UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE
DEPARTAMENTO DE ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

EFEITO DO TAMANHO DO ESTABELECIMENTO SOBRE OS SALÁRIOS: UMA ANÁLISE COM UM PAINEL DE ESTABELECIMENTOS E DE SEUS TRABALHADORES

Alessandro Ribeiro de Carvalho Casalecchi

Orientador: Prof. Dr. Naércio Aquino Menezes Filho

SÃO PAULO
2011

Prof. Dr. João Grandino Rodas
Reitor da Universidade de São Paulo

Prof. Dr. Reinaldo Guerreiro
Diretor da Faculdade de Economia, Administração e Contabilidade

Prof. Dr. Denisard Cneio de Oliveira Alves
Chefe do Departamento de Economia

Prof. Dr. Pedro Garcia Duarte
Coordenador do Programa de Pós-Graduação em Economia

ALESSANDRO RIBEIRO DE CARVALHO CASALECCHI

EFEITO DO TAMANHO DO ESTABELECIMENTO SOBRE OS SALÁRIOS: UMA ANÁLISE COM UM PAINEL DE ESTABELECIMENTOS E DE SEUS TRABALHADORES

Dissertação apresentada ao Departamento de Economia da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo como requisito para a obtenção do título de Mestre em Economia

Orientador: Prof. Dr. Naércio Aquino Menezes Filho

Versão original

SÃO PAULO

2011

FICHA CATALOGRÁFICA

Elaborada pela Seção de Processamento Técnico do SBD/FEA/USP

Casalecchi, Alessandro Ribeiro de Carvalho
Efeito do tamanho do estabelecimento sobre os salários: uma análise com um painel de estabelecimentos e de seus trabalhadores / Alessandro Ribeiro de Carvalho Casalecchi. -- São Paulo, 2011.
86 p.

Dissertação (Mestrado) – Universidade de São Paulo, 2011.
Orientador: Naércio Aquino Menezes Filho.

1. Salários 2. Econometria 3. Empregador 4. Trabalhadores I. Universidade de São Paulo. Faculdade de Economia, Administração e Contabilidade II. Título.

CDD – 331.2

Agradeço ao professor e orientador Naércio Aquino Menezes Filho e a Ricardo Freguglia.

Agradeço também à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo apoio financeiro durante o período de elaboração desta dissertação.

RESUMO

A literatura empírica em economia do trabalho mostra que o tamanho do empregador afeta positivamente o salário de seus trabalhadores. O presente trabalho utiliza uma especificação empírica da equação salarial que inclui efeitos fixos para os trabalhadores, para o empregador e para a qualidade do match *entre eles, de modo que explicações teóricas para a existência do efeito do tamanho do empregador possam ser testadas sem que vieses de variável omitida prejudiquem as conclusões. Esta especificação com três efeitos fixos ainda não foi utilizada na literatura empírica do efeito salarial do tamanho do empregador. Os resultados indicam que mais da metade do efeito do tamanho do empregador é explicado pela correlação do tamanho com as três heterogeneidades não observáveis e que devemos rejeitar a validade empírica das explicações fornecidas pelos modelos teóricos testados de forma geral. Contudo, ressaltamos que tais modelos podem se aplicar em circunstâncias específicas.*

ABSTRACT

Economists have persistently found a positive relationship between the employer size and the wage earned by employees. This paper uses an empirical specification that controls for unobserved heterogeneity at three levels: worker, employer and match quality, for the first time in this literature. Thus, theoretical explanations of the employer size effect can be tested with estimation results that are not biased due to omitted variable problems. We show that more than half of the employer size effect is explained by the correlation between the employer size and the three heterogeneities. The results also suggest that we should reject the tested theoretical models as general explanations of the employer size effect. However, these models may explain the effect in some circumstances.

SUMÁRIO

1 INTRODUÇÃO 3
2 LITERATURA 7
2.1 Literatura empírica 7
2.2 Literatura teórica 8
3 METODOLOGIA 13
3.1 Controle para apenas dois efeitos fixos: trabalhador e empregador 13
3.2 Existência do efeito da qualidade do match 16
3.3 Interpretação errônea 19
3.4 Análise 21
4 DADOS 27
5 RESULTADOS 29
5.1 Especificação básica 29
5.2 Qualificação agregada da mão-de-obra no nível do estabelecimento 35
5.2.1 Avaliação mais rigorosa 39
5.3 População de trabalhadores 40
5.3.1 Avaliação mais rigorosa 49
6 CONCLUSÃO 53
REFERÊNCIAS 54
APÊNDICES 56
APÊNDICE A – Descrição dos dados 57
Tabela 1A – Descrição das variáveis 57
Tabela 2A – Comparação entre o painel não balanceado e o balanceado 62
Tabela 3A – Descrição dos grandes grupos CBO 1994 66
Tabela 4A – Descrição das seções CNAE 1.0 67
APÊNDICE B – Derivações de vieses 68
B.1 Derivação de (4) 68
B.2 Derivação de (9) 68
B.3 Derivação de (I*) 69
B.4 Derivação de (II*) 69
B.5 Derivação de (III*) 70
B.6 Derivação de (IV*) 70
APÊNDICE C – Derivações das equações para estimação *within* 72
C.1 Derivação de (16) 72
C.2 Derivação de (17) 72
C.3 Derivação de (18) 73
APÊNDICE D – Estimações incluindo todas as seções CNAE 1.0 75

1 INTRODUÇÃO

Na teoria básica da firma, o trabalho é contratado por um preço exógeno e se caracteriza por retornos marginais decrescentes. Assim, uma vez que a condição de maximização de lucro exige que o empregador iguale o salário ao valor da produtividade marginal do trabalho, há uma relação *negativa* entre salário e tamanho do empregador (medido em número de trabalhadores), dado o preço do produto produzido. Estudos empíricos têm encontrado, entretanto, uma relação *positiva* ao estimarem a equação salarial. Muito embora alguns destes estudos tenham conseguido diminuir a magnitude desta relação positiva após a utilização de certos controles adicionais – como qualificação da mão-de-obra do empregador e sua razão capital-trabalho, entre outros – ainda não foi encontrado (conforme nosso conhecimento) um conjunto de controles adicionais capaz de tornar o coeficiente do tamanho do empregador condizente com aquela previsão teórica básica.

Pode-se dizer que cada um destes controles adicionais está associado a um modelo teórico que pretende explicar o efeito do tamanho do empregador sobre o salário. A explicação fornecida por um modelo é corroborada quando o coeficiente do tamanho do empregador na equação salarial diminui após o aumento da equação de regressão conforme for recomendado pelo modelo teórico sob avaliação. Este trabalho também testa modelos teóricos que pretendem explicar a existência do efeito positivo do tamanho do empregador, assim como feito anteriormente por diversos outros autores. O principal diferencial em relação aos trabalhos anteriores é que agora levamos em conta a presença de heterogeneidade não observável dos trabalhadores, dos empregadores e da qualidade do *match* entre eles – como propõe Woodcock (2008) – ao introduzirmos efeitos fixos nestes três níveis na especificação da equação salarial. De acordo com nosso conhecimento, estas três fontes de heterogeneidade ainda não foram controladas nas equações salariais estimadas na literatura de efeito do tamanho do empregador sobre os salários, de forma que resultados de estudos anteriores podem ter sofrido de vieses de variável omitida correspondentes a estas heterogeneidades.

Controlar para a heterogeneidade não observável dos trabalhadores se justifica na medida em que elementos como a habilidade e o talento de um trabalhador afetam sua produtividade e, assim, seu salário. Analogamente, a heterogeneidade na qualidade da gestão, no poder de barganha desfrutado pelos empregados de dado empregador, etc., justificam a inserção de um

efeito fixo para cada empregador na equação salarial. Justificar a presença de um efeito fixo para a qualidade do *match* entre o empregado e o empregador na especificação é, contudo, menos trivial.

