁴⁰A existência desse tipo de estabelecimento provavelmente se deve ao fato de que muitos trabalhadores por conta-própria abrem empresas para pagarem menos impostos.

apresenta (menos do que 4%), esse setor também foi excluído da amostra. O resultado da aplicação desses filtros foi uma amostra com 885.637 observações.

Na análise empírica, trabalhou-se com duas amostras distintas. A primeira contém somente os trabalhadores que permaneceram no mesmo estabelecimento entre os anos. Essa amostra possui 796.604 observações, das quais 487.861 (61,2%) se referem a homens e 308.743 (38,8%) a mulheres. A segunda amostra compreende todos os trabalhadores, incluindo os que trocaram de estabelecimento durante o período de análise. Das 885.637 observações dessa amostra completa, 544.051 (61,4%) são para homens e 341.586 (38,6%) para mulheres.

1.4.5 Estatísticas Descritivas

A Tabela 1.3 apresenta estatísticas descritivas para as duas amostras utilizadas na parte de estimação. As linhas 1 e 2 reportam a média do (log) do salário horário real de homens e mulheres. Essas linhas revelam que existe um substancial hiato salarial entre os sexos: em termos de médias geométricas (aritméticas), esse hiato é de 39,1% (41,2%) para amostra de todos os trabalhadores, e 41,9% (43,9%) quando se considera a amostra dos trabalhadores que permaneceram no mesmo estabelecimento. A linha 3 apresenta a média do salário horário no setor industrial. A razão entre os salários médios de homens e mulheres nesse setor pode ser comparada à evidência disponível para um conjunto de 58 países (ver [International Labour Office \(2003, Tab. 5g\)](#)). A média não-ponderada dessa razão para esses países é 77,2%, ao passo que para o Brasil esse número é aproximadamente 50%. Isso indica que, pelo menos para esse setor, o salário relativo das mulheres no Brasil tende a ser menor do que em outros países.⁴¹ É importante assinalar que as médias salariais apresentadas nas linhas 1-3 são bastante similares para as duas amostras utilizadas.

As linhas 4, 5 e 6 reportam respectivamente as médias de experiência geral, de experiência no estabelecimento e de escolaridade dos trabalhadores. Os homens têm em média mais experiência geral e no estabelecimento que as mulheres. Este último grupo tem, entretanto, meio ano a mais de escolaridade que o primeiro grupo. Em conjunto, essas estatísticas revelam a existência de significativas diferenças nos atributos produtivos dos dois sexos. Vale assinalar que a comparação entre as duas amostras utilizadas mostra resultados bastante similares.

A linha 7, que apresenta a média da variável proporção de mulheres entre os estabelecimentos, mostra que os homens tendem a trabalhar em plantas onde as mulheres representam 1/4 da força de trabalho. Em contraste, as mulheres tipicamente trabalham em estabelecimentos nos quais a proporção de mulheres é cerca de 3/5. Esses resultados indicam que as mulheres trabalham em estabelecimentos predominantemente femininos, o oposto ocorrendo com os homens. Tomando-se os dois grupos conjuntamente (não mos-

⁴¹ Vale observar que essas estatísticas não são estritamente comparáveis, já que as coberturas, definições e métodos de cálculo tendem a diferir de país para país.

Tabela 1.3: Estatísticas Descritivas: Município de Belo Horizonte

	Todos os Trabalhadores		Trabalhadores no Mesmo Estabelecimento	
	Homens	Mulheres	Homens	Mulheres
1. Log do Salário Horário	1.30 (0.001)	0.97 (0.002)	1.35 (0.001)	1.00 (0.002)
2. Salário Horário	6.38 (0.02)	4.52 (0.02)	6.56 (0.02)	4.56 (0.02)
3. Salário Horário na Indústria	7.20 (0.08)	3.58 (0.03)	7.33 (0.08)	3.65 (0.03)
4. Experiência Geral	23.2 (0.01)	22.2 (0.02)	23.2 (0.01)	22.2 (0.02)
5. Experiência no Estabelecimento	58.4 (0.1)	48.0 (0.1)	62.8 (0.1)	50.9 (0.1)
6. Escolaridade	8.2 (0.005)	8.7 (0.006)	8.3 (0.005)	8.8 (0.006)
7. Proporção de Mulheres	0.25 (0.0003)	0.58 (0.0004)	0.24 (0.0003)	0.58 (0.0004)
8. Tamanho do Estabelecimento	1270 (3.7)	858 (3.3)	1334 (4.0)	863 (3.6)
9. Ocupação				
9a. Ligados à Produção	0.70 (0.0006)	0.57 (0.0008)	0.68 (0.0007)	0.56 (0.0009)
9b. Administrativos	0.12 (0.0004)	0.28 (0.0008)	0.13 (0.0005)	0.28 (0.0008)
9c. Profissionais	0.11 (0.0004)	0.09 (0.0005)	0.11 (0.0004)	0.10 (0.0005)
9d. Gerentes	0.08 (0.0004)	0.06 (0.0004)	0.08 (0.0004)	0.06 (0.0004)
10. Setor				
10a. Indústria	0.13 (0.0005)	0.11 (0.0005)	0.13 (0.0005)	0.12 (0.0006)
10b. Comércio	0.21 (0.0006)	0.21 (0.0007)	0.21 (0.0006)	0.21 (0.0007)
10c. Serviços	0.66 (0.0006)	0.67 (0.0008)	0.66 (0.0007)	0.67 (0.0008)
Observações	544051	341586	487861	308743
Número de trabalhadores	211284	131001	193477	124194
Número de estabelecimentos	19321	17573	15987	14841

Notas:.

(1) *Erros-padrões entre parênteses.*

(2) *As médias salariais estão expressas em Reais (R\$), tendo sido deflacionadas pelo INPC (base: 2002).*

(3) *Experiência geral e escolaridade são medidas em anos, e experiência no estabelecimento em meses. Tamanho do estabelecimento é a média do número de empregados entre estabelecimentos. Proporção de mulheres é a média da proporção de mulheres entre estabelecimentos.*

(4) *Todos os números da tabela referem-se ao período 2000-2002.*

trado), as médias para a proporção de mulheres é muito semelhante para as amostras completa (37,3%) e dos trabalhadores que permanecem no mesmo estabelecimento (37,1%).

A linha 8, que reporta a média do tamanho entre os estabelecimentos, revela que os homens trabalham em plantas maiores que as mulheres, com a diferença relativa sendo cerca de 48% para a amostra completa e 55% para a amostra dos que continuam no mesmo estabelecimento ao longo do tempo.

As linhas 9a a 9d apresentam a composição ocupacional para homens e mulheres. Essas linhas mostram que ambos os grupos estão concentrados em ocupações ligadas à produção, e que fração de mulheres em ocupações administrativas é superior a dos homens. Relativamente às mulheres, tem-se também que a parcela dos homens em ocupações profissionais e de gerência é ligeiramente maior que a das mulheres. Finalmente, as linhas 10a a 10c, que apresentam a composição setorial para homens e mulheres, mostram que a distribuição de cada grupo entre os setores é similar. Mais uma vez, assinale-se que as diferenças entre as duas amostras é bastante reduzida.

1.5 Medindo Segregação Feminina entre Estabelecimentos

Esta seção está dedicada a medir a magnitude da segregação de mulheres entre estabelecimentos. A fim de avaliar os padrões dessa segregação, empregou-se duas medidas bastante utilizadas na literatura, a saber: o índice de dissimilaridade de Duncan (D) e o coeficiente de segregação de Gini (G). Ambas as medidas apresentam valores no intervalo de zero a um, com o limite inferior correspondendo à situação em que todas as plantas apresentam uma proporção de mulheres igual a observada na força de trabalho (perfeita representatividade), e o limite superior ao caso em que cada planta tem trabalhadores somente de um sexo (perfeita segregação). Formalmente, denotando respectivamente por h_j e m_j a fração de homens e mulheres no estabelecimento j , $j = 1, \dots, J$, pode-se escrever o índice de dissimilaridade de Duncan como:

$$D = \frac{1}{2} \sum_{j=1}^J |m_j - h_j|. \quad (1.54)$$

Esse índice pode ser interpretado como a parcela de mulheres (ou homens) que precisariam trocar de estabelecimentos para fazer com que se chegasse a situação de perfeita representatividade. Embora bastante utilizada, essa medida tem sido criticada pelo fato de que ela é igualmente sensível a pequenos e grandes distanciamentos da situação de perfeita representatividade (ver [Hutchens \(2001\)](#)). Muitos autores sugerem o uso do coeficiente de segregação de Gini, que pode ser definido por:

$$G = 1 - \sum_{j=1}^J m_j \left(h_j + 2 \sum_{i=j+1}^J h_i \right), \quad (1.55)$$

onde se assume que os dados foram inicialmente ordenados pela razão m_j/h_j .⁴²

Como apontado por Carrington & Troske (1997), uma característica pouco atrativa das medidas tradicionais de segregação é que elas assumem valores positivos mesmo se os homens e as mulheres forem aleatoriamente distribuídos entre os estabelecimentos. Carrington & Troske (1997) mostram que esse problema depende essencialmente de dois fatores: a distribuição de classes de tamanho dos estabelecimentos e a fração de mulheres em cada uma dessas classes. Mais especificamente, esses autores mostram que, à medida em que a parcela de estabelecimentos pequenos cresce na amostra, uma alocação aleatória dos trabalhadores entre os estabelecimentos estará associada a maiores valores das medidas de segregação.⁴³

Uma vez que a nossa base de dados contém uma alta proporção de estabelecimentos pequenos, as duas medidas de segregação aqui empregadas serão calculadas seguindo os procedimentos propostos por Carrington & Troske (1997).⁴⁴ A idéia desses procedimentos é separar a parte da segregação que pode ser atribuída à alocação aleatória dos trabalhadores entre os estabelecimentos da parte que é devida a um tipo de segregação mais sistemática.

Sejam D^* e G^* os índices de segregação aleatória de Duncan e Gini, respectivamente. Essas são as medidas que seriam obtidas a partir de uma base de dados na qual os trabalhadores de ambos os sexos são alocados aos estabelecimentos de forma puramente aleatória. Para calcular esses índices, considera-se a distribuição empírica dos sexos na amostra entre diferentes tamanhos de estabelecimentos, e aloca-se aleatoriamente os trabalhadores de acordo com essa distribuição empírica. Esse procedimento mantém a estrutura original da amostra tanto em termos da distribuição dos tamanhos dos estabelecimentos, quanto da composição por gênero da força de trabalho. Na prática, devido ao elevado número de estabelecimentos presentes na base de dados, repetiu-se esse procedimento 100 vezes, com cada replicação contendo uma amostra aleatória de 10% dos estabelecimentos. O índice de

⁴²A diferença entre D e G pode ser vista utilizando-se o exemplo apresentado em Carrington & Troske (1998, p. 450). Suponha-se que existe uma amostra de quatro plantas cujas composições por gênero são as seguintes: cada uma das plantas 1 e 2 empregam 50 mulheres e 50 homens, a planta 3 tem 75 mulheres e 25 homens, e a planta 4 tem 25 mulheres e 75 homens. Para essa situação, $D = 0,25$ e $G = 0,375$, caracterizando essa distribuição como segregada. Suponha-se agora que se quer aumentar a segregação por gênero implementando-se duas trocas diferentes de homens e mulheres entre as plantas: (a) as plantas 1 e 2 trocam um homem por uma mulher, ou (b) a planta 3 troca um homem por uma mulher com a planta 4. O índice de dissimilaridade de Duncan apresenta o mesmo valor ($D = 0,26$) para ambos os casos, enquanto o coeficiente de Gini tem valores respectivamente de 0,38 e 0,39. Portanto, o índice de dissimilaridade trata as duas situações de forma simétrica, ao passo que o coeficiente de Gini trata o caso (b) como aquele que gera o maior incremento na segregação.

⁴³Por exemplo, o índice de dissimilaridade pode ser visto como uma média ponderada da segregação dentro de cada classe de tamanho, somada a alguns termos que captam a dissimilaridade entre essas classes (Carrington & Troske 1997, p. 404). Assim, em casos onde a fração de plantas pequenas constitui uma elevada parcela (ponderada pelo tamanho) da amostra, a segregação devida à alocação aleatória tenderá a ser alta.

⁴⁴A parcela de estabelecimentos com menos de 10 empregados é de 16,6%, com menos de 50 empregados de 44,4% e com menos de 100 empregados de 54,1%.

dissimilaridade e o coeficiente de Gini foram calculados utilizando-se as fórmulas (1.54) e (1.55) para cada replicação. Os valores dos índices D^* e G^* apresentados mais adiante são as médias computadas para todas as replicações.

Os índices D^* e G^* são então utilizados para obter duas medidas de segregação sistemática, denotadas respectivamente por \hat{D} e \hat{G} . A finalidade dessas medidas é transformar os índices originais em medidas que controlem para o efeito da alocação aleatória em gerar segregação entre os estabelecimentos. Elas são definidas por (Carrington & Troske (1997, p. 406)):

$$\hat{S} = \begin{cases} \frac{S-S^*}{1-S^*} & \text{if } S \geq S^* \\ \frac{S-S^*}{S^*} & \text{if } S < S^* \end{cases} \quad (1.56)$$

onde $\hat{S} = \hat{D}, \hat{G}$ e $S^* = D^*, G^*$. A interpretação dessas medidas de segregação sistemática é simples. Se a segregação observada excede a segregação aleatória (i.e. se $S > S^*$), então $\hat{S} > 0$ quantifica o excesso de segregação sobre a que seria esperada pelo simples acaso ($S - S^*$), sendo expresso em termos do máximo de excesso de segregação que poderia ocorrer ($1 - S^*$). Note-se que $\hat{S} = 1$ corresponde ao caso de perfeita segregação (i.e. $S = 1$), porém $\hat{S} = 0$ é equivalente à situação de alocação puramente aleatória (i.e. $S = S^*$). De forma semelhante, se a segregação existente é menor que a segregação aleatória (i.e. $S < S^*$), então $\hat{S} < 0$ mede o excesso de representatividade sob a que seria esperada pelo puro acaso ($S - S^*$), sendo expresso em termos do máximo de excesso de representatividade que poderia ocorrer (S^*). O caso em que $\hat{S} = -1$ corresponde à situação de perfeita representatividade (i.e. $S = 0$).

Em suma, as medidas de segregação sistemática representam um re-mapeamento dos valores originais de D e G do intervalo $[0, 1]$ para o intervalo $[-1, 1]$, onde o segmento $[0, 1]$ ($[-1, 0]$) corresponde ao caso de excesso de segregação (representatividade), no sentido de que a segregação total excede (é menor que) a que seria esperada pela alocação aleatória dos trabalhadores entre os estabelecimentos. Uma vantagem de medir a segregação sistemática é que ela permite comparar a magnitude da segregação entre diferentes bases de dados.⁴⁵

A Tabela 1.4 reporta as três medidas de segregação (total, aleatória e sistemática) para o índice de dissimilaridade de Duncan e o coeficiente de Gini.⁴⁶ As linhas 1a a 1c apresentam os resultados para cada ano disponível nos dados. Ambas as medidas tradicionais de segregação (D e G) mostram a existência de um elevado e estável nível de segregação por gênero entre os estabelecimentos. Como revelam as medidas de segregação aleatória (D^* e G^*), uma considerável parte da segregação total pode ser atribuída à alocação aleatória dos trabalhadores. Dado que essas medidas aleatórias são menores do que as tradicionais,

⁴⁵ A menos que a segregação aleatória seja constante, a comparação da segregação existente entre distintas bases de dados pode levar a conclusões equivocadas sobre qual das bases apresenta maior nível de segregação.

⁴⁶ Os erros-padrões das medidas são calculados por meio de técnicas de *bootstrap*, como propostas por Carrington & Troske (1997).

Tabela 1.4: Medidas de Segregação por Gênero entre Estabelecimentos: Município de Belo Horizonte

	Índice de Dissimilaridade			Coeficiente de Gini		
	Total (D)	Aleatória (D^*)	Sistemática (\hat{D})	Total (G)	Aleatória (G^*)	Sistemática (\hat{G})
1. Todos Trabalhadores						
1a. Ano 2000	0.496 (0.002)	0.154 (0.001)	0.404 (0.003)	0.702 (0.001)	0.517 (0.001)	0.382 (0.002)
1b. Ano 2001	0.496 (0.003)	0.168 (0.001)	0.394 (0.003)	0.703 (0.001)	0.514 (0.001)	0.389 (0.002)
1c. Ano 2002	0.482 (0.003)	0.166 (0.001)	0.379 (0.003)	0.701 (0.001)	0.521 (0.001)	0.375 (0.002)
2. Setor (2000)						
2a. Indústria	0.547 (0.003)	0.216 (0.001)	0.421 (0.005)	0.691 (0.001)	0.471 (0.001)	0.415 (0.002)
2b. Comércio	0.499 (0.001)	0.270 (0.000)	0.313 (0.001)	0.650 (0.001)	0.527 (0.001)	0.259 (0.001)
2c. Serviços	0.474 (0.002)	0.133 (0.001)	0.393 (0.002)	0.741 (0.001)	0.527 (0.001)	0.452 (0.002)

Notas:

(1) Os erros-padrões foram calculados por bootstrap e estão entre parênteses.

(2) Ver texto para a descrição de como essas medidas são computadas.

a segregação sistemática é positiva para todos os anos da amostra. Isso implica que existe excesso de segregação em relação ao máximo que seria esperado pela alocação aleatória dos trabalhadores entre as plantas. Portanto, pode-se concluir que os homens e as mulheres são sistematicamente segregados no mercado de trabalho (formal) de Belo Horizonte. É interessante notar que a segregação sistemática parece ser um pouco menor do que em Portugal, onde as medidas correspondentes são $\hat{D} = 0.49$ e $\hat{G} = 0.67$ (Vieira et al. (2003, Tab. 1)).

A linhas 2a a 2c reportam as mesmas medidas de segregação para a amostra estratificada por setor de atividade para o ano de 2000. Os resultados mostram que a segregação dentro dos setores não é substancialmente distinta da observada para a amostra como um todo. Isso significa que somente uma fração relativamente pequena da segregação no nível agregado pode ser atribuída a segregação entre os setores. Os números obtidos para o setor industrial podem ser comparados aos calculados para o mesmo setor nos Estados Unidos (Carrington & Troske 1998, Tab. 2). Embora o coeficiente de Gini sistemático para a nossa amostra seja mais baixo do que o americano ($\hat{G} = 0,45$), é difícil afirmar qual dos dois casos apresenta menor segregação, já que o valor do índice de dissimilaridade sistemático é superior no primeiro caso em comparação ao observado no segundo ($\hat{D} = 0,33$).

Em suma, a evidência apresentada nesta seção indica que a distribuição dos sexos no nível do estabelecimento é segregada. Entretanto, parte dessa segregação é consistente com a alocação aleatória dos trabalhadores entre os estabelecimentos. Assim mesmo, os

resultados apresentados indicam que o componente sistemático é importante, já que representa uma parte substancial do padrão de segregação por gênero observado no mercado de trabalho (formal) de Belo Horizonte.

1.6 Metodologia Empírica

O fenômeno que queremos estudar se refere a como os salários de homens e mulheres variam com o grau de feminização das firmas. Claramente, o salário de um trabalhador depende de um conjunto amplo de fatores, alguns dos quais são observados (ou pelo menos podem ser aproximados) e outros não (ao menos pelo analista). Entre esses últimos, existem dois que parecem cumprir papéis fundamentais. O primeiro são os atributos não-observáveis dos trabalhadores, tais como suas habilidades, motivações e preferências. O segundo, muitas vezes negligenciado em estudos sobre salários, são as características não-observáveis dos empregadores, tais como seus níveis de eficiência (gerencial) e suas preferências (em particular as associadas à discriminação). Todos esses componentes não-observáveis tendem a influenciar substancialmente a heterogeneidade salarial observada entre trabalhadores e entre firmas. Além disso, essas características não-mensuradas podem estar correlacionadas com as variáveis observadas incluídas no modelo de regressão salarial e, em particular, com nossa variável de interesse, a proporção de mulheres no estabelecimento. Por exemplo, é possível que os indivíduos com menor habilidade não-mensurada estejam alocados em estabelecimentos que possuem uma força de trabalho predominantemente feminina. É possível também que os empregadores com preferências mais discriminadoras contratem relativamente menos mulheres. Assim, idealmente, seria desejável controlar para esses fatores não-observáveis, de forma a tornar possível a avaliação de suas influências sobre as estimativas do parâmetro de interesse.

Um das formas mais utilizadas e potentes para lidar com o problema de componentes não-observáveis é a disponibilidade de dados de painel e o uso de métodos longitudinais. De fato, uma das maiores vantagens de dispor de um painel de dados é que o analista conta com uma dimensão a mais de variação (o tempo), que permite a aplicação de métodos longitudinais para lidar com termos específicos que absorvem as características não-observáveis de trabalhadores e firmas. Em nosso caso, a disponibilidade de um painel de dados combinados de empregadores e empregados nos permite incorporar esses componentes não-observáveis, especificamente o que se costuma denominar de efeito específico do trabalhador e efeito específico do estabelecimento. Tanto quanto sabemos, esse é o primeiro estudo em que se empregam métodos longitudinais para se investigar a conexão entre salários e o grau de feminização das firmas.

Para esse fim, utilizamos um modelo de painel dinâmico, que é o nosso método mais geral. Na seção 1.8, implementamos diversos outros modelos longitudinais para checar a

robustez dos resultados desta seção. Considere, então, a seguinte equação de salários:

$$y_{ijt} = x_{1it}\beta_1 + x_{2i}\beta_2 + z_{1jt}\gamma_1 + z_{2j}\gamma_2 + \theta_i^* + \phi_j^* + \varepsilon_{ijt} \quad (1.57)$$

$$\varepsilon_{ijt} = \rho\varepsilon_{ij,t-1} + v_{ijt}, \quad (1.58)$$

onde y_{ijt} é o log do salário do indivíduo $i = 1, \dots, N$ que está no estabelecimento $j = 1, \dots, J$ no ano $t = 1, \dots, T$. As variáveis x_{1it} e x_{2i} representam as características observadas do indivíduo i que são, respectivamente, variantes e invariantes no tempo, e as variáveis z_{1jt} e z_{2j} compreendem, respectivamente, as características observadas variantes e invariantes no tempo do estabelecimento j . Os termos θ_i^* e ϕ_j^* denotam os efeitos específicos não-mensurados do trabalhador i e do estabelecimento j . O componente ε_{ijt} é um termo de distúrbio aleatório que, por hipótese, segue um processo AR(1), com $|\rho| < 1$. Assume-se que v_{ijt} é um termo de erro com média zero, não correlacionado entre trabalhadores, estabelecimentos e períodos de tempo, mas cuja variância pode ser aglomerada (*clustered*) para cada estabelecimento.

As variáveis incluídas em x_{1it} são a experiência geral dos trabalhadores, o seu valor quadrático e o tempo de experiência dos trabalhadores no estabelecimento. O nível de escolaridade e o grupo ocupacional dos trabalhadores são as variáveis incluídas em x_{2i} . As variáveis proporção de mulheres e tamanho do estabelecimento estão incorporadas em z_{1jt} , e z_{2j} contém o setor de atividade do estabelecimento. Em todas as especificações utilizadas, incluem-se variáveis *dummy* para o ano da observação.

Combinando as equações (1.57) e (1.58), pode-se obter a equação denominada de ‘fatores comuns’

$$\begin{aligned} y_{ijt} &= \rho y_{ij,t-1} + x_{1it}\beta_1 - x_{1i,t-1}\beta_1\rho + x_{2i}\beta_2(1-\rho) + z_{1jt}\gamma_1 - z_{1j,t-1}\gamma_1\rho + \\ &\quad + z_{2j}\gamma_2(1-\rho) + \theta_i^*(1-\rho) + \phi_j^*(1-\rho) + v_{ijt} \\ &= \rho y_{ij,t-1} + x_{1it}\beta_1 + x_{1i,t-1}\pi_1 + x_{2i}\pi_2 + z_{1jt}\gamma_1 + z_{1j,t-1}\varphi_1 + \\ &\quad + z_{2j}\varphi_2 + \theta_i + \phi_j + v_{ijt}, \end{aligned} \quad (1.59)$$

onde $\theta_i = \theta_i^*(1-\rho)$, $\phi_j = \phi_j^*(1-\rho)$, $\pi_1 = -\beta_1\rho$, $\pi_2 = \beta_2(1-\rho)$, $\varphi_1 = -\gamma_1\rho$, e $\varphi_2 = \gamma_2(1-\rho)$. Por causa da presença de variáveis defasadas nesta última equação, perde-se um período de tempo de observação, implicando que, agora, $t = 2, \dots, T$. Assume-se que v_{ijt} é distribuído independentemente de θ_i e de ϕ_j . Seja $w_{ijt} = (x_{1it}, z_{1jt})$. Supõe-se que w_{ijt} pode ser correlacionado com os efeitos individuais θ_i e ϕ_j , mas que a primeira diferença Δw_{ijt} não o é.

Por construção, os efeitos individuais θ_i e ϕ_j são necessariamente correlacionados com a variável dependente defasada $y_{ij,t-1}$, e, sob as hipótese feitas acima, também com w_{ijt} . Isso leva a que o estimador de MQO dos parâmetros da equação (1.59) seja inconsistente. Além disso, quando T é pequeno – que é o nosso caso –, o estimador de Intra-Grupos

(*Within-Groups*) também é inconsistente. Nessa situação, a solução tipicamente adotada na literatura para remover os termos individuais específicos é tomar a primeira diferença das variáveis da equação (1.59).

Entretanto, deve-se notar que, quando se utiliza a amostra para todos os trabalhadores, a aplicação da transformação de primeiras-diferenças à equação (1.59) *não* remove todos os efeitos individuais. De fato, quando os dados contêm trabalhadores que trocam de estabelecimento ao longo do tempo, a primeira diferença não elimina o termo ϕ_j para esses trabalhadores.⁴⁷ Uma possível solução para esse problema é utilizar somente os trabalhadores que permanecem no mesmo estabelecimento entre períodos de tempo. Formalmente, denote-se por $j = j(i, t)$ o estabelecimento no qual o trabalhador i estava empregado no período t . Definindo os trabalhadores que permaneceram no mesmo estabelecimento ao longo do tempo pelo conjunto $Permanecentes = \{i : j(i, t) = j(i, t - 1), t = 2, \dots, T\}$, a transformação de primeiras-diferenças passa a ser capaz de eliminar o efeito específico do estabelecimento da equação (1.59) para esse conjunto de trabalhadores. A amostra que contém os trabalhadores que pertencem a esse conjunto é a que será usada para conduzir as estimações baseadas nos métodos desta seção. A idéia de utilizar informações sobre trabalhadores que permanecem no mesmo estabelecimento para obter estimativas consistentes dos parâmetros foi proposta, num contexto similar ao nosso, por Abowd et al. (1999).

Considere a seguinte transformação de primeiras-diferenças da equação em nível (1.59):

$$\Delta y_{ijt} = \rho \Delta y_{ij,t-1} + \Delta x_{1it} \beta_1 + \Delta z_{1jt} \gamma_1 + \Delta z_{ij,t-1} \varphi_1 + \Delta v_{ijt}, \quad (1.60)$$

para $i = 1, \dots, I^*$, $j = 1, \dots, J^*$ e $t = 3, \dots, T$, onde o asterisco denota os trabalhadores e estabelecimentos pertencentes a amostra de permanentes.⁴⁸ Note que os efeitos específicos do trabalhador e do estabelecimento foram eliminados, o mesmo se aplicando aos termos invariantes no tempo x_{2i} e z_{2j} .

Uma vez que Δv_{ijt} é necessariamente correlacionado com $\Delta y_{ij,t-1}$ (e potencialmente com outros termos), o estimador de MQO é inconsistente. Contudo, estimativas consistentes dos parâmetros podem ser obtidas por meio do uso do método de variáveis instrumentais, com os requisitos fundamentais sendo que os instrumentos sejam correlacionados com as variáveis endógenas porém ortogonais a Δv_{ijt} .

Assume-se que z_{1jt} é pré-determinado no sentido de que z_{1jt} e v_{ijt} são não-correlaciona-

⁴⁷Por exemplo, suponha que existam trabalhadores que saíam do estabelecimento j' num determinado período e entraram no estabelecimento j'' no período seguinte. Nesse caso, a primeira diferença da equação (1.59) gerará o termo $(\phi_{j''} - \phi_{j'})$ para esses trabalhadores, não removendo, portanto, o efeito específico do estabelecimento.

⁴⁸Como a estimação é baseada nesta amostra, os coeficientes associados com as variáveis experiência no estabelecimento e sua defasada não são simultaneamente identificados na presença de um intercepto. O mesmo se aplica aos coeficientes das variáveis experiência geral, seu quadrado e suas respectivas defasadas. Assim, os coeficientes de $x_{1i,t-1}$ não são identificados na equação (1.59) e, por isso, a primeira diferença de $x_{1i,t-1}$ não foi incluída na equação (1.60).

dos, mas z_{1jt} pode ser correlacionado com $v_{ij,t-1}$ e choques anteriores.⁴⁹ Essa suposição leva a que o termo Δz_{1jt} seja correlacionado com Δv_{ijt} , implicando que, juntamente com $\Delta y_{ij,t-1}$, ele precisa ser instrumentalizado. Supõe-se que x_{1it} é estritamente exógeno, o que significa que as variáveis contidas nesse vetor (experiência geral e no estabelecimento) não são correlacionadas nem com v_{ijt} , nem com choques anteriores.

Supõe-se que y_{ij1} e z_{1j1} são não-correlacionados com v_{ijt} para $t = 2, \dots, T$, ou seja, assume-se que as condições iniciais são pré-determinadas. Essa suposição – juntamente com a hipótese de que v_{ijt} não têm correlação serial – implica que $y_{ij,t-2}$ para $t = 3, \dots, T$ são não correlacionados com o termo Δv_{ijt} , o que permite utilizá-los como instrumentos para $\Delta y_{ij,t-1}$ na equação (1.60). Dado que z_{1jt} é assumido como pré-determinado, em princípio o termo Δz_{1jt} na equação (1.60) poderia ser instrumentalizado por $z_{1j,t-1}$ e suas defasagens anteriores para $t = 3, \dots, T$. No entanto, dado que esses instrumentos também podem ser utilizados para instrumentalizar o termo $\Delta z_{1j,t-1}$, utiliza-se $z_{1j,t-1}$ para instrumentalizar Δz_{1jt} e $z_{1j,t-2}$ para instrumentalizar $\Delta z_{1j,t-1}$. Como se assume x_{1it} é estritamente exógeno, o termo Δx_{1it} na equação (1.60) é instrumentalizado por x_{1it} e suas defasagens para $t = 3, \dots, T$.

A estimação da equação de primeiras-diferenças (1.60) é implementada por meio do Método de Momentos Generalizados (MMG), tal como proposto por [Arellano & Bond \(1991\)](#). Nessa estimação, permitimos a presença de aglomeração (*clustering*) ao nível do estabelecimento.

[Blundell & Bond \(1998\)](#) e [Blundell & Bond \(2000\)](#) mostram que, sob certas hipóteses acerca do processo gerador das condições iniciais do modelo, as propriedades do estimador de MMG em primeiras-diferenças podem ser aprimoradas.⁵⁰ Esses autores propõem um estimador de MMG que combina a equação em primeiras-diferenças (1.60) com a equação em nível (1.59), onde as variáveis endógenas nesta última equação são instrumentalizadas utilizando-se suas diferenças defasadas, adequadamente escolhidas.⁵¹ Eles demonstram

⁴⁹ Isto é:

$$E[v_{ij,t+k}|z_{1jt}] \begin{cases} = 0 & \text{se } k \geq 0 \\ \neq 0 & \text{caso contrário.} \end{cases}$$

⁵⁰ Considere o modelo simples: (a) $y_{it} = \rho y_{i,t-1} + \beta x_{it} + \eta_i + \epsilon_{it}$; e, (b) $x_{it} = \alpha x_{i,t-1} + \delta \eta_i + u_{it}$. Por substituição repetida da segunda equação em primeiras-diferenças, pode-se obter: $\Delta x_{it} = \alpha^{t-2} \Delta x_{i2} + \sum_{s=0}^{t-3} \alpha^s \Delta u_{i,t-s}$, implicando que Δx_{it} é não-correlacionado com η_i se e somente se Δx_{i2} for não-correlacionado com η_i . A restrição sobre a condição inicial $E\left[\left(x_{i1} - \frac{\delta \eta_i}{1-\alpha}\right) \delta \eta_i\right] = 0$ garante que $E[\Delta x_{i2} \eta_i] = 0$. Similarmente, pode-se obter que: $\Delta y_{it} = \rho^{t-2} \Delta y_{i2} + \sum_{s=0}^{t-3} \rho^s (\beta \Delta x_{i,t-s} + \Delta \epsilon_{i,t-s})$, implicando que Δy_{it} é não-correlacionado com η_i se e somente se Δy_{i2} for não-correlacionado com η_i . A restrição sobre a condição inicial $E\left[\left(y_{i1} - \frac{\beta \delta \eta_i + \eta_i}{1-\rho}\right) \eta_i\right] = 0$ garante que $E[\Delta y_{i2} \eta_i] = 0$. Essas duas restrições são satisfeitas sob a hipótese de estacionaridade na média para os processos x_{it} e y_{it} . Ver os artigos citados para maiores detalhes sobre essas restrições.

⁵¹ [Arellano & Bover \(1995\)](#) também propuseram o uso de diferenças temporais como instrumentos para variáveis endógenas da equação em nível. Ver as notas de rodapé nas tabelas com os resultados para a especificação dos instrumentos empregados na equação em nível.

que a performance do estimador de MMG do modelo AR(1) em primeiras-diferenças pode ser bastante melhorada com o uso desse estimador estendido, denominado por eles de estimador de MMG em sistema. O uso desse estimador é particularmente apropriado quando o processo gerador dos dados está próximo de um passeio-aleatório, caso este em que as variáveis em nível tornam-se instrumentos fracos para as suas primeiras-diferenças.

Em suma, a estimação do modelo descrito pelas equações (1.59) e (1.60) é realizada empregando-se ambas as abordagens de MMG em diferenças e MMG em sistema. Este último método é implementado por meio dos procedimentos conhecidos como MMG em um-passo e MMG em dois-passos.⁵² A estimação é operacionalizada separadamente por sexo e por sexo e grupo ocupacional. Uma vez que dispomos de apenas três períodos de tempo, o modelo torna-se exatamente identificado, o que impede a aplicação dos testes usuais de sobre-identificação (*e.g.* o teste de Sargan/Hansen) e dos testes de especificação $m1$ e $m2$ propostos por Arellano & Bond (1991). Apesar disso, reportamos o teste de ‘fatores comuns’ para cada parâmetro associado às variáveis em $z_{1j,t-1}$ e para todos os parâmetros conjuntamente.

1.7 Resultados

1.7.1 Por Sexo

A Tabela 1.5 reporta as estimativas obtidas para os parâmetros do modelo especificado pelas equações (1.59) e (1.60). As colunas 1 e 4 apresentam os resultados do MMG em primeiras-diferenças respectivamente para mulheres e homens. As colunas 2 e 5 reportam as estimativas do MMG em sistema seguindo o procedimento em um-passo, enquanto as colunas 3 e 6 mostram as estimativas do MMG em sistema em dois-passos para mulheres e homens, respectivamente.

Como revela a Tabela 1.5, as estimativas do coeficiente associado à variável proporção de mulheres no estabelecimento são negativas e estatisticamente significativas ao nível de 1% para o MMG em sistema. As estimativas pontuais são também negativas para o MMG em diferenças, porém não diferentes de zero do ponto de vista estatístico. Esses resultados indicam que, se existe algum efeito do grau de feminização do estabelecimento sobre os salários de mulheres e homens, então esse efeito parece ser negativo. Tomando os resultados do MMG em sistema, pode-se notar que os salários dos homens são mais negativamente afetados pelo aumento da segregação feminina do que os salários das mulheres. Isso implica que o efeito de interesse tende a reduzir o diferencial salarial por gênero.

⁵²Quando $T = 3$ – que é o nosso caso –, o MMG em primeiras-diferenças é exatamente identificado, o que implica que não há distinção entre as estimativas geradas por esses dois procedimentos. O MMG em dois-passos utiliza estimativas consistentes dos parâmetros obtidas pelo procedimento em um-passo para construir a matriz ótima de pesos que é usada na minimização da função objetivo do MMG.

Tabela 1.5: Estimativas do Modelo Dinâmico por Sexo

Covariadas	Mulheres			Homens		
	DIF-MMG	SIS-MMG	SIS-MMG	DIF-MMG	SIS-MMG	SIS-MMG
	1-Passo (1)	1-Passo (2)	2-Passos (3)	1-Passo (4)	1-Passo (5)	2-Passos (6)
$\log(\text{salário})_{t-1}$	0.1001* (0.0543)	0.1874*** (0.0496)	0.3515*** (0.0456)	0.0052 (0.0417)	0.0728* (0.0389)	0.1691*** (0.0369)
Proporção de mulheres	-0.0886 (0.2042)	-0.0891*** (0.0247)	-0.1056*** (0.0229)	-0.0267 (0.0372)	-0.1975*** (0.0313)	-0.2136*** (0.0271)
Proporção de mulheres $_{t-1}$	-0.0111 (0.0463)	-0.1768*** (0.0257)	-0.1150*** (0.0235)	-0.0213 (0.0367)	-0.1533*** (0.0279)	-0.1039*** (0.0235)
Tamanho estabelecimento (1/10)	-0.0054*** (0.0007)	-0.0004*** (0.0000)	-0.0001** (0.0000)	-0.0067*** (0.0004)	-0.0005*** (0.0001)	-0.0000 (0.0000)
Tamanho estabelecimento $_{t-1}$ (1/10)	0.0004*** (0.0001)	0.0004*** (0.0000)	0.0001 (0.0000)	0.0002** (0.0001)	0.0008*** (0.0001)	0.0002*** (0.0000)
Experiência geral	-0.0299*** (0.0056)	0.0082*** (0.0009)	0.0066*** (0.0009)	-0.0291*** (0.0052)	0.0227*** (0.0013)	0.0204*** (0.0012)
Experiência geral ao quadrado	-0.0001 (0.0001)	-0.0001*** (0.0000)	-0.0001*** (0.0000)	-0.0001 (0.0001)	-0.0003*** (0.0000)	-0.0002*** (0.0000)
Experiência no estabelecimento		0.0034*** (0.0002)	0.0027*** (0.0002)		0.0034*** (0.0001)	0.0030*** (0.0001)
Escolaridade		0.0778*** (0.0048)	0.0617*** (0.0045)		0.0934*** (0.0040)	0.0831*** (0.0038)
Comércio		-0.1271*** (0.0085)	-0.1063*** (0.0081)		-0.2882*** (0.0127)	-0.2599*** (0.0121)
Serviços		0.0052 (0.0051)	0.0015 (0.0051)		-0.0827*** (0.0053)	-0.0798*** (0.0053)
Administrativos		0.1589*** (0.0100)	0.1288*** (0.0093)		0.0861*** (0.0060)	0.0778*** (0.0059)
Profissionais		0.6510*** (0.0396)	0.5246*** (0.0365)		0.5337*** (0.0231)	0.4810*** (0.0219)
Gerentes		0.4905*** (0.0302)	0.3967*** (0.0281)		0.5723*** (0.0252)	0.5151*** (0.0241)
Intercepto		-0.1359*** (0.0166)	-0.0960*** (0.0163)		-0.0379** (0.0130)	-0.0189 (0.0128)
Teste de Fatores Comuns: (P-valor)						
Proporção de mulheres	0.7479	0.0000	0.0000	0.5604	0.0000	0.0000
Tamanho do estabelecimento	0.6445	0.0000	0.3551	0.5914	0.0000	0.0000
Todos os parâmetros	0.8825	0.0000	0.0000	0.7105	0.0000	0.0000
Número de observações	60355	184549	184549	100907	294384	294384

Notas:

1. A variável dependente é o log do salário horário real.
2. Erros-padrões entre parênteses.
3. Níveis de significância: *:10% **:5% ***:1%.
4. Todas as especificações incluem dummies de ano.
5. O teste de 'fatores comuns' é um teste de restrições não-lineares: P-valores reportados.
6. Os instrumentos utilizados são:

DIF-MMG: $y_{ij,t-2}, x_{1it}$ e defasagens, $z_{1j,t-1}$.

SIS-MMG: Eq. em diferenças: $y_{ij,t-2}, x_{1it}, z_{1j,t-1}$; Eq. em nível: $\Delta y_{ij,t-1}, \Delta x_{1it}, x_{2i}, \Delta z_{1jt}, z_{2j}$.

O teste de fatores comuns rejeita as restrições sob os parâmetros para o MMG em sistema, o mesmo não ocorrendo para o MMG em diferenças. Existe, portanto, alguma evidência de que a especificação AR(1) não representa o modelo correto. A inexistência de períodos adicionais de tempo na amostra impede a implementação de outros testes (*e.g.* $m1$ e $m2$) que permitem confirmar se o termo de erro na equação (1.58) está de fato mal especificado.

Em relação às demais covariadas, os resultados mostram que o tamanho do estabelecimento tem um pequeno efeito negativo sobre os salários de ambos os sexos. Como esperado, a experiência geral no mercado de trabalho aumenta os salários porém a uma taxa decrescente.⁵³ O efeito da experiência no estabelecimento é positivo e de magnitude similar para os dois sexos. O impacto da escolaridade tem o sinal positivo esperado, sendo um pouco mais elevado para homens do que para mulheres. Trabalhar na indústria (o setor de referência na regressão) eleva o salário dos homens, porém esse incremento salarial só aparece para as mulheres em comparação com o setor de comércio. Estar em ocupações ligadas à produção (a categoria ocupacional de referência na regressão) tende a reduzir o salário de ambos os sexos. É interessante notar que os homens parecem ter, em média, salários mais altos em ocupações de gerência, enquanto as mulheres em ocupações profissionais.

1.7.2 Por Sexo e Grupo Ocupacional

Nesta sub-seção, apresentamos as estimativas separadas por sexo para os quatro grupos ocupacionais com os quais estamos trabalhando: ocupações ligadas à produção, administrativas, profissionais e de gerência. Obter resultados para distintas categorias ocupacionais é importante, na medida em que o efeito de interesse pode ser bastante diferente entre categorias mais ou menos femininas.⁵⁴

A Tabela 1.6 apresenta os resultados para os trabalhadores ligados à produção. Como no caso da amostra não estratificada por ocupação, as estimativas do coeficiente de interesse são negativas e estatisticamente significativas para o MMG em sistema; para o MMG em diferenças, elas são somente pontualmente negativas. É interessante observar que esse efeito parece ser substancialmente maior sobre os salários dos homens do que das mulheres. As restrições de fatores comuns não são rejeitadas para o MMG em sistema em um-passo para as mulheres e para o MMG em diferenças para os homens.

A Tabela 1.7 contém as estimativas obtidas para os trabalhadores em ocupações administrativas. Para o caso dos salários das mulheres, os resultados mostram que o efeito de trabalhar em estabelecimentos mais femininos é negativo e estatisticamente significativo tanto para o MMG em diferenças, quanto para o MMG em sistema. Embora pontualmente negativos para todos os métodos, esse efeito só se mostra estatisticamente diferente de zero para os salários dos homens no caso do MMG em diferenças. O teste de fatores comuns não rejeita as restrições sob os parâmetros para o MMG em diferenças para ambos os sexos.

A Tabela 1.8, que apresenta os resultados para ocupações profissionais, mostra que o efeito de interesse é negativo e estatisticamente significativo sobre os salários das mulheres

⁵³Dado que a variação na experiência geral resulta da mudança na idade dos indivíduos entre os anos, a primeira diferença dessa variável desempenha o papel do intercepto.

⁵⁴A proporção de mulheres nos estabelecimentos para as categorias ocupacionais são: ligadas à produção, 34,3%; administrativas, 58,2%; profissionais, 35,3%; gerência, 32,2%.