A necessidade de se controlar para a qualidade do *match*, ou para a complementaridade empregado-empregador, na equação salarial se deve ao fato de que a produtividade total de um estabelecimento depende não apenas de características (observáveis e não observáveis) do trabalhador e do empregador, mas também do quanto que as características do trabalhador "combinam" com as do empregador. Em termos mais formais, a produtividade de uma combinação empregado-empregador (*match*) específica pode ser maior ou menor que a soma das produtividades do empregado e do empregador tomadas à parte. Se este componente da produtividade total de um estabelecimento – a complementaridade empregado-empregador, ou qualidade do *match* – não for controlado nas estimações e, ao mesmo tempo, estiver positivamente correlacionado com o tamanho do empregador, haverá viés positivo de variável omitida nos resultados para o coeficiente do tamanho do empregador. Desta forma, a omissão deste controle seria responsável por pelo menos uma parte do efeito positivo do tamanho do empregador sobre o salário observado na literatura empírica.

Por que empregadores maiores tenderiam a realizar *matches* de maior qualidade? Em termos teóricos, uma explicação possível que podemos citar (entre várias outras disponíveis) é de que empregadores maiores seriam capazes de atrair a cada período de tempo mais candidatos a emprego que empregadores menores – talvez por serem mais conhecidos no mercado de trabalho. Neste caso, eles conseguiriam ser mais seletivos na contratação, selecionando apenas aqueles trabalhadores que formariam bons *matches*. Ademais, talvez apenas empregadores maiores é que teriam a capacidade de fazer esta seleção minuciosa, já que teriam tanto departamentos de recursos humanos capacitados para tal seleção refinada quanto departamentos de gestão com bom conhecimento da função de produção do estabelecimento a que o contratado seria alocado e, consequentemente, com bom conhecimento de como melhor aproveitar o novo empregado. Assim, grandes empregadores formariam *matches* de qualidade sistematicamente maior que outros empregadores, dando origem a uma correlação positiva entre tamanho do empregador e qualidade dos *matches* formados. Note que esta correlação, muito embora possa parecer algum tipo de vantagem constante no tempo do lado do empregador, não é captada pelo efeito fixo do empregador se o empregador mudar de

tamanho ao longo do tempo. Como se verá na descrição da base de dados, o tamanho dos empregadores varia no tempo em nossa amostra.

A produtividade atribuível à qualidade do *match* é aqui tratada como um efeito fixo, constante no tempo, e específico à união de cada trabalhador com cada um dos empregadores com quem ele manteve vínculo na amostra. Controlando-se para a qualidade do *match*, além de se controlar para os efeitos fixos do trabalhador e do empregador, devemos eliminar os vieses de variável omitida de que devem ter sofrido estudos anteriores, estimando assim um efeito do tamanho do empregador mais preciso.

Os resultados mostraram que mais da metade do efeito do tamanho do empregador é explicado pela soma destes vieses. Quanto aos testes que realizamos para avaliar a validade empírica de explicações fornecidas por modelos teóricos, todos os testes apontaram para a rejeição dos modelos.

Na seção 2, será feita a revisão das literaturas empírica e teórica sobre o efeito do tamanho do empregador sobre o salário. A seção 3 motiva, explica e fundamenta microeconomicamente, seguindo Woodcock (2008), a especificação usada da equação salarial. Os dados são explicados na seção 4. A seção 5 apresenta os resultados, e a seção 6 conclui o trabalho.

2 LITERATURA

2.1 Literatura empírica

A relação positiva entre salários e tamanho do empregador já foi detectada por diversos estudos. Weiss (1966) encontra que firmas pertencentes a indústrias concentradas – que tendem a reunir firmas grandes – remuneram seus empregados, em média, mais que firmas de indústrias competitivas. Mellow (1982) encontra que o tamanho da firma afeta os salários de forma homogênea entre indústrias, tratando-se portanto, segundo ele, de um efeito generalizado na economia, e não específico a setores. Além disso, o autor mostra que no setor manufatureiro (i) o efeito do tamanho da firma existe mesmo em indústrias pouco concentradas, muito embora seja menor que nas concentradas; (ii) a concentração industrial tem coeficiente significante e positivo na equação salarial mesmo controlando-se para tamanho do empregador (cujo coeficiente também é significante na mesma regressão).

Dunn (1986) constata que firmas grandes pagam salários sistematicamente maiores mesmo após controlar para ocupação e para uma medida de desutilidade do trabalho (o trabalho seria mais desconfortável para o empregado em firmas grandes, o que exigiria um prêmio para mantê-lo no emprego). Brown e Medoff (1989) oferecem uma revisão geral de algumas teorias e hipóteses, juntamente com testes correspondentes, que pretendem explicar o efeito positivo do tamanho do empregador, concluindo que o efeito é explicado, no máximo, apenas parcialmente por qualquer delas, permanecendo sempre um valor positivo e significante para o coeficiente do tamanho do empregador na equação salarial. Troske (1999) consegue reduzir a magnitude da correlação entre salários e tamanho da firma após controlar para a razão capital-trabalho e para a qualificação agregada da mão-de-obra no nível da firma. O autor também consegue reduzir a magnitude da correlação entre salários e tamanho do estabelecimento (unidade produtiva da firma) ao controlar para a qualificação agregada no nível do estabelecimento. Porém, ainda assim a correlação permanece positiva e significante tanto para o tamanho da firma quanto para o tamanho do estabelecimento.

A maioria destes estudos não é capaz de controlar simultaneamente para características observáveis do lado da demanda (empregador) e do lado da oferta (empregados), nem de

controlar para a heterogeneidade não observável do trabalhador, do empregador e da qualidade do *match*. Ademais, a quantidade e a variedade de controles usada por cada autor varia significativamente, muitas vezes por restrições nas suas bases de dados. Estas diferenças limitam a comparabilidade entre seus resultados.

2.2 Literatura teórica

Para explicar a existência de um efeito positivo do tamanho do empregador sobre os salários, autores basearam-se em diferentes hipóteses e teorias, as quais, vale dizer, não necessariamente foram desenvolvidas com o objetivo de explicar aquele efeito. A seguir, apresentaremos algumas delas.

Kremer (1993) propõe uma função de produção que incorpora a possibilidade de erros no processo produtivo, processo este que se dá em etapas. Mais especificamente, cada trabalhador *i* é dotado de uma habilidade q_i (pertencente ao intervalo 0 a 1) que representa a probabilidade de executar a sua tarefa com sucesso e independe do q_j de qualquer colega *j* da mesma firma. Caso um dos trabalhadores falhe na etapa da qual é responsável, o produto perde valor e deve ser descartado. Caso contrário, ocorrendo sucesso na tarefa, o produto segue para a próxima etapa do processo produtivo. Assim, o sucesso de um processo produtivo inteiro, passando por todas as *n* etapas que o compõem, ocorre somente quando todas as *n* etapas são realizadas com sucesso. Assume-se que cada uma das *n* etapas requer apenas um trabalhador e que a oferta de trabalho é inelástica.

A função de produção proposta tem três propriedades de interesse: (i) a produtividade marginal da habilidade do trabalhador *i* é crescente na habilidade dos colegas de trabalho; (ii) a maximização de lucro prevê uma relação positiva entre habilidade e salário; (iii) em equilíbrio há uma relação positiva entre *n* e habilidade do trabalhador. Assim, de (i) temos que em equilíbrio de mercado competitivo a firma contrata trabalhadores de um mesmo nível de habilidade. Note que, em se tratando de equilíbrio competitivo, as firmas são indiferentes entre ter mão-de-obra agregada mais habilidosa ou menos habilidosa, já que em qualquer caso terão lucro econômico zero. De (ii) e (iii), temos que firmas maiores pagam salários maiores.