Tabela 1.6: Estimativas do Modelo Dinâmico por Sexo: Ocupações Ligadas à Produção

Covariadas	Mulheres			Homens		
	DIF-MMG	SIS-MMG	SIS-MMG	DIF-MMG	SIS-MMG	SIS-MMG
	1-Passo (1)	1-Passo (2)	2-Passos (3)	1-Passo (4)	1-Passo (5)	2-Passos (6)
$\log(\text{salário})_{t-1}$	0.1120*** (0.0470)	0.0559 (0.0437)	0.0483 (0.0409)	0.0298 (0.0286)	0.0215 (0.0276)	0.0327 (0.0272)
Proporção de mulheres	-0.3065 (0.2906)	-0.2239*** (0.0399)	-0.2125*** (0.0391)	-0.2437 (0.3631)	-0.4916*** (0.0400)	-0.4978*** (0.0399)
Proporção de mulheres $_{t-1}$	-0.0316 (0.0594)	0.0121 (0.0365)	0.0007 (0.0362)	-0.0895 (0.0636)	-0.1173*** (0.0348)	-0.1119*** (0.0347)
Tamanho estabelecimento (1/10)	-0.0018* (0.0007)	0.0002*** (0.0000)	0.0002*** (0.0000)	-0.0041*** (0.0006)	0.0006*** (0.0001)	0.0008*** (0.0001)
Tamanho estabelecimento $_{t-1}$ (1/10)	-0.0001 (0.0001)	-0.0001 (0.0000)	-0.0001** (0.0000)	-0.0001 (0.0001)	-0.0001 (0.0001)	-0.0003*** (0.0001)
Experiência geral	-0.0364*** (0.0079)	-0.0021 (0.0011)	-0.0017 (0.0011)	-0.0269*** (0.0062)	0.0181*** (0.0011)	0.0181*** (0.0011)
Experiência geral ao quadrado	0.0000 (0.0001)	0.0000* (0.0000)	0.0000 (0.0000)	-0.0002 (0.0001)	-0.0002*** (0.0000)	-0.0002*** (0.0000)
Experiência no estabelecimento		0.0030*** (0.0001)	0.0031*** (0.0001)		0.0034*** (0.0001)	0.0034*** (0.0001)
Escolaridade		0.0532*** (0.0027)	0.0534*** (0.0025)		0.0617*** (0.0020)	0.0608*** (0.0020)
Comércio		-0.0569*** (0.0070)	-0.0573*** (0.0070)		-0.2235*** (0.0085)	-0.2220*** (0.0084)
Serviços		-0.0888*** (0.0071)	-0.0916*** (0.0070)		-0.1343*** (0.0062)	-0.1355*** (0.0061)
Intercepto		0.2942*** (0.0240)	0.3038*** (0.0233)		0.3548*** (0.0180)	0.3623*** (0.0177)
Teste de Fatores Comuns: (P-valor)						
Proporção de mulheres	0.4522	0.9906	0.7852	0.1844	0.0005	0.0004
Tamanho do estabelecimento	0.0276	0.0958	0.0064	0.0706	0.1950	0.0001
Todos os parâmetros	0.0633	0.2479	0.0234	0.1694	0.0005	0.0000

Número de observações	32791	103056	103056	67245	199570	199570
-----------------------	-------	--------	--------	-------	--------	--------

Notas:

1. A variável dependente é o log do salário horário real.
2. Erros-padrões entre parênteses.
3. Níveis de significância: *10% **5% ***1%.
4. Todas as especificações incluem dummies de ano.
5. O teste de 'fatores comuns' é um teste de restrições não-lineares: P-valores reportados.
6. Os instrumentos utilizados são:

DIF-MMG: $y_{ij,t-2}, x_{1it}$ e defasagens, $z_{1j,t-1}$.

SIS-MMG: Eq. em diferenças: $y_{ij,t-2}, x_{1it}, z_{1j,t-1}$; Eq. em nível: $\Delta y_{ij,t-1}, \Delta x_{1it}, x_{2i}, \Delta z_{1jt}, z_{2j}$.

para os três modelos utilizados. No caso dos homens, esse efeito é negativo e estatisticamente diferente de zero para ambos os procedimentos do MMG em sistema. Diferentemente dos resultados anteriores, a segregação feminina parece impactar mais negativamente o salário das mulheres do que o dos homens em ocupações profissionais. O teste de fatores comuns não rejeita a hipótese nula para as regressões das mulheres (marginalmente no MMG em sistema em dois-passos) e para as dos homens no caso do MMG em sistema em um-passo (também marginalmente).

Finalmente, a Tabela 1.9 mostra as estimativas obtidas para ocupações de nível gerencial. Distintamente dos casos anteriores, o efeito da composição feminina dos estabelecimentos, embora pontualmente negativo, não é estatisticamente diferente de zero sobre os

Tabela 1.7: Estimativas do Modelo Dinâmico por Sexo: Ocupações Administrativas

Covariadas	Mulheres			Homens		
	DIF-MMG	SIS-MMG	SIS-MMG	DIF-MMG	SIS-MMG	SIS-MMG
	1-Passo (1)	1-Passo (2)	2-Passos (3)	1-Passo (4)	1-Passo (5)	2-Passos (6)
$\log(\text{salário})_{t-1}$	0.3025*** (0.0681)	0.3502*** (0.0650)	0.3919*** (0.0642)	0.1282 (0.0815)	0.4760*** (0.0889)	0.5450*** (0.0873)
Proporção de mulheres	-0.1467** (0.0736)	-0.3015*** (0.0562)	-0.2641*** (0.0553)	-0.2555*** (0.0774)	-0.0506 (0.0536)	-0.0310 (0.0453)
Proporção de mulheres $_{t-1}$	-0.0331 (0.0589)	0.0111 (0.0498)	-0.0046 (0.0495)	-0.1697 (0.1149)	0.0468 (0.0535)	0.0216 (0.0458)
Tamanho estabelecimento (1/10)	-0.0055*** (0.0012)	0.0004*** (0.0001)	0.0004*** (0.0001)	-0.0063*** (0.0017)	0.0004*** (0.0001)	0.0004*** (0.0001)
Tamanho estabelecimento $_{t-1}$ (1/10)	0.0011*** (0.0003)	-0.0005*** (0.0001)	-0.0004*** (0.0001)	0.0011** (0.0004)	-0.0004*** (0.0001)	-0.0005*** (0.0001)
Experiência geral	-0.0578*** (0.0112)	0.0198*** (0.0026)	0.0173*** (0.0026)	-0.0482*** (0.0125)	0.0198*** (0.0041)	0.0165*** (0.0040)
Experiência geral ao quadrado	0.0005* (0.0002)	-0.0003*** (0.0000)	-0.0003*** (0.0000)	0.0002 (0.0003)	-0.0003*** (0.0001)	-0.0003*** (0.0001)
Experiência no estabelecimento		0.0031*** (0.0003)	0.0029*** (0.0003)		0.0026*** (0.0004)	0.0023*** (0.0004)
Escolaridade		0.0798*** (0.0080)	0.0742*** (0.0079)		0.0633*** (0.0107)	0.0541*** (0.0104)
Comércio		-0.1179*** (0.0151)	-0.1140*** (0.0150)		-0.1221*** (0.0243)	-0.1098*** (0.0239)
Serviços		0.0693*** (0.0138)	0.0625*** (0.0136)		0.0201 (0.0123)	0.0109 (0.0122)
Intercepto		-0.3345*** (0.0462)	-0.2872*** (0.0449)		-0.3082*** (0.0611)	-0.2328*** (0.0595)
Testes de Fatores Comuns: (P-valor)						
Proporção de mulheres	0.2531	0.0040	0.0003	0.0748	0.4721	0.8506
Tamanho do estabelecimento	0.1260	0.0000	0.0000	0.5935	0.0025	0.0014
Todos os parâmetros	0.2007	0.0000	0.0000	0.1958	0.0103	0.0051

Número de observações	17566	52450	52450	13060	37898	37898
-----------------------	-------	-------	-------	-------	-------	-------

Notas:

1. A variável dependente é o log do salário horário real.
2. Erros-padrões entre parênteses.
3. Níveis de significância: *:10% **:5% ***:1%.
4. Todas as especificações incluem dummies de ano.
5. O teste de 'fatores comuns' é um teste de restrições não-lineares: P-valores reportados.
6. Os instrumentos utilizados são:
DIF-MMG: $y_{ij,t-2}, x_{1it}$ e defasagens, $z_{1j,t-1}$.
SIS-MMG: Eq. em diferenças: $y_{ij,t-2}, x_{1it}, z_{1j,t-1}$; Eq. em nível: $\Delta y_{ij,t-1}, \Delta x_{1it}, x_{2i}, \Delta z_{1jt}, z_{2j}$.

salários das mulheres em nenhum dos três métodos empregados. Entretanto, esse efeito é negativo e estatisticamente significativo para os homens em posições de gerência nesses três métodos, o que implica que a segregação feminina entre estabelecimentos tende a reduzir o hiato salarial por gênero para esse tipo de ocupação. As restrições de fatores comuns são aceitas para todos os métodos no caso dos homens e para o MMG em diferenças no caso das mulheres.

Em suma, embora os testes de fatores comuns nem sempre aceitem a especificação AR(1) do modelo empregado, os resultados obtidos acima indicam que o efeito sobre os salários da segregação feminina no nível do estabelecimento é negativo para as mulheres (talvez à exceção de mulheres gerentes) e para os homens. No que se segue, estimamos

Tabela 1.8: Estimativas do Modelo Dinâmico por Sexo: Ocupações Profissionais

Covariadas	Mulheres			Homens		
	DIF-MMG	SIS-MMG	SIS-MMG	DIF-MMG	SIS-MMG	SIS-MMG
	1-Passo (1)	1-Passo (2)	2-Passos (3)	1-Passo (4)	1-Passo (5)	2-Passos (6)
$\log(\text{salário})_{t-1}$	0.0546 (0.0528)	0.1107** (0.0537)	0.1327*** (0.0527)	0.0921 (0.0602)	-0.0082 (0.0572)	0.0079 (0.0501)
Proporção de mulheres	-0.2603*** (0.1113)	-0.4211*** (0.1562)	-0.4928*** (0.1171)	0.0177 (0.1067)	-0.2167** (0.0989)	-0.3488*** (0.0977)
Proporção de mulheres $_{t-1}$	0.1923 (0.1348)	-0.1773 (0.1466)	-0.1043 (0.1074)	0.3539** (0.1152)	0.1922* (0.0919)	0.3226*** (0.0905)
Tamanho estabelecimento (1/10)	-0.0103*** (0.0013)	-0.0007 (0.0006)	-0.0008 (0.0004)	-0.0108*** (0.0009)	0.0003 (0.0002)	-0.0002 (0.0002)
Tamanho estabelecimento $_{t-1}$ (1/10)	0.0008 (0.0006)	0.0006 (0.0006)	0.0006 (0.0004)	-0.0034** (0.0013)	-0.0001 (0.0002)	0.0003* (0.0002)
Experiência geral	0.0056 (0.0151)	0.0087** (0.0043)	0.0088** (0.0041)	-0.0486*** (0.0133)	0.0266*** (0.0034)	0.0242*** (0.0032)
Experiência geral ao quadrado	-0.0009* (0.0004)	0.0000 (0.0001)	0.0000 (0.0001)	0.0002 (0.0003)	-0.0003*** (0.0001)	-0.0002*** (0.0001)
Experiência no estabelecimento		0.0024*** (0.0002)	0.0023*** (0.0002)		0.0025*** (0.0002)	0.0025*** (0.0001)
Escolaridade		0.1238*** (0.0085)	0.1207*** (0.0085)		0.1539*** (0.0089)	0.1515*** (0.0078)
Comércio		-0.2954*** (0.0367)	-0.2850*** (0.0365)		-0.5706*** (0.0420)	-0.5409*** (0.0385)
Serviços		-0.0131 (0.0259)	-0.0074 (0.0258)		-0.1558*** (0.0172)	-0.1533*** (0.0171)
Intercepto		0.3281*** (0.0768)	0.3281*** (0.0738)		-0.0138 (0.0459)	0.0056 (0.0445)
Teste de Fatores Comuns: (P-valor)						
Proporção de mulheres	0.2002	0.0875	0.0726	0.0023	0.0416	0.0007
Tamanho do estabelecimento	0.7664	0.3444	0.1591	0.0060	0.6064	0.0404
Todos os parâmetros	0.4400	0.1507	0.0624	0.0001	0.0836	0.0013

Número de observações	6353	17924	17924	12526	33292	33292
-----------------------	------	-------	-------	-------	-------	-------

Notas:

1. A variável dependente é o log do salário horário real.
2. Erros-padrões entre parênteses.
3. Níveis de significância: *10% **5% ***1%.
4. Todas as especificações incluem dummies de ano.
5. O teste de 'fatores comuns' é um teste de restrições não-lineares: P-valores reportados.
6. Os instrumentos utilizados são:

DIF-MMG: $y_{ij,t-2}, x_{1it}$ e defasagens, $z_{1j,t-1}$.

SIS-MMG: Eq. em diferenças: $y_{ij,t-2}, x_{1it}, z_{1j,t-1}$; Eq. em nível: $\Delta y_{ij,t-1}, \Delta x_{1it}, x_{2i}, \Delta z_{1jt}, z_{2j}$.

diversos outros modelos para a checar a robustez dessas conclusões.

1.8 Robustez dos Resultados

Nosso primeiro teste de robustez foi implementar o modelo dinâmico excluindo os estabelecimentos que eram totalmente masculinos ou femininos. Esse teste não produziu nenhuma mudança qualitativa importante em relação aos resultados apresentados na seção anterior. No segundo teste, excluímos os estabelecimentos com menos de dez empregados, porém resultados similares foram mais uma vez obtidos. Finalmente, empregamos o MMG em sistema utilizando a amostra com o painel balanceado – isto é, trabalhadores com exatamente

Tabela 1.9: Estimativas do Modelo Dinâmico por Sexo: Ocupações Gerenciais

Covariadas	Mulheres			Homens		
	DIF-MMG	SIS-MMG	SIS-MMG	DIF-MMG	SIS-MMG	SIS-MMG
	1-Passo (1)	1-Passo (2)	2-Passos (3)	1-Passo (4)	1-Passo (5)	2-Passos (6)
$\log(\text{salário})_{t-1}$	-0.2930 (0.3601)	-0.0832 (0.2967)	0.1137 (0.1896)	-0.3176 (0.1866)	-0.1391 (0.1751)	0.1216 (0.1375)
Proporção de mulheres	-0.0692 (0.5236)	-0.1281 (0.2300)	-0.0308 (0.1986)	-0.2730** (0.1240)	-0.2994** (0.1484)	-0.1711* (0.1024)
Proporção de mulheres $_{t-1}$	-0.4876** (0.2184)	-0.4431** (0.2107)	-0.4400*** (0.1770)	-0.0187 (0.1325)	0.1386 (0.1387)	0.0667 (0.0928)
Tamanho estabelecimento (1/10)	-0.0141 (0.0292)	0.0000 (0.0003)	0.0001 (0.0003)	-0.0110** (0.0037)	0.0005 (0.0005)	0.0005*** (0.0001)
Tamanho estabelecimento $_{t-1}$ (1/10)	-0.0002 (0.0008)	-0.0004 (0.0002)	-0.0004 (0.0002)	-0.0022** (0.0007)	-0.0002 (0.0005)	-0.0002 (0.0001)
Experiência geral	-0.0170 (0.0269)	0.0124* (0.0066)	0.0094 (0.0058)	-0.0260 (0.0216)	0.0479*** (0.0094)	0.0365*** (0.0079)
Experiência geral ao quadrado	-0.0004 (0.0005)	-0.0001 (0.0001)	-0.0001 (0.0001)	-0.0006 (0.0004)	-0.0005*** (0.0001)	-0.0004*** (0.0001)
Experiência no estabelecimento		0.0043*** (0.0011)	0.0037*** (0.0007)		0.0034*** (0.0005)	0.0027*** (0.0004)
Escolaridade		0.1593*** (0.0429)	0.1314*** (0.0279)		0.2164*** (0.0331)	0.1675*** (0.0262)
Comércio		-0.0960* (0.0435)	-0.0758* (0.0368)		-0.4488*** (0.0756)	-0.3468*** (0.0623)
Serviços		0.2672** (0.0816)	0.2169*** (0.0564)		-0.1673*** (0.0339)	-0.1335*** (0.0309)
Intercepto		-0.1695 (0.0926)	-0.1544 (0.0905)		-0.7112*** (0.1286)	-0.5675*** (0.1086)
Teste de Fatores Comuns: (P-valor)						
Proporção de mulheres	0.0835	0.0496	0.0048	0.5985	0.2718	0.5822
Tamanho do estabelecimento	0.7204	0.0991	0.0420	0.5833	0.6985	0.1386
Todos os parâmetros	0.2226	0.0482	0.0039	0.7905	0.5174	0.2994

Número de observações

3645

11119

11119

8076

23624

23624

Notas:

1. A variável dependente é o log do salário horário real.

2. Erros-padrões entre parênteses.

3. Níveis de significância: *10% **5% ***1%.

4. Todas as especificações incluem dummies de ano.

5. O teste de 'fatores comuns' é um teste de restrições não-lineares: P-valores reportados.

6. Os instrumentos utilizados são:

DIF-MMG: $y_{ij,t-2}, x_{1it}$ e defasagens, $z_{1j,t-1}$.SIS-MMG: Eq. em diferenças: $y_{ij,t-2}, x_{1it}, z_{1j,t-1}$; Eq. em nível: $\Delta y_{ij,t-1}, \Delta x_{1it}, x_{2i}, \Delta z_{1jt}, z_{2j}$.

três observações –, mas as estimativas praticamente não se alteraram.

Dadas as pequenas modificações ocorridas a partir dos testes baseados no modelo dinâmico, implementamos três outros métodos para confirmar nossos resultados anteriores: os modelos Intra-Grupos e em Primeiras-Diferenças (sem dinâmica) e o modelo conhecido como Multi-nível. Embora menos gerais que o modelo dinâmico, esses métodos são largamente empregados em análises com dados de painel, podendo fornecer informações úteis para a checagem de resultados. Além de utilizarmos esses três métodos para a amostra de trabalhadores que permaneceram no mesmo estabelecimento ao longo do tempo, aplicamos também para a amostra com todos os trabalhadores. Na realidade, como detalhado mais adiante, utilizamos uma especificação mais geral para o efeitos específicos não-observáveis

para implementar o modelo de Intra-Grupos para esta última amostra. Apresentaremos primeiramente os resultados baseados na amostra de trabalhadores permanentes e, então, os obtidos com a amostra completa.

1.8.1 Trabalhadores Permanentes

1.8.1.1 Intra-Grupos e Primeiras-Diferenças

Considere a seguinte equação:

$$y_{ijt} = x_{1it}\beta_1 + x_{2i}\beta_2 + z_{1jt}\gamma_1 + z_{2j}\gamma_2 + \theta_i^* + \phi_j^* + \varepsilon_{ijt}, \quad (1.61)$$

a qual é idêntica a equação (1.57). Supõe-se agora que ε_{ijt} é um termo de erro não correlacionado com os regressores dessa equação, ou seja, assume-se que todas as covariadas são estritamente exógenas. Semelhantemente ao assumido anteriormente, permite-se que variância de ε_{ijt} seja aglomerada por estabelecimento. Ainda utilizando a amostra com trabalhadores que permaneceram nos mesmos estabelecimentos entre períodos, se aplicarmos a transformação de Intra-Grupos à equação (1.61), obtemos:

$$\tilde{y}_{ijt} = \tilde{x}_{1it}\beta_1 + \tilde{z}_{1jt}\gamma_1 + \tilde{\varepsilon}_{ijt}, \quad (1.62)$$

onde a notação do ‘til’ representa a variável original subtraída da sua média temporal (*e.g.* $\tilde{y}_{ijt} = y_{ijt} - T^{-1} \sum_{t=1}^T y_{ijt}$). Note-se que todos os termos invariantes no tempo foram eliminados pela transformação de Intra-Grupos, incluindo os efeitos específicos de trabalhadores e estabelecimentos. Dada a hipótese de exogeneidade estrita, o método de Intra-Grupos provê um estimador consistente para os parâmetros β_1 e γ_1 .

Tomando a primeira diferença da equação (1.61), tem-se que:

$$\Delta y_{ijt} = \Delta x_{1it}\beta_1 + \Delta z_{1jt}\gamma_1 + \Delta \varepsilon_{ijt}. \quad (1.63)$$

Com as hipóteses desta seção, o método de Primeiras-Diferenças também fornece um estimador consistente para β_1 e γ_1 . Apresentamos agora os resultados obtidos com base nesses dois métodos por sexo e grupo ocupacional.

A Tabela 1.10 contém as estimativas do coeficiente associado à variável proporção de mulheres no estabelecimento com base nas equações (1.62) e (1.63). Ela também reporta os erros-padrões desse coeficiente e o P-valor do teste-F para a significância de todos os parâmetros incluídos nas regressões. Os resultados mostram que o efeito de trabalhar em estabelecimentos predominantemente femininos é negativo e estatisticamente significativo (pelo menos a 10%) sobre os salários das mulheres em geral, das mulheres em ocupações ligadas à produção e administrativas. As estimativas não são estatisticamente diferentes de zero para as mulheres nas demais categorias ocupacionais, assim como para os homens.

Tabela 1.10: Estimativas dos Modelos de Intra-Grupos e de Primeiras-Diferenças para o Coeficiente da Variável Proporção de Mulheres no Estabelecimento por Sexo e Grupo Ocupacional: Trabalhadores Permanecentes

Ocupação/Sexo	Intra-Grupos				Primeiras-Diferenças			
	Coef. (1)	Erro-Padr. (2)	Teste-F (3)	Observ. (4)	Coef. (5)	Erro-Padr. (6)	Teste-F (7)	Observ. (8)
1.Todos								
1.1 - Mulheres	-0.0526**	0.0268	0.0000	308743	-0.0504*	0.0264	0.0000	184549
1.2 - Homens	0.0109	0.0352	0.0000	487861	0.0172	0.0359	0.0000	294384
2.Ligados à Prod.								
2.1 - Mulheres	-0.0517*	0.0305	0.0000	173321	-0.0563*	0.0319	0.0000	103056
2.2 - Homens	0.0498	0.0442	0.0000	331895	0.0578	0.0456	0.0000	199570
3.Administrativos								
3.1 - Mulheres	-0.0772*	0.0410	0.0033	87334	-0.0599	0.0406	0.0016	52450
3.2 - Homens	-0.0507	0.0599	0.0059	62736	-0.0110	0.0566	0.0002	37898
4.Profissionais								
4.1 - Mulheres	-0.0847	0.1421	0.3347	29495	-0.1609	0.1294	0.4702	17924
4.2 - Homens	-0.0387	0.0917	0.2775	54058	-0.1244	0.0911	0.0904	33292
5.Gerentes								
5.1 - Mulheres	0.0300	0.1121	0.0004	18593	0.0828	0.1256	0.0002	11119
5.2 - Homens	-0.0823	0.0787	0.0000	39172	-0.0768	0.0759	0.0000	23624

Notas:

1. A variável dependente é o log do salário horário real.

2. Níveis de significância: *:10% **:5% ***:1%.

3. P-valores reportados para o teste-F nas colunas (3) e (7).

4. As regressões de também incluem as variáveis experiência geral e seu quadrado, tamanho do estabelecimento, experiência no estabelecimento e dummies de ano.

Esses resultados parecem indicar que, se existe um efeito da segregação feminina no nível do estabelecimento sobre os salários, então, ele é negativo, pelo menos para o caso das mulheres.

1.8.1.2 Modelo Multi-Nível

Uma vez que temos um painel de trabalhadores empregados em diferentes estabelecimentos, pode-se ver a base de dados como estruturada de forma hierárquica em vários níveis. Mais especificamente, a base de dados pode ser vista como uma estrutura em três níveis consistindo de períodos de tempo (Nível 1) nos quais se observam distintos trabalhadores (Nível 2) que estão empregados em diferentes estabelecimentos (Nível 3). Esse formato da base de dados permite, então, a aplicação do que se denomina modelos multi-níveis hierárquicos. Tipicamente, em modelos desse tipo, alguns parâmetros são tomados como fixos, enquanto outros como aleatórios.⁵⁵ Ademais, a especificação desses modelos é geral-

⁵⁵Essa é a razão pela qual esses modelos são também referidos como modelos multi-níveis *mistos*.

mente operacionalizada por meio de relações condicionais entre os níveis, isto é, as relações nos níveis inferiores do modelo são condicionadas a coeficientes aleatórios (geralmente não-observáveis) dos níveis superiores.⁵⁶

A fim de apresentar de forma mais concreta esse modelo, podemos descrever a equação do nível-1 como:

$$y_{i,j,t} = \beta_{0ij} + \beta_{1ij}x_{i,j,t} + \gamma_{0j} + \gamma_{1j}z_{j,t} + \gamma_{2j}pfem_{j,t} + \epsilon_{i,j,t}, \quad (1.64)$$

onde $x_{i,j,t}$ e $z_{j,t}$ são vetores que contêm variáveis observáveis que variam e que não variam no tempo, $pfem_{j,t}$ representa a proporção de mulheres no estabelecimento j no período t , e $\epsilon_{i,j,t}$ é um termo de distúrbio aleatório normalmente distribuído com média zero e variância σ^2 . O vetor de parâmetros β_{0ij} e γ_{0j} são interceptos aleatórios, enquanto β_{1ij} , γ_{1j} e γ_{2j} são coeficientes angulares também aleatórios.

A especificação do nível-2 do modelo (trabalhadores) é a seguinte:

$$\begin{aligned} \beta_{0ij} &= \beta_0 + u_{0ij} \\ \beta_{1ij} &= \beta_1, \end{aligned} \quad (1.65)$$

onde β_0 e β_1 são parâmetros fixos e u_{0ij} é um termo aleatório que varia entre trabalhadores e estabelecimentos, e cujas média e variância são respectivamente zero e τ_0 .

Em relação ao nível-3 do modelo (estabelecimentos), utilizamos duas especificações diferentes. Na primeira, permitimos que somente γ_{0j} seja um termo aleatório, ao passo que na segunda ambos γ_{0j} e γ_{2j} são tratados como aleatórios. A primeira especificação da equação do nível-3 do modelo é dada por

$$\begin{aligned} \gamma_{0j} &= v_{0j} \\ \gamma_{1j} &= \gamma_1 \\ \gamma_{2j} &= \gamma_2, \end{aligned} \quad (1.66)$$

onde γ_1 e γ_2 são parâmetros fixos e v_{0j} é uma variável aleatória com média zero e variância τ_1 . A segunda especificação é dada por

$$\begin{aligned} \gamma_{0j} &= v_{0j}^* \\ \gamma_{1j} &= \gamma_1^* \\ \gamma_{2j} &= \gamma_2^* + v_{2j}, \end{aligned} \quad (1.67)$$

onde γ_1^* e γ_2^* são parâmetros fixos e v_{0j}^* e v_{2j} são variáveis aleatórias com média zero, variâncias τ_1^* e τ_2 , e covariância τ_{12} . Note que, na segunda especificação, permitimos que

⁵⁶Obras amplamente citadas na literatura sobre modelos multi-níveis são Bryk & Raudenbush (1992) e Goldenstein (2002). Uma introdução básica pode ser encontrada em Frees (2004).

o coeficiente aleatório associado à variável proporção de mulheres no estabelecimento (γ_{2j}) varie entre estabelecimentos.

Utilizando as equações (1.64), (1.65) e (1.66) (ou (1.67)), obtemos o modelo combinado:

$$y_{i,j,t} = \beta_0 + \beta_1 x_{i,j,t} + \gamma_1 z_{j,t} + \gamma_2 pfem_{j,t} + u_{0ij} + v_{0j} + \epsilon_{i,j,t} \quad (1.68)$$

ou

$$y_{i,j,t} = \beta_0^* + \beta_1^* x_{i,j,t} + \gamma_1^* z_{j,t} + (\gamma_2^* + v_{2j}) pfem_{j,t} + u_{0ij}^* + v_{0j}^* + \epsilon_{i,j,t}^*, \quad (1.69)$$

onde a equação (1.68) utiliza a primeira especificação do nível-3 e a equação (1.69) a segunda especificação. Utilizamos o asterisco para denotar que os termos da equação (1.69) não são necessariamente os mesmos que os seus correspondentes na equação (1.68). Estamos interessados em estimar os parâmetros fixos das equações acima (i.e. β_0 , β_1 , γ_1 e γ_2 , ou suas versões com asterisco), assim como as variâncias e covariância associadas aos termos aleatórios dessas equações (i.e. σ^2 , τ_0 , τ_1 , τ_2 e τ_{12} , ou suas versões com asterisco).

Menos geral que o arcabouço dinâmico da seção 1.6, o modelo multi-nível assume que os termos não-observáveis (e.g. u_{0ij}^* , v_{0j}^* , v_{2j} e $\epsilon_{i,j,t}^*$) são independentes das variáveis observadas.⁵⁷ Assume-se também que não há correlação entre os termos aleatórios entre os diferentes níveis do modelo.⁵⁸ Ademais, supõe-se que esses termos aleatórios têm uma distribuição normal multi-variada. A estimação é implementada via o método de Máxima Verossimilhança Concentrada, no qual os coeficientes fixos são estimados por mínimos quadrados e as variâncias e covariância dos termos aleatórios são obtidas por meio da maximização da função de verossimilhança em que os coeficientes fixos estimados por mínimos quadrados são substituídos nessa função. Para todas as especificações do modelo de multi-nível, reportamos o P-valor do teste da razão de verossimilhança, onde o modelo não-restrito é comparado ao restrito que impõe variâncias e covariância zero para os termos aleatórios.⁵⁹

A Tabela 1.11 apresenta as estimativas do modelo multi-nível por sexo para a amostra de trabalhadores que mantiveram-se nos mesmos estabelecimentos ao longo do tempo. As colunas (1) e (3) reportam os resultados baseados na equação (1.68), enquanto as colunas (2) e (4) na equação (1.69).

As estimativas do efeito da segregação feminina sobre os salários das mulheres são negativas e estatisticamente significativas para ambas as especificações do modelo multi-nível, com as diferenças entre elas sendo pequenas. Em relação ao modelo no qual as variâncias e covariância dos interceptos e inclinação aleatórios são restritas a zero, o teste

⁵⁷Por essa razão, os modelos multi-níveis podem ser vistos como modelos de efeitos aleatórios.

⁵⁸Entretanto, permite-se que os termos aleatórios em um mesmo nível sejam correlacionados. Nos resultados que apresentamos abaixo, reportamos a estimativa da correlação entre os termos aleatórios associados com a variável $pfem_{j,t}$ (v_{2j}) e o intercepto aleatório (v_{0j}^*).

⁵⁹O teste do tipo Wald não é recomendado para testar se variâncias são iguais a zero, já que pode violar uma das hipóteses básicas de estimações de máxima verossimilhança, a saber, que o (verdadeiro) parâmetro não se encontra na fronteira do espaço de parâmetros.

Tabela 1.11: Estimativas do Modelo Multi-Nível por Sexo: Trabalhadores Permanecentes

Covariadas	Mulheres		Homens	
	Interceptos Aleatórios	Interceptos e Inclinação Aleatórios	Interceptos Aleatórios	Interceptos e Inclinação Aleatórios
	(1)	(2)	(3)	(4)
Proporção de mulheres	-0.2674*** (0.0064)	-0.2788*** (0.0067)	-0.3188*** (0.0064)	-0.2976*** (0.0075)
Experiência geral	0.0105*** (0.0007)	0.0108*** (0.0007)	0.0261*** (0.0007)	0.0267*** (0.0007)
Experiência geral ao quadrado	-0.0001*** (0.0000)	-0.0001*** (0.0000)	-0.0003*** (0.0000)	-0.0003*** (0.0000)
Experiência no estabelecimento	0.0040*** (0.0000)	0.0041*** (0.0000)	0.0037*** (0.0000)	0.0036*** (0.0000)
Escolaridade	0.0960*** (0.0007)	0.0958*** (0.0007)	0.1005*** (0.0006)	0.1002*** (0.0006)
Comércio	-0.1313*** (0.0060)	-0.1279*** (0.0061)	-0.2873*** (0.0050)	-0.2971*** (0.0050)
Serviços	0.0156*** (0.0053)	0.0200*** (0.0054)	-0.0744*** (0.0044)	-0.0805*** (0.0044)
Administrativos	0.1880*** (0.0044)	0.1895*** (0.0044)	0.0877*** (0.0046)	0.0695*** (0.0047)
Profissionais	0.8016*** (0.0070)	0.8073*** (0.0070)	0.5915*** (0.0055)	0.5696*** (0.0056)
Gerentes	0.6170*** (0.0074)	0.6163*** (0.0074)	0.6354*** (0.0057)	0.6298*** (0.0058)
Tamanho estabelecimento (1/10)	-0.0001*** (0.0000)	-0.0001*** (0.0000)	0.0002*** (0.0000)	0.0002*** (0.0000)
Intercepto	-0.1544*** (0.0131)	-0.1523*** (0.0132)	-0.0324** (0.0109)	-0.0309** (0.0110)
Variâncias e Covariância				
σ^2	0.1390	0.1367	0.1568	0.1502
τ_0	0.1365	0.1454	0.1598	0.1435
τ_1	0.1355	0.2636	0.1606	0.1767
τ_2		0.3432		1.4104
τ_{12}		-0.7687		-0.4523
Log da verossimilhança	-242520.2	-242181.8	-414813.1	-415750.7
Teste razão verossimilhança [P-valor]	0.0000	0.0000	0.0000	0.0000
Número de observações	308743	308743	487861	487861

Notas:

1. A variável dependente é o log do salário horário real.
2. Na especificação de interceptos aleatórios, somente os interceptos do nível-2 e do nível-3 são tratados como aleatórios; na especificação de interceptos e inclinação (coeficiente angular) aleatórios, o coeficiente da variável proporção de mulheres no estabelecimento também é tratado como aleatório.
3. Erros-padrões entre parênteses.
4. Níveis de significância: *:10% **:5% ***:1%.
5. Todas as especificações incluem dummies de ano.

da razão de verossimilhança indica que não se pode rejeitar a hipótese de que esses termos aleatórios pertencem de fato ao modelo. É interessante notar que a estimativa da variância do coeficiente associado à variável proporção de mulheres no estabelecimento (τ_2) sugere que o efeito da segregação feminina sobre os salários das mulheres é heterogêneo entre estabelecimentos. Também interessante é que a correlação entre o intercepto aleatório do estabelecimento e a inclinação do coeficiente aleatório da variável proporção de mulheres (τ_{12}) é negativa, indicando que os empregadores com maiores efeitos especí-

ficos não-observáveis (*e.g.* mais altos graus de discriminação e/ou eficiência) tendem a apresentar uma relação mais negativa entre segregação feminina e salários das mulheres. As estimativas dos coeficientes das demais covariadas apresentam os sinais esperados e são qualitativamente similares aos resultados do modelo dinâmico.

Em relação aos homens, as estimativas também revelam um efeito negativo e estatisticamente significativo da segregação feminina sobre os salários. Como no caso das mulheres, o teste de verossimilhança indica que o modelo contém coeficientes aleatórios. Além disso, a estimativa da variância do coeficiente relativo à variável proporção de mulheres (τ_2) sugere a existência de heterogeneidade entre estabelecimentos em termos do efeito da segregação feminina sobre os salários dos homens. A correlação negativa entre o intercepto aleatório do estabelecimento e a inclinação do coeficiente aleatório da variável proporção de mulheres (τ_{12}) indica que a relação entre segregação feminina e o salário dos homens é mais negativa para os empregadores mais discriminadores/eficientes. Os resultados para as demais covariadas têm o sinal esperado e são semelhantes aos do modelo dinâmico.

A Tabela 1.12 contém os resultados das duas especificações do modelo multi-nível para mulheres por grupo ocupacional. Todas as estimativas para o coeficiente da variável proporção de mulheres são negativas e estatisticamente significativas ao nível de 1%. Os testes da razão de verossimilhança indicam que existe heterogeneidade de interceptos e inclinação para todas as categorias ocupacionais. A correlação entre o intercepto aleatório do estabelecimento e a inclinação do coeficiente aleatório da variável proporção de mulheres é negativa para todos os grupos de ocupação.

A Tabela 1.13 apresenta os resultados para homens por categoria ocupacional. As estimativas do coeficiente da variável proporção de mulheres são negativas e estatisticamente significativas para os trabalhadores ligados à produção e de gerência, porém não diferentes de zero do ponto de vista estatístico para os dois demais grupos ocupacionais. Mais uma vez, os testes da razão de verossimilhança confirmam a presença de coeficientes aleatórios no modelo. Observa-se para todas as categorias ocupacionais uma correlação negativa entre o intercepto aleatório do estabelecimento e a inclinação do coeficiente aleatório da variável proporção de mulheres.

Em suma, pode-se concluir com base nos resultados do modelo multi-nível que o efeito da segregação feminina entre estabelecimentos sobre os salários de homens e mulheres é negativo. Esses resultados também mostram que o efeito em questão parece ser heterogêneo entre estabelecimentos e mais negativo para os empregadores com mais elevados graus de discriminação contra a mulher ou de eficiência não-observada.

Tabela 1.12: Estimativas do Modelo Multi-Nível para Mulheres por Grupo Ocupacional: Trabalhadores Permanentes

	Ligados à Produção			Administrativos		Profissionais		Gerência	
	Interceptos Aleatórios (1)	Interceptos e Inclinação Aleatórios (2)	Interceptos Aleatórios (3)	Interceptos e Inclinação Aleatórios (4)	Interceptos Aleatórios (5)	Interceptos e Inclinação Aleatórios (6)	Interceptos Aleatórios (7)	Interceptos e Inclinação Aleatórios (8)	
Covariadas									
Proporção de mulheres	-0.1936*** (0.0108)	-0.1902*** (0.0129)	-0.2613*** (0.0151)	-0.2756*** (0.0164)	-0.3244*** (0.0396)	-0.3424*** (0.0655)	-0.3095*** (0.0349)	-0.3614*** (0.0434)	
Experiência geral	0.0012 (0.0007)	0.0015* (0.0007)	0.0172*** (0.0013)	0.0173*** (0.0013)	0.0078** (0.0024)	0.0089*** (0.0024)	0.0234*** (0.0034)	0.0243*** (0.0035)	
Experiência geral ao quadrado	-0.0000 (0.0000)	-0.0000 (0.0000)	-0.0003*** (0.0000)	-0.0003*** (0.0000)	-0.0000 (0.0001)	-0.0000 (0.0001)	-0.0003*** (0.0001)	-0.0003*** (0.0001)	
Experiência no estabelecimento	0.0022*** (0.0000)	0.0021*** (0.0000)	0.0027*** (0.0001)	0.0027*** (0.0001)	0.0016*** (0.0001)	0.0015*** (0.0001)	0.0021*** (0.0001)	0.0020*** (0.0001)	
Escolaridade	0.0365*** (0.0007)	0.0359*** (0.0007)	0.0623*** (0.0012)	0.0623*** (0.0012)	0.0859*** (0.0023)	0.0833*** (0.0024)	0.0981*** (0.0030)	0.0963*** (0.0031)	
Comércio	-0.0106 (0.0107)	-0.0014 (0.0108)	-0.1241*** (0.0181)	-0.1211*** (0.0186)	-0.0627 (0.0633)	-0.1249 (0.0752)	-0.1061** (0.0381)	-0.0889* (0.0405)	
Serviços	-0.1120*** (0.0107)	-0.1056*** (0.0108)	0.0642*** (0.0177)	0.0731*** (0.0181)	-0.0856 (0.0572)	-0.1239 (0.0690)	0.2398*** (0.0380)	0.2661*** (0.0403)	
Tamanho estabelecimento (1/10)	-0.0001** (0.0000)	-0.0001 (0.0001)	0.0001 (0.0001)	0.0001 (0.0001)	0.0003 (0.0002)	0.0002 (0.0002)	0.0004 (0.0002)	0.0004 (0.0003)	
Intercepto	0.4226*** (0.0166)	0.4128*** (0.0176)	0.1633*** (0.0263)	0.1681*** (0.0269)	0.7344*** (0.0688)	0.8585*** (0.0840)	0.1546* (0.0656)	0.1842*** (0.0692)	
Variação e Covariância									
σ^2	0.1396	0.1360	0.1361	0.1331	0.1350	0.1110	0.1603	0.1419	
τ_0	0.0647	0.0653	0.0998	0.1006	0.1520	0.1570	0.1486	0.1524	
τ_1	0.1107	0.4055	0.1968	0.3196	0.3569	2.5124	0.3730	1.1436	
τ_2		0.5849		0.3591		5.3726		1.9473	
τ_{12}		-0.8958		-0.6549		-0.9363		-0.8458	
Log da verossimilhança	-109984.1	-109425.6	-61862.6	-61741.6	-22056.7	-21109.2	-16345.5	-16130.8	
Teste razão verossimilhança [P-valor]	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
Número de observações	173321	173321	87334	87334	29495	29495	18593	18593	

Notas:

1. A variável dependente é o log do salário horário real.
2. Na especificação de interceptos aleatórios, somente os interceptos do nível-2 e do nível-3 são tratados como aleatórios; na especificação de interceptos e inclinação (coeficiente angular) aleatórios, o coeficiente da variável proporção de mulheres no estabelecimento também é tratado como aleatório.
3. Erros-padrões entre parênteses.
4. Níveis de significância: *.10% **.5% ***.1%.
5. Todas as especificações incluem dummies de ano.

Tabela 1.13: Estimativas do Modelo Multi-Nível para Homens por Grupo Ocupacional: Trabalhadores Permanentes

Covariadas	Ligados à Produção			Administrativos		Profissionais		Gerência	
	Interceptos Aleatórios (1)	Interceptos e Inclinação Aleatórios (2)	Interceptos Aleatórios (3)	Interceptos e Inclinação Aleatórios (4)	Interceptos Aleatórios (5)	Interceptos e Inclinação Aleatórios (6)	Interceptos Aleatórios (7)	Interceptos e Inclinação Aleatórios (8)	
Proporção de mulheres	-0.5288*** (0.0068)	-0.5282*** (0.0068)	0.0171 (0.0236)	0.0230 (0.0276)	-0.0238 (0.0378)	0.0147 (0.0679)	-0.1071*** (0.0369)	-0.1050*** (0.0439)	
Experiência geral	0.0199*** (0.0007)	0.0199*** (0.0007)	0.0238*** (0.0015)	0.0239*** (0.0015)	0.0315*** (0.0017)	0.0324*** (0.0017)	0.0445*** (0.0026)	0.0448*** (0.0026)	
Experiência geral ao quadrado	-0.0003*** (0.0000)	-0.0003*** (0.0000)	-0.0003*** (0.0000)	-0.0003*** (0.0000)	-0.0004*** (0.0000)	-0.0004*** (0.0000)	-0.0005*** (0.0000)	-0.0005*** (0.0000)	
Experiência no estabelecimento	0.0038*** (0.0000)	0.0038*** (0.0000)	0.0024*** (0.0001)	0.0024*** (0.0001)	0.0014*** (0.0001)	0.0013*** (0.0001)	0.0012*** (0.0001)	0.0012*** (0.0001)	
Escolaridade	0.0634*** (0.0006)	0.0634*** (0.0006)	0.0659*** (0.0013)	0.0657*** (0.0013)	0.1060*** (0.0015)	0.1051*** (0.0015)	0.1209*** (0.0021)	0.1213*** (0.0021)	
Comércio	-0.2081*** (0.0051)	-0.2070*** (0.0051)	-0.1858*** (0.0258)	-0.1874*** (0.0262)	-0.3761*** (0.0521)	-0.3622*** (0.0623)	-0.2813*** (0.0352)	-0.2971*** (0.0361)	
Serviços	-0.1156*** (0.0045)	-0.1168*** (0.0045)	0.0241 (0.0251)	0.0326 (0.0255)	-0.2479*** (0.0463)	-0.2276*** (0.0559)	-0.0084 (0.0352)	-0.0098 (0.0363)	
Tamanho estabelecimento (1/10)	0.0005*** (0.0000)	0.0004*** (0.0000)	0.0003*** (0.0001)	0.0002*** (0.0001)	0.0009*** (0.0001)	0.0005*** (0.0002)	0.0009*** (0.0002)	0.0008*** (0.0002)	
Intercepto	0.3598*** (0.0124)	0.3610*** (0.0124)	0.0796* (0.0323)	0.0823* (0.0328)	0.2350*** (0.0506)	0.2336*** (0.0613)	-0.0609 (0.0515)	-0.0507 (0.0526)	
Variância e Covariância									
σ^2	0.1618	0.1617	0.1373	0.1328	0.1092	0.0978	0.2149	0.2052	
τ_0	0.1113	0.0669	0.1137	0.1154	0.1476	0.1504	0.2196	0.2220	
τ_1	0.1123	0.1625	0.2735	0.3131	0.4405	1.0027	0.4711	0.6156	
τ_2		0.0011		0.8894		6.4232		2.1149	
τ_{12}		-0.9986		-0.4625		-0.8216		-0.5585	
Log da verossimilhança	-266945.4	-266936.1	-45566.0	-45388.0	-35021.3	-34359.9	-38947.3	-38843.4	
Teste razão verossimilhança [P-valor]	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
Número de observações	331895	331895	62736	62736	54058	54058	39172	39172	

Notas:

1. A variável dependente é o log do salário horário real.
2. Na especificação de interceptos aleatórios, somente os interceptos do nível-2 e do nível-3 são tratados como aleatórios; na especificação de interceptos e inclinação (coeficiente angular) aleatórios, o coeficiente da variável proporção de mulheres no estabelecimento também é tratado como aleatório.
3. Erros-padrões entre parênteses.
4. Níveis de significância: *.10% **.5% ***.1%.
5. Todas as especificações incluem dummies de ano.