Para Barron *et al* (1987), firmas grandes apresentam maior proporção de trabalhadores qualificados na sua mão-de-obra porque a política de contratação destas firmas é mais seletiva pois, por hipótese, firmas grandes enfrentam altos custos de monitoramento, não do esforço (supõe-se ausência de *moral hazard*), mas sim da *habilidade* dos trabalhadores contratados. Por causa destes altos custos de monitoramento, as firmas de grande porte tendem a contratar o mínimo necessário de trabalhadores, mas que tenham o máximo de habilidade cada. Para isso, controlam a habilidade dos funcionários onde podem, isto é, no instante da *contratação*, já que não conseguem observar a habilidade dos funcionários no dia a dia da firma (*on-the-job*). Mais do que isso, estas firmas tendem a investir mais em treinamento dos novos funcionários (para aumentar a sua habilidade) e na elevação da razão capital-trabalho. Finalmente, como conseqüência de todos estes custos irrecuperáveis do recrutamento seletivo, dos treinamentos e de mais capital, é ótimo para estas firmas pagar um salário maior, com o fim de aumentar a taxa de retenção de trabalhadores e, com isso, minimizar investimentos perdidos com demissões voluntárias.

Bulow e Summers (1986) fornecem um modelo de salário-eficiência no qual empregadores maiores encontram maiores dificuldades para monitorar o esforço de seus trabalhadores do que empregadores menores. Neste cenário, existem apenas dois níveis possíveis de esforço: pleno ou nulo. O monitoramento do esforço é impreciso: o esforço nulo é detectado com probabilidade menor do que 1. Para evitar esforço nulo na mão-de-obra, o empregador pode combinar dois métodos: pagar um salário maior que o de mercado (o de mercado é igual à produtividade marginal do trabalho, que é igual para todo trabalhador), mas punir o empregado com a demissão se detectado o esforço nulo; e/ou gastar mais recursos diretamente em tecnologias de monitoramento. Este segundo método atua no aumento da probabilidade de detecção, enquanto que o primeiro não atua na probabilidade de detecção, mas no incentivo ao empregado. Ademais, o gasto com tecnologias de monitoramento é função crescente da probabilidade de detecção, de maneira que o empregador que desejar cortar mais desta probabilidade terá que despender mais recursos. Diferenças nesta função crescente é que geram diferenças nas práticas salariais entre grandes e pequenos empregadores. Os autores assumem que esta função é “menos favorável” para empregadores maiores. Podemos interpretar esta passagem do texto de Bulow e Summers (1986) como se referindo a uma função que tem, por exemplo, maior derivada, ou seja, para levar a probabilidade de detecção para um dado patamar, empregadores maiores precisam despender mais recursos com tecnologia de monitoramento do que empregadores menores. Neste contexto, os autores

indicam que a condição de primeira ordem do problema do empregador exige que empregadores maiores gastem menos com o método da tecnologia de monitoramento e mais com o método do prêmio salarial. Em suma, todo empregador, grande ou pequeno, enfrenta dificuldades para monitorar seus empregados, mas a dificuldade de monitoramento enfrentada pelos grandes empregadores tem um formato tal que os leva, por decisão ótima, a usar mais do método de prêmio salarial e menos do método de tecnologia de monitoramento.

No modelo de Weiss e Landau (1984), o problema do empregador é minimizar os custos com trabalho dada a sua demanda por trabalho (em unidades de trabalho-eficiência) e dada a distribuição, na população de trabalhadores, da dotação de unidades de trabalho-eficiência. O empregador escolhe os níveis mínimo e máximo de dotação de sua mão-de-obra ao escolher, respectivamente, a dotação medida num exame de recrutamento (*pass-fail test*) e o salário (que será idêntico para toda a mão-de-obra contratada). Assume-se uma distribuição contínua de dotação de unidades de trabalho-eficiência na população de trabalhadores e que o salário de reserva de um trabalhador é função crescente de sua dotação.

Os autores demonstram, entre outras proposições, que

a) Não havendo custo fixo de contratação, empregadores maiores pagam salários maiores e exigem dotação mínima menor;
b) Havendo custo fixo de contratação e sendo os trabalhadores mais homogêneos (em termos de dotação) em níveis suficientemente altos de dotação, a correlação entre tamanho do empregador e salário é negativa para empregadores suficientemente pequenos.

Assim, o modelo de Weiss e Landau (1984) indica as situações em que a relação entre salário e tamanho do empregador é monotonicamente positiva e quando não é.

Em Strand (1987), no primeiro dos três modelos que apresenta, os trabalhadores têm informação imperfeita sobre o salário pago pelos diferentes empregadores do mercado. Existem dois conjuntos de trabalhadores, que se diferenciam pela forma como encontram emprego: aqueles que sorteiam um empregador no mercado e a ele se vinculam; e aqueles que conhecem exatamente dois trabalhadores empregados e comparam o salário que cada um dos dois recebeu no período anterior, vinculando-se ao empregador que pagou mais. Ao contrário de Weiss e Landau (1984), todos os trabalhadores são igualmente produtivos e possuem exatamente o mesmo salário de reserva, enquanto os empregadores é que se diferenciam em

sua produtividade. Em equilíbrio, empregadores mais produtivos empregam mais trabalhadores e, para garantir sua oferta de trabalho, pagam salário médio maior (a distribuição de salário intra-empregador é não colapsada), atraindo para si a quantidade necessária de trabalhadores a cada período de tempo. Nas extensões a este modelo, Strand (1987) permite migrações entre empregos, e não apenas do desemprego para o emprego e vice-versa. Contudo, o resultado geral em equilíbrio sempre é qualitativamente o mesmo do caso mais simples.

Tanto em Weiss e Landau (1984) quanto em Strand (1987), o salário maior é usado para atrair mais trabalhadores. Entretanto, o canal pelo qual o salário maior resulta em mais candidatos ao emprego é diferente nos dois modelos. No primeiro, onde há uma distribuição de salário de reserva na população de trabalhadores, um aumento no salário oferecido por um empregador passa a atrair trabalhadores cujos salários de reserva eram, antes do aumento, maiores que o salário praticado. No segundo, pagar um salário maior a um empregado aumenta a probabilidade de ele atrair um trabalhador conhecido para junto de seu empregador no próximo período. Outra diferença entre os dois artigos diz respeito ao problema do empregador. No primeiro artigo, a quantidade de trabalho contratada (isto é, o tamanho do empregador) é exógena, mas no segundo artigo esta é endógena e a produtividade dos empregadores, que difere entre empregadores por um coeficiente multiplicativo, é o que gera uma distribuição não colapsada de tamanho de empregador na economia.

Burdett e Mortensen (1998) também propõem um modelo no qual existem problemas de informação, nominalmente informação incompleta a respeito das ofertas de salário por parte dos empregadores. Em intervalos aleatórios, os trabalhadores (estejam eles já empregados ou não) sorteiam um empregador e conhecem sua oferta de salário e, com base nesta informação, migram de empregador (ou do desemprego) se o salário realizado superar o recebido até então. Na versão mais simples do modelo, todos os trabalhadores têm salário de reserva idêntico e, em qualquer empregador a que um deles se vincular, sua produtividade é igual a p. Os autores mostram que a distribuição de ofertas de salário não é colapsada neste caso. Nela, os maiores salários são pagos pelos maiores empregadores, os quais são grandes exatamente por pagarem salário maior e, com isso, receberem maior fluxo de trabalhadores a cada período. A distribuição permanece não colapsada nas duas modificações propostas do modelo: uma em que os trabalhadores possuem salário de reserva diferente (existe uma distribuição de salários de reserva na população de trabalhadores); e uma em que o salário de

reserva é idêntico, mas a produtividade do trabalhador varia entre empregadores (o vínculo com o empregador i terá produtividade p_i para todo trabalhador). Notemos que, ao contrário do modelo de Strand (1987), em Burdett e Mortensen (1998) a diferença exógena de produtividade entre empregadores *não* é necessária para gerar dispersão salarial na economia. Ao mesmo tempo, os modelos dos dois artigos se assemelham no sentido de que empregadores mais produtivos pagarão os maiores salários e serão, finalmente, os maiores empregadores da economia.