1.8.2 Todos os Trabalhadores

Finalizaremos nossos testes de robustez utilizando a amostra com todos os trabalhadores, isto é, a base de dados que contém informações sobre os trabalhadores que permanecem no mesmo estabelecimento e os que trocam de empregador ao longo do tempo. Primeiramente, apresentaremos os resultados para essa amostra completa com base no modelo multi-nível da seção anterior. Em seguida, reportaremos as estimativas baseadas no modelos de Intra-Grupos, o qual é obtido a partir de uma especificação distinta da empregada na seção 1.8.1.1 para os efeitos específicos não-observáveis.

1.8.2.1 Modelo Multi-Nível

Uma vez que o modelo multi-nível não faz uso das transformações de Intra-Grupos ou em Primeiras-Diferenças, torna-se possível utilizá-lo com a amostra de todos os trabalhadores. Em outras palavras, como a estimação desse modelo não remove os termos específicos não-observáveis dos trabalhadores e estabelecimentos, ela não sofre do problema de eliminar o termo específico do estabelecimento quando a amostra contém trabalhadores que trocam de estabelecimento entre períodos. A especificação do modelo, assim como sua estimação, são semelhantes às da seção anterior.

A Tabela 1.14 apresenta os resultados do modelo multi-nível por sexo para a amostra com todos os trabalhadores. O formato dessa Tabela é o mesmo da Tabela 1.11.

O principal ponto a se notar da Tabela 1.14 é que os resultados obtidos com base na amostra completa são qualitativamente similares aos estimados com a amostra de trabalhadores que permanecem nos mesmos estabelecimentos. De fato, as estimativas do coeficiente da variável proporção de mulheres são negativas e estatisticamente significativas para homens e mulheres, com magnitudes bastante semelhantes às obtidas anteriormente. Ademais, o teste da razão de verossimilhança confirma a presença de coeficientes aleatórios no modelo. Observa-se também que a correlação entre o intercepto aleatório do estabelecimento e a inclinação do coeficiente aleatório da variável proporção de mulheres é negativa. Essa semelhança de resultados indica que a amostra de trabalhadores permanentes não parece ter nenhum viés importante.

As Tabelas 1.15 e 1.16 apresentam as estimativas por grupo ocupacional para mulheres e homens, respectivamente. Cada uma dessas Tabelas mostra resultados bastante similares – em termos de sinal, significância estatística e magnitude – aos seus correspondentes da amostra de trabalhadores permanentes. Mais uma vez, é possível confirmar a presença de coeficientes aleatórios, e de uma correlação negativa entre interceptos aleatórios e a inclinação do coeficiente da variável proporção de mulheres no estabelecimento.

Tabela 1.14: Estimativas do Modelo Multi-Nível por Sexo: Todos os Trabalhadores

Covariadas	Mulheres		Homens	
	Interceptos Aleatórios	Interceptos e Inclinação Aleatórios	Interceptos Aleatórios	Interceptos e Inclinação Aleatórios
	(1)	(2)	(3)	(4)
Proporção de mulheres	-0.2364*** (0.0065)	-0.2616*** (0.0069)	-0.2868*** (0.0065)	-0.2838*** (0.0075)
Experiência geral	0.0122*** (0.0008)	0.0126*** (0.0008)	0.0256*** (0.0007)	0.0262*** (0.0007)
Experiência geral ao quadrado	-0.0001*** (0.0000)	-0.0001*** (0.0000)	-0.0003*** (0.0000)	-0.0003*** (0.0000)
Experiência no estabelecimento	0.0039*** (0.0000)	0.0039*** (0.0000)	0.0038*** (0.0000)	0.0037*** (0.0000)
Escolaridade	0.0925*** (0.0007)	0.0917*** (0.0007)	0.0947*** (0.0006)	0.0953*** (0.0006)
Comércio	-0.1258*** (0.0061)	-0.1196*** (0.0061)	-0.2672*** (0.0052)	-0.2716*** (0.0053)
Serviços	0.0064 (0.0054)	0.0121* (0.0054)	-0.0960*** (0.0045)	-0.1026*** (0.0046)
Administrativos	0.2122*** (0.0045)	0.2141*** (0.0045)	0.1355*** (0.0050)	0.1258*** (0.0052)
Profissionais	0.8699*** (0.0072)	0.8780*** (0.0073)	0.6721*** (0.0059)	0.6566*** (0.0061)
Gerentes	0.6889*** (0.0079)	0.6840*** (0.0079)	0.7144*** (0.0061)	0.7094*** (0.0063)
Tamanho estabelecimento (1/10)	-0.0000*** (0.0000)	-0.0001*** (0.0000)	0.0002*** (0.0000)	0.0002*** (0.0000)
Intercepto	-0.1949*** (0.0135)	-0.1804*** (0.0136)	-0.0374*** (0.0113)	-0.0419*** (0.0117)
Variância e Covariância				
σ^2	0.1423	0.1380	0.1606	0.1513
τ_0	0.0886	0.0941	0.0980	0.1377
τ_1	0.2582	0.6032	0.3315	0.3323
τ_2		0.7564		1.6416
τ_{12}		-0.8171		-0.4406
Log da verossimilhança	-293761.8	-292415.3	-509583.8	-512140.2
Teste razão verossimilhança [P-valor]	0.0000	0.0000	0.0000	0.0000
Número de observações	341586	341586	544051	544051

Notas:

1. A variável dependente é o log do salário horário real.

2. Na especificação de interceptos aleatórios, somente os interceptos do nível-2 e do nível-3 são tratados como aleatórios; na especificação de interceptos e inclinação (coeficiente angular) aleatórios, o coeficiente da variável proporção de mulheres no estabelecimento também é tratado como aleatório.

3. Erros-padrões entre parênteses.

4. Níveis de significância: *:10% **:5% ***:1%.

5. Todas as especificações incluem dummies de ano.

Tabela 1.15: Estimativas do Modelo Multi-Nível para Mulheres por Grupo Ocupacional: Todos os Trabalhadores

Covariadas	Ligados à Produção			Administrativos			Profissionais			Gerência	
	Interceptos Aleatórios	Interceptos e Inclinação Aleatórios	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Interceptos Aleatórios	Interceptos e Inclinação Aleatórios
Proporção de mulheres	-0.1847*** (0.0111)	-0.1806*** (0.0132)	-0.2656*** (0.0155)	-0.2807*** (0.0170)	-0.2970*** (0.0403)	-0.3130*** (0.0639)	-0.3035*** (0.0375)	-0.3130*** (0.0639)	-0.3035*** (0.0375)	-0.3905*** (0.0483)	-0.3905*** (0.0483)
Experiência geral	0.0027*** (0.0007)	0.0028*** (0.0007)	0.0176*** (0.0014)	0.0176*** (0.0014)	0.0099*** (0.0026)	0.0106*** (0.0026)	0.0275*** (0.0037)	0.0106*** (0.0026)	0.0275*** (0.0037)	0.0283*** (0.0036)	0.0283*** (0.0036)
Experiência geral ao quadrado	-0.0000* (0.0000)	-0.0000* (0.0000)	-0.0003*** (0.0000)	-0.0003*** (0.0000)	-0.0001 (0.0001)	-0.0001 (0.0001)	-0.0004*** (0.0001)	-0.0001 (0.0001)	-0.0004*** (0.0001)	-0.0004*** (0.0001)	-0.0004*** (0.0001)
Experiência no estabelecimento	0.0023*** (0.0000)	0.0023*** (0.0000)	0.0027*** (0.0001)	0.0027*** (0.0001)	0.0015*** (0.0001)	0.0015*** (0.0001)	0.0019*** (0.0001)	0.0015*** (0.0001)	0.0019*** (0.0001)	0.0017*** (0.0001)	0.0017*** (0.0001)
Escolaridade	0.0330*** (0.0007)	0.0325*** (0.0007)	0.0604*** (0.0012)	0.0602*** (0.0013)	0.0844*** (0.0026)	0.0821*** (0.0026)	0.1009*** (0.0032)	0.0821*** (0.0026)	0.1009*** (0.0032)	0.0997*** (0.0032)	0.0997*** (0.0032)
Comércio	-0.0110 (0.0112)	-0.0048 (0.0112)	-0.1261*** (0.0187)	-0.1233*** (0.0192)	-0.1094 (0.0657)	-0.1741* (0.0774)	-0.1462*** (0.0417)	-0.1741* (0.0774)	-0.1462*** (0.0417)	-0.1268** (0.0437)	-0.1268** (0.0437)
Serviços	-0.1182*** (0.0112)	-0.1143*** (0.0112)	0.0763*** (0.0183)	0.0862*** (0.0188)	-0.1242* (0.0602)	-0.1474* (0.0718)	0.2110*** (0.0415)	-0.1474* (0.0718)	0.2110*** (0.0415)	0.2365*** (0.0434)	0.2365*** (0.0434)
Tamanho estabelecimento (1/10)	-0.0001** (0.0000)	-0.0001 (0.0001)	0.0001 (0.0001)	0.0001 (0.0001)	0.0002 (0.0002)	0.0002 (0.0002)	0.0004 (0.0002)	0.0002 (0.0002)	0.0004 (0.0002)	0.0004 (0.0003)	0.0004 (0.0003)
Intercepto	0.3952*** (0.0171)	0.3879*** (0.0182)	0.1602*** (0.0277)	0.1659*** (0.0284)	0.7857*** (0.0732)	0.8973*** (0.0877)	0.1493* (0.0706)	0.7857*** (0.0732)	0.1493* (0.0706)	0.2000** (0.0744)	0.2000** (0.0744)
Variação e Covariância											
σ^2	0.1480	0.1437	0.1420	0.1380	0.1419	0.1156	0.1647	0.1156	0.1647	0.1402	0.1402
τ_0	0.0962	0.0967	0.1579	0.4302	0.2369	0.2485	0.1961	0.2485	0.1961	0.2025	0.2025
τ_1	0.1400	0.5435	0.2248	0.2132	0.4039	2.6620	0.4958	2.6620	0.4958	1.6661	1.6661
τ_2		0.7424		0.3643		5.4780		5.4780		3.0504	3.0504
τ_{12}		-0.9077		-0.7250		-0.9344		-0.9344		-0.8740	-0.8740
Log da verossimilhança	-139549.5	-138743.6	-75329.7	-75180.9	-27097.4	-26169.5	-18232.4	-26169.5	-18232.4	-17899.0	-17899.0
Teste razão verossimilhança [P-valor]	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Número de observações	195779	195779	94707	94707	31965	31965	19135	31965	19135	19135	19135

Notas:

1. A variável dependente é o log do salário horário real.
2. Na especificação de interceptos aleatórios, somente os interceptos do nível-2 e do nível-3 são tratados como aleatórios; na especificação de interceptos e inclinação (coeficiente angular) aleatórios, o coeficiente da variável proporção de mulheres no estabelecimento também é tratado como aleatório.
3. Erros-padrões entre parênteses.
4. Níveis de significância: *.10% **.5% ***.1%.
5. Todas as especificações incluem dummies de ano.

Tabela 1.16: Estimativas do Modelo Multi-Nível para Homens por Grupo Ocupacional: Todos os Trabalhadores

	Ligados à Produção			Administrativos		Profissionais		Gerência	
	Interceptos Aleatórios (1)	Interceptos e Inclinação Aleatórios (2)	Interceptos Aleatórios (3)	Interceptos e Inclinação Aleatórios (4)	Interceptos Aleatórios (5)	Interceptos e Inclinação Aleatórios (6)	Interceptos Aleatórios (7)	Interceptos e Inclinação Aleatórios (8)	
Covariadas									
Proporção de mulheres	-0.4784*** (0.0068)	-0.5100*** (0.0080)	-0.0041 (0.0247)	0.0090 (0.0295)	0.0021 (0.0384)	0.0886 (0.0696)	-0.0563 (0.0384)	-0.0538 (0.0468)	
Experiência geral	0.0215*** (0.0008)	0.0221*** (0.0008)	0.0229*** (0.0016)	0.0231*** (0.0016)	0.0308*** (0.0020)	0.0321*** (0.0020)	0.0440*** (0.0029)	0.0446*** (0.0029)	
Experiência geral ao quadrado	-0.0003*** (0.0000)	-0.0003*** (0.0000)	-0.0003*** (0.0000)	-0.0003*** (0.0000)	-0.0004*** (0.0000)	-0.0004*** (0.0000)	-0.0005*** (0.0001)	-0.0005*** (0.0001)	
Experiência no estabelecimento	0.0042*** (0.0000)	0.0042*** (0.0000)	0.0025*** (0.0001)	0.0024*** (0.0001)	0.0014*** (0.0001)	0.0013*** (0.0001)	0.0012*** (0.0001)	0.0011*** (0.0001)	
Escolaridade	0.0607*** (0.0006)	0.0622*** (0.0007)	0.0632*** (0.0015)	0.0629*** (0.0015)	0.1068*** (0.0018)	0.1050*** (0.0018)	0.1205*** (0.0023)	0.1200*** (0.0024)	
Comércio	-0.1881*** (0.0052)	-0.1892*** (0.0054)	-0.1978*** (0.0276)	-0.1985*** (0.0273)	-0.3603*** (0.0522)	-0.3469*** (0.0623)	-0.3073*** (0.0364)	-0.3238*** (0.0380)	
Serviços	-0.1280*** (0.0046)	-0.1274*** (0.0048)	0.0326 (0.0266)	0.0385 (0.0265)	-0.2278*** (0.0461)	-0.2222*** (0.0556)	-0.0114 (0.0364)	-0.0151 (0.0381)	
Tamanho estabelecimento (1/10)	0.0004*** (0.0000)	0.0004*** (0.0000)	0.0002* (0.0001)	0.0001* (0.0001)	0.0007*** (0.0001)	0.0004*** (0.0002)	0.0008*** (0.0002)	0.0007*** (0.0002)	
Intercepto	0.2881*** (0.0127)	0.2745*** (0.0133)	0.1178*** (0.0352)	0.1198*** (0.0350)	0.2581*** (0.0532)	0.2483*** (0.0627)	-0.0191 (0.0559)	-0.0034 (0.0575)	
Variação e Covariância									
σ^2	0.1676	0.1554	0.1351	0.1297	0.1128	0.1008	0.2264	0.2118	
τ_0	0.0657	0.1105	0.1815	0.1825	0.2928	0.2848	0.3642	0.3676	
τ_1	0.2388	0.2408	0.3227	0.2954	0.4549	1.4755	0.4908	0.7274	
τ_2		1.5323		1.2426		7.5226		2.8608	
τ_{12}		-0.4804		-0.3596		-0.8027		-0.6128	
Log da verossimilhança	-338979.2	-342174.4	-51230.2	-50958.3	-44925.7	-44103.4	-44663.2	-44511.2	
Teste razão verossimilhança [P-valor]	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
Número de observações	381139	381139	64427	64427	57547	57547	40938	40938	

Notas:

1. A variável dependente é o log do salário horário real.
2. Na especificação de interceptos aleatórios, somente os interceptos do nível-2 e do nível-3 são tratados como aleatórios; na especificação de interceptos e inclinação (coeficiente angular) aleatórios, o coeficiente da variável proporção de mulheres no estabelecimento também é tratado como aleatório.
3. Erros-padrões entre parênteses.
4. Níveis de significância: *10% **5% ***1%.
5. Todas as especificações incluem dummies de ano.

1.8.2.2 Intra-Grupos

Como teste final de robustez dos resultados, utilizamos o modelo de Intra-Grupos para a amostra com todos os trabalhadores. Para tanto, re-especificamos a equação (1.61) da seguinte maneira:

$$y_{ijt} = x_{1it}\beta_1 + x_{2it}\beta_2 + z_{1jt}\gamma_1 + z_{2jt}\gamma_2 + \eta_{ij} + \varepsilon_{ijt}. \quad (1.70)$$

Note que, ao invés de especificarmos efeitos fixos separadamente para trabalhadores e estabelecimentos, assumimos agora uma formulação mais geral que contém um termo específico (η_{ij}) que varia entre cada *par* trabalhador-estabelecimento.⁶⁰ Dada essa suposição, a transformação de Intra-Grupos aplicada aos pares trabalhador-estabelecimento passam a ser apropriadas para eliminar completamente o termo η_{ij} da equação (1.70).⁶¹ Como a incorporação de trabalhadores que mudam de estabelecimento entre períodos não impede a remoção do termo específico do par trabalhador-estabelecimento, podemos utilizar aqui a amostra completa. Mantemos as hipóteses feitas na seção 1.8.1.1, a saber: as covariadas são estritamente exógenas e a variância do termo de erro ε_{ijt} pode ser aglomerada por estabelecimento.

Na Tabela 1.17 apresentamos as estimativas de Intra-Grupos para o coeficiente da variável proporção de mulheres no estabelecimento por sexo e grupo ocupacional. De uma forma geral, os resultados são bastante similares aos obtidos com a amostra de trabalhadores permanentes (ver Tabela 1.10). De fato, as estimativas são negativas e estatisticamente significativas (pelo menos a 10%) para as mulheres como um todo e para as que estão em ocupações ligadas à produção e administrativas. No entanto, as estimativas não são estatisticamente significativas para as mulheres nas duas outras categorias ocupacionais, e tampouco para os homens. Embora não estritamente comparáveis, resultados similares para as duas amostras sugerem não haver importante seletividade nos dados dos trabalhadores que permanecem nos mesmos estabelecimentos ao longo do tempo.

⁶⁰Essa especificação é mais geral na medida em que não impomos separabilidade entre os termos não-observáveis do trabalhador e do estabelecimento. Em outras palavras, assumir que o modelo contém $(\theta_i^* + \phi_j^*)$ é um caso especial da postulação que utiliza η_{ij} .

⁶¹Tomando k como um indexador de cada par trabalhador-estabelecimento ij , a transformação Intra-Grupos leva a que: $\eta_{ij} - T^{-1} \sum_{t=1}^T \eta_{ij} = \eta_k - T^{-1} \sum_{t=1}^T \eta_k = \eta_k - \eta_k = 0$. Cabe observar que não utilizamos o modelo em Primeiras-Diferenças nesta sub-seção, pois os resultados são idênticos aos obtidos somente com a amostra de trabalhadores que permanecem no mesmo estabelecimento ao longo do tempo. De fato, uma vez que, para pertencer a essa amostra, o trabalhador deve continuar no mesmo estabelecimento durante (pelo menos) dois períodos adjacentes, os valores das primeiras-diferenças calculados para essa amostra coincidem com os obtidos para o par trabalhador-estabelecimento. Essa situação não se verifica no caso da transformação de Intra-Grupos, já que as observações dos pares trabalhador-estabelecimento associadas somente a um período também são utilizadas na estimação. O mesmo se aplica aos pares trabalhador-estabelecimento correspondentes ao caso em que o trabalhador sai do estabelecimento no segundo período mas retorna ao mesmo estabelecimento no terceiro período.

Tabela 1.17: Estimativas do Modelo de Intra-Grupos para o Coeficiente da Variável Proporção de Mulheres no Estabelecimento por Sexo e Grupo Ocupacional: Todos os Trabalhadores

Ocupação/Sexo	Coeficiente (1)	Erro-Padrão (2)	Teste-F (3)	Observações (4)
1. Todos				
1.1 - Mulheres	-0.0528**	0.0281	0.0000	341586
1.2 - Homens	0.0159	0.0353	0.0000	544051
2. Ligados à Prod.				
2.1 - Mulheres	-0.0517*	0.0333	0.0000	195779
2.2 - Homens	0.0495	0.0438	0.0000	381139
3. Administrativos				
3.1 - Mulheres	-0.0754*	0.0438	0.0058	94707
3.2 - Homens	-0.0530	0.0622	0.0245	64427
4. Profissionais				
4.1 - Mulheres	-0.1039	0.1467	0.4888	31965
4.2 - Homens	-0.0360	0.0956	0.3322	57547
5. Gerentes				
5.1 - Mulheres	0.0389	0.1203	0.0007	19135
5.2 - Homens	-0.0404	0.0891	0.0000	40938

Notas:

1. A variável dependente é o log do salário horário real.

2. Níveis de significância: *:10% **:5% ***:1%.

3. P-valores reportados para o teste-F nas colunas (3) e (7).

4. As regressões também incluem as variáveis experiência geral e seu quadrado, tamanho do estabelecimento, experiência no estabelecimento e dummies de ano.

1.9 Discussão

1.9.1 Sumário dos Resultados Empíricos

Até o momento, apresentamos resultados obtidos por diversos métodos diferentes por sexo e grupo ocupacional. Como uma tentativa de sumariar todos esses resultados, na Tabela 1.18 calculamos a média das estimativas do coeficiente da variável proporção de mulheres no

estabelecimento por grupo amostral e para os distintos métodos/especificações utilizados.⁶² O Painel A dessa Tabela reporta as médias para as mulheres e o Painel B para os homens. Todas as médias foram computadas das estimativas estatisticamente significativas pelo menos ao nível de 10%.⁶³ As primeiras duas colunas apresentam as médias somente para as estimativas obtidas com a amostra de trabalhadores permanentes, enquanto as duas últimas colunas contêm as médias calculadas para ambas as amostras de permanentes e de todos os trabalhadores. Uma vez que o número de observações utilizadas para estimar o coeficiente de interesse variou entre os diferentes métodos/especificações e grupos amostrais, as colunas (1) e (3) reportam as médias não ponderadas, enquanto as colunas (2) e (4) as médias ponderadas. Os pesos usados na ponderação foram a parcela do número de observações para cada par de grupo amostral e método/especificação utilizado. Além das médias, apresentamos também os erros-padrões e o número de estimativas utilizadas para computar essas médias.

O primeiro ponto a se notar da Tabela 1.18 é que todas as médias têm sinal negativo. Isso ocorre independentemente do grupo amostral, da base de dados utilizada ou da forma de ponderação. Em segundo lugar, apesar de o número de estimativas para calcular as médias ser pequeno, os erros-padrões são relativamente baixos. Como consequência desse fato, o limite superior do intervalo de confiança de 95% (não mostrado) é negativo para cada média das estimativas do coeficiente de interesse.⁶⁴

Como observado anteriormente, as estimativas obtidas com a amostra de todos os trabalhadores não são muito diferentes das suas correspondentes para a amostra de trabalhadores permanentes. Isso se confirma pela comparação das médias computadas nas colunas (1) e (3) ou nas colunas (2) e (4). Assim, mais uma vez, não encontramos evidências de que existe um viés de seleção na amostra que contém somente os trabalhadores que permanecem nos mesmos estabelecimentos ao longo do tempo. É interessante notar também que as médias ponderadas e não-ponderadas têm magnitudes bastante similares.

Em suma, os resultados apresentados na Tabela 1.18 indicam que o efeito da segregação feminina no estabelecimento é negativo sobre os salários de ambos os sexos. Na seção que se segue, comparamos esses resultados com as estimativas obtidas com métodos de *cross-section*.

⁶²Em conjunto, empregamos sete diferentes métodos/especificações: 1) MMG em Diferenças, 2) MMG em Sistema em Um-Passo, 3) MMG em Sistema em Dois-Passos, 4) Intra-Grupos, 5) Primeiras-Diferenças, 6) Multi-Nível com Interceptos Aleatórios, e 7) Multi-Nível com Interceptos e Inclinação Aleatórios. Os grupos amostrais são: 1) Mulheres, 2) Mulheres em Ocupações Ligadas à Produção, 3) Mulheres em Ocupações Administrativas, 4) Mulheres em Ocupações Profissionais, 5) Mulheres em Ocupações de Gerência, e os mesmos cinco grupos para os homens.

⁶³Qualitativamente, as conclusões desta sub-seção não se alteram se as estimativas não significativas forem incluídas.

⁶⁴Obviamente, esse limite não pôde ser calculado para o grupo de homens em ocupações administrativas, já que obtivemos somente uma estimativa estatisticamente significativa para esse grupo.

Tabela 1.18: Média das Estimativas do Coeficiente da Variável Proporção de Mulheres no Estabelecimento

Grupo	Permanentes		Permanentes e Todos Trabalhadores	
	Não-Ponderada (1)	Ponderada (2)	Não-Ponderada (3)	Ponderada (4)
A - MULHERES				
A.1 - Todas	-0.1407 (0.0428) 6	-0.1555 (0.0452) 6	-0.1550 (0.0343) 9	-0.1670 (0.0350) 9
A.2 - Ligadas à Prod.	-0.1547 (0.0322) 6	-0.1523 (0.0316) 6	-0.1495 (0.0245) 9	-0.1468 (0.0239) 9
A.3 - Administrativas	-0.2177 (0.0355) 6	-0.2206 (0.0373) 6	-0.2142 (0.0299) 9	-0.2149 (0.0311) 9
A.4 - Profissionais	-0.3682 (0.0403) 5	-0.3726 (0.0343) 5	-0.3501 (0.0302) 7	-0.3464 (0.0258) 7
A.5 - Gerentes	-0.3355 (0.0259) 2	-0.3355 (0.0259) 2	-0.3412 (0.0209) 4	-0.3413 (0.0210) 4
B - HOMENS				
B.1 - Todos	-0.2569 (0.0301) 4	-0.2696 (0.0293) 4	-0.2664 (0.0200) 6	-0.2760 (0.0177) 6
B.2 - Ligadas à Prod.	-0.5116 (0.0098) 4	-0.5158 (0.0095) 4	-0.5058 (0.0083) 6	-0.5068 (0.0087) 6
B.3 - Administrativas	-0.2555 (.) 1	-0.2555 (.) 1	-0.2555 (.) 1	-0.2555 (.) 1
B.4 - Profissionais	-0.2828 (0.0661) 2	-0.2828 (0.0661) 2	-0.2828 (0.0661) 2	-0.2828 (0.0661) 2
B.5 - Gerentes	-0.1911 (0.0408) 5	-0.1618 (0.0385) 5	-0.1911 (0.0408) 5	-0.1618 (0.0385) 5

Notas:

1. 'Permanentes' se refere à amostra de trabalhadores que permaneceram nos mesmos estabelecimentos ao longo do tempo. 'Todos Trabalhadores' corresponde à amostra na qual se incluem os trabalhadores que mudaram de estabelecimento ao longo do tempo.

2. Os pesos usados para obter as médias ponderadas foram a proporção do número de observações utilizadas na estimação do coeficiente de interesse entre os diferentes métodos/especificações econométricos. A soma dos pesos é sempre igual a unidade.

3. Erros-padrões entre parênteses.

4. Os números abaixo dos erros-padrões referem-se ao número de estimativas usadas para computar as médias.

1.9.2 Comparação com Resultados de *Cross-Section*

Na seção 1.2, apresentamos uma resenha dos estudos empíricos que investigaram a relação entre salários e a segregação feminina no nível do estabelecimento. Como os resultados dessa literatura são baseados em dados de *cross-section*, procuramos reproduzir os modelos

utilizados nesses estudos a título de comparação. Para isso, implementamos regressões de *cross-section* com as especificações mais comumente empregadas nesses trabalhos.⁶⁵ Além de contrastarmos nossos resultados de *cross-section* com os da literatura prévia, também os comparamos com os que obtivemos nas seções anteriores.

Implementamos duas especificações distintas: a primeira é baseada em MQO por sexo e a segunda em MQO com dados empilhados (*pooled*) por sexo, onde incluímos um termo de interação entre uma variável *dummy* para mulheres e a variável proporção de mulheres no estabelecimento. Essas especificações são aplicadas para os dados de um ano particular e, para fins de comparação com nossos resultados anteriores, para todos os anos em conjunto. As estimativas estão apresentadas na Tabela 1.19, cujas colunas (1) a (3) se referem aos resultados para o ano de 2000, e as colunas (4) a (6) aos resultados relativos ao conjunto dos anos disponíveis. As colunas (1) e (4) mostram as estimativas para o grupo de mulheres, as colunas (2) e (5) para o grupo de homens, e as colunas (3) e (6) para os dados empilhados por sexo.

Como revela a Tabela 1.19, as estimativas do coeficiente da variável proporção de mulheres são negativas e estatisticamente significativas para ambos os sexos e para as especificações utilizadas. De uma forma geral, esses resultados confirmam o que foi encontrado na literatura prévia. No entanto, em contraste com as evidências apresentadas nessa literatura, nossos resultados de *cross-section* mostram que o efeito de trabalhar em estabelecimentos predominantemente femininos é mais intenso (i.e. mais negativo) sobre o salário dos homens do que o das mulheres.

Comparamos agora nossos resultados de *cross-section* com os obtidos nas seções anteriores. Analisando primeiramente o caso feminino, se tomarmos as médias calculadas para os métodos/especificações que empregamos anteriormente (ver linha (A.1) da Tabela 1.18), é possível constatar uma substancial redução (em termos absolutos) do efeito da segregação feminina sobre o salário das mulheres quando se passa das estimativas por MQO de *cross-section* para as obtidas com dados de painel e métodos longitudinais. De fato, enquanto as estimativas de *cross-section* estão em torno de $-0,30$, a média obtida com base em métodos longitudinais é aproximadamente $-0,15$. Essa constatação também pode ser confirmada pela comparação com a maior parte dos métodos longitudinais individuais. A única exceção são as estimativas obtidas com o modelo multi-nível, cujas magnitudes para o coeficiente de interesse são ligeiramente menos negativas que as de MQO.

Em relação aos resultados para homens, a comparação também revela uma importante redução (em termos absolutos) nas estimativas. De fato, tomando a média para os métodos/especificações de painel – cerca de $-0,26$ (ver linha (B.1) na Tabela 1.18) –, tem-se uma significativa atenuação do efeito de interesse em relação às estimativas de MQO – aproximadamente $-0,40$. No caso dos homens, todas as estimativas individuais

⁶⁵Naturalmente, como esses estudos utilizam conjuntos distintos de variáveis de controle, nossos resultados não são estritamente comparáveis aos deles.

Tabela 1.19: Estimativas por Mínimos Quadrados Ordinários: Dados de Cross-Section

Covariadas	Ano 2000			Todos os Anos		
	Mulheres (1)	Homens (2)	Empilhado (3)	Mulheres (4)	Homens (5)	Empilhado (6)
Proporção de mulheres	-0.3020*** (0.0085)	-0.3856*** (0.0084)	-0.4248*** (0.0081)	-0.3094*** (0.0050)	-0.3719*** (0.0049)	-0.4103*** (0.0047)
Tamanho estabelecimento (1/10)	-0.0000*** (0.0000)	0.0002*** (0.0000)	0.0001*** (0.0000)	-0.0001*** (0.0000)	0.0002*** (0.0000)	0.0001*** (0.0000)
Experiência geral	0.0112*** (0.0009)	0.0267*** (0.0008)	0.0190*** (0.0006)	0.0097*** (0.0005)	0.0252*** (0.0005)	0.0174*** (0.0004)
Experiência geral ao quadrado	-0.0001*** (0.0000)	-0.0003*** (0.0000)	-0.0002*** (0.0000)	-0.0001*** (0.0000)	-0.0003*** (0.0000)	-0.0002*** (0.0000)
Experiência no estabelecimento	0.0039*** (0.0000)	0.0035*** (0.0000)	0.0037*** (0.0000)	0.0042*** (0.0000)	0.0037*** (0.0000)	0.0039*** (0.0000)
Comércio	-0.1406*** (0.0073)	-0.2924*** (0.0059)	-0.2328*** (0.0046)	-0.1356*** (0.0042)	-0.2924*** (0.0035)	-0.2312*** (0.0027)
Serviços	0.0208** (0.0064)	-0.0472*** (0.0051)	-0.0214*** (0.0040)	0.0156*** (0.0038)	-0.0718*** (0.0030)	-0.0390*** (0.0024)
Administrativos	0.1890*** (0.0054)	0.0844*** (0.0054)	0.1396*** (0.0038)	0.1827*** (0.0031)	0.0878*** (0.0032)	0.1392*** (0.0022)
Profissionais	0.7718*** (0.0085)	0.5451*** (0.0064)	0.6236*** (0.0051)	0.7898*** (0.0049)	0.5759*** (0.0038)	0.6523*** (0.0030)
Gerentes	0.6329*** (0.0090)	0.6600*** (0.0067)	0.6529*** (0.0053)	0.6067*** (0.0052)	0.6294*** (0.0040)	0.6231*** (0.0032)
Intercepto	-0.1656*** (0.0160)	-0.0578*** (0.0128)	0.0257** (0.0099)	-0.1152*** (0.0096)	-0.0013 (0.0079)	0.0741*** (0.0061)
Mulher			-0.2842*** (0.0060)			-0.2684*** (0.0035)
Mulher × Proporção mulheres			0.1350*** (0.0118)			0.1158*** (0.0069)
R ²	0.5323	0.5244	0.5409	0.4988	0.4836	0.5022
Número de observações	92907	148701	241608	308743	487861	796604

Notas:

1. A variável dependente é o log do salário horário real.
2. Erros-padrões entre parênteses.
3. Níveis de significância: *:10% **:5% ***:1%.
4. A especificação com todos os anos inclui dummies de ano.

dos métodos longitudinais corroboram essa constatação.

As comparações acima sugerem que a estimação da relação entre salários e segregação feminina no nível do estabelecimento é afetada pela inclusão de efeitos específicos não-observáveis para trabalhadores e estabelecimentos. Em outras palavras, o fato de que o parâmetro de interesse é sensível à inclusão desses efeitos indica que as características não-observáveis (invariantes no tempo) de trabalhadores e empregadores capturam parte da correlação entre a composição por gênero dos estabelecimentos e os salários de homens e mulheres. Essa evidência permite concluir que os fatores específicos não-observados relativos aos trabalhadores (*e.g.* preferências e habilidades produtivas) e aos empregadores (*e.g.* discriminação e eficiência) realmente importam para a estimação do efeito de interesse.

Outra constatação que surge das comparações acima é que as estimativas longitudinais são menos negativas que as de *cross-section*. Nesse sentido, esse resultado revela a existência de uma correlação negativa entre a variável proporção de mulheres no estabelecimento e as características não-observáveis de trabalhadores e estabelecimentos. Uma possível expli-

cação para essa correlação negativa é que trabalhadores (observacionalmente equivalentes) com menos habilidades não-mensuradas tendem a estar empregados em estabelecimentos com maior densidade de mulheres. Alternativamente, empregadores menos eficientes e/ou discriminadores tendem a contratar relativamente mais mulheres.

Em termos dos dois modelos teóricos que desenvolvemos anteriormente, as previsões do modelo de *quality sorting* (seção 1.3.3) são compatíveis com as evidências de um efeito negativo da segregação feminina sobre os salários de homens e mulheres. O modelo monopsonístico de discriminação pelo empregador (seção 1.3.2) é mais flexível no sentido de que prevê sinais distintos para o efeito em questão. Em particular, existe um conjunto de valores dos parâmetros do modelo que levam ao resultado de que o grau de feminização das firmas afeta negativamente os salários de ambos os sexos (ver caso C na Tabela 1.2).

1.9.3 Efeito sobre o Hiato Salarial por Gênero

Nesta seção, apresentamos os resultados da amplamente utilizada decomposição de Oaxaca-Blinder para o hiato salarial por gênero.⁶⁶ Fazendo uso de regressões de salários para cada sexo, essa metodologia busca decompor o hiato salarial em duas partes. A primeira está associada ao diferencial de “premiação” das características dos trabalhadores, sendo medida pelas diferenças dos correspondentes coeficientes das duas regressões. Esse componente é normalmente considerado a parcela não-justificada ou discriminatória do hiato salarial. A segunda parte é atribuída ao diferencial de “dotações” entre homens e mulheres, sendo medida pelas diferenças das características (médias) dos dois grupos. Esse componente é usualmente visto como a parcela justificada ou não-discriminatória do hiato salarial.⁶⁷ Idealmente, essa decomposição procura comparar a estrutura salarial observada com a que prevaleceria na ausência de discriminação.

Formalmente, a decomposição utiliza inicialmente as seguintes regressões (omitindo sub-escritos):

$$\ln(w_h) = X_h\beta_h + e_h \quad (1.71)$$

$$\ln(w_m) = X_m\beta_m + e_m, \quad (1.72)$$

onde h e m denotam respectivamente homens e mulheres, X , que inclui uma constante, representa o vetor de características de trabalhadores e estabelecimentos, β é um vetor de parâmetros, e e um termo de erro. Dada a propriedade de que os resíduos de uma regressão

⁶⁶Oaxaca (1973) e Blinder (1973).

⁶⁷Na realidade, diferenças de dotações entre os dois grupos podem ser uma consequência da existência de discriminação passada ou presente contra as mulheres. Por exemplo, percebendo que são discriminadas no mercado de trabalho, as mulheres podem ter menos incentivos para investir em capital humano, o que leva a que parte das diferenças de características produtivas entre os sexos possa ser considerada discriminatória.

têm média igual a zero, pode-se utilizar as equações (1.71) e (1.72) para obter:

$$\overline{\ln(w_h)} - \overline{\ln(w_m)} = \overline{X_h}b_h - \overline{X_m}b_m, \quad (1.73)$$

onde as barras superiores denotam médias e b_g , $g = h, m$, representam os coeficientes estimados. Adicionando e subtraindo da equação (1.73) o termo $\overline{X_m}b_h$ ou o termo $\overline{X_h}b_m$, obtêm-se duas decomposições distintas para o diferencial de salários médios entre os sexos:

$$\overline{\ln(w_h)} - \overline{\ln(w_m)} = (b_h - b_m)\overline{X_m} + (\overline{X_h} - \overline{X_m})b_h \quad (1.74)$$

$$= (b_h - b_m)\overline{X_h} + (\overline{X_h} - \overline{X_m})b_m. \quad (1.75)$$

O primeiro termo do lado direito dessas duas equações mede o quanto do hiato salarial se deve à diferença entre os coeficientes, onde essa diferença é ponderada por $\overline{X_m}$ ou $\overline{X_h}$. Esse é o componente do diferencial que é considerado discriminatório. O segundo termo do lado direito captura o quanto do hiato pode ser atribuído à diferença de características médias entre os sexos, sendo esta ponderada pelos coeficientes da regressão para homens (b_h) ou para mulheres (b_m). Esse componente se refere à parte do diferencial considerada não discriminatória. Denominaremos a equação (1.74) de decomposição “masculina”, uma vez que o vetor b_h é utilizado como referência (contra-factual) da estrutura salarial considerada não-discriminatória. A equação (1.75) será denominada de decomposição “feminina”.

Claramente, as equações (1.74) e (1.75) não fornecem necessariamente os mesmos resultados. Esse é um problema comum dos números índices. Uma possível maneira de contornar esse problema é definir o vetor de coeficientes $b = Wb_h + (I - W)b_m$, onde W é uma matriz de pesos e I a matriz identidade. Utilizando essa definição de b , se somarmos e subtrairmos os termos $\overline{X_h}b$ e $\overline{X_m}b$ da equação (1.73), podemos obter após alguns re-arranjos algébricos a seguinte expressão:

$$\overline{\ln(w_h)} - \overline{\ln(w_m)} = (\overline{X_h}(I - W) + \overline{X_m}W)(b_h - b_m) + (Wb_h + (I - W)b_m)(\overline{X_h} - \overline{X_m}). \quad (1.76)$$

Esta equação é mais flexível que as equações (1.74) e (1.75), já que permite a escolha do peso que se quer dar a cada uma das estruturas salariais dos sexos. Apresentaremos os resultados para três diferentes escolhas de W : (i) $W = I$, que conduz à decomposição masculina; (ii) $W = 0$, que fornece a decomposição feminina; e, (iii) $W = 0,5I$, que produz uma média situada no meio do caminho entre os dois casos anteriores.

Uma outra possível solução foi proposta por Neumark (1988) e Oaxaca & Ransom (1988), que sugeriram que os coeficientes da regressão empilhada por sexo podem ser usados como uma *proxy* para a estrutura salarial não-discriminatória. Isso leva a:

$$\overline{\ln(w_h)} - \overline{\ln(w_m)} = \overline{X_h}(b_h - b^*) + \overline{X_m}(b_m - b^*) + (\overline{X_h} - \overline{X_m})b^*, \quad (1.77)$$

onde b^* é o vetor de coeficientes estimados da regressão empilhada. A soma dos primeiros dois termos do lado direito dessa equação representa as diferenças nos coeficientes, enquanto o último termo corresponde às diferenças nas características.

O principal objetivo dessa análise de decomposição é investigar o quanto do hiato salarial por gênero pode ser atribuído à segregação feminina no nível do estabelecimento. Assim, focaremos nossa atenção nas contribuições dos dois componentes da decomposição (características e coeficientes) que se referem à variável proporção de mulheres no estabelecimento. Para isso, utilizaremos o nosso modelo econométrico mais geral, especificamente as estimativas obtidas do MMG em sistema em um-passo (Tabela 1.5). Deve-se observar, todavia, que resultados relativamente distintos emergiriam se utilizássemos os coeficientes estimados de outros métodos empregados anteriormente.

A Tabela 1.20 reporta os resultados das quatro especificações da decomposição acima apresentadas. As colunas (A) e (B) contêm, respectivamente, os resultados da decomposição masculina e feminina, a coluna (C) os da especificação equiponderada (i.e. $W = 0,5I$ na equação (1.76)) e a coluna (D) os baseados na regressão empilhada. A primeira linha da Tabela informa o valor do hiato do logaritmo do salário horário entre homens e mulheres. O Painel A apresenta as contribuições dos dois componentes (características e coeficientes) como um todo, enquanto o Painel B se refere somente às contribuições relacionadas especificamente à variável proporção de mulheres. A contribuição percentual de cada componente como um todo para o hiato salarial é apresentada nas linhas (2.a) e (3.a) do Painel A. O Painel B contém a contribuição percentual de cada componente da variável proporção de mulheres para o hiato salarial total (linhas (4.a) e (5.a)), assim como para os seus respectivos componentes específicos (linhas (4.b) e (4.c)).

Iniciando pelo Painel A, vemos que as três primeiras especificações fornecem resultados semelhantes em termos das contribuições dos dois componentes da decomposição: cerca de 46% do hiato salarial pode ser atribuído às diferenças de características entre homens e mulheres, com os outros 54% sendo absorvidos pelas diferenças de coeficientes. Distintamente, a decomposição baseada na regressão empilhada revela que o primeiro componente responde por cerca de 2/3 do hiato, ao passo que o segundo pelo 1/3 restante. Os resultados desta última especificação estão mais em linha com os disponíveis na literatura internacional.⁶⁸

Os resultados apresentados no Painel B mostram que o sinal da contribuição da diferença nos coeficientes associados à variável proporção de mulheres é negativo e oposto ao da contribuição da diferença na média dessa variável para homens e mulheres. Isso implica que o efeito do primeiro componente é de reduzir o hiato salarial, ao passo que o do segundo é de aumentá-lo. Vale observar, no entanto, que esses dois efeitos tendem a se cancelar, o que significa que a contribuição conjunta da variável de interesse é praticamente insignificante. Um ponto importante a se destacar é que a contribuição da diferença

⁶⁸Ver, por exemplo, Neumark (1988), Oaxaca & Ransom (1988) e Reilly & Wirjanto (1999a).

Tabela 1.20: Decomposições de Oaxaca-Blinder para o Hiato Salarial por Gênero

	Decomposição Masculina (A)	Decomposição Feminina (B)	Decomposição Eqüiponderada (C)	Regressão Empilhada (D)
(1) - Hiato Salarial	0.3571	0.3571	0.3571	0.3571
A - Todas as Características				
(2) - Diferenças Características	0.1702	0.1578	0.1640	0.2416
(2.a) - Percentual (2)/(1)	47.7	44.2	45.9	67.7
(3) - Diferenças Coeficientes	0.1869	0.1993	0.1931	0.1155
(3.a) - Percentual (3)/(1)	52.3	55.8	54.1	32.3
B - Proporção de Mulheres				
(4) - Diferenças Características	0.0657	0.0296	0.0477	0.0787
(4.a) - Percentual (4)/(1)	18.4	8.3	13.3	22.0
(4.b) - Percentual (4)/(2)	38.6	18.8	29.1	32.6
(5) - Diferenças Coeficientes	-0.0622	-0.0262	-0.0442	-0.0752
(5.a) - Percentual (5)/(1)	-17.4	-7.3	-12.4	-21.1
(5.b) - Percentual (5)/(3)	-33.3	-13.1	-22.9	-65.1

Notas:

1. A decomposição masculina (feminina) utiliza os coeficientes da regressão para homens (mulheres) como ponderadores. A decomposição eqüiponderada utiliza os coeficientes das regressões por sexo dando pesos iguais a eles. A decomposição da regressão empilhada utiliza os coeficientes da regressão em conjunto para homens e mulheres.