3 METODOLOGIA

O avanço introduzido neste trabalho em relação à literatura empírica já existente consiste em realizar testes de modelos teóricos que pretendem explicar a origem do efeito do tamanho do empregador utilizando uma especificação diferente da equação salarial, que controla para a heterogeneidade não observável dos trabalhadores, dos empregadores e da qualidade dos *matches* entre eles. Sem que se controle para estas três fontes de heterogeneidade, as estimativas do efeito do tamanho do empregador estarão viesadas se o tamanho do empregador estiver correlacionado com alguma delas. De acordo com nosso conhecimento, apenas equações salariais sem nenhum efeito fixo, ou apenas com o efeito fixo do trabalhador, foram estimadas nesta literatura específica. Na literatura mais geral de economia do trabalho, especificações com dois efeitos fixos (do trabalhador e do empregador) e com três efeitos fixos (os anteriores mais o da qualidade do *match*) já foram utilizadas. Woodcock (2008) é um exemplo de trabalho recente que usa a especificação com os três efeitos. Entretanto, estas especificações não foram utilizadas para analisar especificamente o efeito do tamanho do empregador sobre o salário.

A seguir, na primeira subseção, argumentamos que o coeficiente do tamanho do empregador na equação salarial é estimado de forma consistente tanto na especificação com dois efeitos fixos quanto na especificação com três efeitos. Na segunda subseção, motivamos detalhadamente a necessidade de se controlar para o efeito fixo da qualidade do *match*. Na terceira subseção, explicamos como a interpretação dos resultados obtidos da especificação com três efeitos fixos difere da interpretação dos resultados da especificação com dois efeitos fixos, muito embora o estimador do coeficiente do tamanho do empregador seja consistente em ambos os casos. A quarta e última subseção fundamenta microeconomicamente a especificação com três efeitos fixos, recorrendo à modelagem de Woodcock (2008).

3.1 Controle para apenas dois efeitos fixos: trabalhador e empregador

Os coeficientes da equação salarial que não controla para nenhum dos três efeitos fixos sofrem de no mínimo dois vieses de variável omitida: um referente à omissão do efeito fixo

do trabalhador e outro à omissão do efeito fixo do empregador. Para sermos mais explícitos, imaginemos uma base de dados em painel, balanceada, com N trabalhadores, J empregadores, T períodos de tempo e P variáveis explicativas observáveis que variam no tempo. Consideremos por ora que o modelo verdadeiro é

$$y_{ijt} = \boldsymbol{x}'_{ijt}\boldsymbol{\beta} + \theta_i + \psi_j + \varepsilon_{ijt} \tag{1}$$

onde y_{ijt} é o logaritmo natural do salário do trabalhador i no empregador j no instante t; $\boldsymbol{x}'_{ijt}$ é o vetor $(1 \times P)$ de variáveis observáveis que variam no tempo, referentes tanto ao trabalhador i quanto ao empregador j, no instante t; $\boldsymbol{\beta}$ é o vetor $(P \times 1)$ de coeficientes correspondente a $\boldsymbol{x}'_{ijt}$; θ_i e ψ_j são os efeitos fixos do trabalhador i e do empregador j, respectivamente; e ε_{ijt} é o termo de erro aleatório. Em notação matricial, o modelo pode ser reescrito como

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{D\theta} + \boldsymbol{F\psi} + \boldsymbol{\varepsilon} \tag{2}$$

onde $\boldsymbol{y}$ é $(NT \times 1)$; $\boldsymbol{X}$ tem dimensão $(NT \times P)$; $\boldsymbol{D}$ tem dimensão $(NT \times N)$ e contém os indicadores (*dummies*) dos N trabalhadores; $\boldsymbol{F}$ tem dimensão $(NT \times J)$ e contém os indicadores (*dummies*) dos empregadores para os quais os indivíduos trabalham em cada instante t; $\boldsymbol{\varepsilon}$ tem dimensão $(NT \times 1)$; $\boldsymbol{\beta}$ tem dimensão $(P \times 1)$; $\boldsymbol{\theta}$ tem dimensão $(N \times 1)$; e $\boldsymbol{\psi}$ tem dimensão $(J \times 1)$.

Para calcular explicitamente o viés citado acima, suponhamos que o seguinte modelo seja estimado:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{v} \tag{3}$$

Neste caso, teremos[1]

$$E(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta} + E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{D\theta}] + E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{F\psi}] \tag{4}$$

isto é, $\widehat{\boldsymbol{\beta}}$ será um estimador viesado de $\boldsymbol{\beta}$, onde o viés tem um componente que podemos atribuir à correlação das variáveis $\boldsymbol{X}$ com o efeito fixo do trabalhador, e um componente

[1] Para detalhes da derivação desta e das outras fórmulas desta seção, ver o Apêndice B.

atribuível à correlação das variáveis $\boldsymbol{X}$ com o efeito fixo do empregador. A introdução dos dois efeitos fixos entre as variáveis de controle da equação salarial, como em (2), leva à eliminação destes dois componentes do viés. Podemos mostrar que a estimação do coeficiente do tamanho do empregador no modelo (2), que contém os dois efeitos fixos, pode ser feita usando-se apenas a variação que os dados sofrem dentro da história de um mesmo trabalhador num mesmo empregador. Para isso, primeiro calculemos a média de (1) para um dado trabalhador i num dado empregador j:

$$y_{ijt} = \boldsymbol{x}'_{ijt}\boldsymbol{\beta} + \theta_i + \psi_j + \varepsilon_{ijt} \Longrightarrow$$

$$\Longrightarrow \frac{\sum_t [y_{ijt}]}{T} = \frac{\sum_t [\boldsymbol{x}'_{ijt}]}{T}\boldsymbol{\beta} + \frac{\sum_t [\theta_i]}{T} + \frac{\sum_t [\psi_j]}{T} + \frac{\sum_t [\varepsilon_{ijt}]}{T} \Longrightarrow$$

$$\Longrightarrow \bar{y}_{ij} = \bar{\boldsymbol{x}}'_{ij}\boldsymbol{\beta} + \theta_i + \psi_j + \bar{\varepsilon}_{ij}$$

Agora, subtraindo esta equação em média da equação (1) original, temos

$$y_{ijt} - \bar{y}_{ij} = \left(\boldsymbol{x}'_{ijt} - \bar{\boldsymbol{x}}'_{ij}\right)\boldsymbol{\beta} + (\theta_i - \theta_i) + \left(\psi_j - \psi_j\right) + \left(\varepsilon_{ijt} - \bar{\varepsilon}_{ij}\right) \Longrightarrow$$

$$\Longrightarrow y_{ijt} - \bar{y}_{ij} = \left(\boldsymbol{x}'_{ijt} - \bar{\boldsymbol{x}}'_{ij}\right)\boldsymbol{\beta} + \left(\varepsilon_{ijt} - \bar{\varepsilon}_{ij}\right) \tag{5}$$

de forma que a estimação de (5) por mínimos quadrados ordinários não sofre do viés de variável omitida proveniente da omissão dos efeitos fixos do trabalhador e do empregador, uma vez que eles não estão mais presentes na equação de regressão. Este procedimento garante consistência e ausência de viés para o estimador de $\boldsymbol{\beta}$. Entretanto, a interpretação dos resultados fica comprometida, pelo seguinte motivo.

Suponhamos que a produtividade do vínculo do trabalhador i com o empregador j seja função não apenas dos fatores produtivos entregues pelo trabalhador, de um lado, e pelo empregador, do outro lado. Isto é, suponhamos que a complementaridade entre estes fatores produtivos também afete a produtividade e afete de maneira constante no tempo, enquanto durar o vínculo de i com j. Denotemos a medida desta complementaridade por ϕ_{ij} , um efeito fixo, em que o subscrito indica o trabalhador e o empregador em questão. Sob estas hipóteses, note

que a derivação de (5) teria eliminado ϕ_{ij} da equação salarial da mesma forma que eliminou θ_i e ψ_j. Tal eliminação não intencional garante consistência de $\boldsymbol{\beta}$ no caso de ϕ_{ij} estar correlacionado com alguma variável de $\boldsymbol{X}$. Contudo, o problema é que a interpretação dos resultados finais, ou seja, a atribuição final do peso que cada heterogeneidade não observável tem na explicação do efeito positivo do tamanho do empregador sobre o salário, pode acabar sendo feita erroneamente. A explicação de como esta interpretação errônea acontece é feita na subseção 3.3. Antes, na subseção 3.2 a seguir, motivaremos a existência do efeito da qualidade do *match* (ϕ_{ij}).