2. O hiato salarial se refere ao diferencial das médias do logaritmo do salário horário entre homens e mulheres.

na média da variável que mede a segregação feminina é relativamente alta, representando de 8% a 22% do hiato salarial total (linha (4.a)), e de 19% a 39% da contribuição das diferenças em todas as características (linha (4.b)).⁶⁹ Nesse sentido, mudanças no grau de segregação de mulheres entre os estabelecimentos podem ter importantes impactos para reduzir o diferencial de salários por gênero. Na próxima seção, apresentamos uma tentativa simples de medir esses impactos.

⁶⁹Para a especificação baseada na regressão empilhada, a diferença na média da variável proporção de mulheres constitui a contribuição mais importante tanto para o hiato salarial total, quanto para a contribuição conjunta de todas as características.

1.9.4 Considerações de Política

Várias políticas têm sido propostas para reduzir o hiato salarial entre homens e mulheres. Em linhas gerais, podemos dividi-las em dois grupos.⁷⁰ A primeira é baseada em “preços”, no sentido de que objetiva equalizar diretamente os salários de homens e mulheres que estão em ocupações similares num mesmo estabelecimento. O exemplo mais conhecido desse tipo de iniciativa é a política denominada de “valor comparável” (*comparable worth*).⁷¹ Já o segundo grupo de políticas centra-se em “quantidades”, uma vez que busca garantir igualdade de oportunidades de emprego para as mulheres nos estabelecimentos. Exemplos desse segundo grupo são programas de equidade no emprego e iniciativas de ação afirmativa. Observe-se que somente esse segundo tipo de política busca atacar diretamente a questão segregação por gênero no nível do estabelecimento.

Embora ambas as formas de política sejam potencialmente capazes de reduzir o diferencial de salários por sexo, centraremos nossa atenção somente nos efeitos das que procuram reduzir diretamente a segregação por gênero. Para isso, adotaremos a metodologia proposta por Reilly & Wirjanto (1999a) para avaliar o impacto de um programa hipotético de equidade no emprego. A idéia dessa metodologia é computar a magnitude da mudança no hiato salarial que decorreria da imposição de um nível pré-especificado para a proporção de mulheres em todos os estabelecimentos. Nesse exercício, assume-se que o programa modifica apenas “quantidades” (i.e. a composição do emprego por gênero dos estabelecimentos), mas não “preços” (i.e. o coeficiente da regressão associado com a composição feminina dos estabelecimentos).

Para computar o impacto de um programa hipotético como esse, utilizaremos as decomposições de Oaxaca-Blinder apresentadas na seção precedente. Especificamente, temos inicialmente:

$$\overline{\ln(w_h)} - \overline{\ln(w_m)} = \overline{p_h} b_{h,pfem} - \overline{p_m} b_{m,pfem} + (\overline{X_h^*} b_h^* - \overline{X_m^*} b_m^*), \quad (1.78)$$

onde $\overline{p_h}$ é a proporção média de mulheres para o grupo de homens, $\overline{p_m}$ a proporção média de mulheres para o grupo de mulheres, $b_{g,pfem}$, $g = h, m$, são os coeficientes estimados da variável proporção de mulheres nas regressões para homens e mulheres, e $\overline{X_g^*} b_g^*$, $g = h, m$, representa o produto de todas as demais características médias com os seus respectivos coeficientes nessas regressões (incluindo os interceptos).

Suponha que o programa restrinja todos os estabelecimentos a terem uma proporção de mulheres pré-fixada, denotada por $\overline{p_p}$. Isso levaria a que os salários que o homem e a

⁷⁰Consideramos apenas aquelas políticas que tratam da questão diretamente no nível do estabelecimento (ou firma). Existem outras iniciativas que não são estritamente focalizadas nesse nível, tais como políticas de treinamento e educacional voltadas para as mulheres.

⁷¹Por exemplo, se existem duas ocupações consideradas de valor comparável num estabelecimento, com uma delas predominantemente feminina e a outra masculina (*e.g.* secretárias e motoristas), a política de valor comparável estabelece que o empregador deve pagar o mesmo salário para os trabalhadores dessas duas ocupações.

mulher (médios) receberiam com esse programa fossem, respectivamente:

$$\overline{\ln(w_h)}^p = \overline{p_p} b_{h,pfem} + \overline{X_h^*} b_h^* \quad (1.79)$$

$$\overline{\ln(w_m)}^p = \overline{p_p} b_{m,pfem} + \overline{X_m^*} b_m^*. \quad (1.80)$$

Somando e subtraindo as equações (1.79) e (1.80) da expressão (1.78), pode-se obter após algumas manipulações algébricas que:

$$\overline{\ln(w_h)} - \overline{\ln(w_m)} = (\overline{p_h} - \overline{p_p}) b_{h,pfem} + (\overline{p_p} - \overline{p_m}) b_{m,pfem} + (b_{h,pfem} - b_{m,pfem}) \overline{p_p} + \Delta, \quad (1.81)$$

onde $\Delta = (\overline{X_h^*} b_h^* - \overline{X_m^*} b_m^*)$. Essa equação pode ser usada para analisar o impacto desse programa hipotético sobre o hiato salarial por gênero. O primeiro (segundo) termo do lado direito da equação capta a mudança no salário médio de homens (mulheres) que resultaria da modificação na proporção de mulheres existente nos estabelecimentos para o nível pré-fixado. O terceiro termo reflete a parte do hiato salarial correspondente à diferença nos coeficientes da variável proporção de mulheres entre as regressões para os dois sexos. Cabe assinalar que, com a hipótese de que a intervenção modifica apenas quantidades, esse terceiro termo não é afetado pelo programa. O último termo da equação permanece fixo no exercício, podendo ser então visto como um “intercepto”.

Para calcular o efeito da política por meio da equação (1.81), é necessário escolher o valor de seu único parâmetro livre: $\overline{p_p}$. Utilizaremos quatro diferentes níveis para esse parâmetro. O primeiro corresponde à menor proporção de mulheres observada entre as categorias ocupacionais com que trabalhamos (32% para ocupações de gerência). O segundo simplesmente respeita a proporção de mulheres existente no mercado de trabalho formal (39%). O terceiro representa a proporção de mulheres na força de trabalho como um todo, incluindo os trabalhadores por conta-própria e sem carteira assinada (44%).⁷² O quarto nível corresponde ao caso em que todos os estabelecimentos teriam 50% de mulheres (e de homens).

A Tabela 1.21 reporta tanto os resultados da decomposição apresentada na equação (1.81), quanto os impactos potenciais do programa que impõe níveis pré-fixados para a composição feminina dos estabelecimentos. Cada coluna contém os resultados correspondentes aos quatro níveis mencionados no parágrafo anterior, os quais estão apresentados na linha (1) da Tabela. A linha (2) informa o hiato do log do salário horário por gênero e a linha (3) a parcela desse hiato atribuída a todas as variáveis da regressão, à exceção da referente à proporção de mulheres no estabelecimento (i.e. o termo Δ da equação (1.81)). As linhas (4), (5) e (6) apresentam respectivamente as partes do hiato salarial associadas a cada um dos três primeiros termos da equação (1.81). A contribuição percentual de cada termo está apresentada abaixo das respectivas linhas. A linha (7) reporta o impacto poten-

⁷²Esse número foi calculado com base na Pesquisa Nacional por Amostra de Domicílios (PNAD) para a área metropolitana de Belo Horizonte no ano de 2001.

Tabela 1.21: Decomposição do Impacto de um Programa que Fixa a Proporção de Mulheres nos Estabelecimentos

Componentes	Programa 1 (A)	Programa 2 (B)	Programa 3 (C)	Programa 4 (D)
(1) - Nível Pré-Fixado	32%	39%	44%	50%
(2) - Hiato Salarial	0.3571	0.3571	0.3571	0.3571
(3) - Delta	0.3536	0.3536	0.3536	0.3536
(4) - Primeiro Componente	0.0153	0.0291	0.0380	0.0508
(4.a) - Percentual (4)/(2)	4.3	8.2	10.6	14.2
(5) - Segundo Componente	0.0227	0.0165	0.0125	0.0067
(5.a) - Percentual (5)/(2)	6.4	4.6	3.5	1.9
(6) - Terceiro Componente	-0.0347	-0.0423	-0.0471	-0.0542
(6.a) - Percentual (6)/(2)	-9.7	-11.8	-13.2	-15.2
(7) - Impacto (4)+(5)	0.0380	0.0456	0.0505	0.0575
(7.a) - Percentual (7)/(2)	10.7	12.8	14.1	16.1

Notas:

Definições das linhas: Linha (1) Nível pré-fixado para a proporção de mulheres nos estabelecimentos; Linha (2) Hiato do logaritmo do salário entre homens e mulheres; Linha (3) Parcela do hiato devida a todas as covariadas (incluindo o intercepto) com exceção da variável proporção de mulheres no estabelecimento; Linha (4) Efeito sobre o hiato da mudança na proporção média de mulheres para o grupo de homens; Linha (4.a) Contribuição percentual da linha (4); Linha (5) Efeito sobre o hiato da mudança na proporção média de mulheres para o grupo de mulheres; Linha (5.a) Contribuição percentual da linha (5); Linha (6) Efeito sobre o hiato da diferença nos coeficientes da variável proporção de mulheres no estabelecimento; Linha (7) Soma das linhas (4) e (5); Linha (7.a) Contribuição percentual da linha (7).

cial da intervenção, correspondendo a soma das linhas (4) e (5).⁷³ A linha (7.a) apresenta a contribuição percentual do programa para modificar o diferencial de salários por gênero.

O primeiro ponto a se notar da Tabela 1.21 é que elevações no nível pré-fixado para a proporção de mulheres aumentam o impacto do programa (linha (7.a)). O segundo ponto é que as elevações dos níveis pré-fixados aumentam (reduzem) a contribuição do primeiro (segundo) componente. Esses dois pontos são uma consequência dos seguintes resultados: (i) a relação entre salários e proporção de mulheres é menos negativa para mulheres do que para homens; e, (ii) a expressiva distância entre as proporções médias de mulheres para os grupos de homens e de mulheres.⁷⁴ Outro ponto a se observar é que as magnitudes

⁷³Cabe lembrar que a suposição de que os coeficientes não se alteram com a implementação da política leva a que os números da linha (6) não contribuam para o cálculo do impacto total.

⁷⁴Os valores de $\overline{p_h}$ e $\overline{p_m}$ são cerca de 0,24 e 0,58 respectivamente, com $(b_{m,fem}, b_{h,fem}) =$

do impacto não são pequenas. Por exemplo, tomando o nível pré-fixado de 44% – que representa a proporção de mulheres na força de trabalho como um todo –, o programa seria capaz de reduzir o hiato salarial por gênero em 14,1%, ou seja, em cerca de 1/7 desse hiato. É interessante assinalar que essa magnitude é similar à encontrada por [Reilly & Wirjanto \(1999a, Tab. 4\)](#). De fato, ao imporem um nível pré-fixado de 43%, esses autores obtêm uma redução no hiato salarial por sexo no Canadá de 18,1%.

Podemos concluir com base nesse exercício simples que parece existir espaço para reduzir o diferencial de salário por gênero no Brasil por meio de políticas que busquem diminuir diretamente a segregação feminina no nível do estabelecimento. Isso não significa que outras intervenções públicas não possam atingir reduções similares ou mais altas (*e.g.* política de valor comparável) do que as aqui calculadas. No entanto, essa investigação requer uma análise de custo-benéfico apropriada, uma tarefa que está além dos objetivos desse capítulo.

1.10 Conclusões

Neste capítulo, procuramos examinar como a segregação por gênero nos estabelecimentos afeta os salários de homens e mulheres. Para investigar empiricamente esse tema, utilizamos um painel de dados brasileiros com informações combinadas de empregadores e empregados para o período de 2000 a 2002. Essa base de dados reúne informações prestadas anualmente por todos os estabelecimentos registrados no país, sendo rica em termos das informações que contém sobre as características dos trabalhadores e estabelecimentos. Metodologicamente, fizemos uso de vários modelos longitudinais para analisar o efeito da segregação feminina no nível do estabelecimento sobre os salários de homens e mulheres. Também obtivemos estimativas de *cross-section* desse efeito, as quais são comparadas com os resultados longitudinais.

As principais contribuições deste capítulo para a literatura são de duas ordens. Como toda a literatura anterior é baseada em dados de *cross-section*, a primeira contribuição é que este é o primeiro estudo que utiliza um painel de dados combinados de empregadores e empregados para avaliar se a estimação da relação entre salários e segregação feminina no nível do estabelecimento é afetada pela inclusão de fatores específicos não-observáveis de trabalhadores e estabelecimentos. A segunda contribuição se refere ao desenvolvimento de dois modelos teóricos que procuram explicar os sinais do efeito da segregação por gênero nas firmas sobre os salários de homens e de mulheres.

Nossa análise longitudinal indica que o efeito de interesse é negativo tanto para homens quanto para mulheres, sendo mais intenso (i.e. mais negativo) sobre os salários do primeiro

($-0,0891$, $-0,1975$). Assim, níveis pré-fixados mais altos tendem a aumentar o impacto do programa, com a contribuição do primeiro componente se tornando mais baixa (alta) que a do segundo a medida em que esses níveis se aproximam de $\overline{p_h}$ ($\overline{p_m}$).

grupo. A obtenção de sinais negativos está em linha tanto com os encontrados na literatura anterior, quanto com os nossos próprios resultados de *cross-section*. Todavia, a evidência de um efeito mais intenso para homens não corrobora o que foi previamente encontrado.

Embora ambas apresentem sinal negativo, as estimativas longitudinais são bastante menores em termos absolutos que as nossas estimativas de *cross-section*. Esse resultado nos permite extrair duas conclusões. Em primeiro lugar, a existência de uma diferença de magnitude entre os dois tipos de estimativas revela que os fatores não-observáveis de trabalhadores e estabelecimentos são correlacionados com a variável proporção de mulheres nos estabelecimentos. Portanto, podemos concluir que as estimativas tradicionais de *cross-section* são enviesadas.

Em segundo lugar, o fato de que estimativas longitudinais são menores em módulo que as de *cross-section* indica que a correlação entre os fatores não-observáveis e a variável que mede a segregação feminina no nível do estabelecimento é negativa. Uma possível explicação para isso é que trabalhadores de ambos os sexos com menores níveis de habilidade não-mensurada tendem a ocupar empregos em estabelecimentos com maior densidade de mulheres. Alternativamente, pode ser também que empregadores menos discriminadores/eficientes tendam a contratar relativamente mais trabalhadores do sexo feminino.

O conhecido modelo de Becker/Arrow de discriminação pelo empregador foi inserido num arcabouço de concorrência monopsonística de forma a permitir às firmas pagarem salários distintos. Assume-se que as firmas são heterogêneas tanto em termos de discriminação quanto de produtividade. Sob a hipótese de que existe uma correlação entre os níveis de produtividade e de discriminação dos empregadores, o modelo prediz várias direções para os sinais que a relação entre o grau de feminização das firmas e o salário de homens e mulheres pode ter. Em particular, as predições desse modelo são compatíveis com os resultados empíricos de que o efeito da segregação feminina é negativo sobre os salários de ambos os sexos.

O segundo modelo que desenvolvemos formaliza pela primeira vez na literatura a combinação de algumas idéias propostas pelas teorias de discriminação estatística e de *quality sorting*. Assume-se que as firmas são heterogêneas em suas demandas relativas por trabalho qualificado, e que os trabalhadores incorrem em custos para adquirir qualificações. Os empregadores não sabem *a priori* a verdadeira produtividade dos trabalhadores e formam crenças de que as mulheres precisam adquirir um maior nível de qualificação do que os homens para trabalhar em postos qualificados. O modelo prediz que as mulheres (homens) tenderão a estar sobre-representados em firmas cujas demandas relativas por qualificação são menores (maiores), o que leva a que o efeito da segregação feminina seja negativo sobre os salários dos dois sexos. Essas predições também são compatíveis tanto com os nossos resultados empíricos, quanto os da literatura prévia.

Como em outros estudos (*e.g.* Johnson & Solon (1986), Carrington & Troske (1998) e Reilly & Wirjanto (1999a)), os resultados empíricos deste capítulo sugerem que a segre-

gação por gênero nos estabelecimentos responde por uma parcela importante do diferencial de salários entre homens e mulheres. Isso é confirmado pelas decomposições desse diferencial aqui implementadas, as quais indicam que a contribuição da diferença da proporção (média) de mulheres nos estabelecimentos entre os grupos de homens e mulheres é relativamente alta. Um exercício simples de simulação de um programa de igualdade de oportunidades de emprego entre os sexos mostra que o hiato salarial por gênero poderia ser reduzido em cerca de $1/7$. Deve-se observar, entretanto, que outras políticas poderiam ser mais custo-efetivas para reduzir esse hiato. Contudo, esse tipo de análise requer investigações comparativas mais profundas, o que está além do escopo deste capítulo.

Capítulo 2

Efeitos do Tempo, Efeitos Econômicos e Impactos de Programas Sociais

2.1 Introdução

Existe uma extensa literatura empírica que vem buscando avaliar os impactos de programas sociais.¹ As evidências mostram que os impactos médios dos programas tendem a variar nas seguintes dimensões: (i) os grupos demográficos (ou de qualificação) enfocados; (ii) as regiões do país onde o programa foi implementado; (iii) as coortes de participantes do programa; e, (iv) o horizonte temporal dos impactos. Neste capítulo, focaremos nossa atenção nesta última dimensão.

Suponhamos que o grupo demográfico (ou de qualificação) seja fixado, e que sejamos capazes de observar o impacto médio de uma intervenção (*e.g.* um programa de treinamento) para uma de suas coortes de participantes em uma região de um país. Suponhamos também que esse impacto seja observado para um certo intervalo de tempo. Dado esse contexto, uma questão natural que se pode levantar é a seguinte: É possível afirmar que a trajetória temporal do impacto médio do programa representa o *verdadeiro* efeito da intervenção ao longo do tempo? Em princípio, a resposta a esta pergunta é negativa, uma vez que podem existir outros fatores que influenciam a trajetória temporal observada. Em particular, um fator potencial são as mudanças nas condições econômicas que, direta ou indiretamente, afetam os resultados do programa ao longo do tempo. Por exemplo, parece natural supor que pelo menos uma parte do padrão temporal do impacto médio de um programa de treinamento seja explicado por deslocamentos na oferta ou na demanda relativa por trabalhadores treinados.

¹Para resenhas sobre essa literatura, ver, por exemplo, Barnow (1987), Gueron (1990), LaLonde (1995) e Heckman et al. (1999).

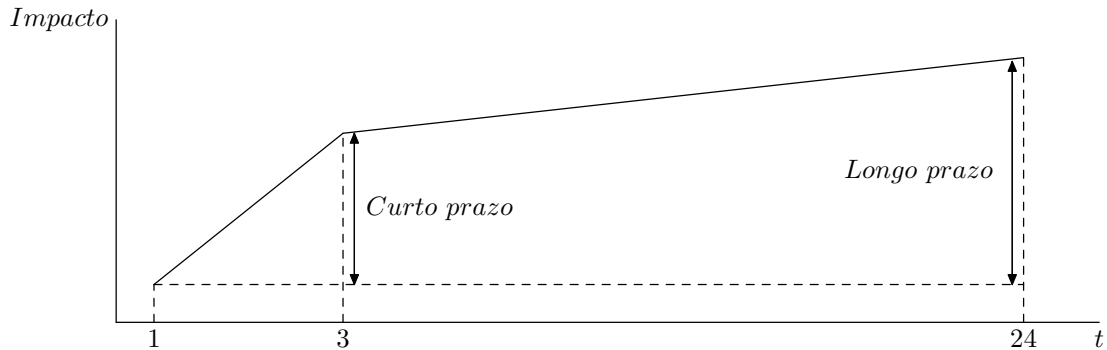
A fim de distinguir o verdadeiro efeito temporal de uma intervenção das possíveis influências externas que afetam a trajetória dos impactos médio dessa intervenção, introduzimos um conceito que denominamos *efeito interno temporal* de um programa. Esse efeito pode ser entendido como um tipo de efeito de “duração”, que é disparado e está associado somente à própria intervenção. Por exemplo, *ceteris paribus*, um programa de treinamento poderia gerar impactos que são negativos nos primeiros meses após início da intervenção, mas que se tornam positivos desse ponto em diante (ou vice-versa). Introduzimos também um outro conceito que denominamos *efeito econômico externo*. Esse efeito capta somente a influência de mudanças nas variáveis econômicas (agregadas) que afetam os impactos médios de um programa. Por exemplo, se pudéssemos manter tudo o mais constante, é possível que melhores condições econômicas estejam associadas a impactos mais elevados do programa, ou vice-versa. Mais adiante, definiremos esses conceitos formalmente.²

A Figura 2.1 procura ilustrar essas idéias de uma maneira simples. Ela representa uma situação hipotética de um programa de treinamento cujos impactos médios sobre o emprego foram consistentemente estimados para três meses distintos após o início do programa: $t = 1, 3$ e 24 . Denominemos as diferenças nos impactos médios entre os meses 1 e 3 e 1 e 24 de impactos de curto e longo prazos do programa, respectivamente. De imediato, poder-se-ia pensar que essas diferenças representam o verdadeiro impacto do programa para esses dois horizontes de tempo. Contudo, essa constatação pode não ser correta. Considere o seguinte contexto. Suponha que, logo após o início do programa, seus participantes se tornem “enclausurados” (“*locked-in*”), em decorrência do fato de que o programa absorve parte do tempo disponível desses participantes para buscar emprego. Assuma também que, após o seu final, o programa torne seus participantes efetivamente treinados e mais produtivos. Adicionalmente, suponha que a demanda relativa por trabalhadores treinados esteja se deslocando para cima (baixo) no curto (longo) prazo, com a oferta relativa mantida constante durante todo o período. Assim, dado esse contexto, poderíamos dizer que o impacto positivo que observamos no curto prazo se deve à melhora nas condições econômicas (i.e. ao aumento na demanda relativa), que mais que compensou os efeitos negativos de *lock-in* do programa. No longo prazo, entretanto, a situação é revertida: o efeito positivo observado se deve mais ao programa em si mesmo do que a piora nas condições econômicas (i.e. na redução da demanda relativa). Portanto, poderíamos concluir que o efeito interno temporal é negativo (positivo) no curto (longo) prazo, com o oposto se aplicando ao efeito econômico externo.

Um ponto importante a se assinalar é que tanto o sinal quanto a magnitude do efeito interno temporal podem variar de acordo com a situação econômica vigente. Por exemplo,

²A influência das condições econômicas sobre os impactos médios de uma intervenção já foi reconhecida na literatura de avaliação de programas (*e.g.* Heckman et al. (1997)). Entretanto, tanto quanto sabemos, este capítulo é o primeiro estudo que procura analisar de forma geral e sistemática as diferenças entre os efeitos do programa em si mesmo e os efeitos relacionados às mudanças nas condições econômicas.

Figura 2.1: Impactos Hipotéticos Médios de um Programa



é possível que esse efeito seja negativo quando as condições econômicas são ruins, ou vice-versa. De modo semelhante, o sinal e a magnitude do efeito econômico externo pode depender do horizonte de tempo para o qual ele é calculado. Por exemplo, é possível que o efeito econômico externo seja positivo no curto prazo e negativo num horizonte mais longo, ou vice-versa.

Nosso interesse centra-se na identificação desses efeitos. As possibilidades de que: (i) o efeito interno temporal interaja com as condições econômicas, e (ii) o efeito econômico externo dependa do horizonte temporal, precisam ser levadas em conta na análise de identificação. Em princípio, a teoria econômica não fornece *a priori* a forma exata de como se dão essas interações. Assim, a análise de identificação de ambos os efeitos deve ser conduzida com base no uso de estratégias flexíveis, entre as quais a não-paramétrica é claramente a mais geral.³ Algumas vantagens adicionais da busca pela identificação não-paramétrica desses efeitos são que ela simplifica bastante a análise e revela mais claramente as forças por trás da identificação.

O principal objetivo deste capítulo é investigar quais aspectos desses dois efeitos podem ser identificados não-parametricamente. Para tanto, partimos de uma especificação não-paramétrica para os impactos médios de um programa e, impondo certas restrições ao modelo, analisamos que partes desses efeitos podem ser identificadas de forma não-paramétrica.

Nossa análise de identificação é conduzida para dois contextos distintos em relação à disponibilidade de informações que um analista tipicamente tem. No primeiro, assume-se que existem dados somente de uma coorte de participantes do programa para pelo menos duas regiões de um país. No segundo, supõe-se que os dados estão disponíveis somente para uma região do país (ou país como um todo), para pelo menos duas coortes distintas de participantes do programa.

³Existem diversos trabalhos na literatura empírica de avaliação de programas que procuram controlar para a influência de variáveis econômicas sobre os impactos médios de um programa (*e.g.* Sianesi (2001), Zhang (2003) e Røed & Zhang (2003)). Entretanto, isso é tipicamente implementado por meio da imposição de formas funcionais específicas na estimação dos impactos médios dos programas sendo avaliado.

Ao longo deste capítulo, abstraímos de considerações sobre como os impactos médios do programa foram estimados. Para isso, mantemos a hipótese de que o analista conhece (ou foi capaz de estimar consistentemente) esses impactos médios. Supomos também que o analista sabe o conjunto completo de variáveis que determinam os níveis médios dos impactos, sendo essas variáveis medidas sem erro. Adicionalmente, não levamos em consideração possíveis aspectos relacionados a efeitos de equilíbrio geral. Isso pode ser justificado se o impacto do programa sobre a economia é insignificante, uma situação que tende a ocorrer quando a escala da intervenção é pequena relativamente ao tamanho da economia (ou mercado de trabalho).

Acreditamos que a investigação sobre os dois efeitos de interesse pode ser importante para o aprimoramento das políticas públicas. Por exemplo, para aqueles programas que podem ser (ou de fato são) replicados continuamente no tempo, o conhecimento desses efeitos pode ajudar os responsáveis pelo programa a prever melhor os impactos da intervenção e, portanto, permitir o aperfeiçoamento de seu desenho.

Além desta introdução, este capítulo contém quatro seções. Nas seções 2 e 3, definimos os arcabouços analíticos e derivamos os resultados de identificação dos efeitos de interesse respectivamente para os contextos de múltiplas regiões e múltiplas coortes. Na seção 4, apresentamos um exercício empírico para o qual utilizamos dados experimentais que foram coletados para avaliar a efetividade de um programa nacional de treinamento de mão-de-obra no Brasil (Plano Nacional de Qualificação do Trabalhador - PLANFOR). Uma vez que esses dados estão disponíveis para duas áreas metropolitanas do país e uma coorte de participantes, esse exercício ilustra a análise de identificação apresentada na seção 2. A seção 5 contém as conclusões.

2.2 Múltiplas Regiões e Coorte Única

Considere uma situação na qual o analista conhece o impacto médio de um programa sobre uma variável de interesse para uma coorte de participantes em pelo menos duas regiões de um país. Suponha que esse impacto médio é conhecido para vários períodos de tempo e para diferentes tipos de indivíduos. Denotando o impacto médio do programa para os indivíduos do tipo $i \in \{1, \dots, I\}$, na região $r \in \{1, \dots, R\}$ e no período $t \in \{1, \dots, T\}$ por α_{irt} , pode-se escrever:⁴

$$\alpha_{irt} = m(Z_{1ir}, Z_{2irt}, X_{irt}, d_t, r) \quad (2.1)$$

⁴O impacto médio pode corresponder a distintos parâmetros de interesse utilizados na literatura de avaliação de programas. Por exemplo, se ele significa o efeito médio do tratamento (*Average Treatment Effect - ATE*): $\alpha_{irt} = E[Y_{irt}^1 - Y_{irt}^0 | Z_{1ir}, Z_{2irt}, X_{irt}, d_t, r]$, onde Y^1 representa a variável de resultado para aqueles que recebem o tratamento, e Y^0 a variável de resultado caso os tratados não tivessem recebido o programa. Para uma discussão sobre definições e métodos de estimação dos parâmetros de interesse, ver, por exemplo: Heckman & Robb (1985), Heckman et al. (1999) e Blundell & Dias (2000).

onde m é uma função não-paramétrica, Z_{1ir} representa o vetor de características invariantes no tempo (*e.g.* sexo e cor) dos indivíduos do tipo i na região r , Z_{2irt} o vetor de características variantes no tempo (*e.g.* idade) dos indivíduos do tipo i na região r no período t , X_{irt} é um vetor de dimensão K de variáveis econômicas externas que afetam α_{irt} , e d_t denota o número de períodos após o final (ou início) do programa pela coorte de participantes observada.⁵ Para simplificar a análise, assumiremos que $K = 1$. Todavia, após a derivação dos resultados da identificação dos efeitos de interesse, discutiremos o caso em que $K > 1$.

O efeito interno temporal está associado a variações em d_t quando os valores dos demais argumentos de m são mantidos constantes em um nível pré-determinado. Similarmente, o efeito econômico externo está relacionado a variações em X_{irt} mantendo-se os valores dos outros argumentos de m fixos num certo nível. Uma vez que não restringimos nem os argumentos de m a serem variáveis contínuas, nem a própria função m a ser diferenciável, as suas derivadas parciais não são necessariamente definidas. Portanto, a fim de estudar a identificação dos efeitos de interesse, trabalhamos com o conceito denominado de *diferenças parciais*, isto é, as diferenças (finitas) que são obtidas quando se varia um dos argumentos de m mantendo-se todos os demais constantes em um nível pré-especificado.

Definição 2.1 *O efeito interno temporal (EIT) é definido pela diferença parcial:*

$$\Delta d(z_1, z_2, x, d'', d', r^\diamond) = m(z_1, z_2, x, d'', r^\diamond) - m(z_1, z_2, x, d', r^\diamond),$$

onde z_1, z_2, x, r^\diamond e $d'' > d'$ são valores específicos assumidos respectivamente por Z_{1ir} , Z_{2irt} , X_{irt} , r e d_t .

Definição 2.2 *O efeito econômico externo (EEE) é definido pela diferença parcial:*

$$\Delta x(z_1, z_2, x'', x', d, r^\diamond) = m(z_1, z_2, x'', d, r^\diamond) - m(z_1, z_2, x', d, r^\diamond),$$

onde z_1, z_2, d e r^\diamond são valores específicos assumidos respectivamente por Z_{1ir} , Z_{2irt} , d_t e r , e x' e x'' são valores distintos assumidos por X_{irt} .

2.2.1 Identificação do EIT: Múltiplas Regiões

Consideramos inicialmente a identificação do EIT no contexto de múltiplas regiões e uma única coorte de participantes. Sejam d' e d'' dois valores de d_t para os quais o analista está interessado em identificar o EIT. Utilizamos as seguintes condições de identificação.⁶

R1: Para cada grupo $i \in \{1, \dots, I\}$ e região $r \in \{1, \dots, R\}$ para os quais o analista está interessado em identificar o EIT, existem valores comuns de Z_{2irt} quando t muda de valor.

⁵Se que \bar{l} representa o período em que essa coorte finalizou (ou iniciou) o programa: $d_t = t - \bar{l}$.

⁶Empregaremos uma notação em que o valor de uma variável qualquer é definido pelos seus sub-escritos.

Assume-se que $z_{2irt'} = z_{2irt''}$, onde $i \in \{1, \dots, I\}$, $r \in \{1, \dots, R\}$, $t' \in \{1, \dots, T-1\}$ e $t' < t''$ representam o grupo, a região e os períodos de tempo de interesse, respectivamente.

R2: Para cada grupo $i \in \{1, \dots, I\}$ e região $r \in \{1, \dots, R\}$ da condição R1, existem valores iguais de X_{irt} quando t muda de valor. Assume-se que $x_{irt'} = x_{irt''}$, onde $i \in \{1, \dots, I\}$, $r \in \{1, \dots, R\}$, $t' \in \{1, \dots, T-1\}$ e $t' < t''$ são os mesmos grupo, região e períodos de tempo da condição R1, respectivamente.

Proposição 2.1 *Para o modelo especificado pela equação (2.1), o efeito interno temporal (EIT) é identificado se as condições R1 e R2 forem satisfeitas.*

Demonstração. *A demonstração é simples. Dos dados, o analista pode observar:*

$$\begin{aligned}\Delta &= \alpha_{irt''} - \alpha_{irt'} \\ &= m(z_{1ir}, z_{2irt''}, x_{irt''}, d_{t''}, r) - m(z_{1ir}, z_{2irt'}, x_{irt'}, d_{t'}, r),\end{aligned}$$

onde $i \in \{1, \dots, I\}$, $r \in \{1, \dots, R\}$, $t' \in \{1, \dots, T-1\}$ e $t' < t''$ representam o grupo, a região e os períodos de tempo de interesse, respectivamente.

Utilizando as condições R1 e R2, tem-se que:

$$\begin{aligned}\Delta &= m(z_{1ir}, z_{2irt'}, x_{irt''}, d_{t''}, r) - m(z_{1ir}, z_{2irt'}, x_{irt'}, d_{t'}, r) \\ &= m(z_{1ir}, z_{2irt'}, x_{irt'}, d_{t''}, r) - m(z_{1ir}, z_{2irt'}, x_{irt'}, d_{t'}, r) \\ &= \Delta d(z_{1ir}, z_{2irt'}, x_{irt'}, d_{t''}, d_{t'}, r),\end{aligned}$$

onde a primeira igualdade segue da condição R1 e a segunda de R2. ■

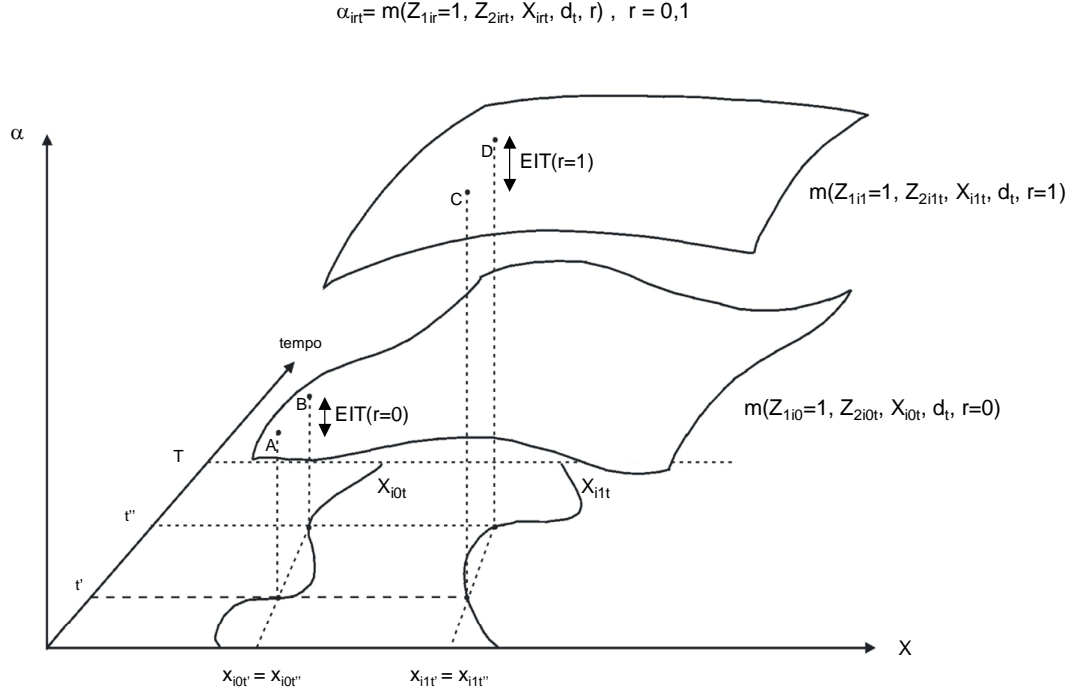
Para visualizar a identificação do EIT no contexto de múltiplas regiões e uma única coorte de participantes, a Figura 2.2 apresenta uma função hipotética m para duas regiões distintas ($r = 0, 1$) e um dado tipo de indivíduos ($Z_{1ir} = 1$, e.g. homens). O eixo vertical mede o impacto médio do programa (α), enquanto os eixos horizontais representam as condições econômicas (X) e o tempo decorrido após o início (ou final) do programa ($d_t = t - \bar{t}$).⁷ Note que, como não impomos restrições sobre a forma funcional de m , as duas superfícies apresentadas na Figura não necessariamente têm o mesmo formato. No plano horizontal $\text{tempo} \times X$, aparecem duas curvas representando a evolução hipotética das condições econômicas para cada região. Observe que as condições econômicas coincidem em cada região nos períodos t' e t'' .⁸

Implicitamente na Figura 2.2, assume-se que o impacto médio do programa é conhecido para indivíduos do tipo $Z_{1ir} = 1$ com as mesmas características variantes no tempo nos

⁷Devido à correspondência direta entre d_t e t , nesta e nas demais Figuras abaixo, um dos eixos horizontais será representado pelo tempo-calendário.

⁸Apesar de não se requerer que a coincidência das condições econômicas ocorra para os mesmos períodos de tempo em cada região, a Figura está desenhada dessa maneira para facilitar a sua visualização.

Figura 2.2: Superfícies da Função Hipotética m para Duas Regiões e o Mesmo Tipo de Indivíduos



períodos t' e t'' (por exemplo, mesma idade). Isso implica que a restrição R1 é satisfeita. O fato de que as condições econômicas são as mesmas em cada região nos períodos t' e t'' leva a que a restrição R2 também seja atendida entre esses dois períodos. Assim, as distâncias verticais entre os pontos A e B para a região $r = 0$, e entre C e D para a região $r = 1$, identificam o EIT para cada região considerada.

A condição R1 requer que o impacto médio do programa seja conhecido para indivíduos com as mesmas características variantes no tempo (i.e. Z_{2irt}) entre os períodos t' e t'' para cada região de interesse. Por exemplo, se Z_{2irt} representa a idade (a) dos indivíduos no ano t , então o analista deve conhecer o impacto médio do programa para indivíduos do tipo i na região r com idade igual a a' no período t' e igual a $a'' = a'$ no período t'' . Em princípio, essa condição não tende a ser muito difícil de ser respeitada para a maior parte das bases de dados utilizadas na estimação dos impactos médios de um programa.

A restrição R2 requer que as condições econômicas sejam as mesmas em cada região entre os períodos para quais se busca a identificação do EIT. Isso implica que a trajetória temporal da variável econômica X_{irt} deve exibir algum tipo de comportamento cíclico (ou pelo menos não ser estritamente monotônica), de tal forma a que se possa encontrar

condições econômicas similares entre dois períodos em geral, e entre t' e t'' em particular. Cabe notar que esse pode ser um requerimento forte, especialmente se o analista estiver interessado em calcular o EIT para diversos períodos de tempo. Ademais, se a dimensão de X_{irt} for maior que a unidade – isto é, se $K > 1$ –, mais difícil será satisfazer a condição R2. De fato, observar *cada* variável em X_{irt} assumindo o mesmo valor entre os mesmos dois períodos de tempo pode ser difícil de se encontrar na prática.

Apesar dessas dificuldades, deve-se assinalar que, sob as condições R1 e R2, a identificação do EIT pode ser obtida de forma *totalmente* não-paramétrica. Esse resultado é importante, uma vez que mostra que o efeito de interesse pode ser estimado sem a imposição de formas funcionais ao modelo. Cumpre observar que R1 e R2 são condições *locais*, isto é, elas só precisam ser atendidas para certos valores das variáveis a elas associadas. Se, no entanto, elas forem satisfeitas para diversos valores dessas variáveis, o efeito de interesse pode ser identificado para um conjunto mais amplo de pontos.

2.2.2 Identificação do EEE: Múltiplas Regiões

Consideramos agora a identificação do EEE. Sejam τ' e τ'' dois períodos de tempo disponíveis nos dados. As condições de identificação são as seguintes:

R3: Seja $Z_{irt} = (Z_{1ir}, Z_{2irt})$. Denote por Ω_r o domínio de Z_{irt} na região $r \in \{1, \dots, R\}$. Defina S como a interseção de $\Omega_{r'}$ e $\Omega_{r''}$, $r' \in \{1, \dots, R\}$, $r'' \in \{1, \dots, R\}$, com $r' \neq r''$. Assume-se que S é não-vazio em $\tau' \in \{1, \dots, T\}$, $\tau'' \in \{1, \dots, T\}$, com $\tau' \neq \tau''$.

R4: Existem pelo menos um período de tempo em $\{1, \dots, T\}$ e duas regiões em $\{1, \dots, R\}$ para o quais X_{irt} assume o mesmo valor. Assume-se que $x_{ir'\tau'} = x_{ir''\tau'}$, onde tanto o período τ' quanto as regiões $\{r', r''\}$ estão associados com a condição R3.

R5: Existem pelo menos um período de tempo em $\{1, \dots, T\}$ e duas regiões em $\{1, \dots, R\}$ para o quais X_{irt} assume valores distintos. Assume-se que $x_{ir'\tau''} \neq x_{ir''\tau''}$, onde tanto o período τ'' quanto as regiões $\{r', r''\}$ estão associados com a condição R3.

R6: Assume-se que:

$$m(Z_{1ir}, Z_{2irt}, X_{irt}, d_t, r) = m_1(Z_{1ir}, Z_{2irt}, X_{irt}, d_t) + m_2(r),$$

onde m_1 e m_2 são funções não-paramétricas.

Proposição 2.2 *Para o modelo especificado pela equação (2.1), o efeito econômico externo (EEE) é identificado se as condições R3-R6 forem satisfeitas.*

Demonstração. *Os dados permitem que o analista possa observar os seguintes objetos:*

$$\begin{aligned} \Delta' &= \alpha_{ir''\tau'} - \alpha_{ir'\tau'} \\ &= m(z_{1ir''}, z_{2ir''\tau'}, x_{ir''\tau'}, d_{\tau'}, r'') - m(z_{1ir'}, z_{2ir'\tau'}, x_{ir'\tau'}, d_{\tau'}, r'), \end{aligned}$$

e

$$\begin{aligned}\Delta'' &= \alpha_{ir''\tau''} - \alpha_{ir'\tau''} \\ &= m(z_{1ir''}, z_{2ir''\tau''}, x_{ir''\tau''}, d_{\tau''}, r'') - m(z_{1ir'}, z_{2ir'\tau''}, x_{ir'\tau''}, d_{\tau''}, r'),\end{aligned}$$

onde $\{r', r''\}$ e $\{\tau', \tau''\}$ são as regiões e os períodos de tempo da condição R3.

A identificação procede em dois passos. Primeiramente, pode-se escrever:

$$\begin{aligned}\Delta' &= m(z_{1ir''}, z_{2ir''\tau'}, x_{ir''\tau'}, d_{\tau'}, r'') - m(z_{1ir''}, z_{2ir''\tau'}, x_{ir'\tau'}, d_{\tau'}, r') \\ &= m(z_{1ir''}, z_{2ir''\tau'}, x_{ir''\tau'}, d_{\tau'}, r'') - m(z_{1ir''}, z_{2ir''\tau'}, x_{ir''\tau'}, d_{\tau'}, r'),\end{aligned}$$

onde a primeira igualdade segue da condição R3 e a segunda da condição R4. Da restrição R6, tem-se que:

$$\begin{aligned}\Delta' &= [m_1(z_{1ir''}, z_{2ir''\tau'}, x_{ir''\tau'}, d_{\tau'}) - m_1(z_{1ir''}, z_{2ir''\tau'}, x_{ir'\tau'}, d_{\tau'})] \\ &\quad + [m_2(r'') - m_2(r')] \\ &= m_2(r'') - m_2(r').\end{aligned}$$

Em segundo lugar, pode-se escrever que:

$$\begin{aligned}\Delta'' &= m(z_{1ir''}, z_{2ir''\tau''}, x_{ir''\tau''}, d_{\tau''}, r'') - m(z_{1ir''}, z_{2ir''\tau''}, x_{ir'\tau''}, d_{\tau''}, r') \\ &= [m_1(z_{1ir''}, z_{2ir''\tau''}, x_{ir''\tau''}, d_{\tau''}) - m_1(z_{1ir''}, z_{2ir''\tau''}, x_{ir'\tau''}, d_{\tau''})] \\ &\quad + [m_2(r'') - m_2(r')],\end{aligned}$$

onde a primeira igualdade segue da condição R3 e a segunda da condição R6. Note que o segundo termo em colchetes no lado direito dessa expressão já foi identificado no primeiro passo por Δ' . Assim, re-arranjando essa última expressão, obtém-se:

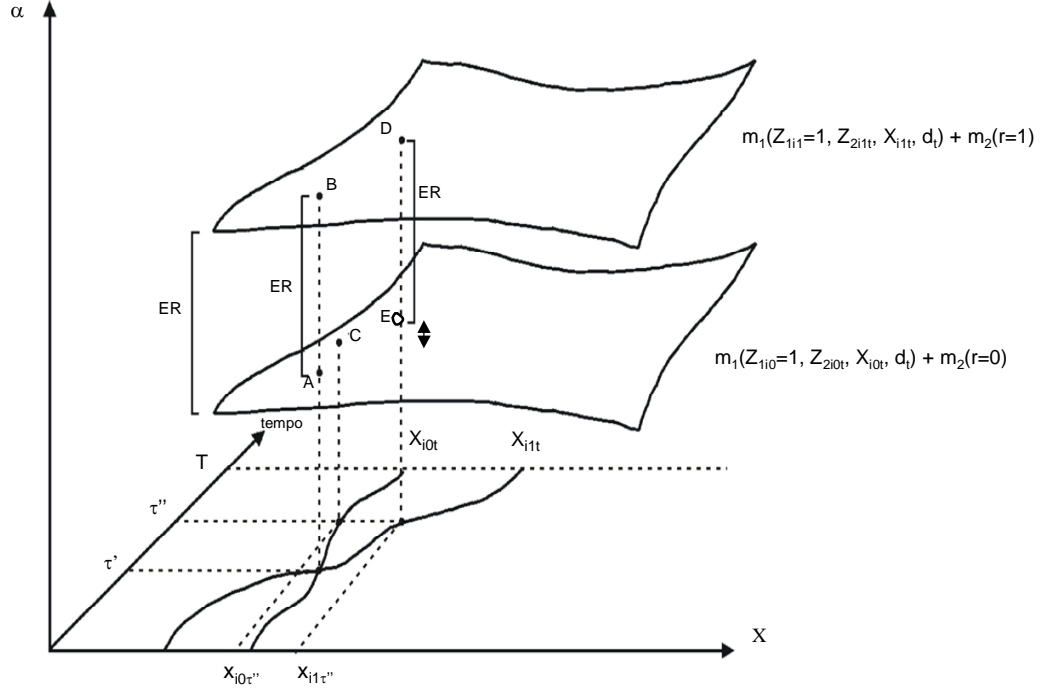
$$\begin{aligned}\Delta'' - [m_2(r'') - m_2(r')] &= \Delta'' - \Delta' \\ &= m_1(z_{1ir''}, z_{2ir''\tau''}, x_{ir''\tau''}, d_{\tau''}) - \\ &\quad m_1(z_{1ir''}, z_{2ir''\tau''}, x_{ir'\tau''}, d_{\tau''}),\end{aligned}$$

que identifica $\Delta x(z_{1ir''}, z_{2ir''\tau''}, x_{ir''\tau''}, x_{ir'\tau''}, d_{\tau''})$ quando a condição R5 é satisfeita. ■

A Figura 2.3 procura ilustrar a identificação do EEE. Como na Figura 2.2, ela contém uma função hipotética m para duas regiões ($r = 0, 1$) e um determinado tipo de indivíduos ($Z_{1ir} = 1$). Os eixos vertical e horizontais medem as mesmas quantidades que os seus correspondentes na Figura 2.2. Distintamente daquela Figura, a condição de separabilidade R6 impõe agora que as duas superfícies tenham o mesmo formato (isto é, elas são paralelas). Note que as condições econômicas nas duas regiões coincidem no período τ' , porém são

Figura 2.3: Superfícies da Função Hipotética m para Duas Regiões e o Mesmo Tipo de Indivíduos

$$\alpha_{irt} = m(Z_{1ir}=1, Z_{2irt}, X_{irt}, d_t, r), \quad r=0,1$$



diferentes em τ'' .

Assume-se implicitamente na Figura 2.3 que o impacto médio do programa é conhecido para indivíduos nas duas regiões com as mesmas características variantes e invariantes no tempo nos períodos τ' e τ'' . Isso implica que a restrição R3 é satisfeita. A condição de “cruzamento” R4 é atendida em τ' , uma vez que a evolução das condições econômicas entre as duas regiões se cruzam nesse período. Assim, dada a imposição da condição de separabilidade R6, a distância vertical entre os pontos A e B identifica o que denominaremos de *efeito regional* - *ER* (ver abaixo). Isso corresponde ao primeiro passo na demonstração da Proposição 2.2. Se tomarmos agora a distância vertical entre os pontos C e D e subtrairmos dela a distância entre D e E (que é idêntica a distância entre A e B por força da condição R6), obtemos a distância vertical entre C e E. Uma vez que as condições econômicas são diferentes entre as regiões em τ'' (i.e. a condição R5 é satisfeita nesse período), temos então que a distância vertical entre C e E identifica alguns aspectos do EEE.

Seguem algumas importantes considerações. Em primeiro lugar, cabe assinalar que não é possível variar X_{irt} sem modificar o valor de pelo menos um dos argumentos da função m . De fato, o valor da variável X_{irt} é unicamente determinado quando a região (r), o

período de tempo (t) e o tipo de indivíduos (i) são especificados. Assim, como estamos interessados em identificar o efeito de variações independentes nas condições econômicas sobre o impacto médio do programa, torna-se necessário impor restrições ao modelo. Uma dessas restrições é a condição de separabilidade R6, a qual, no entanto, não é suficiente para gerar isoladamente a identificação do EEE no presente contexto. Isso decorre do fato de que o termo aditivo $m_2(r)$ muda de valor quando se utilizam informações de outras regiões no processo de identificação. Especificamente, o uso de informações de r'' e r' faz aparecer o termo $[m_2(r'') - m_2(r')]$, o qual precisa ser identificado pela imposição de outra restrição ao modelo. Ao requerer que as condições econômicas sejam as mesmas entre as regiões para algum período de tempo, a condição R4 cumpre esse papel.

Em segundo lugar, vale a pena discutir um efeito que também pode ser de interesse. Defina

$$\Delta r(z_1, z_2, x, d, r', r'') = m(z_1, z_2, x, d, r'') - m(z_1, z_2, x, d, r')$$

como o efeito regional (ER), onde z_1 , z_2 , x e d representam valores específicos assumidos respectivamente por Z_{1ir} , Z_{2irt} , X_{irt} e d_t , e $r' \neq r''$. Pode-se provar facilmente que o ER é identificado se as condições R3 e R4 são satisfeitas.⁹ Observe-se que, como a condição R4 não pode ser atendida quando condição R5 é satisfeita, não é possível identificar o EEE quando o ER é identificado. De fato, enquanto a restrição R5 requer que as condições econômicas sejam diferentes entre as regiões utilizadas na análise, a restrição R4 demanda que essas condições sejam as mesmas. É importante notar que isso cria um problema cuja principal consequência é que EEE não é identificado para todos os pontos que se deseje.

As duas considerações feitas acima mostram que somente alguns aspectos do EEE podem ser identificados sob as restrições R3-R6. Claramente, a condição de separabilidade R6 impede a identificação totalmente não-paramétrica do EEE. Como resultado, ela previne que o efeito de interesse seja identificado incluindo-se interações potencialmente importantes entre o componente regional r e os demais argumentos da função m .¹⁰ Ademais, uma vez que as condições R4 e R5 não podem ser satisfeitas simultaneamente, a identificação do EEE não ocorre para aqueles períodos em que a condição R4 é atendida.

Se o verdadeiro modelo não apresenta aditividade separável, então não se estará identificando o EEE, mas sim uma mistura de efeitos. Para ver isso, vale relembrar que a identificação do EEE é alcançada por meio da subtração de dois objetos observados: $\Delta'' - \Delta'$.

⁹Omitindo a condição R6 no primeiro passo da demonstração da Proposição 2.2, é possível ver que esse efeito é identificado de forma não-paramétrica. Claramente, se a condição R6 é imposta ao modelo, esse efeito se reduz a $[m_2(r'') - m_2(r')]$.

¹⁰Deve-se assinalar que outro argumento poderia ter sido escolhido para ser separado de m . Por exemplo, poder-se-ia postular a condição R6 em termos de Z_{1ir} , isto é: $m(Z_{1ir}, Z_{2irt}, X_{irt}, d_t, r) = m_1(Z_{2irt}, X_{irt}, d_t, r) + m_2(Z_{1ir})$. Nesse caso, pequenas modificações nas restrições R3, R4 e R5 permitiriam a identificação de alguns aspectos do EEE.

Se a aditividade separável não for uma característica do verdadeiro modelo, tem-se que:¹¹

$$\begin{aligned}\Delta'' - \Delta' &= \{m(z_{1ir''}, z_{2ir''\tau''}, x_{ir''\tau''}, d_{\tau''}, r'') - m(z_{1ir''}, z_{2ir''\tau'}, x_{ir'\tau'}, d_{\tau'}, r')\} \\ &\quad - \{m(z_{1ir''}, z_{2ir''\tau'}, x_{ir''\tau'}, d_{\tau'}, r') - m(z_{1ir''}, z_{2ir''\tau''}, x_{ir''\tau''}, d_{\tau''}, r'')\},\end{aligned}$$

que é uma expressão que captura uma combinação do EEE e do ER no período τ'' (o primeiro termo entre chaves no lado direito da expressão), menos o ER no período τ' (o segundo termo entre chaves).

Se a dimensão de X_{irt} , K , for maior do que a unidade, a identificação do efeito econômico externo da k -ésima variável de X_{irt} requer uma restrição adicional: cada uma das demais variáveis em X_{irt} deve apresentar o mesmo valor entre as regiões r' e r'' em $t = \tau''$. Especificamente, definindo

$$X_{-k,irt} \equiv (X_{1irt}, \dots, X_{k-1,irt}, X_{k+1,irt}, \dots, X_{Kirt}),$$

essa restrição adicional pode ser escrita como: $x_{-k,ir'\tau''} = x_{-k,ir''\tau''}$. Além dessa nova restrição, a condição R4 precisa ser modificada, acrescentando-se a exigência de que cada variável em $X_{-k,irt}$ apresente o mesmo valor entre r' e r'' em $t = \tau'$. Note-se que essas novas restrições não são necessariamente triviais. De fato, pode ser difícil encontrar situações empíricas em que cada variável econômica do modelo apresenta o mesmo valor entre as regiões num determinado período de tempo, porém somente uma delas assume valores diferentes entre as mesmas regiões num outro período.

2.3 Múltiplas Coortes e Região Única

Considere agora um contexto no qual o analista conhece o impacto médio do programa para pelo menos duas coortes de participantes e somente uma região do país (ou o país inteiro). Suponha que essa informação está disponível para diferentes tipos de indivíduos. Denotando o impacto médio do programa para os indivíduos do tipo $i \in \{1, \dots, I\}$ no período $t \in \{1, \dots, T\}$ e que pertencem a coorte $c \in \{1, \dots, C\}$ por α_{ict} , pode-se escrever:

$$\alpha_{ict} = h(Z_{1ic}, Z_{2ict}, X_{it}, d_{ct}, c), \quad (2.2)$$

onde h é uma função não-paramétrica; Z_{1ic} e Z_{2ict} representam respectivamente vetores de características invariantes e variantes dos indivíduos do tipo i no período t da coorte c ; X_{it} é um vetor de dimensão K de variáveis econômicas externas que afetam α_{ict} ; e, d_{ct} denota o número de períodos após o final (ou início) do programa pela coorte de participantes

¹¹ Assume-se na expressão abaixo que as condições R3 e R4 são atendidas.

c .¹² Note que X_{it} não possui o sub-escrito c , o que implica que as condições econômicas prevalentes em qualquer período de tempo t são comuns a todas as coortes existentes nesse período. Como antes, assumiremos que $K = 1$. Após a derivação do resultados da identificação dos efeitos de interesse, analisaremos o caso em que $K > 1$.

As definições do EIT e do EEE para o presente contexto são as seguintes:

Definição 2.3 *O efeito interno temporal (EIT) é definido pela diferença parcial:*

$$\Delta d(z_1, z_2, x, d'', d', c^\diamond) = h(z_1, z_2, x, d'', c^\diamond) - h(z_1, z_2, x, d', c^\diamond),$$

onde z_1, z_2, x, c^\diamond e $d'' > d'$ são valores específicos assumidos respectivamente por $Z_{1ct}, Z_{2ict}, X_{it}, c$ e d_{ct} .

Definição 2.4 *O efeito econômico externo (EEE) é definido pela diferença parcial:*

$$\Delta x(z_1, z_2, x'', x', d^\diamond, c^\diamond) = h(z_1, z_2, x'', d^\diamond, c^\diamond) - h(z_1, z_2, x', d^\diamond, c^\diamond),$$

onde z_1, z_2, d^\diamond e c^\diamond são valores específicos assumidos respectivamente por $Z_{1ic}, Z_{2ict}, d_{ct}$ e c , e x' e x'' são valores distintos assumidos por X_{it} .

2.3.1 Identificação do EIT: Múltiplas Coortes

Sejam d' e d'' dois valores de d_{ct} para os quais se deseja identificar o EIT. Utilizaremos as seguintes condições de identificação.^{13,14}

A1: Para cada grupo $i \in \{1, \dots, I\}$ e coorte $c \in \{1, \dots, C\}$ para os quais o analista está interessado em identificar o EIT, existem valores comuns de Z_{2ict} quando t muda de valor. Assume-se que $z_{2ict'} = z_{2ict''}$, onde $i \in \{1, \dots, I\}$, $c \in \{1, \dots, C\}$, $t' \in \{1, \dots, T-1\}$ e $t' < t''$ são o grupo, a coorte e os períodos de tempo de interesse, respectivamente.

A2: Para cada grupo $i \in \{1, \dots, I\}$ e coorte $c \in \{1, \dots, C\}$ da condição A1, existem valores iguais de X_{it} quando t muda de valor. Assume-se que $x_{it'} = x_{it''}$, onde $i \in \{1, \dots, I\}$, $c \in \{1, \dots, C\}$, $t' \in \{1, \dots, T-1\}$ e $t' < t''$ são os mesmos grupo, coorte e períodos de tempo da condição A1, respectivamente.

Proposição 2.3 *Para o modelo especificado pela equação (2.2), o efeito interno temporal (EIT) é identificado se as condições A1 e A2 forem satisfeitas.*

¹²O sub-escrito c foi acrescentado a d para indicar que, para um período qualquer do tempo-calendário $t \in \{1, \dots, T\}$, o tempo transcorrido desde o final (ou início) do programa por duas coortes distintas é diferente.

¹³Mais uma vez, empregaremos uma notação em que o valor de uma variável é especificado pelos seus sub-escritos.

¹⁴No que se segue, quando nos referirmos a distintos períodos do tempo-calendário, estaremos admitindo que é possível observar a coorte de participantes $c \in \{1, \dots, C\}$.

Demonstração. A demonstração é trivial e análoga a da Proposição 2.1. Dos dados, o analista pode observar:

$$\begin{aligned}\Delta &= \alpha_{ict''} - \alpha_{ict'} \\ &= h(z_{1ic}, z_{2ict''}, x_{it''}, d_{ct''}, c) - h(z_{1ic}, z_{2ict'}, x_{it'}, d_{ct'}, c),\end{aligned}$$

onde $i \in \{1, \dots, I\}$, $c \in \{1, \dots, C\}$, $t' \in \{1, \dots, T-1\}$ e $t' < t''$ representam o grupo, a coorte e os períodos de tempo de interesse, respectivamente.

Assim, das condições A1 e A2, tem-se que:

$$\begin{aligned}\Delta &= h(z_{1ic}, z_{2ict'}, x_{it''}, d_{ct''}, c) - h(z_{1ic}, z_{2ict''}, x_{it'}, d_{ct'}, c) \\ &= h(z_{1ic}, z_{2ict'}, x_{it'}, d_{ct''}, c) - h(z_{1ic}, z_{2ict''}, x_{it''}, d_{ct'}, c) \\ &= \Delta d(z_{1ic}, z_{2ict'}, x_{it'}, d_{ct''}, d_{ct'}, c),\end{aligned}$$

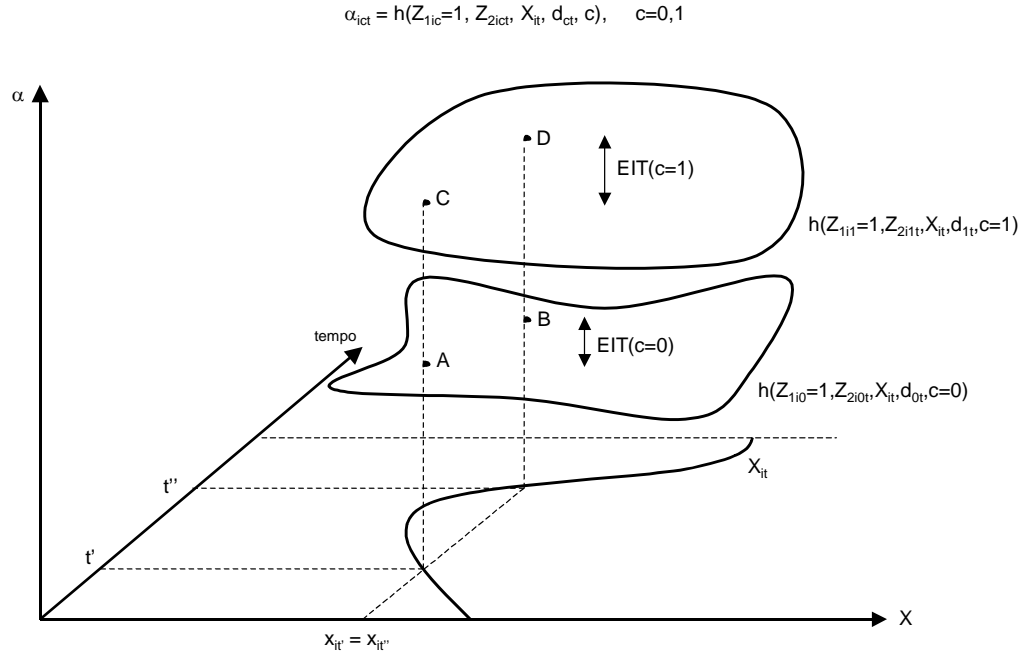
onde a primeira igualdade segue de A1 e a segunda de A2. ■

Analogamente ao contexto de múltiplas regiões, a restrição A1 demanda que o impacto médio do programa seja conhecido para indivíduos com as mesmas características variantes no tempo (*e.g.* idade) entre os períodos t' e t'' para cada coorte de participantes de interesse. Também de forma similar ao contexto de múltiplas regiões, a condição A2 requer que a evolução das condições econômicas seja não-monotônica, o que permite, por exemplo, que X_{it} exiba o mesmo valor em t' e t'' . Entretanto, enquanto no contexto de múltiplas regiões as condições econômicas são específicas da região considerada, no presente contexto elas são comuns para todas as coortes de participantes. Como resultado, o requisito de que $x_{it'} = x_{it''}$ em A2 torna possível a identificação do EIT para todas as coortes existentes nos períodos t' e t'' . Claramente, como as coortes finalizaram (ou iniciaram) o programa em momentos distintos do calendário, os valores de d_{ct} serão diferentes para cada coorte em t' e t'' . Mais uma vez, cabe observar que, quanto maior K , menor a probabilidade de que a condição A2 seja atendida na prática.

A Figura 2.4 ilustra a identificação do EIT para duas coortes de participantes ($c = 0, 1$) e um dado tipo de indivíduos ($Z_{1ic} = 1$). Similarmente à Figura 2.2, não impusemos nenhuma restrição sobre o forma da função h , implicando que as duas superfícies não necessariamente têm o mesmo formato. Está implicitamente assumido na Figura 2.4 que se conhece o impacto médio do programa para indivíduos de cada coorte com as mesmas características variantes no tempo nos períodos t' e t'' . Assume-se, portanto, que a condição A1 é satisfeita. A evolução das condições econômicas está representada no plano horizontal $\text{tempo} \times X$. Note que, em t' e t'' , as condições econômicas são as mesmas, o que faz com que a condição A2 seja atendida entre esses períodos. Assim, a distâncias verticais entre os pontos A e B e entre C e D correspondem ao EIT para as coortes $c = 0$ e $c = 1$, respectivamente. Como mencionado anteriormente, deve-se notar que o EIT não está sendo

identificado para o mesmo par de valores de d_{ct} para cada coorte, já que, por construção, elas completaram (começaram) o programa em momentos distintos do calendário.

Figura 2.4: Superfícies da Função Hipotética h para Duas Coortes e o Mesmo Tipo de Indivíduos



As condições A1 e A2 são suficientes para que o EIT seja identificado de maneira *totalmente* não-paramétrica. Esse resultado é semelhante ao encontrado anteriormente no contexto de múltiplas regiões e somente uma coorte de participantes. A principal diferença é que, enquanto no presente contexto requer-se apenas que as condições econômicas sejam as mesmas entre períodos de tempo, na situação anterior exigia-se que as condições econômicas fossem as mesmas entre períodos de tempo para *cada* região disponível nos dados.

2.3.2 Identificação do EEE: Múltiplas Coortes

Em relação à identificação do EEE, sejam τ' , τ'' , τ''' e τ'''' quatro períodos de tempo distintos para os quais se assume que: $d_{c'\tau'} = d_{c''\tau''}$ e $d_{c'\tau'''} = d_{c''\tau''''}$, onde c' e c'' representam duas coortes de participantes pertencentes a $\{1, \dots, C\}$. Utilizamos o seguinte conjunto de restrições.

A3: Seja $Z_{ict} = (Z_{1ic}, Z_{2ict})$. Denote por Ω_c o domínio de Z_{ict} para a coorte $c \in \{1, \dots, C\}$. Defina S como a interseção de $\Omega_{c'}$ e $\Omega_{c''}$, $c' \in \{1, \dots, C\}$, $c'' \in \{1, \dots, C\}$, com $c' \neq c''$. Assume-se que S é não-vazio em $\tau' \in \{1, \dots, T\}$ e $\tau'' \in \{1, \dots, T\}$, assim como em $\tau''' \in \{1, \dots, T\}$ e $\tau'''' \in \{1, \dots, T\}$, com $\tau' < \tau'' < \tau''' < \tau''''$.

A4: Existem pelo menos dois períodos de tempo em $\{1, \dots, T\}$ para os quais X_{it} assume o mesmo valor. Assume-se que $x_{i\tau'} = x_{i\tau''}$, onde os períodos $\{\tau', \tau''\}$ estão associados com a condição A3.

A5: Existem pelo menos dois períodos de tempo em $\{1, \dots, T\}$ para os quais X_{it} assume valores distintos. Assume-se que $x_{i\tau'''} \neq x_{i\tau''''}$, onde os períodos $\{\tau''', \tau''''\}$ estão associados com a condição A3.

A6: Assume-se que

$$h(Z_{1ic}, Z_{2ict}, X_{it}, d_{ct}, c) = h_1(Z_{1ic}, Z_{2ict}, X_{it}, d_{ct}) + h_2(c),$$

onde h_1 e h_2 são funções não-paramétricas.

Proposição 2.4 *Para o modelo especificado pela equação (2.2), o efeito econômico externo (EEE) é identificado se as condições A3-A6 forem satisfeitas.*

Demonstração. *Dos dados, o analista pode observar os seguintes objetos:*

$$\begin{aligned} \Delta' &= \alpha_{ic'\tau''} - \alpha_{ic'\tau'} \\ &= h(z_{1ic'}, z_{2ic'\tau''}, x_{i\tau''}, d_{c'\tau''}, c'') - h(z_{1ic'}, z_{2ic'\tau'}, x_{i\tau'}, d_{c'\tau'}, c'), \end{aligned}$$

e

$$\begin{aligned} \Delta'' &= \alpha_{ic''\tau''''} - \alpha_{ic''\tau'''} \\ &= h(z_{1ic'}, z_{2ic''\tau''''}, x_{i\tau''''}, d_{c''\tau''''}, c'') - h(z_{1ic'}, z_{2ic''\tau'''}, x_{i\tau'''}, d_{c''\tau'''}, c'), \end{aligned}$$

onde $\{c', c''\}$ e $\{\tau', \tau'', \tau''', \tau''''\}$ são as coortes e períodos de tempo da condição A3.

Como na demonstração da Proposição 2.2, procederemos em dois passos. Primeiramente, pode-se escrever:

$$\begin{aligned} \Delta' &= h(z_{1ic'}, z_{2ic'\tau''}, x_{i\tau''}, d_{c'\tau''}, c'') - h(z_{1ic'}, z_{2ic'\tau'}, x_{i\tau'}, d_{c'\tau'}, c') \\ &= h(z_{1ic'}, z_{2ic'\tau''}, x_{i\tau''}, d_{c''\tau''}, c'') - h(z_{1ic'}, z_{2ic'\tau'}, x_{i\tau'}, d_{c'\tau'}, c') \\ &= h(z_{1ic'}, z_{2ic'\tau''}, x_{i\tau''}, d_{c''\tau''}, c'') - h(z_{1ic'}, z_{2ic'\tau''}, x_{i\tau''}, d_{c''\tau''}, c'), \end{aligned}$$

onde a primeira e a segunda igualdades seguem respectivamente das condições A3 and A4, e a terceira da suposição de que, em τ' e τ'' , $d_{c'\tau'} = d_{c''\tau''}$. Utilizando agora a restrição

A6, essa última expressão pode ser re-escrita como:

$$\begin{aligned}\Delta' &= [h_1(z_{1ic''}, z_{2ic''\tau''}, x_{i\tau''}, d_{c''\tau''}) - h_1(z_{1ic''}, z_{2ic''\tau''}, x_{i\tau''}, d_{c''\tau''})] \\ &\quad + [h_2(c'') - h_2(c')] \\ &= h_2(c'') - h_2(c').\end{aligned}$$

Em segundo lugar, pode-se escrever que

$$\begin{aligned}\Delta'' &= h(z_{1ic''}, z_{2ic''\tau''''}, x_{i\tau''''}, d_{c''\tau''''}, c'') - h(z_{1ic''}, z_{2ic''\tau''''}, x_{i\tau''''}, d_{c''\tau''''}, c') \\ &= [h_1(z_{1ic''}, z_{2ic''\tau''''}, x_{i\tau''''}, d_{c''\tau''''}) - h_1(z_{1ic''}, z_{2ic''\tau''''}, x_{i\tau''''}, d_{c''\tau''''})] \\ &\quad + [h_2(c'') - h_2(c')],\end{aligned}$$

onde a primeira igualdade segue da condição A3 e da suposição de que, em τ''' e τ'''' , $d_{c'\tau'''} = d_{c''\tau''''}$, e a última igualdade segue de A6.

Uma vez que o segundo termo em colchetes no lado direito dessa última expressão já foi identificado no primeiro passo por Δ' , pode-se re-arranjá-la para obter:

$$\begin{aligned}\Delta'' - [h_2(c'') - h_2(c')] &= \Delta'' - \Delta' \\ &= h_1(z_{1ic''}, z_{2ic''\tau''''}, x_{i\tau''''}, d_{c''\tau''''}) - \\ &\quad h_1(z_{1ic''}, z_{2ic''\tau''''}, x_{i\tau''''}, d_{c''\tau''''}),\end{aligned}$$

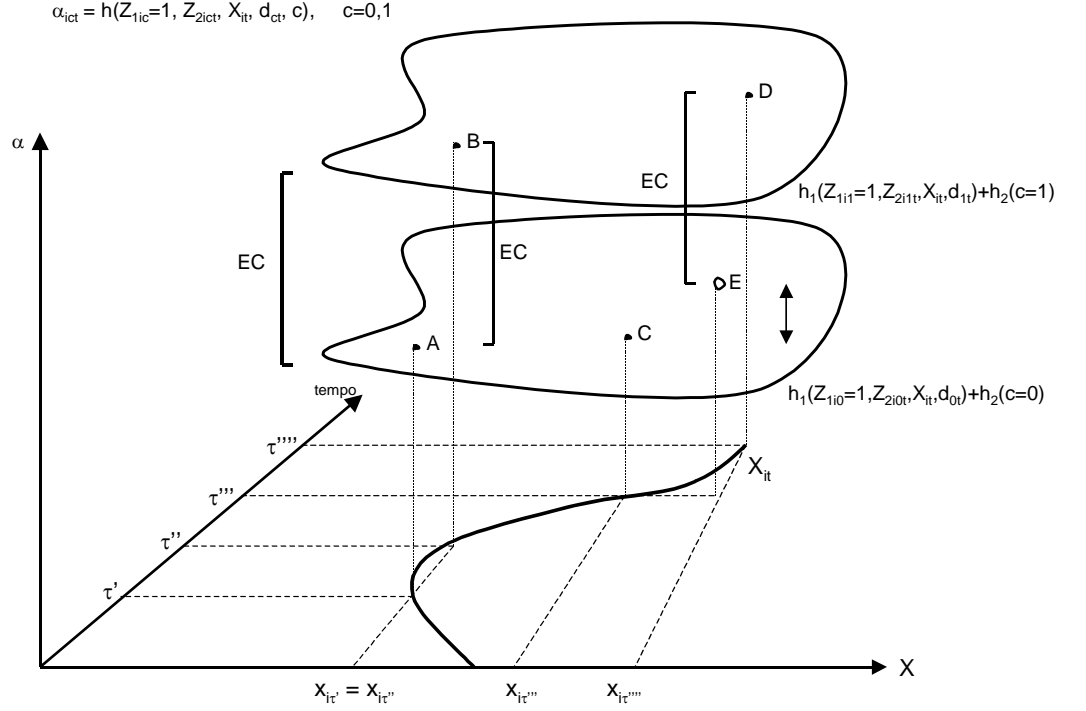
que identifica $\Delta x(z_{1ic''}, z_{2ic''\tau''''}, x_{i\tau''''}, x_{i\tau''''}, d_{c''\tau''''})$ quando a condição A5 é satisfeita. ■

A Figura 2.5 permite visualizar a identificação do EEE no presente contexto. Há duas suposições implícitas nessa Figura. A primeira é que o valor de d_{ct} é o mesmo em τ' e τ'' (digamos d^+), e em τ''' e τ'''' (digamos d^{++}) para as coortes $c = 0$ e $c = 1$. A segunda é que o impacto médio do programa é conhecido para indivíduos com as mesmas características invariantes (Z_{1ic}) e variantes (Z_{2ict}) no tempo nas duas coortes em τ' e τ'' , assim como em τ''' e τ'''' . Isso garante que a condição A3 é atendida nesses períodos de tempo.

Por força da condição de separabilidade A6, as duas superfícies desenhadas na Figura 2.5 possuem o mesmo formato (i.e. elas são paralelas), o que implica que a distância entre elas é constante. Dado que as condições econômicas são as mesmas em τ' e τ'' (i.e. a condição A4 é atendida nesses períodos), a distância vertical entre os pontos A e B mede a distância fixa entre as duas superfícies (i.e. mede a diferença $[h_2(c = 1) - h_2(c = 0)]$). Isso corresponde ao primeiro passo da demonstração da Proposição 2.4. Tomando agora a distância vertical entre os pontos C e D e subtraindo desta a distância entre D e E (que é igual a distância entre A e B devido a A6), obtém-se a distância vertical entre os pontos C e E, a qual identifica alguns aspectos do EEE. Note que a restrição A5 é satisfeita em τ''' e τ'''' , uma vez que as condições econômicas são diferentes nesses dois períodos.

Considerações semelhantes àsquelas do contexto de múltiplas regiões podem ser feitas.

Figura 2.5: Superfícies da Função Hipotética h para Duas Coortes e o Mesmo Tipo de Indivíduos



Primeiramente, X_{it} não pode variar de forma totalmente independente do valor assumido por alguns argumentos da função h : se compararmos as condições econômicas para diferentes tipos de indivíduos (períodos de tempo), o valor do argumento Z_{1ic} (d_{ct}) também muda. Assim, e como no contexto anterior, invocamos uma condição de separabilidade (A6), a qual, no entanto, não é suficiente para produzir isoladamente a identificação do EEE. De fato, essa condição gera o termo $[h_2(c'') - h_2(c')]$, o qual emerge da diferença entre os impactos médios do programa para essas duas coortes entre τ' e τ'' . A condição A4 desempenha o papel da condição anterior R4, no sentido de que ela permite a identificação desse termo.

A segunda consideração tem a ver com o efeito coorte (EC), o qual pode ser definido por:

$$\Delta c(z_1, z_2, x, d, c', c'') = h(z_1, z_2, x, d, c'') - h(z_1, z_2, x, d, c'),$$

onde z_1 , z_2 , x e d representam valores específicos assumidos respectivamente por Z_{1ic} , Z_{2ic} , X_{it} e d_{ct} , e $c' \neq c''$. O EC pode ser identificado de forma totalmente não-paramétrica se as condições A3 e A4 forem satisfeitas.¹⁵ Cabe assinalar que, como as condições A4 e A5

¹⁵Esse resultado pode ser provado sem o uso da restrição A6 no primeiro passo da demonstração

não podem ser satisfeitas ao mesmo tempo, não se pode identificar o EEE quando o EC é identificado. Esse é o mesmo tipo de problema mencionado anteriormente no contexto de múltiplas regiões.

Essas considerações mostram que somente alguns aspectos do EEE são identificados no presente contexto. Isso se deve ao fato de que a restrição de separabilidade aditiva A6 impede a existência de possíveis interações interessantes entre c e os demais argumentos da função h .¹⁶ Além disso, a identificação do EEE só pode ocorrer para aqueles pares de períodos nos quais a condição A4 não é satisfeita.

Se a separabilidade da condição A6 não for uma característica do verdadeiro modelo, então o objeto $\Delta'' - \Delta'$ estará identificando uma combinação do EEE com efeitos de coorte. De fato, nesse caso, teríamos que:¹⁷

$$\begin{aligned} \Delta'' - \Delta' = & \{h(z_{1ic''}, z_{2ic''\tau''''}, x_{i\tau''''}, d_{c''\tau''''}, c'') - h(z_{1ic''}, z_{2ic''\tau''''}, x_{i\tau''}, d_{c''\tau''''}, c')\} \\ & - \{h(z_{1ic''}, z_{2ic''\tau''}, x_{i\tau''}, d_{c''\tau''}, c'') - h(z_{1ic''}, z_{2ic''\tau''}, x_{i\tau''}, d_{c''\tau''}, c')\}, \end{aligned}$$

que é uma expressão que captura uma combinação do EEE e do EC entre τ''' e τ'''' (o primeiro termo entre chaves no lado direito da expressão), menos o EC entre τ' e τ'' (o segundo termo entre chaves).

Se $K > 1$, há que se introduzir uma condição adicional que requeira que cada variável em $X_{-k,it} = (X_{1it}, \dots, X_{k-1,it}, X_{k+1,it}, \dots, X_{Kit})$ assumam o mesmo valor em τ''' e τ'''' , isto é: $x_{-k,i\tau'''} = x_{-k,i\tau''''}$. Ademais, a restrição A4 precisa ser modificada de forma a que cada variável em X_{it} possua um valor idêntico em τ' e τ'' .

Ao requerer que as condições econômicas sejam as mesmas entre dois períodos de tempo, a restrição A4 pode ser vista como um tipo de condição de não-monotonicidade assemelhada à restrição A2. Observe-se, entretanto, que essa semelhança não ocorre no contexto de múltiplas regiões, já que as restrições R4 e R2 têm naturezas distintas naquele contexto. De fato, enquanto a restrição R2 requer não-monotonicidade das condições econômicas para *cada* região presente nos dados, a restrição R4 demanda que as condições econômicas sejam as mesmas *entre* as regiões.

da Proposição 2.4. Quando essa restrição é imposta, o EC se reduz a $[h_2(c'') - h_2(c')]$, que corresponde a distância (fixa) entre as duas superfícies apresentadas na Figura 2.5.

¹⁶Se o componente de coorte captura potenciais diferenças da qualidade de um programa entre as coortes, então a restrição A6 previne o cálculo do EEE para distintos níveis de qualidade desse programa.

¹⁷Assume-se que as condições A3 e A4 são atendidas.

2.4 Exercício Empírico

2.4.1 Descrição do Programa

Para implementar o exercício empírico, fazemos uso de dados experimentais que foram coletados para avaliar a efetividade de um programa público de qualificação de mão-de-obra no Brasil (Plano Nacional de Qualificação do Trabalhador - PLANFOR). Iniciado em 1996, esse programa oferece treinamento profissional em sala de aula, consistindo na provisão de habilidades básicas necessárias para certas ocupações (*e.g.* garçons, cabeleireiros e empregos administrativos). O programa opera continuamente durante o ano, com uma nova coorte de participantes iniciando-o em cada mês. A provisão dos serviços do programa é descentralizada, sendo o seu conteúdo baseado, em larga medida, nas escolhas feitas pelas autoridades estaduais e conselhos locais. O governo federal provê uma parte do financiamento do programa aos estados do país, os quais sub-contratam entidades da sociedade civil (*e.g.* escolas comunitárias, igrejas e ONGs) para entregar os serviços de treinamento propriamente ditos. A população-alvo consiste basicamente de indivíduos considerados em situação vulnerável, abrangendo principalmente os desempregados (incluindo os que estão procurando o primeiro emprego) e o trabalhadores com níveis baixos de educação e/ou renda (Ministério do Trabalho 1998). A inscrição no programa é voluntária, sendo a aceitação do inscrito geralmente concedida com base nos critérios de elegibilidade. A escala do programa em 1998 é relativamente pequena, representando menos de 1,5% da força de trabalho para o conjunto das regiões metropolitanas do país.¹⁸

Em 1998/99, o Ministério do Trabalho brasileiro financiou uma avaliação dos impactos do PLANFOR sobre os rendimentos do trabalho e empregabilidade dos seus participantes. Essa avaliação foi baseada em dados coletados para amostras experimentais em duas regiões metropolitanas do país, a saber: Rio de Janeiro e Fortaleza.¹⁹ O processo de aleatorização dos indivíduos para “dentro” e para “fora” do programa ocorreu em agosto de 1998, sendo que quase todos os selecionados para participar do programa começaram as atividades de treinamento em setembro do mesmo ano. Nesse mês, os participantes e não-participantes foram entrevistados em ambas as regiões por meio da aplicação de um mesmo questionário, que incluiu questões retrospectivas sobre a história dos respondentes no mercado de trabalho nos doze meses anteriores ao começo das atividades do programa. Um *follow-up* ocorreu em novembro de 1999, com questões retrospectivas remontando ao mês de setembro de 1998. Isso fornece, portanto, quinze pontos no tempo para os quais se pode calcular os impactos médios do programa em cada região utilizada na análise.

¹⁸Esse ponto é importante, uma vez que estamos ignorando efeitos de equilíbrio geral em nossa análise.

¹⁹Não existem informações específicas sobre o número de participantes do PLANFOR por região metropolitana do país. No entanto, os gestores do programa informaram em conversa pessoal com esse autor que o valor de 1,5% pode ser considerado uma boa estimativa para a escala do programa nas regiões do Rio de Janeiro e de Fortaleza.

2.4.2 Base de Dados

A amostra inicial contém 4622 indivíduos, dos quais 2256 (49%) são da região metropolitana do Rio de Janeiro. Uma vez que 97% dos indivíduos da amostra dessa região têm entre 14 e 26 anos de idade e somente 57% dos indivíduos da amostra de Fortaleza estão nessa faixa etária, restringimos a amostra final aos 3511 indivíduos das duas regiões que se encontram dentro desse intervalo de idade. Note que essa restrição procura atender, pelo menos em parte, as exigências demandadas pelas condições R1 e R3 da seção 2.2. Também como parte dessas exigências, o exemplo empírico será apresentado separadamente por sexo. A Tabela 2.1 mostra o tamanho da amostra final desagregada por região, sexo e status de tratamento.

Tabela 2.1: Tamanho da Amostra por Região, Sexo e Status de Tratamento

	Tratamento	Controle	Total
Rio de Janeiro	1139	1031	2170
Homens	381	318	699
Mulheres	758	713	1471
Fortaleza	612	729	1341
Homens	249	321	570
Mulheres	363	408	771

A Tabela 2.2 apresenta estatísticas descritivas para algumas características dos grupos de tratamento e controle em cada região utilizada no exercício. Em linhas gerais, essa Tabela revela que o procedimento de aleatorização dos dois grupos experimentais foi relativamente bem implementado. De fato, quase nenhuma diferença de características médias entre o grupo de tratamento e controle em cada região é estatisticamente significativa. A única exceção é a diferença da proporção de brancos entre os grupos na região do Rio de Janeiro.²⁰

2.4.3 Metodologia

No exercício empírico, utilizamos apenas o impacto médio do programa sobre o rendimento horário do trabalho.²¹ Essa variável foi calculada como a razão entre o rendimento mensal

²⁰O processo de aleatorização foi operacionalizado para cada turma de alunos do programa, e não para o conjunto global destes em cada região. Assim, a fim de minimizar esse desbalanceamento por cor no Rio de Janeiro, tentamos re-ponderar a amostra criando pesos específicos por turma de alunos. Contudo, como essa re-ponderação não alterou significativamente os resultados, optamos por utilizar a amostra sem o uso desses pesos.

²¹Cabe observar que os indivíduos com rendimento do trabalho nulo também foram incluídos no cômputo desses impactos médios. Nesse sentido, parte do impacto do programa sobre o emprego está incorporado em nossa análise.

Tabela 2.2: Estatísticas Descritivas da Amostra por Região e Status de Tratamento

	Rio de Janeiro		Fortaleza	
	Tratamento	Controle	Tratamento	Controle
1. Proporção de homens	0.33 (0.01)	0.31 (0.01)	0.41 (0.02)	0.44 (0.02)
2. Proporção de brancos	0.37 (0.01)	0.42 (0.02)	0.35 (0.02)	0.31 (0.02)
3. Idade (anos)	17.6 (0.1)	17.3 (0.1)	20.1 (0.1)	20.5 (0.1)
4. Proporção de chefes de família	0.01 (0.00)	0.01 (0.00)	0.06 (0.01)	0.07 (0.01)
5. Proporção de solteiros	0.96 (0.01)	0.97 (0.01)	0.85 (0.01)	0.85 (0.01)
6. Escolaridade (anos)	8.5 (0.1)	8.3 (0.1)	9.1 (0.1)	9.1 (0.1)
7. Escolaridade do pai (anos)	3.7 (0.1)	3.9 (0.1)	3.1 (0.1)	3.4 (0.1)
8. Escolaridade da mãe (anos)	4.3 (0.1)	4.6 (0.1)	3.8 (0.1)	4.0 (0.1)
9. Número de filhos	0.1 (0.0)	0.1 (0.0)	0.2 (0.0)	0.2 (0.0)
10. Idade começou a trabalhar (anos)	15.7 (0.2)	15.4 (0.2)	16.3 (0.2)	16.6 (0.2)
11. Salário horário				
11a. Junho/1998	0.13 (0.01)	0.12 (0.01)	0.29 (0.04)	0.26 (0.02)
11b. Julho/1998	0.12 (0.01)	0.14 (0.01)	0.27 (0.03)	0.25 (0.02)
11c. Agosto/1998	0.12 (0.01)	0.18 (0.02)	0.26 (0.03)	0.24 (0.02)
12. Taxa de ocupação				
12a. Junho/1998	0.14 (0.01)	0.14 (0.01)	0.25 (0.02)	0.26 (0.02)
12b. Julho/1998	0.14 (0.01)	0.15 (0.01)	0.24 (0.02)	0.25 (0.02)
12c. Agosto/1998	0.14 (0.01)	0.17 (0.01)	0.22 (0.02)	0.24 (0.02)

Notas:.

(1) Erros-padrões entre parênteses.

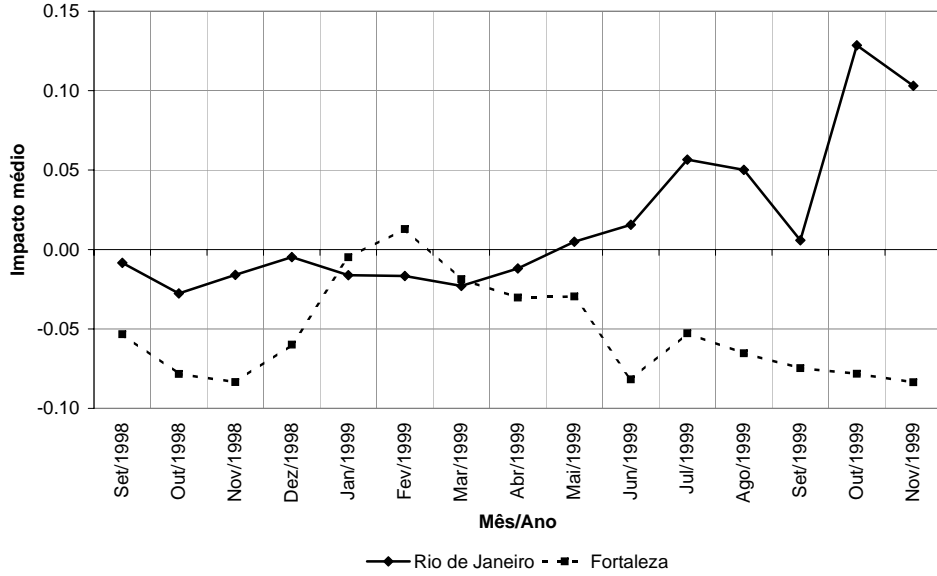
(2) O salário horário representa a média do salário horário dos indivíduos nos mês correspondente dividida pela média do salário horário do período entre setembro de 1998 e novembro de 1999.

do trabalho e o número de horas trabalhadas no mês correspondente (horas semanais vezes 4,3). A fim expressar o impacto do programa em termos relativos, o rendimento horário em cada mês foi dividido pelo rendimento horário médio entre o período de setembro de 1998 e novembro de 1999.