3.2 Existência do efeito da qualidade do *match*

Para começar a explicação, podemos apontar para uma limitação do uso do efeito fixo do trabalhador: ele é frequentemente usado para medir algo que, na prática, pode não medir corretamente. Este efeito fixo (θ_i) é usualmente interpretado como a habilidade do trabalhador[2], a qual incorpora diversas capacidades e talentos constantes no tempo que afetam sua produtividade. Entretanto, tal efeito é medido em uma única dimensão, algo como uma "habilidade geral" que é transferível e aproveitável em todo e qualquer emprego a que o trabalhador venha a se vincular. Mais realista seria considerar que a habilidade tem duas dimensões: uma "geral", aproveitável pelo trabalhador em qualquer emprego; e uma "específica", aproveitável apenas em determinados empregos.

Um exemplo de habilidade "geral" é a inteligência. Todo emprego necessita de um desempenho intelectual mínimo do trabalhador, dependendo da função a ser preenchida, e trabalhadores mais inteligentes tenderão a ter remuneração maior em qualquer emprego. Em outras palavras, a inteligência é útil, aproveitável em qualquer emprego. Do outro lado, temos como exemplos de habilidade "específica" a capacidade de comunicação e a de organização metódica. Nem todo emprego necessita de trabalhadores que se comunicam bem ou que são metódicos. Mesmo que um trabalhador tenha vantagens nesses quesitos, elas podem ser desnecessárias ao processo produtivo de determinados empregadores, e, portanto, não serão

[2] Rigorosamente, o efeito fixo do trabalhador também controla para as variáveis observáveis constantes no tempo, e não apenas para as não observáveis constantes no tempo. Contudo, para fins de clareza de exposição, a partir daqui restringiremos nossa atenção apenas ao componente não observável do efeito fixo do trabalhador, sem prejuízo da interpretação mais rigorosa deste parâmetro. O mesmo vale para os efeitos fixos do empregador e da qualidade do *match*.

remuneradas, já que não afetam a produtividade do trabalho. Entretanto, outros empregadores, aqueles que necessitarem destas habilidades em sua função de produção, tenderão a remunerá-las, além de já remunerar a inteligência como fazem os empregadores anteriores. Note que o trabalhador carrega suas capacidades consigo para onde for, mas mesmo que elas sejam, em si, constantes no tempo, a utilização delas e seu conseqüente impacto na produtividade do trabalho varia no tempo. O que determina se as habilidades "específicas" de um trabalhador serão aproveitadas é tanto o trabalhador, que deve ter estas habilidades consigo, quanto o empregador, que deve necessitar delas em sua função de produção. Em outras palavras, o impacto das habilidades "específicas" é determinado pela combinação (*match*) empregado-empregador.

Mesmo empregos de uma mesma indústria, uma mesma localidade, uma mesma ocupação, podem aproveitar diferentemente as habilidades "específicas" de um mesmo trabalhador, dependendo da natureza da função de produção do empregador. Por exemplo, um estabelecimento que fabrica produtos químicos de maior risco de manuseio pode exigir um funcionário mais metódico, enquanto que um estabelecimento na mesma rua (localidade) e que fabrica produtos químicos (mesma indústria) menos perigosos, estaria disposto a empregar este mesmo trabalhador na mesma função (ocupação), mas não remuneraria sua habilidade "específica" (ser mais metódico que a média dos trabalhadores) tanto quanto o estabelecimento anterior, já que não necessita dela. Entretanto, ambos os estabelecimentos provavelmente remunerariam a sua inteligência igualmente, pois a inteligência do trabalhador provavelmente é variável importante nas duas funções de produção.

Para consolidar a motivação da existência de um efeito da qualidade do *match*, podemos tornar a intuição econômica anterior mais concreta, através da seguinte ilustração algébrica. Imagine que o modelo verdadeiro seja

$$y_{ijt} = \theta_i + \phi_{ij} + \eta_{ijt} \tag{6}$$

onde θ_i é um efeito fixo e mede a habilidade "geral" do trabalhador i; ϕ_{ij} é um efeito fixo e mede quanto das habilidades "específicas" do mesmo trabalhador são aproveitadas pelo empregador j na sua função de produção (qualidade do *match*); e η_{ijt} é o termo de erro aleatório. Em notação matricial, o modelo fica

$$\boldsymbol{y} = \boldsymbol{D\theta} + \boldsymbol{G\phi} + \boldsymbol{\eta} \tag{7}$$

onde $\boldsymbol{G}$ contém os indicadores (*dummies*) das diferentes combinações empregado-empregador (*matches*) e tem dimensão $(NT \times C)$, sendo C o número total de combinações empregado-empregador (*matches*) diferentes na amostra; $\boldsymbol{\phi}$ é $(C \times 1)$ e $\boldsymbol{\eta}$ é $(NT \times 1)$. Se estimarmos um modelo que omite os efeitos da qualidade dos *matches*, isto é

$$\boldsymbol{y} = \boldsymbol{D\theta} + \boldsymbol{\xi} \tag{8}$$

teremos que o estimador de $\boldsymbol{\theta}$ será tal que

$$E(\widehat{\boldsymbol{\theta}}) = \boldsymbol{\theta} + E[(\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}'\boldsymbol{G\phi}] \tag{9}$$

Cada linha de (9) é dada por

$$E(\hat{\theta}_i) = \theta_i + E\left(\frac{\sum_j \sum_t \phi_{ij} 1(J(i,t)=j)}{T}\right) \tag{10}$$

onde $J(i,t)$ é uma função que retorna o empregador j em que i trabalhou no instante t e $1(\cdot)$ é a função indicadora, que vale 1 se seu conteúdo for verdadeiro, e 0 se for falso. Ou seja, o valor esperado do estimador da habilidade "geral" do trabalhador i é igual à sua habilidade "geral" verdadeira mais o valor esperado da média ponderada de ϕ_{ij}, em que os pesos são o número de períodos da amostra que ele passou com cada empregador. Em outras palavras, o coeficiente que é usualmente tido na literatura como constante no tempo (o efeito fixo do trabalhador) seria estimado, na verdade, como a soma do componente da habilidade que tem um impacto sobre a produtividade que é, de fato, constante no tempo (habilidade "geral") e do componente cujo impacto sobre a produtividade varia no tempo (qualidade do *match*, ou o quanto das habilidades "específicas" é aproveitado pelo empregador de então na sua função de produção). Vale lembrar novamente que as habilidades, tanto a "geral" quanto as "específicas", são *constantes* no tempo, mas o quanto das "específicas" é efetivamente aproveitado pode variar no tempo conforme o trabalhador muda de empregador.