As Figuras 2.6 e 2.7 apresentam a evolução temporal do impacto médio do programa sobre os rendimentos do trabalho respectivamente para homens e mulheres em cada região. A Figura 2.6 mostra que trajetória do impacto foi bastante diferente entre as duas regiões: enquanto o impacto em Fortaleza foi negativo e oscilante ao longo de quase todo o período de análise, o impacto no Rio de Janeiro foi negativo e constante até o oitavo mês após o início do programa, passando a ser positivo desse momento em diante. A Figura 2.7 mostra que o impacto médio do programa também foi bastante diferente entre as duas regiões para o caso das mulheres. De fato, o impacto foi negativo e decrescente em Fortaleza durante

todo o período de análise, ao passo que no Rio de Janeiro esse impacto foi positivo e suavemente crescente ao longo desse intervalo.

Figura 2.6: Impacto Médio sobre o Rendimento do Trabalho por Região Metropolitana: Homens



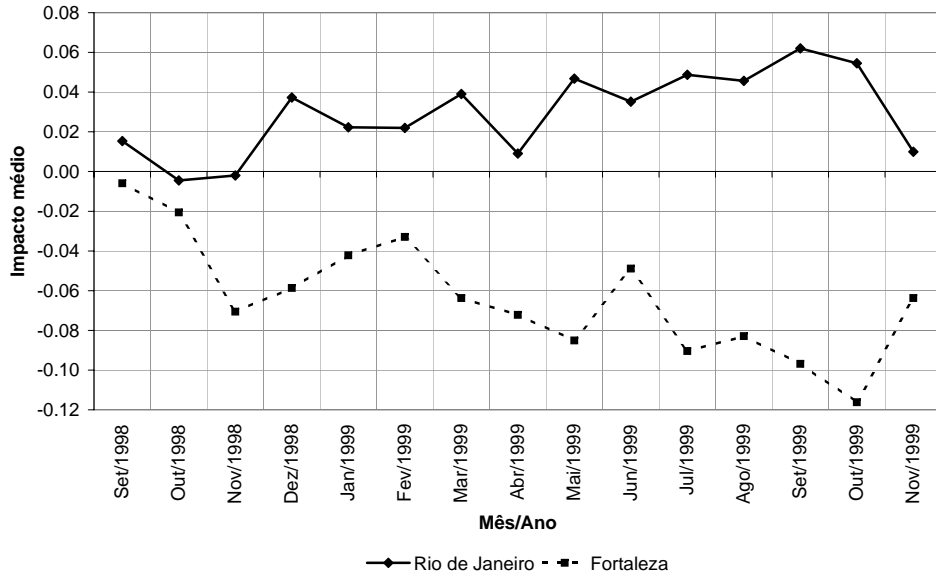
Desconsiderando os erros amostrais, as estimativas dos impactos médios apresentadas nas Figuras 2.6 e 2.7 deveriam incorporar as influências dos argumentos da função m apresentados na seção 2.2. Por exemplo, tomando o caso da região do Rio de Janeiro (digamos $r = 1$) para o grupo de mulheres (digamos $i = 0$) na faixa etária de 14-26 anos em setembro de 1998 ($t = S98$), podemos escrever com base na equação (2.1) que:

$$\alpha_{1,1,S98} = m(Z_{1,0,1} = 0, Z_{2,0,1,S98} = 1, X_{0,1,S98}, d_{S98} = 1, r = 1),$$

onde $Z_{1,0,1} = 0$ denota o grupo de mulheres no Rio de Janeiro, $Z_{2,0,1,S98} = 1$ denota a faixa etária de 14-26 anos em setembro de 1998 para esse mesmo grupo e mesma região, $X_{0,1,S98}$ representa as variáveis econômicas em setembro de 1998 que afetam o impacto médio do programa para os mesmos grupo, região e mês, e $d_{S98} = 1$ denota o primeiro período após o início do programa.

Estamos interessados em estimar o EEE e o EIT com base nos resultados obtidos para o contexto de múltiplas regiões e uma única coorte de participantes do programa. Como discutido na seção 2.2, alguns aspectos do EEE são identificados de forma não-paramétrica se as condições R3-R6 forem satisfeitas. Viu-se também na referida seção que a identificação

Figura 2.7: Impacto Médio sobre o Rendimento do Trabalho por Região Metropolitana: Mulheres



totalmente não-paramétrica do EIT é alcançada se as condições R1 e R2 forem atendidas.

As condições R1 e R3 requerem que existam indivíduos com as mesmas características variantes e invariantes no tempo nas duas regiões para certos períodos de tempo. No que tange às características que não mudam no tempo, poder-se-ia trabalhar com estratos da amostra definidos, por exemplo, pelo sexo e cor dos indivíduos. Entretanto, devido ao tamanho relativamente pequeno da amostra com a qual estamos trabalhando, restringiremos nossa atenção somente aos estratos definidos pelo sexo dos indivíduos. Em relação às características que mudam no tempo, uma vez que no questionário do *follow-up* não há perguntas sobre essas características, no que se segue ignoraremos o requerimento de comparar indivíduos com as mesmas características variantes no tempo entre dois períodos quaisquer.²²

Em princípio, a escolha das variáveis que entram em X_{irt} deveria ser ditada por um modelo econômico que indicasse quais são as variáveis econômicas relevantes que afetam os impactos médios do programa em análise. Em nosso caso, por simplicidade, assumimos que uma medida de demanda por trabalho em cada região representa uma escolha adequada.²³ Temos também que levar em consideração o requerimento imposto pela condição

²²Como o intervalo de tempo amostrado é relativamente curto, acreditamos que ignorar esse requerimento não é demasiadamente problemático.

²³Idealmente, essa medida deveria se referir à demanda *relativa* por trabalhadores treinados.

de “cruzamento” R4, a qual exige que as condições econômicas sejam as mesmas entre as regiões para pelo menos um ponto no tempo. Adicionalmente, a condição R2 requer que as condições econômicas em cada região não sigam uma trajetória temporal estritamente monotônica.

Após analisar a evolução de diversas medidas de demanda por trabalho para as duas regiões amostradas, fomos capazes de encontrar somente uma que atendeu as exigências das condições R2 e R4 para o período em foco. Trata-se do que denominaremos *número líquido de contratações*, o qual é calculado pela subtração do número total de contratações pelo número total de desligamentos. Essa medida foi construída com base em dados administrativos que contém informações sobre o número de admissões e desligamentos mensais de todos os estabelecimentos registrados no Brasil (Cadastro Geral de Empregados e Desempregados - CAGED). Esse número líquido de contratações foi computado separadamente para homens e mulheres na faixa etária de 14 a 26 anos para cada região metropolitana de nossa análise. Isso foi implementado para cada um dos 15 meses para os quais estimamos o impacto médio do programa. As Figuras 2.8 e 2.9 apresentam a evolução temporal dessa variável respectivamente para homens e mulheres nas duas regiões metropolitanas. Como revelam essas Figuras, existe um cruzamento da nossa medida de demanda por trabalho em dezembro de 1998 no caso dos homens, e em janeiro de 1999 no caso das mulheres. Pode-se observar também nessas Figuras que a condição de não-monotonicidade R2 é (aproximadamente) atendida em cada região durante o período de análise.

Apresentaremos os resultados de duas formas distintas. Na primeira, computamos ambos o EIT e o EEE localmente, isto é, somente naqueles pontos em que suas respectivas condições de identificação são satisfeitas. Especificamente, o EIT é calculado para aqueles pares de meses nos quais o número de contratações líquidas é o mesmo em cada região metropolitana.²⁴ Uma vez que a identificação do EIT pode ser obtida de maneira totalmente não-paramétrica, esse efeito foi computado separadamente para cada região. Em relação ao EEE, este só foi calculado para os meses em que o número de contratações líquidas é *diferente* entre as regiões. Como discutido na seção 2.2, isso se deve ao fato de que a condição de cruzamento R4 impede que o EEE seja identificado naqueles períodos em que as condições econômicas são as *mesmas* entre as regiões. Nos períodos em que esses cruzamentos ocorrem – isto é, em dezembro de 1998 para os homens e em janeiro de 1999 para as mulheres –, calculamos o que denominamos de efeito regional (ver seção 2.2.2).

A segunda forma de apresentação dos resultados é baseada em uma regressão semi-paramétrica.²⁵ Seja U_{irt} o erro amostral das estimativas do impacto médio do programa

Contudo, não encontramos nenhuma variável que se aproximasse com o mínimo de exatidão dessa medida relativa.

²⁴A fim de aumentar o número de pontos em que o EIT é identificado, arredondamos o número de contratações líquidas para múltiplos de 100.

²⁵A hipótese implícita aqui é que existiriam diversas regiões cujas condições econômicas estariam dentro dos intervalos delimitados pelo número de contratações líquidas mensais entre as duas regiões metropolitanas amostradas. Em termos das Figuras 2.8 e 2.9, isso significa que área entre as curvas

Figura 2.8: Contrações Líquidas por Região Metropolitana: Homens

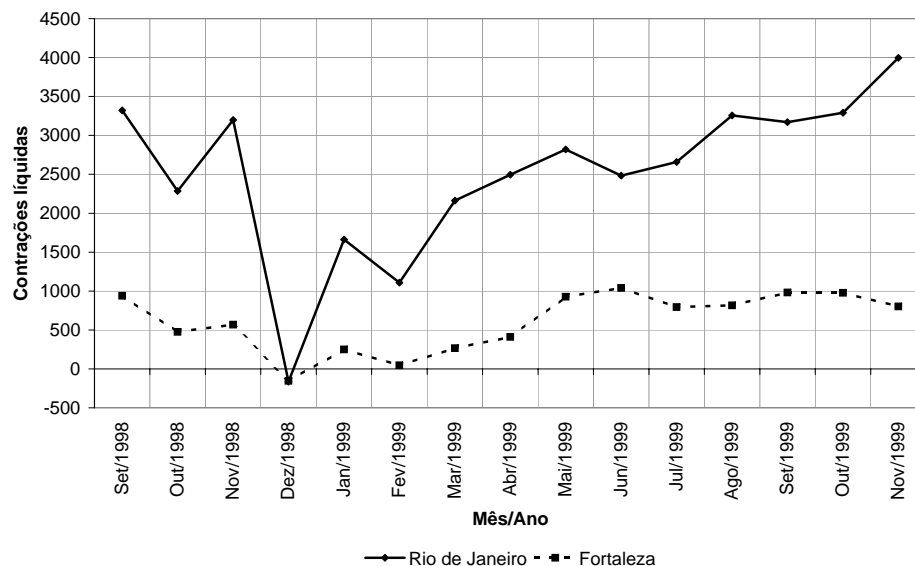
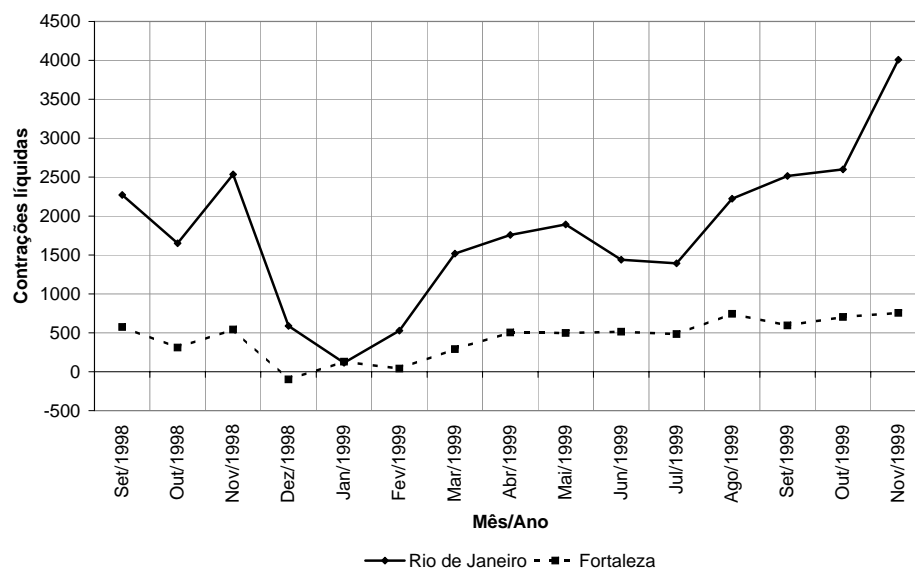


Figura 2.9: Contrações Líquidas por Região Metropolitana: Mulheres



do Rio de Janeiro e Fortaleza estaria preenchida com várias outras curvas correspondentes a essas regiões hipotéticas. O ponto de cruzamento em cada uma dessas Figuras seria mantido, entretanto,

(denotado por $\hat{\alpha}_{irt}$) para o grupo de indivíduos i na região r no período t . Utilizando a condição de separabilidade R6, a regressão pode ser especificada pelo seguinte modelo parcial-linear:²⁶

$$\hat{\alpha}_{irt} = m_1(Z_{1ir}, X_{irt}, d_t) + m_2(r) + U_{irt} = m_1(Z_{1ir}, X_{irt}, d_t) + a_r + U_{irt}, \quad (2.3)$$

onde $Z_{1ir} = i = \{1 = \text{homens}, 0 = \text{mulheres}\}$, $r = \{1 = \text{Rio de Janeiro}, 0 = \text{Fortaleza}\}$, $t = \{\text{Set}/98, \dots, \text{Nov}/99\}$, $d_t = \{1, \dots, 15\}$, e X_{irt} é o número de contratações líquidas de homens ou mulheres para o Rio de Janeiro ou Fortaleza no mês t . A função m_2 foi especificada por meio de uma variável *dummy* de região, cujo coeficiente associado, a_r , fornece o efeito regional.²⁷ Essa regressão foi estimada separadamente por sexo.

Embora não seja estritamente compatível com os resultados (locais) de identificação que apresentamos na seção 2.2, uma das vantagens da regressão semi-paramétrica é que ela permite a comparação do EIT para diversos níveis de contratações líquidas e do EEE para vários períodos de tempo.²⁸ No entanto, essa regressão apresenta uma importante desvantagem: a imposição da condição de separabilidade na equação (2.3) leva a que o EIT não seja identificado/estimado de forma totalmente não-paramétrica. Especificamente, tal como o EEE, o EIT torna-se igual para as duas regiões.

2.4.4 Resultados

2.4.4.1 Estimativas Locais

A Tabela 2.3 apresenta as estimativas locais do EIT para homens (Painel A) e mulheres (Painel B). As primeiras quatro colunas se referem aos resultados para o Rio de Janeiro, enquanto as quatro últimas para Fortaleza. Cada linha contém o EIT estimado para pares de períodos de tempo distintos e para o respectivo valor específico do número de contratações líquidas em cada região. Uma vez que não há muitos pares de períodos em que o número de contratações líquidas coincide para cada sexo e região, o EIT só pode ser estimado localmente para um número relativamente pequeno de pontos.

Começando pelos resultados referentes ao Rio de Janeiro, o EIT parece ser positivo

²⁶Esse modelo foi analisado por *e.g.* Robinson (1988) e Speckman (1988). Ver Pagan & Ullah (1999, cap. 5) para uma apresentação geral desse modelo.

²⁷A função m_1 foi estimada mediante o uso de métodos de Kernel. Os erros-padrões (*point-wise* para a função m_1) foram obtidos por meio do método de *bootstrap* com 100 replicações.

²⁸É importante observar que essa regressão é estimada levando-se em consideração a evolução de fato observada do número de contratações líquidas para cada sexo em cada região. Em outras palavras, a regressão respeita os limites dos suportes dos níveis de contratações líquidas entre as regiões em cada mês e para cada sexo. Caso esses limites não fossem respeitados – por exemplo, caso a regressão fosse implementada como se número de contratações líquidas preenchessem um *retângulo* no plano Cartesiano –, estaríamos ignorando completamente o fato de que o valor de X_{irt} não pode ser variado de forma independente dos valores assumidos pelos demais argumentos da função m .

Tabela 2.3: O Efeito Interno Temporal por Sexo e Região Metropolitana:
Estimativas Locais

Rio de Janeiro				Fortaleza			
Contratações	Períodos		EIT	Contratações	Períodos		EIT
Líquidas	de Tempo			Líquidas	de Tempo		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. HOMENS							
3300	1	12	0.059 (0.002)	900	1	9	0.024 (0.004)
3300	1	14	0.137 (0.003)	300	5	7	-0.014 (0.001)
3200	3	13	0.022 (0.002)	1000	10	13	0.007 (0.002)
2500	8	10	0.028 (0.002)	1000	10	14	0.004 (0.003)
				800	11	12	-0.013 (0.001)
				800	11	15	-0.031 (0.003)
B. MULHERES							
2500	3	13	0.064 (0.001)	600	1	13	-0.091 (0.003)
1400	10	11	0.014 (0.000)	300	2	7	-0.043 (0.003)
				500	3	8	-0.002 (0.001)
				500	3	9	-0.015 (0.001)
				500	3	10	0.022 (0.001)
				500	3	11	-0.020 (0.001)
				700	12	14	-0.033 (0.001)

Notas:.

(1) Erros-padrões entre parênteses.

para cada par de períodos de tempo e níveis de contratações líquidas apresentados na Tabela 2.3. Isso é observado para ambos os sexos nessa região. Uma vez que não é possível identificar não-parametricamente o EIT entre os mesmos períodos para diferentes condições econômicas, é difícil dizer se esse efeito apresentaria o mesmo sinal (ou magnitude) para distintos níveis de contratações líquidas. Adicionalmente, como o valor da nossa medida das condições econômicas não coincide para períodos de tempo logo após o início do programa (digamos, para os primeiros três meses), não se pode avaliar (localmente) se o EIT é positivo ou não no curto prazo.

As estimativas para Fortaleza indicam que o EIT é negativo para as mulheres para

quase todos os pares de períodos e níveis de contratações líquidas apresentados na Tabela 2.3. Em particular, o EIT é bastante negativo entre o 1º e o 13º meses após o começo do programa para um nível de contratações líquidas igual a 600. Os resultados do EIT para os homens em Fortaleza oscilam em termos do sinal. Mais uma vez, como não se pode identificar o EIT entre os mesmos pares de períodos para condições econômicas diferentes, não é possível checar se o padrão observado para ambos os sexos se manteria se os níveis de contratações líquidas fossem distintos. Tal como no caso do Rio de Janeiro, não é possível calcular estimativas locais do EIT no curto prazo para Fortaleza.

A Tabela 2.4 contém as estimativas do EEE para homens e mulheres. As colunas (1) e (2) apresentam respectivamente o mês e o correspondente período de tempo após o início de programa; as colunas (3)-(4) e (6)-(7) informam o nível de contratações líquidas para os quais o EEE foi calculado.²⁹ As estimativas desse efeito estão apresentadas nas colunas (5) e (8), respectivamente para homens e mulheres. Como discutido na seção 2.2, sempre que as condições econômicas coincidem entre as regiões, o EEE não é identificado. Assim, esse efeito não foi estimado no período 4 (dezembro de 1998) para os homens e no período 5 (janeiro de 1999) para as mulheres. Como nossa medida de condições econômicas é diferente entre as regiões para os demais períodos, o EEE pôde ser computado para cada um desses períodos.

Em princípio, se os diferentes níveis das condições econômicas fossem os mesmos entre as regiões em distintos pontos no tempo, seria possível identificar o EEE para esses níveis em cada um desses períodos.³⁰ Isso permitiria avaliar como o EEE varia no tempo. Em nosso caso, entretanto, não ocorre essa coincidência de diferentes níveis entre as regiões, o que implica que os resultados para o EEE são específicos para cada período e par de valores das contratações líquidas.

Da observação dos resultados da Tabela 2.4, não encontramos um padrão claro para o efeito de variações nas contratações líquidas sobre o impacto médio do programa. Uma explicação para isso é que o EEE depende do período para o qual ele é estimado. Por exemplo, apesar do EEE não ser estritamente comparável em diferentes períodos de tempo, é possível notar que ele é negativo até o 9º mês após o início do programa no caso dos homens, e até o 2º mês no caso das mulheres. Também notável é que o EEE é relativamente pequeno nos primeiros três meses para os homens, tornando-se mais pronunciado (em termos absolutos) desse ponto em diante. Essa mudança na magnitude do EEE ao longo do tempo não parece ocorrer para as mulheres, pelo menos até o 11º período após o início do programa.

O (constante) efeito regional entre o Rio de Janeiro e Fortaleza é apresentado nas notas abaixo da Tabela 2.4. Suas estimativas têm sinal positivo para homens e mulheres,

²⁹Para manter consistência com o que fizemos no caso do EIT, os níveis de contratações líquidas estão apresentados como múltiplos de 100.

³⁰Isso ocorreria, por exemplo, se as trajetórias temporais das condições econômicas das regiões tivessem um formato cruzado de “X”.

Tabela 2.4: O Efeito Econômico Externo por Sexo: Estimativas Locais

Mês Calendário (1)	Período de Tempo (2)	Homens			Mulheres		
		Contratações		EEE (5)	Contratações		EEE (8)
		Líquidas			Líquidas		
		(3)	(4)		(6)	(7)	
Setembro 98	1	900	3300	-0.010 (0.004)	600	2300	-0.043 (0.002)
Outubro 98	2	500	2300	-0.004 (0.004)	300	1700	-0.048 (0.002)
Novembro 98	3	600	3200	0.012 (0.004)	500	2500	0.004 (0.003)
Dezembro 98	4	-200	-200	- (0.003)	-100	2300	0.031 (0.003)
Janeiro 99	5	300	1700	-0.066 (0.003)	100	100	- (0.003)
Fevereiro 99	6	0	1100	-0.085 (0.003)	0	500	-0.010 (0.003)
Março 99	7	300	2200	-0.059 (0.003)	300	1500	0.038 (0.003)
Abril 99	8	400	2500	-0.037 (0.003)	500	1800	0.017 (0.003)
Maio 99	9	900	2800	-0.021 (0.003)	500	1900	0.067 (0.003)
Junho 99	10	1000	2500	0.042 (0.004)	500	1400	0.020 (0.003)
Julho 99	11	800	2700	0.054 (0.004)	500	1400	0.075 (0.003)
Agosto 99	12	800	3300	0.060 (0.004)	700	2200	0.064 (0.003)
Setembro 99	13	1000	3200	0.026 (0.004)	600	2500	0.094 (0.003)
Outubro 99	14	1000	3300	0.152 (0.005)	700	2600	0.106 (0.003)
Novembro 99	15	800	4000	0.132 (0.005)	800	4000	0.009 (0.003)

Notas:.

(1) Erros-padrões entre parênteses.

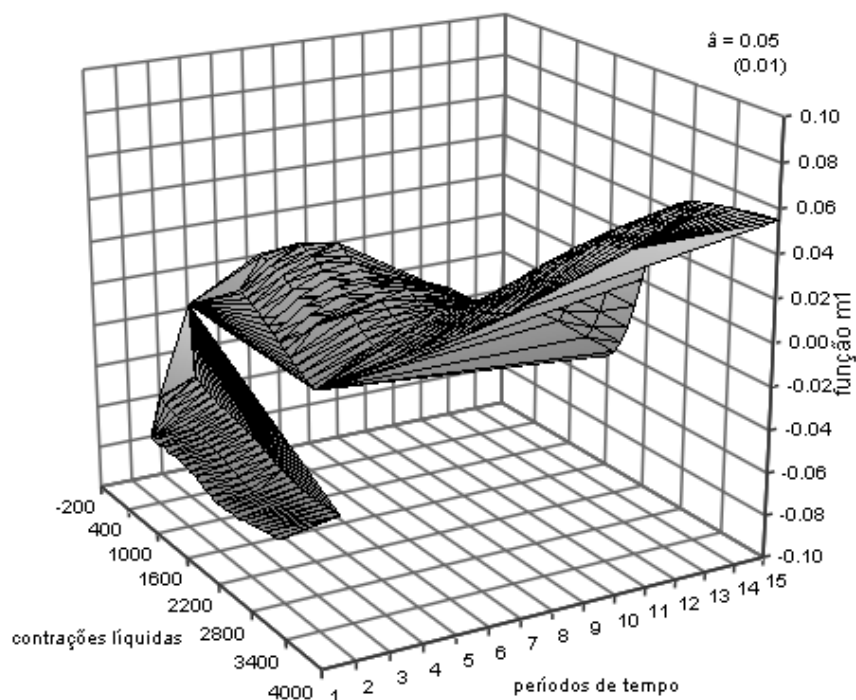
(2) A estimativa (erro-padrão) do efeito regional é 0,055 (0,004) para homens e 0,064 (0,003) para mulheres.

significando que existe uma vantagem do Rio de Janeiro em relação à Fortaleza para ambos os sexos.

2.4.4.2 Estimativas Semi-paramétricas

A equação 2.3 foi estimada separadamente para homens e mulheres. As estimativas da função m_1 estão apresentadas nas Figuras 2.10 e 2.11, respectivamente para homens e mulheres. A estimativa do coeficiente associado a *dummy* de região (erro-padrão entre parênteses) também está apresentada nessas Figuras. O primeiro ponto a se notar é que o “cruzamento” das condições econômicas entre as regiões (ver Figuras 2.8 e 2.9) cria um

Figura 2.10: Estimativa da Função m_1 para Homens

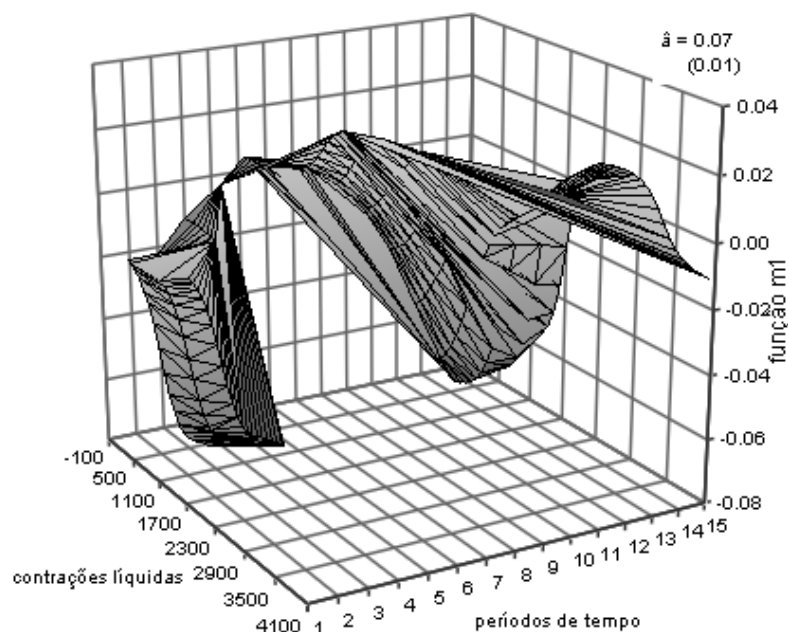


estreitamento na superfície estimada para m_1 no ponto correspondente a esse “cruzamento”. Outro ponto a se observar é que o valor estimado de a_r é positivo para ambos os sexos, um resultado também encontrado com base nas estimativas locais da sub-seção anterior. Graficamente, isso significa que a superfície de m_1 para o Rio de Janeiro situa-se (de maneira paralela) acima da superfície correspondente à Fortaleza.

Para se computar o EIT e o EEE, há que se tirar “fatias” das superfícies apresentadas nas Figuras 2.10 e 2.11, sendo o primeiro (segundo) efeito associado às fatias correspondentes a determinados valores das contratações líquidas (períodos de tempo). A Figura 2.12 mostra as fatias referentes a cinco diferentes valores assumidos pela variável contratações líquidas, ao passo que a Figura 2.13 apresenta as fatias correspondentes a três períodos distintos após o início do programa. Essas duas Figuras se referem aos homens, enquanto as Figuras 2.14 e 2.15 são as análogas para as mulheres.³¹ As áreas hachuradas ao longo das curvas presentes nessas Figuras correspondem ao intervalo de confiança de 95%. Para se calcular o EIT, deve-se obter a distância vertical entre quaisquer dois pontos que pertencem a uma curva particular nas Figuras 2.12 ou 2.14. O mesmo se aplica ao cálculo do EEE nas Figuras 2.13 ou 2.15.

³¹Cabe observar que, devido à restrição de cruzamento R4 e à existência de diferentes suportes para o número de contratações líquidas em cada período de tempo, as fatias tiradas das Figuras 2.10 e 2.11 não são nem necessariamente contínuas nem compartilham exatamente do mesmo suporte.

Figura 2.11: Estimativa da Função m_1 para Mulheres

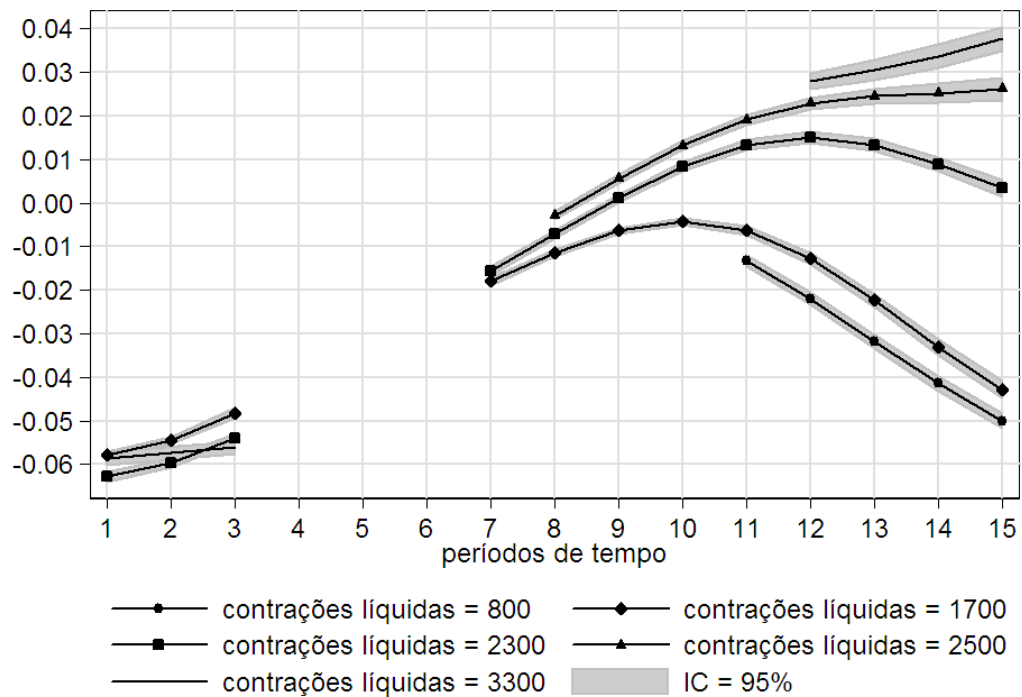


Como se pode observar na Figura 2.12, o EIT é relativamente pequeno nos primeiros três meses após o começo do programa, para diferentes níveis de contratações líquidas.³² A Figura também revela que, a medida em que o tempo decorre, o EIT torna-se mais sensível às mudanças nas condições econômicas. Por exemplo, tomando o intervalo entre o 11º e o 15º meses, o EIT é negativo para valores mais baixos das contratações líquidas (*e.g.* 800), porém positivo para valores mais altos (*e.g.* 3300). Cabe assinalar que os sinais do EIT na Figura 2.12 são os mesmos que os obtidos com as estimativas locais não-paramétricas apresentadas na Tabela 2.3.

O EEE para homens em três períodos específicos pode ser visto pela Figura 2.13. O primeiro ponto a se notar é que esse efeito parece variar para distintos períodos de tempo. Isso pode ser visto pela comparação do formato das três curvas apresentadas na Figura, para o suporte comum de contratações líquidas (*i.e.* entre 1000 e 2200). O segundo ponto a se observar é que as mudanças nas condições econômicas não parecem afetar muito o impacto médio do programa em seu começo (período 1). Entretanto, isso deixa de ser assim após alguns meses, como indicado pela curva correspondente ao 7º período, a qual mostra um EEE negativo. Ademais, no 14º período, o EEE muda de sinal (comparado ao 7º período) e torna-se elevado, pelo menos para o intervalo compreendido entre 1400 e 2600 contratações líquidas. Vale notar que esses resultados estão em linha com as estimativas

³² Isso pode se dever a efeitos de “enclausuramento” (*lock-in effects*) do programa, que retira tempo dos treinandos para a busca de (melhores) empregos durante a vigência do programa.

Figura 2.12: Efeito Interno Temporal para Homens: Estimativas Semi-Paramétricas



locais apresentadas na Tabela 2.4.

O EIT para mulheres pode ser observado a partir da Figura 2.14, a qual apresenta fatias da Figura 2.11, correspondendo a cinco níveis distintos de contratações líquidas. Como no caso dos homens, o EIT parece ser sensível a mudanças nas condições econômicas, especialmente quando nos movemos para o final do período de análise. Isso pode ser confirmado, por exemplo, pela comparação das curvas associadas aos níveis de contratações líquidas de 1200 e 2300 após o 12º período. Com uma única exceção, os sinais do EIT obtidos da Figura 2.14 são os mesmos dos apresentados na Tabela 2.3.³³

Em relação ao EEE para mulheres, a Figura 2.15 mostra que esse efeito parece variar bastante para distintos períodos de tempo. A comparação dos formatos das curvas dentro do suporte comum das contratações líquidas (aqui entre 700 e 1500) permite confirmar essa constatação. É possível observar também que o sinal do EEE muda entre os períodos, sendo essencialmente negativo no início do programa (período 1), porém positivo nos períodos seguintes (períodos 7 e 14). Mais uma vez, observamos uma similaridade entre os resultados do EEE obtidos da Figura 2.15 e os apresentados na Tabela 2.4.

Em suma, ambas as estimativas locais e semi-paramétricas apresentadas nesta seção indicam a existência de interações entre os fatores que determinam o impacto médio do

³³A exceção refere-se a estimativa entre os períodos 3 e 10 no nível de contratações líquidas igual a 500.

Figura 2.13: Efeito Econômico Externo para Homens: Estimativas Semi-Paramétricas

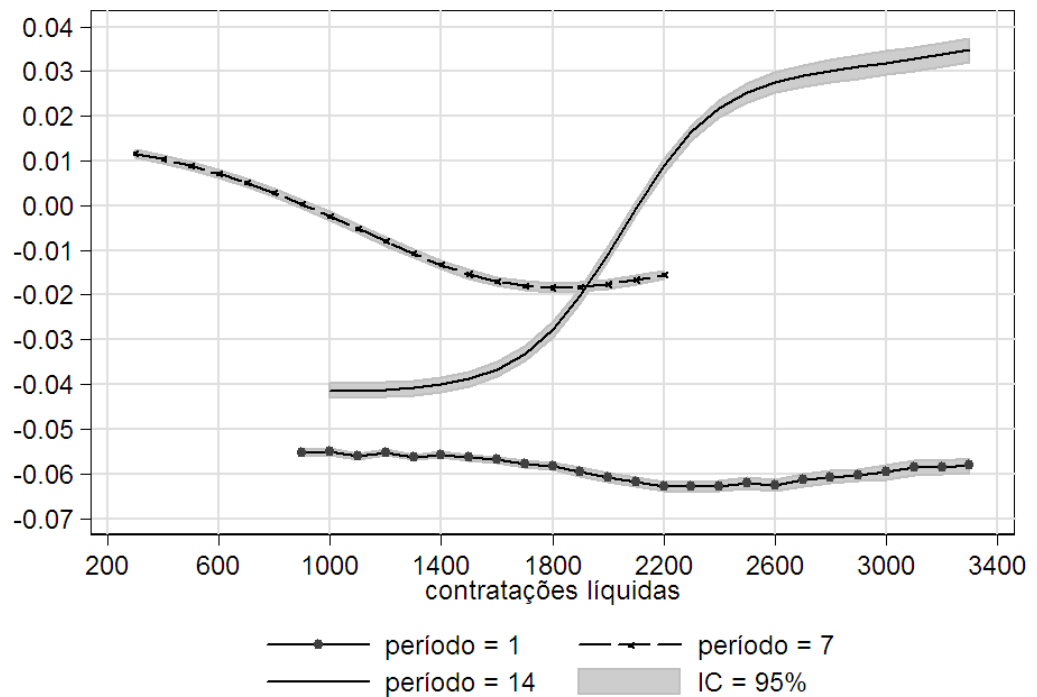


Figura 2.14: Efeito Interno Temporal para Mulheres: Estimativas Semi-Paramétricas

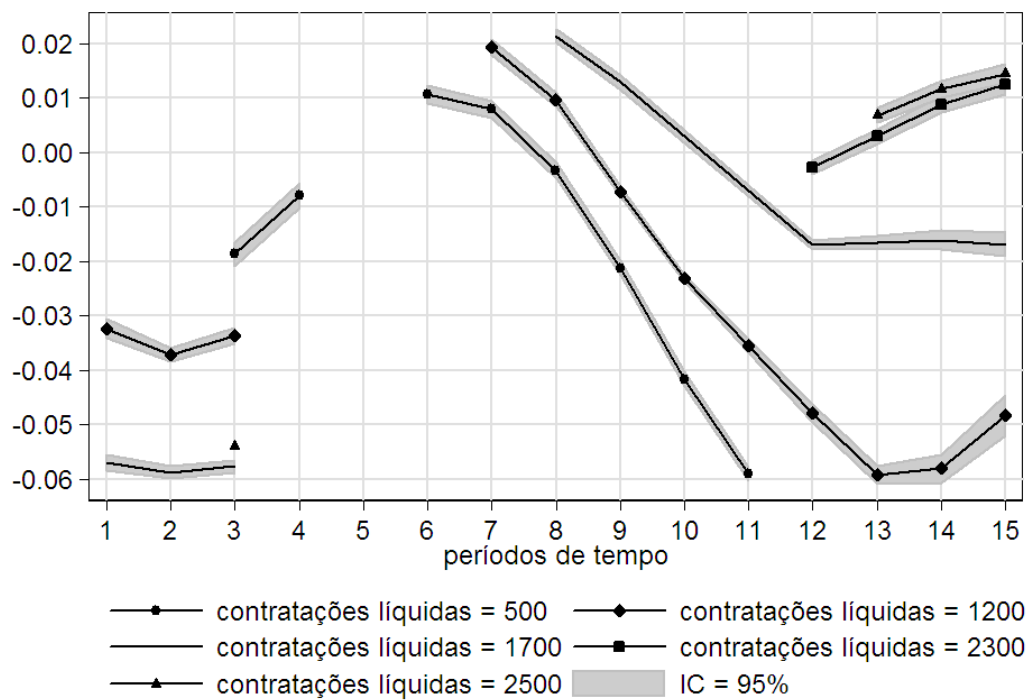
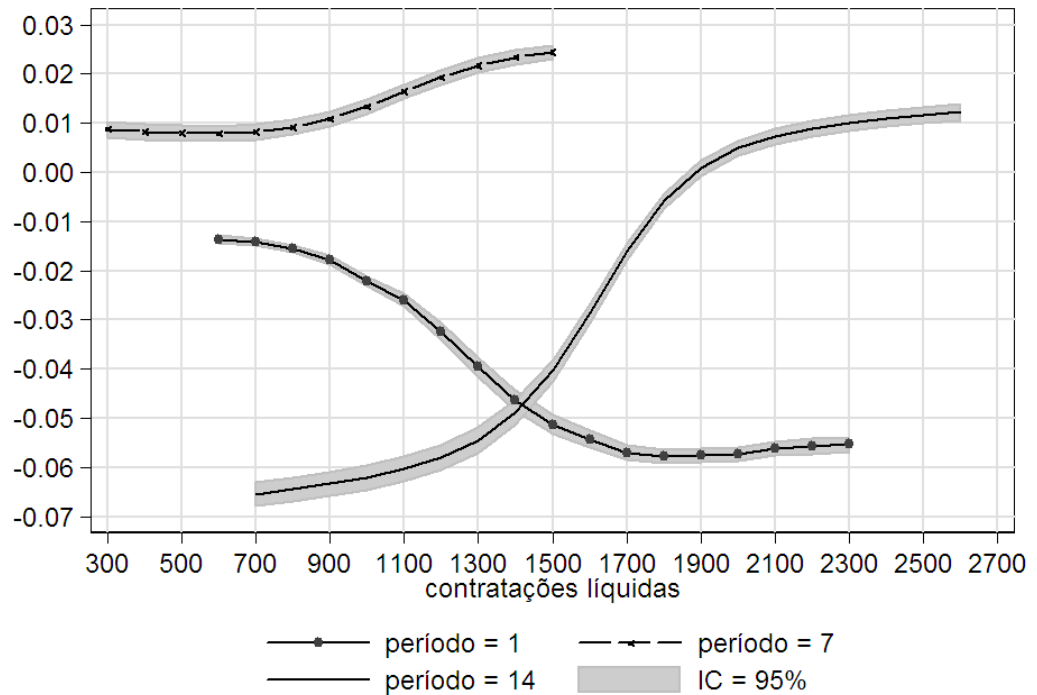


Figura 2.15: Efeito Econômico Externo para Mulheres: Estimativas Semi-Paramétricas



programa, particularmente X_{irt} e d_t . Nesse sentido, como o formato dessas interações não é conhecido *a priori*, concluímos que pode haver um significativo benefício em identificar/estimar o EIT e o EEE de forma não-paramétrica.

2.5 Conclusões

As evidências disponíveis sobre avaliação de programas sociais mostram que o impacto médio dessas intervenções tende a variar ao longo do tempo. Neste capítulo, procuramos analisar que parte dessas variações se deve aos efeitos do programa em si mesmo e que parte está associada a alterações nas condições econômicas. Para tanto, introduzimos dois conceitos que procuram distinguir o efeito que pode ser atribuído somente a existência do programa (efeito interno temporal) do efeito que advém das mudanças no ambiente econômico (efeito econômico externo).

O principal objetivo deste capítulo foi estudar a identificação desses dois efeitos. Para isso, partimos de uma especificação não-paramétrica dos impactos médios de um programa social qualquer, e propusemos um conjunto de condições que permitem identificar não-parametricamente alguns aspectos desses efeitos. A análise de identificação mostrou que somente o efeito interno temporal é identificado de forma totalmente não-paramétrica. Esse resultado foi estabelecido para dois contextos diferentes: um no qual se assume que o

analista dispõe de informações para múltiplas regiões e somente uma coorte de participantes do programa; e outro no qual existem dados para múltiplas coortes de participantes e somente uma região.

A razão que impossibilita a identificação totalmente não-paramétrica do efeito econômico externo é que existe um tipo de dependência funcional entre as variáveis que medem as condições econômicas e os demais argumentos da função não-paramétrica que determina os impactos médios do programa. De fato, tanto no contexto de múltiplas regiões quanto no de múltiplas coortes, só é possível variar as medidas das condições econômicas em dimensões que já aparecem como argumentos da referida função não-paramétrica. Por exemplo, no contexto de múltiplas regiões, não é possível variar o valor das medidas das condições econômicas sem alterar o grupo de indivíduos, a região ou o período de tempo para os quais o impacto médio do programa foi calculado. É essa impossibilidade de variar independentemente o valor das medidas das condições econômicas que exige a imposição de certas restrições de natureza paramétrica.³⁴

Entre as condições utilizadas na análise de identificação do efeito econômico externo, três merecem ser destacadas. A primeira é uma restrição de separabilidade aditiva, que impõe a separação de somente um dos argumentos da função que determina os impactos médios do programa. A segunda é uma condição de “cruzamento” (“não-monotonicidade”) no contexto de múltiplas regiões (múltiplas coortes). No primeiro contexto, esta condição requer que as condições econômicas sejam as *mesmas* entre as regiões para um período de tempo; no segundo, ela exige *iguais* condições econômicas entre dois períodos de tempo. A terceira condição demanda que as condições econômicas sejam *diferentes*, o que é um requerimento inerente à própria definição do efeito econômico externo. Claramente, a exigência desta última condição cria uma incompatibilidade com o demandado pela segunda condição.

A principal consequência da imposição dessas três condições é que somente alguns aspectos do efeito econômico externo são identificados de forma não-paramétrica. Em primeiro lugar, a condição de separabilidade impede a existência de interações potencialmente importantes entre alguns dos argumentos da função não-paramétrica inicial. Por exemplo, no contexto de múltiplas regiões, a condição de separabilidade restringe o efeito econômico externo a ser idêntico entre as regiões. Em segundo lugar, a incompatibilidade acima mencionada entre a segunda e a terceira condições impossibilita que esse efeito possa ser identificado não-parametricamente para todos os períodos de tempo.

Apesar das condições aqui propostas não produzirem a identificação totalmente não-paramétrica do efeito econômico externo, elas são menos restritivas do que a imposição *a priori* de um modelo completamente parametrizado. Esse fato pode ser importante, na medida em que permite ao analista estimar o efeito de interesse sem impor formas

³⁴As consequências desse tipo de dependência funcional para a identificação não-paramétrica em geral são o objeto do próximo capítulo da presente tese.

funcionais demasiadamente fortes.

Na parte final deste capítulo, apresentamos um exemplo empírico que procura ilustrar a identificação/estimação do efeito interno temporal e do efeito econômico externo para o contexto de múltiplas regiões e uma coorte de participantes. Na implementação desse exemplo, utilizamos dados experimentais de duas regiões metropolitanas brasileiras para estimar (consistentemente) os impactos médios de um programa público de treinamento de mão-de-obra. Também fizemos uso de dados administrativos para obter uma medida de demanda por trabalho (número de contratações líquidas) para essas duas regiões ao longo do tempo. O valor dessa variável coincide para um único ponto no tempo entre as regiões, o que permite satisfazer a condição de “cruzamento”. Os resultados empíricos da identificação foram apresentados num sentido estritamente local, e também com o uso de uma regressão semi-paramétrica (modelo parcial-linear). Em linhas gerais, os resultados do exercício indicam que o efeito interno temporal tende a variar com as condições da economia (mercado de trabalho), e o efeito econômico externo com o período de tempo após o início do programa. Como não impusemos formas funcionais para essas interações, esses resultados confirmam os ganhos associados à identificação/estimação (parcialmente) não-paramétrica dos efeitos de interesse.

Capítulo 3

Identificação de Diferenças Parciais com Dependência Funcional entre Covariadas

3.1 Introdução

O estudo de identificação tem uma extensa história na área de econometria, com contribuições fundamentais de, por exemplo, [Koopmans & Reiersøl \(1950\)](#), [Koopmans et al. \(1950\)](#) e [Hurwicz \(1950\)](#). Embora os primeiros autores tenham concebido os principais conceitos da análise de identificação para o contexto não-paramétrico, até o final dos anos 80 a maior parte dos trabalhos nessa área esteve concentrada em identificação a partir de um arcabouço paramétrico. Em um importante artigo, [Roehrig \(1988\)](#) renovou o interesse em identificação não-paramétrica, e nos últimos vinte anos tem aparecido uma crescente literatura voltada para a análise de identificação sob essa ótica.

Em essência, existem duas vantagens em estudar identificação a partir de um arcabouço não-paramétrico. A primeira está relacionada ao fato de que a teoria econômica tipicamente não fornece formas funcionais específicas para as relações postuladas pelo modelo que baseia a análise dos dados. Nesse sentido, o contexto não-paramétrico é claramente o mais flexível para se estudar a identificação dos objetos de interesse do modelo. A segunda vantagem é que a generalidade da identificação não-paramétrica normalmente permite um entendimento mais claro dos mecanismos que estão por trás da identificação desses objetos. Por essas razões, impor formas funcionais paramétricas desde o princípio da análise pode ser uma estratégia um tanto enganosa para os propósitos de identificação.¹

¹ Isso não quer dizer que parametrizações não possam ser empregadas na estimação dos modelos. Em geral, tendo assegurado a identificação não-paramétrica dos objetos de interesse, o analista pode escolher especificações menos flexíveis na fase de estimação. Essa é uma das principais mensagens de [Roehrig \(1988\)](#).

Este capítulo procura investigar a identificação não-paramétrica dos efeitos de variações no valor das covariadas que determinam a média condicional de uma variável de resultado. Nosso interesse recai sobre um tipo de situação em que os valores de algumas covariadas não podem ser modificados sem alterar os valores das demais covariadas do modelo. Em outras palavras, estamos interessados num tipo de caso em que algumas covariadas podem ser consideradas funcionalmente dependentes das demais.

Mais especificamente, seja Y uma variável de resultado cuja média, α , é completamente determinada por dois conjuntos de variáveis: $X = (X_1, \dots, X_K)$ e $Z = (Z_1, \dots, Z_L)$. Seja $m : \Omega_W \rightarrow \Re$ uma função não-paramétrica tal que $\alpha = E[Y|X, Z] = m(X_1, \dots, X_K, Z_1, \dots, Z_L)$, onde Ω_W denota o conjunto de valores que $W = (X, Z)$ pode assumir. Suponha que os valores de Z possam ser variados livremente e que os valores de X sejam unicamente determinados quando os valores de Z são especificados. Note que essa última suposição cria restrições funcionais que conectam as variações em X às variações de Z . Isso leva a que X seja funcionalmente dependente de Z , o que permite expressá-lo como uma função (vetorial) de Z : $f : D_Z \rightarrow D_X$, $X = f(Z)$, onde D_Z e D_X representam os conjuntos de valores que Z e X podem assumir, respectivamente.

Tipicamente, em economia, o interesse centra-se no efeito de uma determinada covariada sobre a média condicional da variável de resultado quando se varia *somente* o valor dessa covariada.² Em nosso caso, devido à dependência funcional de X em relação a Z , claramente não é possível obter o efeito de qualquer elemento de X sobre α sem impor restrições ao modelo. O principal objetivo deste capítulo é estudar quais aspectos desse efeito podem ser identificados de forma não-paramétrica. Para tanto, propomos um conjunto de condições suficientes que é capaz de produzir a identificação não-paramétrica de partes desse efeito. Embora não seja o nosso foco aqui, investigamos também a identificação do efeito de variar o valor de um dos elementos de Z sobre α .

Definimos os efeitos de interesse com base no que se costuma denominar *diferenças parciais* de uma função. Especificamente, esses efeitos são definidos por variações *finitas* que ocorrem no valor da função m quando somente um de seus argumentos muda de valor. A principal razão para o uso desse objeto é que não restringimos as covariadas a serem contínuas, nem as funções m e f a serem diferenciáveis. Distintamente das derivadas parciais, esse objeto é factível de ser identificado em situações em que as covariadas apresentam variações discretas, ou não se quer impor diferenciabilidade das funções do modelo.

A identificação dos efeitos de interesse é estudada num sentido local, ou seja, para valores particulares assumidos pelos argumentos de m . Optamos por essa abordagem porque, na ausência de restrições paramétricas globais, os dados podem ser informativos somente para identificar esses efeitos para certos valores dos argumentos de m . Se res-

²Por exemplo, assumindo que as variáveis que determinam o valor de uma variável de resultado podem variar livremente, Heckman (2000, p.53) define *efeito causal* como a mudança na variável de resultado que é produzida pela variação na variável determinante de interesse quando os valores das demais variáveis determinantes são mantidos constantes.

trições paramétricas forem impostas *a priori* ao modelo, então é possível que o valor local desses efeitos seja identificado globalmente (por exemplo, os parâmetros associados aos argumentos de m numa especificação linear desta função). Nesse sentido, a força das restrições paramétricas é somente permitir a identificação global dos efeitos de interesse.³

Nossa análise de identificação se restringe a situações em que o analista conhece (ou consegue estimar consistentemente) a média condicional da variável de resultado. Assumimos também que o analista sabe o conjunto completo de variáveis que determinam a média condicional, e que essas variáveis são medidas sem erro. A principal implicação dessas suposições é que, na análise de identificação, abstraímos da presença de componentes (endógenos) não-observáveis que normalmente aparecem na literatura de identificação. Mais formalmente, o nosso caso pode ser visto como o modelo no qual existe um componente não-observável que é aditivo e independente das covariadas, isto é: $Y = m(X, Z) + U$, onde U denota o termo não-observável e $U \perp X, Z$.

Por identificação entenderemos a situação em que se pode inferir sem incerteza qual o valor do objeto de interesse a partir dos dados observados e de um conjunto de restrições impostas ao modelo. Em outras palavras, diremos que o objeto de interesse é identificado se ele pode ser unicamente determinado com base nos dados disponíveis e nas hipóteses adotadas.⁴

Acreditamos que os resultados de identificação deste capítulo podem ser úteis para diferentes áreas da pesquisa aplicada. Por exemplo, no capítulo anterior, apresentamos uma aplicação para a área de avaliação de programas, sendo o principal objeto de interesse o efeito de variações nas condições econômicas sobre o impacto médio de um programa social qualquer. Como apontado naquele capítulo, o problema de dependência funcional entre covariadas aparece porque os valores das variáveis que medem as condições econômicas não podiam variar sem alterar os valores das demais variáveis que determinam o impacto médio do programa. Uma outra possível aplicação se refere à literatura que utiliza dados em que um conjunto de unidades de *cross-section* (*e.g.* países, regiões ou indústrias) é observado por um certo número de períodos. Um exemplo típico dessa literatura são os estudos que fazem uso de informações temporais de estados de um país para estimar os efeitos do

³Assinale-se, no entanto, que, se a especificação paramétrica não representar uma boa aproximação do processo gerador dos dados, então a identificação dos verdadeiros efeitos não estará demonstrada.

⁴Tradicionalmente, um problema de identificação aparece porque existem diversos modelos alternativos que são compatíveis com os dados disponíveis. Uma das maiores contribuições dos fundadores da área de identificação em econometria foi mostrar que o problema de identificação pode ser resolvido por meio da imposição de restrições aos modelos em consideração. Geralmente, essas restrições envolvem um conjunto de hipóteses sobre diversos aspectos dos modelos, por exemplo sobre as formas funcionais das equações estruturais (*e.g.* lineares, semi-paramétricas, etc.) ou sobre a relação estatística entre os termos observáveis e não-observáveis (*e.g.* independência ou média condicional fixa). Ver Koopmans & Reiersøl (1950) e Hurwicz (1950) para uma apresentação formal dos conceitos tipicamente utilizados na área de identificação. Note-se que as nossas hipóteses de que o termo não-observável é aditivo e independente das covariadas já representam restrições ao modelo.

salário mínimo sobre o nível de emprego. O tipo de restrição funcional considerada neste capítulo aparece nesse exemplo, pois não é possível variar o valor do mínimo sem modificar os valores associados com as dimensões seccional e temporal dos dados.

Este capítulo está organizado da seguinte maneira. Na seção 2, fazemos um breve resumo da literatura relacionada. Na seção 3, definimos o nosso arcabouço analítico e apresentamos os resultados de identificação dos efeitos de interesse. Na seção 4, ilustramos os resultados de identificação da seção anterior por meio de um simples exemplo analítico baseado na literatura que utiliza dados de *cross-section* ao longo do tempo. A seção 5 contém as conclusões.

3.2 Literatura Relacionada

Embora os conceitos básicos da área de identificação tenham sido concebidos a partir de um arcabouço não-paramétrico (ver Koopmans & Reiersøl (1950) e Hurwicz (1950)), durante algumas décadas o estudo de identificação tratou essencialmente de modelos em que se assumia que as relações entre as variáveis era paramétrica e linear. Em larga medida, isso se deveu ao fato de que a análise econométrica esteve focada na estimação de modelos macroeconômicos de extração Keynesiana, os quais não eram tipicamente caracterizados por estruturas (altamente) não-lineares.⁵ No entanto, a “revolução” da micro-econometria que se inicia nos anos 70 originou uma importante mudança no foco de atenção. De fato, a partir dessa época, o interesse passou progressivamente a se voltar cada vez mais para o comportamento microeconômico dos agentes, o qual está sujeito a presença de uma série de não-linearidades.⁶

Estendendo o trabalho de Brown (1983), Roehrig (1988) re-estimulou o interesse em identificação não-paramétrica. De fato, utilizando os conceitos elaborados pelos fundadores da área de identificação, este autor estabelece condições para a identificação não-paramétrica de sistemas de equações não-lineares quando as variáveis observáveis e não-observáveis são independentes.⁷ A maior parte dos resultados de Roehrig (1988) se baseia em restrições de exclusão (*exclusion restrictions*), ou seja, na hipótese de que existem variáveis que pertencem exclusivamente a certas equações do sistema.

Newey & Powell (1988) estudam a identificação não-paramétrica de um modelo com uma única equação, na qual o termo não-observável aparece de forma aditiva e é potencialmente correlacionado com as covariadas. Os autores estabelecem condições de identificação

⁵Esses modelos, os quais ainda podem ser encontrados na maioria dos manuais de econometria, normalmente recebem a denominação de Modelos de Equações Simultâneas Lineares.

⁶Ver Chesher (2002) para uma breve descrição da história da área de identificação em econometria.

⁷Em um artigo recente, Benkhard & Berry (2006) apontam um erro em Brown (1983) e Roehrig (1988). Matzkin (2005) provê uma correção e estabelece novas condições de identificação não-paramétrica desses sistemas.

da função não-paramétrica de interesse, as quais são baseadas na existência de variáveis instrumentais. Newey et al. (1999) estendem o arcabouço considerado em Newey & Powell (1988) para um sistema de equações não-paramétricas, nas quais os termos não-observáveis também são aditivos e potencialmente correlacionados com as covariadas. Tal como em Newey & Powell (1988), a identificação é baseada na existência de variáveis instrumentais.

Imbens & Newey (2001) consideram um sistema de equações sem aditividade, cujos termos não-observáveis podem ser correlacionados com as covariadas. Além de imporem certas condições de monotonicidade sobre as funções não-paramétricas estruturais, esses autores também fazem uso da existência de variáveis instrumentais para obter resultados de identificação dessas funções e de outros objetos de interesse.⁸

Matzkin (2003) considera um modelo com uma única equação em que o termo não-observável é não-aditivo, porém independente das covariadas. Utilizando restrições de monotonicidade da função não-paramétrica em relação ao termo não-observável, esse trabalho estabelece condições de identificação baseadas no uso de quantis condicionais desse termo não-observável.

Chesher (2003) estuda a identificação local de derivadas parciais de sistemas de funções não-separáveis, permitindo que os termos não-observáveis sejam correlacionados com as covariadas. Como em Matzkin (2003), restrições de monotonicidade sobre as funções não-paramétricas em relação aos termos não-observáveis são utilizadas, e condições de identificação de quantis condicionais são estabelecidas. Adaptando o arcabouço de Chesher (2003) para o caso em que as covariadas exibem variações discretas, Chesher (2007) considera a identificação do mesmo tipo de objeto de interesse deste capítulo: diferenças parciais. Esse trabalho, que utiliza um arcabouço mais geral que o nosso, propõe algumas condições de identificação que podem ser de alguma forma comparadas com as que aqui apresentamos. Na próxima seção, procuraremos discutir as diferenças existentes.

Cabe assinalar que, tanto quanto sabemos, nenhum dos trabalhos da literatura de identificação não-paramétrica trata explicitamente da existência do tipo de dependência funcional que consideramos neste capítulo.

3.3 Modelo e Resultados de Identificação

3.3.1 Modelo

Seja Y uma variável escalar de resultado cujo valor é completamente determinado por um vetor de variáveis observáveis W de dimensão $K+L$ e um termo de distúrbio aleatório não-observável U . Seja $m^* : \Omega \rightarrow \Re$ uma função não-paramétrica com domínio $\Omega \subset \Re^{K+L+1}$,

⁸O principal objeto considerado por esses autores é a média condicional que se obtém quando se integra a função não-paramétrica associada a uma variável de resultado em relação ao termo não-observável correspondente. Este, por sua vez, é condicionado aos demais termos não-observáveis do sistema.

tal que: $Y = m^*(W, U)$. Ao longo do capítulo, mantemos as seguintes hipóteses:

H1: A função m^* satisfaz: $m^*(W, U) = m(W) + U$, onde $m : \Omega_W \rightarrow \mathbb{R}$ é uma função não-paramétrica com domínio $\Omega_W \subset \mathbb{R}^{K+L}$.

H2: U e W são distribuídos independentemente.

Estamos interessados em estudar a identificação não-paramétrica dos efeitos das variáveis de W sobre a média condicional de Y , quando há uma restrição que faz com que os valores de um sub-conjunto de elementos de W não possam ser variados independentemente dos valores assumidos pelo sub-conjunto complementar. Uma maneira de pensar essa restrição é que existe um tipo de dependência funcional do primeiro sub-conjunto de elementos de W em relação ao segundo sub-conjunto. Assim, particionando W em $W = (X, Z)$, onde X e Z são vetores de variáveis respectivamente com dimensões $K \times 1$ e $L \times 1$, iremos supor que os elementos de X são funcionalmente relacionados aos elementos de Z . Podemos escrever, então, que:⁹

$$X = f(Z) = f(Z_1, \dots, Z_L), \quad (3.1)$$

onde $f : D_Z \rightarrow D_X$ é uma função vetorial de dimensão K , $f = \{f_i\}_{i=1}^K$, $D_Z \subset \mathbb{R}^L$ representa o domínio de Z e $D_X \subset \mathbb{R}^K$ o conjunto de valores assumidos por X .¹⁰

Dadas as hipóteses acima, podemos escrever a média condicional de Y como:

$$\alpha = E[Y|W] = m(X, Z) = m(X_1, \dots, X_K, Z_1, \dots, Z_L). \quad (3.2)$$

Assumimos que o analista conhece (ou é capaz de estimar consistentemente) $E[Y|W] = E[Y|X, Z]$ a partir dos dados observados.^{11,12} Supomos também que os valores de $X = f(Z)$ são perfeitamente conhecidos para todo o domínio de Z , D_Z .

⁹Na expressão a seguir, estamos assumindo que cada elemento de X depende de todos os elementos de Z . Essa suposição é feita somente por simplicidade notacional, uma vez que o problema de dependência funcional do qual estamos tratando aparece mesmo quando X depende apenas de um elemento de Z . Note-se também que estamos supondo que não existem outras variáveis que afetam o valor de X além daquelas contidas no vetor Z .

¹⁰Cabe observar que, devido à dependência funcional de X em relação a Z , o domínio da função m , Ω_W , passa a ser dependente dos valores produzidos pela função f .

¹¹É importante assinalar que, como $E[Y|X, Z]$ é simplesmente a projeção de Y sobre (X, Z) , esse objeto está bem definido mesmo quando $X = Z$ ou $X = f(Z)$. Observe-se, no entanto, que essa projeção como função de X e Z tem múltiplas representações. Por exemplo, se $m(X, Z)$ é uma representação, então, para qualquer escalar λ , $h(X, Z) = m(X, Z) + \lambda[X - f(Z)]$ também representa a mesma projeção. Isso se deve ao fato de que, para todos os pontos no domínio de m , tem-se que $X - f(Z) = 0$. Cabe notar que a possibilidade de representar essa projeção de múltiplas formas não constitui uma limitação. Ao contrário, é exatamente essa possibilidade que cria o problema de identificação que queremos estudar.

¹²Embora possamos escrever o nosso modelo em “forma reduzida” (i.e. $E[Y|X, Z] = E[Y|f(Z), Z] = E[Y|Z]$), isso não impede que o efeito parcial de X permaneça sendo um objeto de interesse. Ademais, é importante observar que só estamos interessados nos casos em que as variáveis em X e Z têm “naturezas” distintas. Por exemplo, não estamos estudando situações em que $Z = idade$ e $X = idade^2$.

Não restringimos as variáveis em $W = (X, Z)$ a serem contínuas, nem as funções m e f a serem diferenciáveis. Nesse caso, as derivadas parciais de m e f em relação aos seus argumentos não estão necessariamente definidas. Assim, para estudar os efeitos de variações em X e Z sobre α , trabalhamos com o conceito de *diferenças parciais*, isto é, diferenças finitas em m que são obtidas quando se varia o valor de um dos argumentos dessa função mantendo-se os valores dos demais argumentos constantes. Mais especificamente, seja $X_{-k} = (X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_K)$ e $Z_{-l} = (Z_1, \dots, Z_{l-1}, Z_{l+1}, \dots, Z_L)$. Utilizamos as seguintes definições.

Definição 3.1 *O efeito de Z_l sobre α é definido pela diferença parcial*

$$\Delta_{Z_l}(x, z_l'', z_l', z_{-l}) = m(x, z_l'', z_{-l}) - m(x, z_l', z_{-l}), \quad (3.3)$$

onde z_l'' e z_l' são dois valores distintos assumidos por Z_l , e x e z_{-l} são valores particulares assumidos por X e Z_{-l} , respectivamente.

Definição 3.2 *O efeito de X_k sobre α é definido pela diferença parcial*

$$\Delta_{X_k}(x_k'', x_k', x_{-k}, z) = m(x_k'', x_{-k}, z) - m(x_k', x_{-k}, z), \quad (3.4)$$

onde x_k'' e x_k' são dois valores distintos assumidos por X_k , e x_{-k} e z são valores particulares assumidos por X_{-k} e Z , respectivamente.

3.3.2 Resultados de Identificação

Nesta sub-seção, discutimos a identificação dos dois tipos de efeitos definidos na sub-seção anterior. Nosso objetivo é investigar que aspectos desses dois efeitos podem ser identificados de forma não-paramétrica. Para tanto, propomos um conjunto de condições suficientes que são capazes de gerar a identificação desses aspectos.

3.3.2.1 Identificação do Efeito de Z_l

Consideramos primeiramente a identificação de Δ_{Z_l} . Suponha que o analista está interessado em identificar Δ_{Z_l} para os pontos $(X = x', Z_l = z_l'', Z_{-l} = z_{-l}')$ e $(X = x', Z_l = z_l', Z_{-l} = z_{-l}')$, isto é, suponha que se busque a identificação de $\Delta_{Z_l}(x', z_l'', z_l', z_{-l}')$. Utilizamos o seguinte conjunto de condições.

C1: Existe um sub-conjunto não-vazio no domínio de Z , $D_Z^* \subseteq D_Z$, no qual se pode fixar o valor de Z_{-l} quando Z_l muda de valor. Assume-se que $(z_l', z_{-l}') \in D_Z^*$ e $(z_l'', z_{-l}') \in D_Z^*$, com $z_l' \neq z_l''$.

Sejam $x' = f(z_l', z_{-l}')$ e $x'' = f(z_l'', z_{-l}')$, onde (z_l', z_{-l}') e (z_l'', z_{-l}') foram definidos na condição C1.

C2: A função f não é estritamente monotônica em D_Z^* (como definido em C1). Assume-se que $x' = f(z'_l, z'_{-l}) = f(z''_l, z'_{-l}) = x''$, com $(z'_l, z'_{-l}) \in D_Z^*$ e $(z''_l, z'_{-l}) \in D_Z^*$.

Proposição 3.1 *Para o modelo especificado pelas equações (3.1) e (3.2), se as condições (C1) e (C2) forem satisfeitas, então o efeito de Z_l sobre α , Δ_{Z_l} , é identificado.*

Demonstração. A prova é trivial. A condição C1 permite que se possa observar dos dados o seguinte objeto:

$$\begin{aligned}\Delta &= E[Y|X = x'', Z_l = z''_l, Z_{-l} = z'_{-l}] - E[Y|X = x', Z_l = z'_l, Z_{-l} = z'_{-l}] \\ &= m(x'', z''_l, z'_{-l}) - m(x', z'_l, z'_{-l}),\end{aligned}$$

onde $x'' = f(z''_l, z'_{-l})$ e $x' = f(z'_l, z'_{-l})$. Utilizando agora a condição C2, temos que:

$$\Delta = m(x', z''_l, z'_{-l}) - m(x', z'_l, z'_{-l}) = \Delta_{Z_l}(x', z''_l, z'_l, z'_{-l}).$$

■

A condição C1 é uma restrição de natureza local, uma vez que ela só requer que Z_{-l} permaneça fixo quando Z_l assume certos valores particulares. No entanto, como não existe dependência funcional entre Z_l e Z_{-l} , é possível que haja situações em que o valor de Z_l possa ser variado de forma independente do valor assumido por Z_{-l} . Nesse caso, a condição C1 torna-se dispensável, pois sempre seria factível fixar o valor de Z_{-l} (por exemplo, em $Z_{-l} = z'_{-l}$) para todos os valores assumidos por Z_l (por exemplo, $Z_l = z'_l$ e $Z_l = z''_l$).¹³

A restrição C2 representa uma condição de não-monotonicidade estrita, no sentido de que ela requer valores iguais para função f quando seus argumentos mudam de valor. Tal como a condição C1, o requerimento exigido pela condição C2 também tem natureza local: esses valores iguais de f podem ocorrer somente para certos valores particulares de $Z = (Z_l, Z_{-l})$. Caso existam “áreas planas” geradas por f – isto é, se para diversos valores de $Z = (Z_l, Z_{-l})$ a função f produz os mesmos valores de X –, então Δ_{Z_l} pode ser identificado para todos os valores de $Z = (Z_l, Z_{-l})$ que geram essas áreas (assumindo-se que $Z = (Z_l, Z_{-l}) \in D_Z^*$). Em outras palavras, caso f seja fracamente monotônica, é possível identificar Δ_{Z_l} para um conjunto mais amplo de pontos.

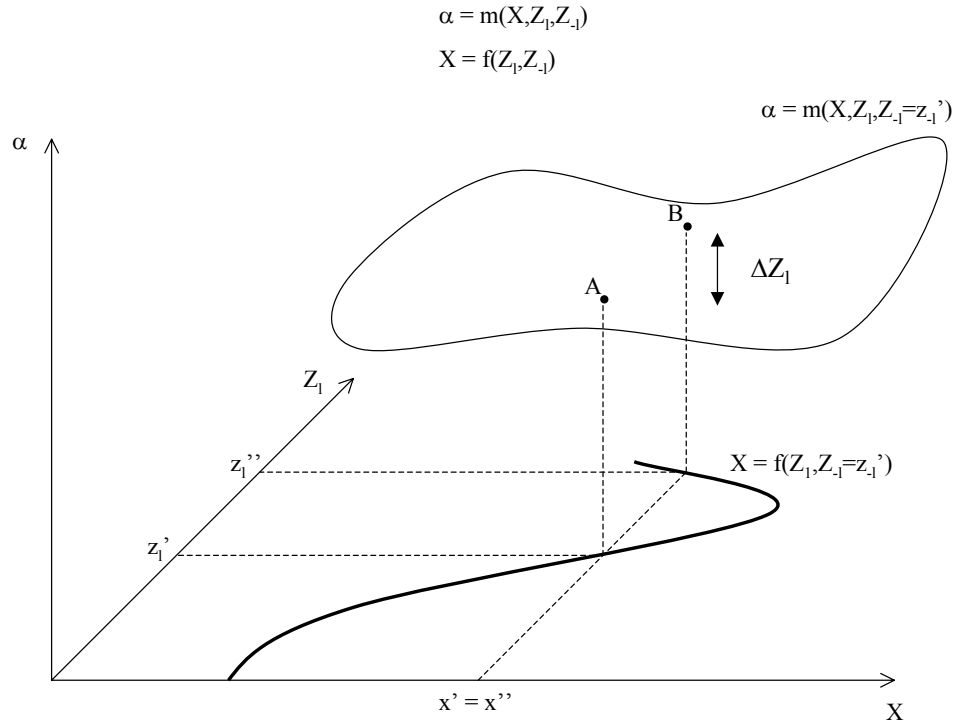
Uma dificuldade em satisfazer a condição C2 pode ocorrer quando a dimensão do vetor X for razoavelmente grande. De fato, como a condição C2 requer que *cada* variável em X assumo o mesmo valor quando Z_l varia, quanto maior K , mais dificilmente essa condição tenderá a ser atendida.

A Figura 3.1 procura ilustrar a identificação de Δ_{Z_l} para o caso em que $K = 1$. Por simplicidade, assumimos na Figura que podemos fixar o valor de Z_{-l} em z'_{-l} quando Z_l varia ao longo de todo o seu domínio. Note que essa suposição garante que a condição

¹³Obviamente, caso $L = 1$, a condição C1 também pode ser dispensada.

C1 é satisfeita. A média condicional de Y é medida no eixo vertical, enquanto os dois demais eixos medem as variáveis X e Z_l , assumidas escalares. A superfície apresentada na Figura representa a função m quando $Z_{-l} = z'_{-l}$, e a curva que aparece no plano $X \times Z_l$ descreve a função f para o mesmo valor fixo $Z_{-l} = z'_{-l}$. Observe que $x' = f(z'_l, z'_{-l}) = f(z''_l, z'_{-l}) = x''$, o que implica que a condição C2 é atendida nesses pontos. Portanto, a distância vertical entre os pontos A e B – que correspondem respectivamente aos objetos observados $E[Y|X = x', Z_l = z'_l, Z_{-l} = z'_{-l}]$ e $E[Y|X = x'', Z_l = z''_l, Z_{-l} = z'_{-l}]$ – identifica $\Delta_{Z_l}(x', z'_l, z'_l, z'_{-l})$.

Figura 3.1: Funções Hipotéticas m e f : Identificação de Δ_{Z_l}



O resultado importante desta sub-seção é que as condições C1 e C2 são suficientes para identificar Δ_{Z_l} de forma *totalmente* não-paramétrica. Isso mostra que o efeito de interesse pode ser estimado sem a necessidade de impor formas funcionais ao modelo. Entretanto, como as condições C1 e C2 têm natureza local, é possível que esse efeito não possa ser estimado para todos os pontos de interesse. Neste caso, se restrições paramétricas forem impostas ao modelo, então os valores locais do efeito de interesse podem se tornar o valor global desse efeito (por exemplo, o parâmetro associado a Z_l num modelo linear paramétrico). Observe-se, no entanto, que, embora a opção pela parametrização do mo-

delo possibilite a identificação global do efeito de interesse, ela não é a fonte que permite identificá-lo.

3.3.2.2 Identificação do Efeito de X_k

Consideramos agora a identificação de Δ_{X_k} . Suponha que o analista está interessado em identificar Δ_{X_k} nos pontos $(X_k = x''_k, X_{-k} = x'_{-k}, Z_l = z'_l, Z_{-l} = z'_{-l})$ e $(X_k = x'_k, X_{-k} = x'_{-k}, Z_l = z'_l, Z_{-l} = z'_{-l})$.¹⁴ Utilizamos o seguinte conjunto de condições.

C3: Existe um sub-conjunto não-vazio no domínio de Z , $D_Z^{**} \subset D_Z$, no qual existem os mesmos valores distintos de Z_l quando Z_{-l} assume diferentes valores. Assume-se que os pares $\{(z'_l, z^*_{-l}), (z''_l, z^*_{-l})\} \in D_Z^{**}$ e $\{(z'_l, z'_{-l}), (z''_l, z'_{-l})\} \in D_Z^{**}$, com $z'_l \neq z''_l$ e $z^*_{-l} \neq z'_{-l}$.

Sejam $x^* = f(z'_l, z^*_{-l})$, $x'' = f(z''_l, z^*_{-l})$, $x' = f(z'_l, z'_{-l})$ e $x'' = f(z''_l, z'_{-l})$, onde (z'_l, z^*_{-l}) , (z''_l, z^*_{-l}) , (z'_l, z'_{-l}) e (z''_l, z'_{-l}) foram definidos na condição C3.

C4: Existe um sub-conjunto não-vazio de valores que X pode assumir, $D_X^* \subset D_X$, no qual existe pelo menos um valor comum de X . Assume-se que $x^* = f(z'_l, z^*_{-l}) = f(z''_l, z^*_{-l}) = x''$, com $(z'_l, z^*_{-l}) \in D_Z^{**}$ e $(z''_l, z^*_{-l}) \in D_Z^{**}$.

C5: Existe um sub-conjunto não-vazio de valores que X pode assumir, $D_X^{**} \subset D_X$, no qual existem valores comuns de X_{-k} . Assume-se que $x'_{-k} = f_{-k}(z'_l, z'_{-l}) = f_{-k}(z''_l, z'_{-l}) = x''_{-k}$, com $(z'_l, z'_{-l}) \in D_Z^{**}$ e $(z''_l, z'_{-l}) \in D_Z^{**}$.

C6: No sub-conjunto D_X^{**} definido em C5, existem valores distintos de X_k . Assume-se que $x'_k = f_k(z'_l, z'_{-l}) \neq f_k(z''_l, z'_{-l}) = x''_k$, com $(z'_l, z'_{-l}) \in D_Z^{**}$ e $(z''_l, z'_{-l}) \in D_Z^{**}$.

C7: Assume-se que

$$m(X, Z) = m_1(X_k, X_{-k}, Z_{-l}) + m_2(Z_l),$$

onde m_1 e m_2 são funções não-paramétricas.

Proposição 3.2 *Para o modelo especificado pelas equações (3.1) e (3.2), se as condições (C3), (C4), (C5), (C6) e (C7) forem satisfeitas, então o efeito de X_k sobre α , Δ_{X_k} , é identificado.*

Demonstração. A condição C3 permite que se observe dos dados os seguintes objetos:

(a)

$$\begin{aligned} \Delta &= E[Y|X_k = x''_k, X_{-k} = x''_{-k}, Z_l = z''_l, Z_{-l} = z'_{-l}] - \\ &\quad E[Y|X_k = x'_k, X_{-k} = x'_{-k}, Z_l = z'_l, Z_{-l} = z'_{-l}] \\ &= m(x''_k, x''_{-k}, z''_l, z'_{-l}) - m(x'_k, x'_{-k}, z'_l, z'_{-l}), \end{aligned}$$

onde $x'_k = f_k(z'_l, z'_{-l})$, $x'_{-k} = f_{-k}(z'_l, z'_{-l})$, $x''_k = f_k(z''_l, z'_{-l})$ e $x''_{-k} = f_{-k}(z''_l, z'_{-l})$; e

¹⁴Esses pontos não são necessariamente os mesmos que os definidos na sub-seção anterior.

(b)

$$\begin{aligned}\Delta^* &= E[Y|X_k = x_k^{*''}, X_{-k} = x_{-k}^{*''}, Z_l = z_l'', Z_{-l} = z_{-l}^*] - \\ &\quad E[Y|X_k = x_k^{*'}, X_{-k} = x_{-k}^{*'}, Z_l = z_l', Z_{-l} = z_{-l}^*] \\ &= m(x_k^{*''}, x_{-k}^{*''}, z_l'', z_{-l}^*) - m(x_k^{*'}, x_{-k}^{*'}, z_l', z_{-l}^*),\end{aligned}$$

onde $x_k^{*'} = f_k(z_l', z_{-l}^*)$, $x_{-k}^{*'} = f_{-k}(z_l', z_{-l}^*)$, $x_k^{*''} = f_k(z_l'', z_{-l}^*)$ e $x_{-k}^{*''} = f_{-k}(z_l'', z_{-l}^*)$.

A identificação procede em dois passos. Primeiramente, utilizando Δ^* , temos que:

$$\begin{aligned}\Delta^* &= [m_1(x_k^{*''}, x_{-k}^{*''}, z_{-l}^*) - m_1(x_k^{*'}, x_{-k}^{*'}, z_{-l}^*)] + [m_2(z_l'') - m_2(z_l')] \\ &= [m_1(x_k^{*'}, x_{-k}^{*'}, z_{-l}^*) - m_1(x_k^{*'}, x_{-k}^{*'}, z_{-l}^*)] + [m_2(z_l'') - m_2(z_l')] \\ &= m_2(z_l'') - m_2(z_l'),\end{aligned}$$

onde a primeira igualdade decorre da imposição da condição C7, e a segunda segue da condição C4. Portanto, Δ^* identifica $m_2(z_l'') - m_2(z_l')$.

No segundo passo, utilizando Δ , temos que:

$$\begin{aligned}\Delta &= [m_1(x_k'', x_{-k}'', z_{-l}') - m_1(x_k', x_{-k}', z_{-l}')] + [m_2(z_l'') - m_2(z_l')] \\ &= [m_1(x_k'', x_{-k}'', z_{-l}') - m_1(x_k', x_{-k}', z_{-l}')] + [m_2(z_l'') - m_2(z_l')],\end{aligned}$$

onde a primeira igualdade decorre mais uma vez da imposição da condição C7, e a segunda segue da condição C5. Notando-se que o segundo termo em colchetes do lado direito dessa última expressão, $[m_2(z_l'') - m_2(z_l')]$, já foi identificado no primeiro passo por Δ^* , podemos escrever:

$$\begin{aligned}\Delta - [m_2(z_l'') - m_2(z_l')] &= \Delta - \Delta^* \\ &= m_1(x_k'', x_{-k}'', z_{-l}') - m_1(x_k', x_{-k}', z_{-l}'),\end{aligned}$$

que identifica $\Delta_{X_k}(x_k'', x_k', x_{-k}', z_{-l}')$ quando a condição C6 é satisfeita. ■

A condição C3 requer que os dados contenham os mesmos valores distintos de Z_l quando Z_{-l} assume diferentes valores. Especificamente, C3 demanda que o domínio de Z contenha pelo menos dois sub-conjuntos de pontos: $\{(z_l', z_{-l}'), (z_l'', z_{-l}')\}$ e $\{(z_l', z_{-l}^*), (z_l'', z_{-l}^*)\}$. Tal como a condição C1 da sub-seção anterior, a condição C3 também é uma restrição local. Todavia, uma vez que não há dependência funcional entre Z_l e Z_{-l} , é possível que se possa variar o valor de Z_l independentemente do valor assumido por Z_{-l} . Neste caso, a condição C3 pode ser dispensada.

A condição C4 exige que os dados possuam pelo menos um ponto cujo valor de cada elemento do vetor $X = (X_k, X_{-k})$ seja o mesmo quando Z_{-l} está fixo e Z_l muda de valor. Em particular, C4 requer que $X = f(Z_l, Z_{-l})$ assumam o mesmo valor para um dos sub-conjuntos de pontos associados com a condição C3, nomeadamente $\{(z_l', z_{-l}^*), (z_l'', z_{-l}^*)\}$.

Como na sub-seção anterior, caso a função f gere “áreas planas”, existirá um conjunto mais amplo de valores de X com o mesmo valor.

A condição C5 é menos restrita que a condição C4, na medida em que exige que somente o valor de X_{-k} seja o mesmo quando Z_{-l} está fixo e Z_l muda de valor. Especificamente, ela requer que $X_{-k} = f_{-k}(Z_l, Z_{-l})$ assumam o mesmo valor para o outro sub-conjunto de pontos associado com a condição C3, isto é: $\{(z'_l, z'_{-l}), (z''_l, z'_{-l})\}$. A condição C6 demanda que X_k assumam diferentes valores para esse mesmo sub-conjunto de pontos.

Caso $K = 1$, pode-se dispensar integralmente a condição C5, e a condição C4 passa a requerer que apenas o valor de X_k seja igual para o sub-conjunto de pontos a ela associado [i.e. $\{(z'_l, z^*_{-l}), (z''_l, z^*_{-l})\}$]. Contudo, se a magnitude de K for relativamente elevada, o atendimento das condições C4 e C5 tornar-se-á mais complicada. De fato, nesse caso, os dados devem apresentar pelo menos um ponto para o qual todo o vetor X assume o mesmo valor e um outro ponto para o qual cada argumento de X_{-k} exibe o mesmo valor. Como essa situação pode ser difícil de se encontrar na prática, os resultados de identificação aqui apresentados tendem a ser mais úteis para modelos em que a magnitude de K é relativamente baixa.

A restrição C7 é uma condição de separabilidade aditiva: ela impõe que um dos elementos de Z , Z_l , apareça aditivamente na função m .¹⁵ É importante assinalar que outros elementos de Z poderiam também aparecer de forma aditiva em m , mas isso seria uma imposição mais forte do que necessária.

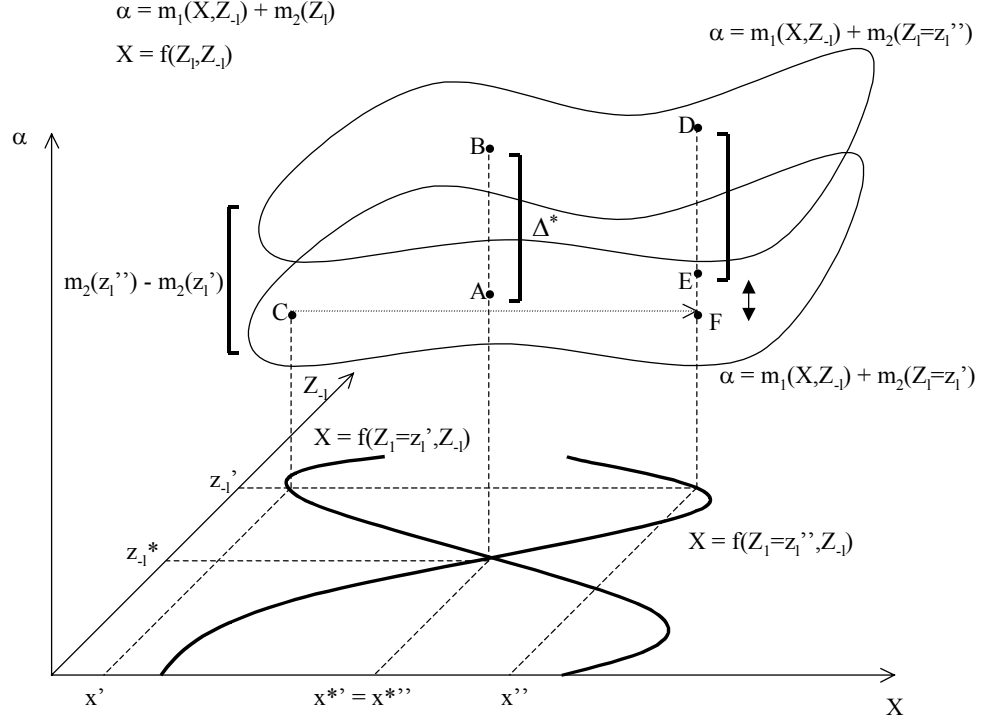
A Figura 3.2 procura ilustrar a identificação de Δ_{X_k} para $K = 1$ e $L = 2$. Para facilitar a visualização, assumimos que Z_l pode ser variado independentemente de Z_{-l} , o que garante que a condição C3 é atendida. A média condicional de Y é medida no eixo vertical, enquanto os dois demais eixos medem as variáveis X e Z_{-l} , assumidas escalares. As duas superfícies apresentadas na Figura representam a função m para dois valores distintos assumidos por Z_l : z'_l e z''_l . A condição de separabilidade C7 impõe que essas duas superfícies sejam paralelas, sendo a distância (fixa) entre elas igual a $m_2(z''_l) - m_2(z'_l)$. No plano $X \times Z_{-l}$, aparecem duas curvas correspondentes à função f quando Z_l assume os dois mesmos valores (i.e. z'_l e z''_l). Cabe observar que essas duas curvas se cruzam quando $Z_{-l} = z^*_{-l}$, implicando que X assume o mesmo valor para este ponto. Note, no entanto, que, quando $Z_{-l} = z'_{-l}$, X assume valores distintos.

O primeiro passo na demonstração da Proposição 3.2 está associado ao ponto onde as curvas no plano $X \times Z_{-l}$ se cruzam. Note que neste ponto: $Z_{-l} = z^*_{-l}$, $X = x^* = x^{**}$, mas Z_l assume os valores z'_l e z''_l . Isso implica que a condição C4 é satisfeita nesse cruzamento das curvas, o qual está associado aos pontos A e B na Figura. Assim, por causa da condição de separabilidade C7, o comprimento do segmento entre A e B mede a distância fixa entre as duas superfícies, especificamente $\Delta^* = m_2(z''_l) - m_2(z'_l)$.

Note agora que, quando $Z_{-l} = z'_{-l}$, $x' = f(z'_l, z'_{-l}) \neq f(z''_l, z'_{-l}) = x''$, o que implica

¹⁵Cabe notar que esse Z_l não é necessariamente o mesmo que o da sub-seção anterior.

Figura 3.2: Funções Hipotéticas m e f : Identificação de Δ_{X_k}



que a condição C6 é satisfeita neste ponto.¹⁶ Os valores (x', z_l', z_{-l}') e (x'', z_l'', z_{-l}'') estão associados aos pontos C e D na Figura, respectivamente. A distância vertical entre esses dois pontos corresponde ao valor de Δ na demonstração da Proposição 3.2. Utilizando a condição de separabilidade C7 e o fato de que $\Delta^* = m_2(z_l'') - m_2(z_l')$, se subtrairmos a distância vertical entre C e D da distância entre A e B (que é igual a distância entre os pontos D e E por força da condição C7), obtemos a distância entre E e F (este último correspondendo ao ponto C). Em outras palavras, a diferença $\Delta - \Delta^*$ resulta na distância entre os pontos E e F, a qual identifica Δ_{X_k} no ponto de interesse.