3.3 Interpretação errônea

Suponhamos agora que o modelo verdadeiro seja dado por

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{D\theta} + \boldsymbol{F\psi} + \boldsymbol{G\phi} + \boldsymbol{\omega} \tag{11}$$

ou seja, o modelo (2), mais controles para a qualidade dos *matches*. Se o modelo estimado for o (2), então teremos que

$$E(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta} \tag{12}$$

isto é, o estimador de $\boldsymbol{\beta}$ é não viesado, como argumentado anteriormente. Se estimássemos um *pooled* OLS (*POLS*), uma especificação somente com o efeito fixo do trabalhador (*EFt*), uma somente com o efeito fixo do empregador (*EFe*), e uma com os dois efeitos como em (2) (chamaremos de *EFte*), isto é, respectivamente

POLS: $$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon_I} \tag{I}$$

EFt: $$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{D\theta} + \boldsymbol{\varepsilon_{II}} \tag{II}$$

EFe: $$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{F\psi} + \boldsymbol{\varepsilon_{III}} \tag{III}$$

EFte: $$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{D\theta} + \boldsymbol{F\psi} + \boldsymbol{\varepsilon_{IV}} \tag{IV}$$

teríamos, respectivamente[3],

$$E(\widehat{\boldsymbol{\beta}}_{POLS}) = \boldsymbol{\beta} + E[(\boldsymbol{X'X})^{-1}\boldsymbol{X'D\theta}] + E[(\boldsymbol{X'X})^{-1}\boldsymbol{X'F\psi}] + E[(\boldsymbol{X'X})^{-1}\boldsymbol{X'G\phi}] \tag{I*}$$

$$E(\widehat{\boldsymbol{\beta}}_{EFt}) = \boldsymbol{\beta} + E[(\boldsymbol{X'M_DX})^{-1}\boldsymbol{X'M_DF\psi}] + E[(\boldsymbol{X'M_DX})^{-1}\boldsymbol{X'M_DG\phi}] \tag{II*}$$

$$E(\widehat{\boldsymbol{\beta}}_{EFe}) = \boldsymbol{\beta} + E[(\boldsymbol{X'M_FX})^{-1}\boldsymbol{X'M_FD\theta}] + E[(\boldsymbol{X'M_FX})^{-1}\boldsymbol{X'M_FG\phi}] \tag{III*}$$

$$E(\widehat{\boldsymbol{\beta}}_{EFte}) = \boldsymbol{\beta} \tag{IV*}$$

Contudo, se tivéssemos assumido que o modelo verdadeiro era o (2), e não o (11), teríamos escrito as esperanças acima erroneamente, assim:

---

[3] As expressões a seguir podem ser facilmente derivadas através do procedimento de Frisch-Waugh-Lovell, onde $\boldsymbol{M_A} = \boldsymbol{I} - \boldsymbol{A}(\boldsymbol{A'A})^{-1}\boldsymbol{A'}$ é a matriz que gera os resíduos da regressão do vetor por ela pré-multiplicado contra as variáveis contidas na matriz $\boldsymbol{A}$. Detalhes da derivação estão no Apêndice B.

$$E(\widehat{\boldsymbol{\beta}}_{POLS}) = \boldsymbol{\beta} + E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{D}\boldsymbol{\theta}] + E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{F}\boldsymbol{\psi}] \quad \text{(I**)}$$

$$E(\widehat{\boldsymbol{\beta}}_{EFt}) = \boldsymbol{\beta} + E[(\boldsymbol{X}'\boldsymbol{M_D}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M_D}\boldsymbol{F}\boldsymbol{\psi}] \quad \text{(II**)}$$

$$E(\widehat{\boldsymbol{\beta}}_{EFe}) = \boldsymbol{\beta} + E[(\boldsymbol{X}'\boldsymbol{M_F}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M_F}\boldsymbol{D}\boldsymbol{\theta}] \quad \text{(III**)}$$

$$E(\widehat{\boldsymbol{\beta}}_{EFte}) = \boldsymbol{\beta} \quad \text{(IV**)}$$

Neste contexto, para comparar os estimadores das especificações de (I) a (IV) entre si, poderíamos fazer, por exemplo,

$$E(\widehat{\boldsymbol{\beta}}_{EFt} - \widehat{\boldsymbol{\beta}}_{EFte}) = E[(\boldsymbol{X}'\boldsymbol{M_D}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M_D}\boldsymbol{F}\boldsymbol{\psi}] \quad (13)$$

$$E(\widehat{\boldsymbol{\beta}}_{EFe} - \widehat{\boldsymbol{\beta}}_{EFte}) = E[(\boldsymbol{X}'\boldsymbol{M_F}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M_F}\boldsymbol{D}\boldsymbol{\theta}] \quad (14)$$

Ou seja, diríamos que o valor esperado da diferença entre os estimadores de $\boldsymbol{\beta}$ no *EFt* e no *EFte* seria igual ao valor esperado da correlação de $\boldsymbol{X}$ com $\boldsymbol{\psi}$, condicional na identidade dos trabalhadores. Em outras palavras, se o coeficiente de $\widehat{\boldsymbol{\beta}}_{EFt}$ correspondente ao tamanho do empregador fosse estimado maior que sua contrapartida em $\widehat{\boldsymbol{\beta}}_{EFte}$, atribuiríamos tal diferença ao fato de que a correlação do tamanho do empregador com o efeito fixo do empregador tende a ser positiva ao longo da história de um dado trabalhador da amostra, conforme ele migra entre empregadores (pois é zero quando ele não migra). Analogamente, se o coeficiente de $\widehat{\boldsymbol{\beta}}_{EFe}$ correspondente ao tamanho do empregador fosse estimado maior que sua contrapartida em $\widehat{\boldsymbol{\beta}}_{EFte}$, diríamos, com base em (14), que a correlação, condicional na identidade do empregador, do tamanho do empregador com o efeito fixo do trabalhador tende a ser positiva através dos trabalhadores que eventualmente se vincularam ao empregador em qualquer período. Toda esta interpretação é errônea porque ignora a existência dos efeitos fixos da qualidade do *match*.

A interpretação correta, feita a partir das esperanças de (I*) a (IV*), é entretanto mais complicada, pois envolve mais correlações – as relativas à qualidade dos *matches*. Saber qual das correlações – do tamanho do empregador com cada efeito fixo – tende a ser mais importante na geração dos vieses no coeficiente do tamanho do empregador presentes na literatura, ou mesmo qual tende a ser o sinal destas correlações, não é fácil como antes. Mesmo assim, é possível extrair conclusões relevantes, como se verá na seção de resultados.

3.4 Análise

A especificação empírica a ser estimada pode ser fundamentada microeconomicamente como segue. Woodcock (2008) expõe seu modelo referindo-se a firmas, mas aqui usaremos o modelo pensando em estabelecimentos, já que o empregador na nossa amostra é o estabelecimento de trabalho, e não a firma. O autor supõe a seguinte função de produção para uma combinação empregado-empregador (*match*) específica:

$$Q_{ij} = \mu L_i^{\theta} K_j^{\psi} M_{ij}^{\phi}$$

onde Q_{ij} é o produto do trabalhador *i* no estabelecimento *j*; $L_i > 0$ mede o conjunto dos insumos entregues pelo trabalhador *i* (trabalho em si, habilidade, qualificação, etc.); $K_j > 0$ mede o conjunto dos insumos entregues pelo estabelecimento *j* (capital, sofisticação do capital, capital organizacional, etc.); e $M_{ij} > 0$ mede o conjunto dos insumos específicos à combinação de *i* com *j*, que pode ser interpretado como a complementaridade peculiar entre o trabalhador *i* e o estabelecimento *j*. Os termos escalares μ, θ, ψ e ϕ são parâmetros do modelo.

Woodcock (2008) assume também que a determinação do salário se dá por uma barganha de Nash em que cada trabalhador do estabelecimento *j* tem poder de barganha γ_j. Sendo p_j o preço do produto de *j* e w_{ij} o salário de *i* em *j*, e sob as hipóteses adicionais de que a *outside option* é zero para o trabalhador e para o estabelecimento, de que o trabalhador maximiza w_{ij} e de que o estabelecimento maximiza $(p_j Q_{ij} - w_{ij})$, a solução da barganha[4] é

$$w_{ij} = \gamma_j p_j Q_{ij}$$

Substituindo-se a função de produção e aplicando-se o logaritmo natural em ambos os lados, teremos:

$$\ln w_{ij} = \ln \gamma\mu + \theta \ln L_i + \left(\psi \ln K_j + \ln\left(\gamma_j/\gamma\right) + \ln p_j\right) + \phi \ln M_{ij}$$

[4] Apesar de Woodcock (2008) não explicitar, este resultado pode ser derivado, por exemplo, do problema $\max_{w_{ij}} \left(p_j Q_{ij} - w_{ij}\right)^{1-\gamma_j} \left(w_{ij}\right)^{\gamma_j}$.