Devido à dependência funcional entre X e Z , não se pode variar X_k mantendo-se o valor de Z fixo. Assim, para identificar o efeito de X_k , torna-se necessário impor restrições ao modelo. Uma dessas restrições é a condição de separabilidade aditiva C7, a qual requer que uma das variáveis em Z , Z_l , não interaja com os demais argumentos da função m . Contudo, isoladamente, a condição de separabilidade não é suficiente para gerar a identificação do efeito de interesse, o que se deve ao fato de que o termo aditivo $m_2(Z_l)$ muda de valor nos pontos $x' = f(z_l', z_{-l}') \neq f(z_l'', z_{-l}'') = x''$. Em decorrência disso, uma outra restrição foi

¹⁶Assinale-se que, como $K = 1$, a condição C5 não desempenha nenhum papel na Figura.

introduzida no modelo. Essa restrição é desempenhada pela condição C4, a qual permite identificar a mudança no valor do termo aditivo, especificamente: $m_2(z_l'') - m_2(z_l')$. É importante observar que a condição C4 só é satisfeita nos pontos em que $X = (X_k, X_{-k})$ assume o mesmo valor. Entretanto, nesses pontos, a condição C6 não pode ser atendida, implicando que Δ_{X_k} não pode ser identificado quando a condição C4 é satisfeita.

Essas observações mostram que somente alguns aspectos de Δ_{X_k} são identificados não-parametricamente. Primeiramente, ao impedir a identificação totalmente não-paramétrica de Δ_{X_k} , a condição de separabilidade C7 impossibilita a exploração de interações potencialmente interessantes entre Z_l e os demais argumentos de m (em particular X_k). Em segundo lugar, como as condições C4 e C6 não podem ser satisfeitas simultaneamente, a identificação de Δ_{X_k} só pode ocorrer nos pontos nos quais a primeira dessas condições não é atendida. Como resultado, o efeito de interesse não pode ser identificado em qualquer ponto que se deseje.

Se o verdadeiro modelo não apresenta separabilidade aditiva, então, não estaremos identificando Δ_{X_k} . Para ver isso, vale re-lembrar que, na demonstração da Proposição 3.2, a identificação de Δ_{X_k} foi obtida tirando-se a diferença entre dois objetos observados: $\Delta - \Delta^*$. Se a separabilidade aditiva não for uma característica do processo gerador dos dados, teremos que:¹⁷

$$\begin{aligned} \Delta - \Delta^* &= \{m(x_k'', x_{-k}', z_l'', z_{-l}') - m(x_k', x_{-k}', z_l', z_{-l}')\} \\ &\quad - \{m(x_k^*, x_{-k}', z_l'', z_{-l}^*) - m(x_k^*, x_{-k}', z_l', z_{-l}^*)\}, \end{aligned}$$

o qual é um objeto que não identificará Δ_{X_k} , mas sim uma combinação de diferentes efeitos: o primeiro termo entre chaves capta o efeito de mudar os valores de X_k entre (x_k', x_k'') e de Z_l entre (z_l', z_l'') , e o segundo termo entre chaves mede o efeito de variar somente o valor de Z_l entre z_l' e z_l'' . Assim, claramente, não estaremos identificando o efeito de interesse.

É importante salientar que as restrições C3-C7 são condições locais: elas são capazes de produzir a identificação de (alguns aspectos de) Δ_{X_k} para valores específicos assumidos pelos argumentos da função m . Se o analista não está interessado em identificação local, restrições paramétricas podem ser impostas para permitir a identificação de características globais do modelo. Deve-se notar, entretanto, que a identificação de Δ_{X_k} não requer a adoção de um modelo completamente aditivo e parametrizado. De fato, como observamos anteriormente, o que a condição de separabilidade C7 demanda é que somente um elemento de Z apareça de forma aditiva no modelo.

Vale frisar que as restrições que aqui propomos são condições suficientes para identificar o efeito de interesse. Embora utilizando um arcabouço diferente do nosso, Chesher (2007) propõe outras condições de identificação.¹⁸ Em linhas gerais, a principal condição

¹⁷Na próxima expressão, assumimos que as condições C3-C6 são satisfeitas.

¹⁸O arcabouço de Chesher (2007) incorpora componentes não-observáveis que são não-aditivos e potencialmente correlacionados às covariadas do modelo.

apresentada por esse autor requer que exista um sub-conjunto no domínio de Z em que as variações no valor desse vetor não alterem *diretamente* a variável de resultado, porém sejam capazes de modificar o valor de X (assumindo $K = 1$). Adaptada para o nosso contexto, essa condição impõe que a função m só mude de valor por meio das mudanças em $X = f(Z)$ oriundas de variações nos valores de Z que pertençam a esse sub-conjunto de valores.¹⁹ Observe-se que, com essa condição, é possível identificar (localmente) o efeito de X de forma totalmente não-paramétrica. No entanto, para isso, é preciso que o analista esteja preparado para delimitar *a priori* os valores de Z pertencentes a esse sub-conjunto.²⁰ Dado que a delimitação exata desse sub-conjunto pode ser difícil, se o analista estiver preparado para assumir algum tipo de separabilidade aditiva, então as condições aqui propostas podem ser utilizadas para identificar não-parametricamente alguns aspectos do efeito de interesse.

3.4 Um Exemplo Analítico

Existe um amplo número de estudos empíricos que utilizam bases de dados com observações de unidades de *cross-section* repetidas ao longo do tempo. Exemplos típicos dessas unidades são regiões, países ou setores de atividade, sendo a frequência temporal de observação mensal, trimestral ou anual. A esse tipo de base de dados, dá-se usualmente o nome de séries de tempo de *cross-sections*.

A maior parte dos trabalhos dessa literatura está interessada em estimar o efeito de uma covariada particular sobre um certa variável de resposta. Um exemplo são os diversos estudos que utilizam informações temporais de estados de um país para estimar o efeito do salário mínimo sobre o nível de emprego (*e.g.* Neumark & Wascher (1992), Williams (1993), Card & Krueger (1995) e Baker et al. (1999)). Outros exemplos incluem estudos de demanda por trabalho que fazem uso de séries de tempo de setores de atividade (ver resenha em Hamermesh (1993, chapter 3)), trabalhos sobre a taxa de criminalidade que utilizam dados repetidos de regiões de um país (*e.g.* Wolpin (1980)), e estudos na área de desenvolvimento econômico que empregam bases de dados com informações anuais de diferentes países (*e.g.* Dasgupta et al. (2001)).

Nessa literatura, as variáveis que entram no modelo apresentam variações tanto na dimensão temporal quanto na seccional. A equação a ser estimada normalmente inclui conjuntos separados de variáveis *dummies* para cada uma dessas dimensões. A equação

¹⁹Mais especificamente, seja $\Omega_Z \subset D_Z$ esse sub-conjunto de valores. Expressando $\alpha = m(X, Z) = m(f(Z), Z)$ e assumindo que $Z \in \Omega_Z$, a condição demanda que função m seja sensível à primeira aparição de Z em m (i.e. via $X = f(Z)$), mas insensível a sua segunda aparição.

²⁰Claramente, esse sub-conjunto não pode ser idêntico ao domínio de Z , pois, neste caso, a função m não conteria a segunda aparição Z .

típica encontrada nessa literatura pode ser expressa da seguinte forma:

$$Y_{ct} = \eta + \beta X_{1ct} + X_{2ct}\gamma + T_t\tau_t + S_c\delta_c + \epsilon_{ct}, \quad (3.5)$$

onde Y_{ct} representa a variável de resposta (*e.g.* nível de emprego) para a unidade de *cross-section* c (*e.g.* estado do país) no período t (*e.g.* mês). A variável X_{1ct} é a covariada (*e.g.* o salário mínimo) cujo efeito o analista está interessado em estimar, e X_{2ct} contém um conjunto de variáveis de controle (*e.g.* a taxa de desemprego). As variáveis T_t e S_c denotam as *dummies* para o período t e a unidade de *cross-section* c , respectivamente. A variável ϵ_{ct} representa um termo de distúrbio aleatório, para o qual se assume independência estatística em relação às covariadas do modelo. O vetor $(\eta, \beta, \gamma, \tau_t, \delta_c)$ denota os coeficientes a serem estimados, sendo β o parâmetro de interesse.

Seja $W = (X_1, X_2, T, S)$ a matriz que contém as covariadas apropriadamente arranjadas por unidades de *cross-section* e períodos de tempo. Dadas as hipóteses acima adotadas, sabe-se que os parâmetros do modelo são identificados se a matriz $(W'W)$ tiver posto completo. Esse requisito não é satisfeito se houver perfeita multi-colinearidade entre as covariadas do modelo, ou seja, se uma ou mais covariadas forem *linearmente dependentes* de outras. Nesse sentido, a existência de multi-colinearidade num modelo linear é análoga ao nosso problema de dependência funcional entre as covariadas num contexto não-paramétrico.

O modelo acima pode ser transladado de forma direta para o nosso arcabouço. Primeiramente, note que os valores de X_{1ct} e X_{2ct} são unicamente determinados quando c e t são especificados. Essa restrição cria um tipo de dependência funcional, que pode ser expressa por:

$$X_{1ct} = f_1(t, c) \quad (3.6)$$

$$X_{2ct} = f_2(t, c). \quad (3.7)$$

Essas expressões correspondem às equações representadas em (3.1), onde $X_{1ct} = X_k$, $X_{2ct} = X_{-k}$ e $Z = (t, c)$.

Em segundo lugar, utilizando a hipótese de que o termo de erro é independente das covariadas, podemos escrever:

$$\alpha_{ct} = E[Y_{ct}|X_{1ct}, X_{2ct}, t, c] = m(X_{1ct}, X_{2ct}, t, c). \quad (3.8)$$

Esta expressão corresponde à equação (3.2).

Pela Proposição 3.2, sabemos que, se as condições C3-C7 forem satisfeitas, alguns aspectos do efeito de X_{1ct} são identificados de forma não-paramétrica. Como discutido na seção 3.3.2.2, os resultados da Proposição 3.2 têm natureza local. Assim, suponhamos que o analista esteja buscando identificar o efeito de interesse para certos valores das

covariadas do modelo. Sejam esses valores dados por: $(x'_{1ct}, x'_{2ct}, t', c')$ e $(x''_{1ct}, x''_{2ct}, t', c')$. A seguir, analisamos o conteúdo das condições C3-C7 utilizando um exemplo em que Y_{ct} é a proporção de ocupados e X_{1ct} e X_{2ct} representam, respectivamente, o salário mínimo (real) e a taxa de desemprego no estado c no mês t .

Iniciando pela condição C7, esta requer que somente um dos elementos de Z não interaja com os demais argumentos da função m . Suponhamos que este elemento seja $Z_l = t$, o que implica que $Z_{-l} = c$ e, portanto, que: $m(X_{1ct}, X_{2ct}, t, c) = m_1(X_{1ct}, X_{2ct}, c) + m_2(t)$. Note que a imposição da separabilidade em t impossibilita que o efeito de interesse seja identificado não-parametricamente nessa dimensão.²¹

A condição C3 demanda que os dados contenham os mesmos valores de t quando c assume diferentes valores. Dado que o analista está interessado na identificação do efeito do salário mínimo para o estado c' no mês t' , é necessário existam informações sobre esse e outros estados para o mesmo conjunto de meses (incluindo t'). Essa exigência não tende a ser muito severa no presente contexto, uma vez que ela costuma ser satisfeita para a maior parte das bases de dados com formato de *cross-sections* repetidas. Assim, sejam $\{c', c^*, c_3, c_4, \dots, C\}$ e $\{t', t'', t_3, t_4, \dots, T\}$ os conjuntos de estados e meses presentes nos dados, respectivamente. Assumiremos que todos os estados são observados em todos os meses

A condição C4 requer que exista pelo menos um ponto nos dados para o qual o vetor $X_{ct} = (X_{1ct}, X_{2ct})$ assume o mesmo valor. Suponhamos que isso ocorra no estado c^* entre os meses t' e t'' . Especificamente, isso significa que tanto o salário mínimo quanto a taxa de desemprego no estado c^* assumem o mesmo valor entre esses dois meses, ou seja: $x'_{1ct} = f_1(c^*, t') = f_1(c^*, t'') = x''_{1ct}$ e $x'_{2ct} = f_2(c^*, t') = f_2(c^*, t'') = x''_{2ct}$. Uma vez que o analista conhece o valor da média condicional de Y para todos os pontos disponíveis nos dados, a aplicação das condições C4 e C7 garante que:²²

$$\begin{aligned} \Delta^* &= E[Y|X_{1ct} = x'_{1ct}, X_{2ct} = x'_{2ct}, t = t'', c = c^*] - \\ &\quad E[Y|X_{1ct} = x'_{1ct}, X_{2ct} = x'_{2ct}, t = t', c = c^*] \\ &= m_2(t'') - m_2(t'). \end{aligned}$$

Portanto, a simples subtração das médias condicionais de Y nesses pontos identifica a diferença da função m_2 entre os meses t'' e t' .

A condição C5 demanda que o valor de X_{2ct} seja igual no estado em que o analista deseja identificar o efeito de interesse para o mesmo par de meses associado com a condição C4. No nosso exemplo, isso significa que a taxa de desemprego tem que ser a mesma no estado c' entre os meses t' e t'' , ou seja: $x'_{2ct} = f_2(c', t') = f_2(c', t'') = x''_{2ct}$. Já condição C6

²¹Nossa escolha por $(Z_l, Z_{-l}) = (t, c)$ é arbitrária. *Mutatis mutandis*, a análise de identificação seria análoga se o modelo fosse separável em c .

²²Essa expressão pode ser obtida com base no primeiro passo da demonstração da Proposição 3.2.

requer que o valor de X_{1ct} seja diferente para esses mesmos estado e par de meses, isto é: $x'_{1ct} = f_1(c', t') \neq f_1(c', t'') = x''_{1ct}$. A aplicação das condições C5, C6 e C7 nesses pontos leva a que:

$$\begin{aligned}\Delta &= E[Y|X_{1ct} = x''_{1ct}, X_{2ct} = x''_{2ct}, t = t'', c = c'] - \\ &\quad E[Y|X_{1ct} = x'_{1ct}, X_{2ct} = x'_{2ct}, t = t', c = c'] \\ &= m(x''_{1ct}, x'_{2ct}, t'', c') - m(x'_{1ct}, x'_{2ct}, t', c') \\ &= [m_1(x''_{1ct}, x'_{2ct}, c') - m_1(x'_{1ct}, x'_{2ct}, c')] + [m_2(t'') - m_2(t')],\end{aligned}$$

sendo a última igualdade decorrente da condição C7. Note que, caso essa condição não fosse imposta, ficaríamos somente com a segunda igualdade, a qual não identifica o efeito de X_{1ct} , já que t também assume valores diferentes. A condição C7 gera, portanto, a presença do termo $[m_2(t'') - m_2(t')]$, o qual foi identificado por Δ^* , ou seja, pela diferença das médias condicionais de Y no estado c^* entre t'' e t' . Assim, tomando a diferença entre Δ e Δ^* , obtemos:

$$\begin{aligned}\Delta - \Delta^* &= \{E[Y|X_{1ct} = x''_{1ct}, X_{2ct} = x''_{2ct}, t = t'', c = c'] - \\ &\quad E[Y|X_{1ct} = x'_{1ct}, X_{2ct} = x'_{2ct}, t = t', c = c']\} \\ &\quad - \{E[Y|X_{1ct} = x''_{1ct}, X_{2ct} = x''_{2ct}, t = t'', c = c^*] - \\ &\quad E[Y|X_{1ct} = x'_{1ct}, X_{2ct} = x'_{2ct}, t = t', c = c^*]\} \\ &= m_1(x''_{1ct}, x'_{2ct}, c') - m_1(x'_{1ct}, x'_{2ct}, c'),\end{aligned}\tag{3.9}$$

o qual é um objeto que identifica não-parametricamente alguns aspectos do efeito de interesse no estado c' , quando o salário mínimo varia entre os valores x'_{1ct} e x''_{1ct} , e a taxa de desemprego é igual a x'_{2ct} .

Cabe observar que a identificação do efeito do mínimo foi alcançada para valores particulares assumidos por X_{1ct} , X_{2ct} e c . No entanto, se existirem outros estados na base de dados cujos valores do salário mínimo foram distintos e cujas taxas de desemprego foram iguais entre t' e t'' , então, será possível identificar o efeito de interesse nesses estados para os valores assumidos por essas variáveis. Ademais, caso esses pares de valores do salário mínimo e da taxa de desemprego forem diferentes entre os estados, o analista poderá identificar o efeito do mínimo para um conjunto (potencialmente) amplo de estados, de valores do salário mínimo e da taxa de desemprego. É importante notar que isso não se aplica ao estado c^* , já que, por hipótese, o valor do salário mínimo não variou neste estado entre o par de meses em questão.

Esta última observação confirma o que discutimos na seção 3.3.2.2, a saber: não é possível satisfazer as condições C4 e C6 simultaneamente. Portanto, sob as condições aqui propostas, aparece uma limitação que impede à identificação do efeito

de interesse em qualquer ponto que se deseje. Mais especificamente, esse efeito não poderá ser identificado naqueles pontos em que a condição C4 for satisfeita.

Apesar de ser (potencialmente) possível identificar o efeito de interesse para distintos estados e diferentes valores do salário mínimo e da taxa de desemprego, a aplicação da condição de separabilidade C7 ao nosso exemplo torna esse efeito invariante com respeito a t . Nesse sentido, se o efeito do mínimo for de fato distinto para cada mês t' e t'' , essa condição não permite que se identifique esse efeito separadamente para cada um desses meses.²³ Note, no entanto, que, em contraste com o modelo totalmente aditivo e parametrizado da equação (3.5), a condição C7 demanda separabilidade somente de um argumento da função m . Como consequência, à exceção de t , o efeito de interesse pode conter interações completamente flexíveis entre as outras covariadas do modelo.

Em suma, o que esse exemplo procura mostrar é que é possível aplicar os resultados deste capítulo para identificar (localmente) alguns aspectos do efeito interesse de forma não-paramétrica. Isso representa uma vantagem, na medida em que o analista pode explorar algumas interações entre as covariadas sem impor formas funcionais demasiadamente restritas às equações do modelo.

3.5 Conclusões

Neste capítulo, estudamos a identificação de diferenças parciais de uma função não-paramétrica quando um sub-conjunto de seus argumentos são funcionalmente relacionados aos demais argumentos dessa função. Essas restrições funcionais impedem que se possa variar livremente o valor do primeiro sub-conjunto de variáveis, criando, assim, dificuldades para se alcançar a identificação totalmente não-paramétrica das diferenças parciais relativas a essas variáveis. Em modelos lineares totalmente parametrizados, essas dificuldades são equivalentes ao conhecido problema de perfeita multi-colinearidade.

Utilizamos diferenças parciais para definir os efeitos de interesse, uma vez que não assumimos que a função não-paramétrica é diferenciável, ou que seus argumentos são variáveis contínuas. O principal benefício de usar esse tipo de objeto é que os resultados de identificação aqui obtidos podem ser aplicados para situações em que as covariadas apresentam variações discretas.

Nosso principal objetivo foi investigar que aspectos das diferenças parciais de interesse podem ser identificados não-parametricamente. A busca pela identificação

²³Em outras palavras, a condição de separabilidade impõe que esse efeito seja o mesmo tanto para t' quanto para t'' .

não-paramétrica é útil por pelo menos duas razões. Em primeiro lugar, como a teoria econômica normalmente não provê formas funcionais específicas para as equações de um modelo, o arcabouço não-paramétrico é claramente o mais adequado para se iniciar a análise de identificação. Em segundo lugar, estudar identificação sob a perspectiva não-paramétrica tende a revelar de forma mais transparente as forças que estão por trás da identificação das características de interesse do modelo. Nesse sentido, e como apontado por Roehrig (1988), julgar a identificação dessas características a partir da imposição *a priori* de restrições paramétricas pode dar respostas enganosas.

Propusemos um conjunto de condições suficientes para gerar a identificação não-paramétrica de alguns aspectos do efeito de variações no valor de uma variável funcionalmente dependente do modelo. Uma dessas condições foi uma restrição de separabilidade, a qual requer que somente *um* dos argumentos apareça aditivamente na função (inicialmente) não-paramétrica. Embora impeça a exploração de interações potencialmente importantes entre o argumento separado e os demais argumentos dessa função, essa condição traz, claramente, uma vantagem em relação aos modelos *totalmente* aditivos que encontramos na maior parte dos estudos empíricos em economia. Uma outra condição utilizada exige que cada variável funcionalmente dependente assuma o *mesmo* valor pelo menos uma vez. Como o efeito de interesse só faz sentido quando a variável funcionalmente dependente assume *diferentes* valores, essa condição impossibilita que esse efeito seja identificado em qualquer ponto que se deseje.

Propusemos também condições para identificar o efeito associado a mudanças no valor de uma variável não-funcionalmente dependente no modelo. Essas condições são capazes de produzir a identificação totalmente não-paramétrica desse efeito. A principal condição requer que as variáveis funcionalmente dependentes assumam os mesmos valores nos pontos onde se busca a identificação desse efeito.

Todos os resultados de identificação aqui apresentados são locais no sentido de que os efeitos de interesse são identificados para certos valores assumidos pelas covariadas do modelo. Embora esses efeitos possam ser potencialmente identificados para um conjunto amplo de valores, caso não se esteja interessado somente nesse conjunto de valores, pode-se optar por impor restrições paramétricas ao modelo para permitir a identificação global desses efeitos. Entretanto, é importante assinalar que essa opção está associada com a identificação global dos efeitos de interesse, mas não representa a verdadeira fonte que permite identificá-los.

Acreditamos que os resultados deste capítulo podem ser úteis para diversos tipos de aplicações. O capítulo anterior contém um exemplo empírico para a área de

avaliação de programas sociais. No presente capítulo, apresentamos um exemplo analítico que procura ilustrar como esses resultados podem ser utilizados para ampla literatura que faz uso de dados de *cross-sections* repetidas.

Conclusões e Pesquisas Futuras

Apresentamos aqui as principais conclusões da tese e algumas possíveis direções para pesquisas futuras.

No primeiro capítulo, utilizamos um painel de dados combinados de empregadores e empregados (Relação Anual de Informações Sociais - RAIS) para estimar o efeito da segregação feminina no nível do estabelecimento sobre os salários de homens e mulheres. Esse efeito foi estimado com base em vários modelos longitudinais, sendo o modelo auto-regressivo de primeira ordem (AR(1)) o mais geral. Como todos os demais estudos da literatura relacionada utilizam dados de *cross-section*, um dos principais objetivos do primeiro capítulo foi checar se as estimativas obtidas com esse tipo de dados são alteradas quando se incluem efeitos específicos não-observáveis de trabalhadores (*e.g.* habilidade inata e preferências) e estabelecimentos (*e.g.* eficiência gerencial e grau de discriminação).

Nossas estimativas longitudinais indicam que o efeito de interesse é negativo, o que está em linha com os resultados que obtemos com uma análise de *cross-section*. Todavia, embora ambos os tipos de estimativas apresentem o mesmo sinal, os resultados longitudinais são menores em módulo que os de *cross-section*. Essa comparação de estimativas leva a duas conclusões. A primeira é que a inserção no modelo dos efeitos específicos não-observáveis de trabalhadores e estabelecimentos realmente importa para a estimar corretamente o efeito de interesse. Em outras palavras, as estimativas de *cross-section* tenderão a ser enviesadas sempre que não se levar em consideração que os fatores não-observáveis são correlacionados com a composição por gênero dos estabelecimentos. A segunda conclusão é que o grau de segregação feminina nos estabelecimentos é negativamente correlacionado com esses fatores específicos não-observáveis. Algumas possíveis explicações para essa correlação negativa são que: (a) os empregadores com níveis mais elevados de produtividade e/ou de discriminação tendem a contratar relativamente mais homens do que mulheres; e, (b) os trabalhadores de ambos os sexos com menor produtividade não-observada tendem a ser alocados em estabelecimentos predominantemente

femininos.

Embora bastante rica em termos das informações que contém, a base de dados que utilizamos cobre somente três anos (2000-2002). Por ser um painel relativamente curto, não pudemos empregar diferentes especificações para o modelo dinâmico (por exemplo, uma especificação AR(2)). Na realidade, alguns testes de ‘fatores comuns’ que implementamos indicam que a hipótese de um modelo AR(1) não parece adequada. Assim, uma possibilidade de pesquisa no futuro é obter um painel de dados mais longo com informações combinadas de empregadores e empregados.

Uma das principais contribuições do segundo capítulo da tese foi procurar distinguir de forma geral e sistemática dois efeitos relevantes para a área de avaliação de programas sociais. O primeiro efeito procura captar as variações temporais no verdadeiro impacto médio dos programas, ao passo que o segundo se refere a influência que as condições econômicas externas aos programas podem ter sobre esse impacto médio. Denominamos o primeiro efeito de *efeito interno temporal* e o segundo de *efeito econômico externo*.

O principal objetivo do segundo capítulo foi analisar que aspectos desses dois efeitos podem ser identificados de forma não-paramétrica. Essa análise foi operacionalizada para um contexto não-paramétrico, já que não dispomos de conhecimento prévio sobre como os impactos médios dos programas são determinados. A principal vantagem do uso dessa abordagem é que evitamos impor *a priori* formas funcionais específicas ao modelo, as quais podem induzir a uma avaliação equivocada da identificação dos verdadeiros efeitos de interesse.

A análise de identificação foi implementada para duas situações empíricas distintas: na primeira, supomos que o analista possui informações para múltiplas regiões de um país e somente uma coorte de participantes do programa; na segunda, assumimos que o investigador dispõe de dados para múltiplas coortes de participantes e somente uma região do país. Para ambas as situações, assumimos que o analista conhece (ou foi capaz de estimar consistentemente) os impactos médios do programa ao longo do tempo. Supomos também que a escala do programa é relativamente pequena, o que permite abstrair de possíveis efeitos de equilíbrio geral.

Propusemos um conjunto de condições suficientes que é capaz de produzir a identificação totalmente não-paramétrica do efeito interno temporal. Já o efeito econômico externo só é identificado de forma parcialmente não-paramétrica. Isso se deve ao fato de que existe um tipo de dependência funcional no modelo que impede variações independentes nas covariadas que medem as condições econômicas externas ao programa.

O segundo capítulo também contém um exemplo empírico dos resultados de identificação que lá obtivemos. Utilizamos dados experimentais que foram coletados para avaliar a efetividade de um programa público de treinamento de mão-de-obra para uma coorte de participantes em duas regiões metropolitanas brasileiras. Como esse exemplo ilustra os resultados de identificação correspondentes ao primeiro tipo de situação empírica mencionada anteriormente, seria interessante obter dados de múltiplas coortes de participantes de um programa qualquer para estimar os dois efeitos de interesse para essa outra situação empírica.

No terceiro capítulo da tese, generalizamos os resultados de identificação do segundo capítulo para um contexto não-paramétrico geral que apresenta dependência funcional entre as covariadas do modelo. Num arcabouço não-paramétrico, essa dependência funcional é equivalente à existência de perfeita multi-colinearidade entre as covariadas de um modelo de regressão linear. Das condições que propusemos, demonstramos que o efeito de covariadas não funcionalmente dependentes são identificados de forma completamente não-paramétrica. Entretanto, somente alguns aspectos do efeito das covariadas funcionalmente dependentes são identificados não-parametricamente. Isso decorre da imposição de uma restrição de separabilidade que requer que apenas *uma* covariada apareça de forma aditiva no modelo. Esse requerimento mostra que não é necessário impor formas funcionais demasiadamente restritivas para identificar os efeitos de interesse.

Referências Bibliográficas

- Abowd, J. M. & Kramarz, F. (1999), The Analysis of Labor Markets Using Matched Employer-Employee Data, *in* O. Ashenfelter & D. Card, eds, 'Handbook of Labor Economics', Vol. 3B, North Holland, Amsterdam, pp. 2629–2710.
- Abowd, J. M., Kramarz, F. & Margolis, D. N. (1999), 'High Wage Workers and High Wage Firms', *Econometrica* **67**, 251–333.
- Aigner, D. J. & Cain, G. G. (1977), 'Statistical Theories of Discrimination in Labor Markets', *Industrial and Labor Relations Review* **30**, 175–187.
- Akerlof, G. A. & Dickens, W. T. (1982), 'The Economic Consequences of Cognitive Dissonance', *American Economic Review* **72**, 307–319.
- Altonji, J. G. & Blank, R. M. (1999), Race and Gender in the Labor Market, *in* O. Ashenfelter & D. Card, eds, 'Handbook of Labor Economics', Vol. 3C, North Holland, Amsterdam, pp. 3143–3259.
- Anker, R. (1997), 'Theories of Occupational Segregation by Sex: An Overview', *International Labour Review* **136**, 315–339.
- Arellano, M. & Bond, S. R. (1991), 'Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations', *Review of Economic Studies* **58**, 277–297.
- Arellano, M. & Bover, O. (1995), 'Another Look at the Instrumental Variable Estimation of Error-Components Models', *Journal of Econometrics* **68**, 29–51.
- Arrow, K. (1973a), Some Mathematical Models of Race Discrimination in the Labor Market, *in* O. Ashenfelter & A. Rees, eds, 'Discrimination in Labor Markets', Princeton University Press, New Jersey, pp. 112–129.
- Arrow, K. (1973b), The Theory of Discrimination, *in* O. Ashenfelter & A. Rees, eds, 'Discrimination in Labor Markets', Princeton University Press, New Jersey, pp. 143–164.

- Ashenfelter, O. & Hannan, T. (1986), 'Sex Discrimination and Product Market Competition: The Case of the Banking Industry', *Quarterly Journal of Economics* **101**, 149–173.
- Baker, M., Benjamin, D. & Stanger, S. (1999), 'The Highs and Lows of the Minimum Wage Effect: A Time-Series Cross-Section Study of Canadian Law', *Journal of Labor Economics* **17**, 318–350.
- Barnow, B. S. (1987), 'The Impact of CETA Programs on Earnings: A Review of the Literature', *Journal of Human Resources* **22**, 157–193.
- Barros, R. P., Machado, A. F. & Mendonça, R. S. (1997), A Desigualdade da Pobreza: Estratégias Ocupacionais e Diferenciais por Gênero, Texto para Discussão 453, Instituto de Pesquisa Econômica Aplicada, IPEA.
- Bayard, K., Hellerstein, J., Neumark, D. & Troske, K. (1999), New Evidence on Sex Segregation and Sex Differences in Wages from Matched Employee-Employer Data, Working Paper 7003, National Bureau of Economic Research, NBER.
- Becker, G. S. (1971), *The Economics of Discrimination*, 2^a edn, University of Chicago Press, Chicago.
- Benkhard, C. L. & Berry, S. (2006), 'On the Nonparametric Identification of Nonlinear Simultaneous Equations Models: Comment on B. Brown (1983) and Roehrig (1988)', *Econometrica* **74**, 1429–1440.
- Bergmann, B. R. (1974), 'Occupational Segregation, Wages and Profits When Employers Discriminate by Race or Sex', *Eastern Economic Journal* **1**, 103–110.
- Bhaskar, V., Manning, A. & To, T. (2002), 'Oligopsony and Monopsonistic Competition in Labour Markets', *Journal of Economic Perspectives* **16**, 155–174.
- Bhaskar, V. & To, T. (2003), 'Oligopsony and the Distribution of Wages', *European Economic Review* **47**, 371–399.
- Blau, F. (1977), *Equal Pay in the Office*, D.C. Heath and Co, Lexington, Massachusetts.
- Blau, F. D. (1984), Occupational Segregation and Labor Market Discrimination, in B. F. Reskin, ed., 'Sex Segregation in the Workplace: Trends, Explanations, Remedies', National Academy Press, Washington, D.C.

- Blau, F. D. & Ferber, M. A. (1992), *The Economics of Women, Men, and Work*, 2^a edn, Prentice-Hall, New Jersey.
- Blau, F. & Jusenius, C. (1976), 'Economists' Approaches to Sex Segregation in the Labor Market: An Appraisal', *Signs* **1**(3), 181–199.
- Blinder, A. S. (1973), 'Wage Discrimination: Reduced Form and Structural Estimates', *Journal of Human Resources* **8**, 436–455.
- Blundell, R. & Bond, S. (1998), 'Initial Conditions and Moment Restrictions in Dynamic Panel Data Models', *Journal of Econometrics* **87**, 115–143.
- Blundell, R. & Bond, S. (2000), 'GMM Estimation with Persistent Panel Data: An Application to Production Functions', *Econometric Reviews* **19**, 321–340.
- Blundell, R. & Dias, M. C. (2000), 'Evaluation Methods for Non-Experimental Data', *Fiscal Studies* **21**, 427–468.
- Brown, B. W. (1983), 'The Identification Problem in Systems Nonlinear in the Variables', *Econometrica* **51**, 175–196.
- Bryk, A. S. & Raudenbush, S. W. (1992), *Hierarchical Linear Models: Applications and Data Analysis Methods*, 1^a edn, Sage, London.
- Buckley, J. (1971), 'Pay Differences Between Men and Women in the Same Job', *Monthly Labor Review* **94**, 36–39.
- Burdett, K. & Mortensen, D. T. (1998), 'Wage Differentials, Employer Size, and Unemployment', *International Economic Review* **39**, 257–273.
- Cain, G. J. (1986), The Economic Analysis of Labor Market Discrimination: A Survey, in O. Ashenfelter & R. Layard, eds, 'Handbook of Labor Economics', Vol. 1, North Holland, Amsterdam, pp. 693–785.
- Card, D. & Krueger, A. B. (1995), *Myth and Measurement: The New Economics of the Minimum Wage*, Princeton University Press, New Jersey.
- Carrington, W. J. & Troske, K. R. (1995), 'Gender Segregation in Small Firms', *Journal of Human Resources* **30**, 503–533.
- Carrington, W. J. & Troske, K. R. (1997), 'On Measuring Segregation in Samples With Small Units', *Journal of Business and Economic Statistics* **15**, 402–409.

- Carrington, W. J. & Troske, K. R. (1998), 'Sex Segregation in U.S. Manufacturing', *Industrial and Labor Relations Review* **51**, 445–464.
- Chesher, A. (2002), 'Instrumental Values', *Cemmap Working Paper* (17).
- Chesher, A. (2003), 'Identification in Nonseparable Models', *Econometrica* **71**, 1405–1441.
- Chesher, A. (2007), 'Instrumental Values', *Journal of Econometrics* **139**, 15–34.
- Christ, C. F. (1994), 'The Cowles Commission's Contributions to Econometrics at Chicago, 1939-1955', *Journal of Economic Literature* **32**, 30–59.
- Coate, S. & Loury, G. (1993), 'Will Affirmative-Action Policies Eliminate Negative Stereotypes?', *American Economic Review* **83**, 1220–1240.
- Dasgupta, S., Mody, A., Roy, S. & Wheeler, D. (2001), 'Environmental Regulation and Development: A Cross-country Empirical Analysis', *Oxford Development Studies* **29**, 173–187.
- England, P. (1992), *Comparable Worth: Theories and Evidence*, Aldine de Gruyter, New York.
- Fields, J. & Wolff, E. N. (1995), 'Interindustry Wage Differentials and the Gender Wage Gap', *Industrial and Labor Relations Review* **49**, 105–120.
- Filer, R. K. (1989), Occupational Segregation, Compensating Differentials, and Comparable Worth, in R. T. Michael, H. Hartmann & B. O'Farrell, eds, 'Pay Equity: Empirical Inquiries', National Academy Press, Washington, D.C.
- Frees, E. W. (2004), *Longitudinal and Panel Data: Analysis and Applications in the Social Sciences*, Cambridge University Press, Cambridge.
- Goldenstein, H. (2002), *Multilevel Statistical Models*, 3^a edn, Edward Arnold, London.
- Groshen, E. L. (1991), 'The Structure of the Female/Male Wage Differential: Is It Who You Are, What You Do, or Where You Work?', *Journal of Human Resources* **26**, 457–472.
- Gueron, J. M. (1990), 'Work and Welfare: Lessons on Employment Programs', *Journal of Economic Perspectives* **4**, 79–98.

- Hamermesh, D. S. (1993), *Labor Demand*, Princeton University Press, New Jersey.
- Heckman, J. J. (2000), 'Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective', *Quarterly Journal of Economics* **115**, 45–97.
- Heckman, J. J., Ichimura, H. & Todd, P. E. (1997), 'Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program', *Review of Economic Studies* **64**, 605–654.
- Heckman, J. J., LaLonde, R. J. & Smith, J. A. (1999), The Economics and Econometrics of Active Labor Market Programs, in O. Ashenfelter & D. Card, eds, 'Handbook of Labor Economics', Vol. 3A, North Holland, Amsterdam, pp. 1865–2097.
- Heckman, J. J. & Robb, R. (1985), Alternative Methods for Evaluating the Impact of Interventions, in J. J. Heckman & B. Singer, eds, 'Longitudinal Analysis of Labour Market Data', Cambridge University Press, pp. 156–246.
- Hurwicz, L. (1950), Generalization of the Concepts of Identification, in T. C. Koopmans, ed., 'Statistical Inference in Dynamic Economic Models', Cowles Commission Monograph 10, John Wiley and Sons, New York.
- Hutchens, R. M. (2001), 'Numerical Measures of Segregation: Desirable Properties and Their Implications', *Mathematical Social Sciences* **58**, 709–804.
- Imbens, G. W. & Newey, W. K. (2001), Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity. Mimeo (revisado em 2003), Department of Economics, MIT.
- International Labour Office (2003), *Yearbook of Labour Statistics*, International Labour Office, Geneva.
- Johnson, G. & Solon, G. (1986), 'Estimates of Direct Effects of Comparable Worth Policy', *American Economic Review* **76**, 1117–1125.
- Killingsworth, M. R. (1990), *The Economics of Comparable Worth*, W.E. Upjohn Institute for Employment Research, Kalamazoo, Michigan.
- Koopmans, T. C. & Reiersøl, O. (1950), 'The Identification of Structural Characteristics', *Annals of Mathematical Statistics* **21**, 165–181.

- Koopmans, T. C., Rubin, H. & Leipnik, R. B. (1950), Measuring the Equation Systems of Dynamic Components, *in* T. C. Koopmans, ed., ‘Statistical Inference in Dynamic Economic Models’, Cowles Commission Monograph 10, John Wiley and Sons, New York.
- LaLonde, R. J. (1995), ‘The Promise of Public Sector-Sponsored Training Programs’, *Journal of Economic Perspectives* **9**, 149–168.
- Lundberg, S. J. & Startz, R. (1983), ‘Private Discrimination and Social Intervention in Competitive Labor Markets’, *American Economic Review* **73**, 340–347.
- Macpherson, D. A. & Hirsch, B. T. (1995), ‘Wages and Gender Composition: Why Do Women’s Jobs Pay Less?’, *Journal of Labor Economics* **13**, 426–471.
- Madden, J. F. (1973), *The Economics of Sex Discrimination*, Lexington Books, Lexington, Massachusetts.
- Manning, A. (2003), *Monopsony in Motion: Imperfect Competition in Labor Markets*, Princeton University Press, Princeton, New Jersey.
- Matzkin, R. L. (2003), ‘Nonparametric Estimation of Nonadditive Random Functions’, *Econometrica* **71**, 1339–1376.
- Matzkin, R. L. (2005), Identification in Nonparametric Simultaneous Equations. Mimeo, Northwestern University.
- McNulty, D. (1967), ‘Differences in Pay Between Men and Women Workers’, *Monthly Labor Review* **90**, 40–43.
- Mincer, J. & Polachek, S. (1974), ‘Family Investments in Human Capital: Earnings of Women’, *Journal of Political Economy* **82**, S76–S108.
- Ministério do Trabalho (1998), Avaliação Gerencial do Plano Nacional de Qualificação Profissional, Relatório Técnico, Secretaria de Formação Profissional.
- Neumark, D. (1988), ‘Employers Discriminatory Behaviour and the Estimation of Wage Discrimination’, *Journal of Human Resources* **23**, 279–295.
- Neumark, D. & Wascher, W. (1992), ‘Employment Effects of Minimum Wages and Subminimum Wages: Panel Data on State Minimum Wage Laws’, *Industrial and Labor Relations Review* **46**, 55–81.

- Newey, W. K. & Powell, J. L. (1988), Nonparametric Instrumental Variables Estimation. Mimeo (revisado em 2002), Department of Economics, UC Berkeley.
- Newey, W. K., Powell, J. L. & Vella, F. (1999), ‘Nonparametric Estimation of Triangular Simultaneous Equations Models’, *Econometrica* **67**, 565–603.
- Oaxaca, R. L. (1973), ‘Male-Female Wage Differentials in Urban Labor Markets’, *International Economic Review* **14**, 693–709.
- Oaxaca, R. L. & Ransom, M. R. (1988), ‘Searching for the Effect of Unionism on the Wages of Union and Nonunion Workers’, *Journal of Labor Research* **9**, 139–148.
- Oliveira, A. M. H. (2001), Occupational Gender Segregation and Effects on Wages in Brazil, in X. G. P. Conference, ed., ‘Proceeding of XXIV General Population Conference’.
- Ometto, A. M., Hoffmann, R. & Alves, M. C. (1997), ‘A Segregação por Gênero no Mercado de Trabalho nos Estados de São Paulo e Pernambuco’, *Economia Aplicada* **1**, 393–423.
- Ometto, A. M., Hoffmann, R. & Alves, M. C. (1999), ‘Participação da Mulher no Mercado de Trabalho: Discriminação em Pernambuco e São Paulo’, *Revista Brasileira Econometria* **53**, 287–322.
- Pagan, A. & Ullah, A. (1999), *Nonparametric Econometrics*, Cambridge University Press.
- Phelps, E. S. (1972), ‘The Statistical Theory of Racism and Sexism’, *American Economic Review* **62**, 659–661.
- Polachek, S. W. (1979), Occupational Segregation Among Women: Theory, Evidence and a Prognosis, in C. B. Lloyd, E. Andrews & C. Gilroy, eds, ‘Women in the Labor Market’, Columbia University Press, New York.
- Polachek, S. W. (1981), ‘Occupational Self-Selection: A Human Capital Approach to Sex Differences in Occupational Structure’, *Review of Economics and Statistics* **63**, 60–69.
- Reilly, K. T. & Wirjanto, T. S. (1999a), ‘Does More Mean Less? The Male/Female Wage Gap and the Proportion of Females at the Establishment Level’, *Canadian Journal of Economics* **32**, 906–929.

- Reilly, K. T. & Wirjanto, T. S. (1999*b*), ‘The Proportion of Females in the Establishment: Discrimination, Preferences and Technology’, *Canadian Public Policy* **25**, S73–S94.
- Robinson, P. M. (1988), ‘Root-N Consistent Semiparametric Regression’, *Econometrica* **56**, 931–954.
- Røed, K. & Zhang, T. (2003), ‘Does Unemployment Compensation Affect Unemployment Duration?’, *Economic Journal* **113**, 190–206.
- Roehrig, C. S. (1988), ‘Conditions for Identification in Nonparametric and Parametric Models’, *Econometrica* **56**, 433–447.
- Sianesi, B. (2001), ‘An Evaluation of Swedish System of Active Labour Market Programmes’, *Cemmap Working Paper* (02).
- Sorensen, E. (1989), Measuring the Effect of Occupational Sex and Race Composition on Earnings, in R. T. Michael, H. Hartmann & B. O’Farrell, eds, ‘Pay Equity: Empirical Inquiries’, National Academy Press, Washington, D.C.
- Speckman, P. (1988), ‘Kernel Smoothing in Partial Linear Models’, *Journal of the Royal Statistical Society* **50**, 413–436.
- Treiman, D. J. & Hartmann, H. I. (1981), *Women, Work, and Wages: Equal Pay for Jobs of Equal Value*, National Academy Press, Washington, D.C.
- Vieira, J. A. C., Cardoso, A. R. & Portela, M. (2003), Recruitment and Pay at the Establishment Level: Gender Segregation and the Wage Gap in Portugal, Discussion Paper 789, Institute for the Study of Labor, IZA.
- Williams, N. (1993), ‘Regional Effects of the Minimum Wage on Teenage Employment’, *Applied Economics* **25**, 1517–1528.
- Wolpin, K. I. (1980), ‘A Time Series-Cross Section Analysis of International Variation in Crime and Punishment’, *Review of Economics and Statistics* **62**, 417–423.
- Zhang, T. (2003), ‘Identifying Treatment Effects of Active Labour Market Programmes for Norwegian Adults’, *Memoranda, University of Oslo* (26).