onde γ é a média geométrica de γ_j. Como (i) θ, ψ e ϕ não são identificáveis separadamente de L_i , K_j e M_{ij} e (ii) não observamos preços – p_j – nem política salarial – γ_j – dos estabelecimentos, o autor propõe a especificação empírica:

$$\ln w_{ijt} = \mu^* + \boldsymbol{x}'_{it}\beta + \theta_i + \psi_j + \phi_{ij} + \varepsilon_{ijt}$$

onde

$\mu^* = \ln \gamma\mu$
$\boldsymbol{x}'_{\boldsymbol{it}}\beta + \theta_i = \theta \ln L_i$
$\psi_j = \psi \ln K_j + \ln(\gamma_j/\gamma) + \ln p_j$
$\phi_{ij} = \phi \ln M_{ij}$

Serão testados alguns modelos teóricos que tentam explicar a existência de um efeito positivo do tamanho do empregador sobre os salários. Para testá-los, será usado o seguinte modelo sobre dados em painel, adaptado da especificação de Woodcock (2008):

$$\ln w_{ijt} = \boldsymbol{x}'_{ijt}\boldsymbol{\beta} + \theta_i + \psi_j + \phi_{ij} + \varepsilon_{ijt} \tag{15}$$

onde w_{ijt} é o salário do trabalhador *i* empregado no estabelecimento *j* no instante *t*; $\boldsymbol{x}'_{ijt}$ é um vetor $(1 \times P)$ que contém uma coluna de "1" e as variáveis observáveis que variam no tempo do empregado *i* e do estabelecimento *j* em que ele trabalha, ambos no instante *t*; θ_i é o efeito fixo do trabalhador *i* e agrega o componente não observável e o efeito das variáveis constantes no tempo do indivíduo *i* [5]; ψ_j é o efeito fixo do estabelecimento e agrega o componente não observável e o efeito das variáveis constantes no tempo do estabelecimento *j*; ϕ_{ij} é o efeito não observável da complementaridade produtiva entre o empregado *i* e o estabelecimento *j* que o emprega; e ε_{ijt} é o termo de erro aleatório. Suponhamos que existam na amostra N trabalhadores, J firmas, C diferentes combinações empregado-empregador (*matches*) e T períodos de tempo. Em notação matricial a equação (15) fica como (11).

[5] Ver nota de rodapé número 2.

A Tabela 1 abaixo expõe os modelos teóricos a serem testados. A tabela também indica o procedimento de teste que usamos para testar o poder explicativo de cada modelo. Restringimo-nos a testar modelos, e não as hipóteses que os sustentam, porque, na prática, foi possível testar apenas proposições derivadas, e não premissas. Metodologicamente, tal fato significa que se rejeitarmos alguma destas proposições derivadas teremos evidência de que o modelo sendo avaliado não explica a origem do efeito do tamanho do empregador sobre o salário, sem que possamos saber quais de suas premissas são falsas. Contudo, ao mesmo tempo, a não rejeição não significa que o modelo explica a realidade, mas apenas que ele é compatível, ou consistente, com ela. Em suma, rejeitar um modelo é um resultado mais conclusivo que não rejeitar. Na Tabela 1 e em outros trechos do texto, usaremos a sigla ETE para nos referir à expressão “efeito do tamanho do estabelecimento”.

Tabela 1 – Modelos teóricos testados

Explicação contida no modelo	Literatura	Procedimento de teste
O ETE existe porque estabelecimentos maiores tendem a contratar trabalhadores mais habilidosos, e/ou a praticar políticas salariais mais vantajosas aos empregados, e/ou a formar *matches* de maior qualidade.		Comparação de estimativas entre especificações que omitem e que incluem os três efeitos fixos ($\boldsymbol{\theta}$, $\boldsymbol{\psi}$ e $\boldsymbol{\phi}$).
O ETE existe porque estabelecimentos grandes tendem a concentrar uma proporção de mão-de-obra qualificada maior que outros estabelecimentos, a qual tem um efeito adicional sobre a produtividade.	Kremer (1993)	Comparação de estimativas entre a amostra total e uma sub-amostra com menor variação da qualificação agregada no nível do estabelecimento.
O ETE existe porque estabelecimentos maiores enfrentam maiores dificuldades para monitorar o esforço dos trabalhadores, pagando salário maior para estimular o esforço.	Bulow e Summers (1986)	Comparação de estimativas entre mercados de trabalho que diferem na quantidade de trabalhadores.
O ETE existe porque os estabelecimentos que são maiores pagam salários maiores como forma de atrair mão-de-obra suficiente para manter este seu tamanho ótimo.	Weiss e Landau (1984) Strand (1987) Burdett e Mortensen (1998)	

Neste trabalho as estimações serão feitas usando-se efeitos fixos, que permitem correlação entre a heterogeneidade não observável e as variáveis observáveis. Em outras palavras, garantimos robustez de resultados frente a qualquer tipo de elemento constante no tempo e peculiar a algum setor, região geográfica, tipo de trabalhador, arcabouço institucional, etc. que nos tenha fugido ao alcance em nossos dados. A abordagem de efeito fixo também atende melhor a desenvolvimentos consolidados da teoria microeconômica, ao permitir correlação entre habilidade e educação, por exemplo. Não é difícil pensar, ainda, em outras fontes de correlação entre heterogeneidade não observável e variáveis observáveis: se assumirmos que os jovens tendem a ter maior disposição para o trabalho intenso e desgastante, desempenhando-o melhor que os mais velhos, então há correlação entre o componente de "disposição" do efeito não observável do trabalhador e a idade do mesmo, que é observável e varia no tempo. Outra possibilidade: podemos assumir que trabalhadores muito inábeis

tendem a permanecer num mesmo emprego somente até que sua verdadeira habilidade seja revelada ao empregador, que então o demite. Assim, teremos correlação positiva entre o efeito não observável do trabalhador e a *tenure* ("tempo de casa" ou antiguidade no emprego) do mesmo.

Para fins de comparação, serão estimados quatro tipos de especificações: *pooled* OLS (*POLS*); efeito fixo apenas para o trabalhador (*EFt*); efeito fixo apenas para o estabelecimento (*EFe*); e efeitos fixos para o trabalhador, o estabelecimento e a qualidade do *match* – ou complementaridade empregado-empregador – (*EFtec*). Todos os estimadores por efeitos fixos serão do tipo *within estimators*, isto é, *não* serão obtidos pelo uso de *dummies* correspondentes aos efeitos fixos. Explicitamente[6], para o *EFt*, o modelo

$$\ln w_{ijt} - \overline{\ln w_{\iota}} = \left(\boldsymbol{x}'_{ijt} - \overline{\boldsymbol{x}}'_{i}\right)\boldsymbol{\beta} + \left(\varepsilon_{ijt} - \bar{\varepsilon}_{i}\right) \tag{16}$$

será estimado por mínimos quadrados ordinários. Para o *EFe* e o *EFtec* serão estimados por mínimos quadrados ordinários, respectivamente, os modelos

$$\ln w_{ijt} - \overline{\ln w_{J}} = \left(\boldsymbol{x}'_{ijt} - \overline{\boldsymbol{x}}'_{j}\right)\boldsymbol{\beta} + \left(\varepsilon_{ijt} - \bar{\varepsilon}_{j}\right) \tag{17}$$

$$\ln w_{ijt} - \overline{\ln w_{\iota J}} = \left(\boldsymbol{x}'_{ijt} - \overline{\boldsymbol{x}}'_{ij}\right)\boldsymbol{\beta} + \left(\varepsilon_{ijt} - \bar{\varepsilon}_{ij}\right) \tag{18}$$

Nestas equações em desvio, o subscrito da média corresponde a qual dimensão está fixa no cálculo da média. Desse modo, em (16) a média é calculada para o mesmo *i*, variando *j* e *t*. Em (17) a média é calculada para o mesmo *j*, variando *i* e *t*. Em (18), a média é calculada para o mesmo par (*i*, *j*), variando *t*. Neste último caso, o objetivo é restringir a regressão apenas à variação que ocorre dentro de um mesmo *match*. Em todas as estimações, a matriz de variância-covariância dos estimadores será robusta à correlação entre as observações dentro do grupo representado por cada trabalhador. Isto é, o *cluster* usado é o trabalhador.

[6] Para as derivações detalhadas das equações a seguir, ver o Apêndice C.

4 DADOS

A base de dados original é uma amostra de 2% da RAISMIGRA[7], com abrangência para todo o Brasil no período 1995 a 2002. Entretanto, a base em sua versão final, que foi a efetivamente utilizada nas análises, é uma sub-amostra que consiste em um painel balanceado[8] de 8 anos (1995 a 2002) para 68.547 trabalhadores (total de 68.547 · 8 = 548.376 pares *it*), que trabalharam para 68.226 estabelecimentos diferentes, formando um total de 99.907 *matches* diferentes. Cada trabalhador aparece os oito anos na amostra, mas um estabelecimento aparece somente enquanto algum trabalhador da amostra permanece empregado nele. Apenas o último vínculo empregatício de cada ano é registrado para o trabalhador. No caso de, ao final do ano, o trabalhador estar com dois empregos ativos, o emprego registrado é aquele com data de admissão mais antiga. O valor assumido por cada variável corresponde ao valor que ela assume ao final do ano relevante. Por exemplo, o salário em 1995 é dado pela remuneração do trabalhador em dezembro de 1995, medida em número de salários mínimos. O salário mínimo usado é aquele em vigor no ano relevante. Se o trabalhador não estiver com o emprego ativo em dezembro, o salário registrado é aquele do mês de desligamento. Nas nossas estimações o salário foi transformado de número de salários mínimos para salário real em reais (R$), a preços de 1995. O índice de inflação usado no cálculo do salário real foi o IPCA (Índice de Preços ao Consumidor Amplo).

O que identifica um *match* na amostra é o par (i , j) e não um vínculo empregatício ininterrupto no tempo. Assim, se um trabalhador esteve empregado no estabelecimento A de 1995 a 1997, mudou para o estabelecimento B e lá ficou de 1998 a 2000, e voltou a estar empregado por A em 2001 e em 2002, os dois vínculos que este trabalhador teve com A, embora separados no tempo, são tratados como um único vínculo. A escolha de tal abordagem é conveniente e condizente com as estimações. Primeiro, porque o efeito fixo da qualidade do *match* é constante no tempo, e deste ponto de vista não há por que tratar os dois vínculos como diferentes. Segundo, uma vez que a legislação brasileira impede o reajuste salarial nominal negativo durante o vínculo, mas não impede no caso de demissão e posterior

[7] A RAISMIGRA é obtida a partir da RAIS (Relação Anual de Informações Sociais). A RAIS é uma base de dados administrativa organizada pelo Ministério do Trabalho e Emprego, formada a partir de declarações enviadas a ele pelas empresas formais do país. Todos os empregadores, urbanos e rurais, e inclusive os estabelecimentos dotados de CNPJ sem empregados, devem declarar a RAIS.

[8] A Tabela 2A no Apêndice A indica que a amostra balanceada não é muito diferente da sua contrapartida não balanceada.

readmissão do trabalhador, a variação salarial no tempo para um mesmo *match* será mais fiel à noção teórica de otimização pelo empregador.

Cerca de 50% dos *matches* diferentes da amostra usada duram todos os oito anos. Os outros 50% não apresentam um padrão marcante de duração: o segundo padrão mais comum, formado pelos *matches* que ocorrem apenas em 1995 e não reaparecem pelo menos nos sete anos seguintes que a amostra cobre, representa apenas 5,02% da amostra. Há no mínimo 100 diferentes padrões destes relacionamentos de um mesmo empregado com um mesmo empregador ao longo dos oito anos da amostra.

Para a formação desta base de dados, foram descartados da base original trabalhadores das seguintes seções CNAE 1.0: indústrias extrativas; produção e distribuição de eletricidade, gás e água; transporte, armazenagem e comunicações; administração pública, defesa e seguridade social; educação; saúde e serviços sociais; outros serviços coletivos, sociais e pessoais; serviços domésticos; organismos internacionais e outras instituições extraterritoriais. Também foram descartados trabalhadores cuja divisão CNAE 1.0 era a 651, correspondente ao Banco Central do Brasil, localizada dentro da seção "Intermediação financeira, seguros, previdência complementar e serviços relacionados", cujas outras divisões permaneceram na amostra. O objetivo desta eliminação de observações foi restringir a amostra a trabalhadores de estabelecimentos voltados unicamente à maximização de lucro. Assim, seções CNAE 1.0 com potencialmente maior participação de empregados no setor público e em organizações que não necessariamente maximizam lucro, como Organizações Não Governamentais, entre outros tipos de empregador, foram descartadas. Todas as outras seções e divisões CNAE 1.0 não descartadas estão representadas na amostra utilizada. A Tabela 1A no Apêndice A resume características das variáveis utilizadas.

A base de dados não possui variáveis que descrevem o capital do empregador. Por isso, não foi possível controlar para este fator de produção, seja na forma de estoque de capital físico, na forma de intensidade de capital na produção, etc. Consequentemente, vieses de variável omitida decorrentes da ausência deste tipo de controle estarão presentes nos resultados sem que possamos, no entanto, evitar tal problema. Vale salientar que os trabalhos empíricos da literatura de efeito do tamanho do empregador sobre o salário em geral sofrem de limitações parecidas, sendo incapazes de controlar para variáveis potencialmente relevantes mas ausentes de suas bases de dados.

5 RESULTADOS

Os resultados estão organizados em três subseções. A primeira subseção apresenta resultados para as quatro especificações (*POLS*, *EFt*, *EFe* e *EFtec*), com o fim de compararmos os seus resultados entre si antes de iniciarmos os testes dos modelos teóricos. Chamaremos os resultados desta primeira subseção de resultados para a "especificação básica". Esta especificação possui os seguintes controles: *dummies* para oito categorias de tamanho do estabelecimento, *dummies* para grau de instrução do trabalhador, *dummies* para faixa etária do trabalhador, tempo no emprego (*tenure*), *dummies* para unidade federativa do estabelecimento empregador, *dummies* para região metropolitana do estabelecimento empregador, *dummies* para indústria, *dummies* para ocupação, *dummies* para ano e *dummy* para emprego temporário. Nenhuma das variáveis explicativas está em logaritmo, somente a variável dependente (salário) está. A categoria omitida de tamanho do estabelecimento é "até 4 empregados". As duas outras subseções realizam os testes da teoria.

O Apêndice D, ao final do trabalho, contém tabelas e gráficos de resultados correspondentes aos presentes nesta seção, mas obtidos para a amostra que inclui todas as seções CNAE 1.0 (incluindo setor público, ONGs, etc.)

5.1 Especificação básica

A Tabela 2 a seguir contém os resultados da especificação básica. O Gráfico 1 contém os coeficientes para cada tamanho de estabelecimento, para cada tipo de modelo estimado.

Tabela 2 – Especificação básica

Variável dependente: logaritmo do salário

	POLS	*EFt*	*EFe*	*EFtec*
Número de empregados				
5 a 9	0.1780***	0.0743***	0.0291***	0.0316***
	(0.0080)	(0.0061)	(0.0060)	(0.0052)
10 a 19	0.2712***	0.1248***	0.0623***	0.0615***
	(0.0082)	(0.0078)	(0.0077)	(0.0068)
20 a 49	0.3667***	0.1746***	0.0836***	0.0898***
	(0.0080)	(0.0087)	(0.0095)	(0.0081)
50 a 99	0.4549***	0.2044***	0.0961***	0.1079***
	(0.0089)	(0.0096)	(0.0117)	(0.0095)
100 a 249	0.5586***	0.2396***	0.1047***	0.1212***
	(0.0086)	(0.0099)	(0.0133)	(0.0105)
250 a 499	0.6203***	0.2723***	0.1167***	0.1483***
	(0.0092)	(0.0104)	(0.0154)	(0.0114)
500 a 999	0.6395***	0.2730***	0.0998***	0.1543***
	(0.0096)	(0.0107)	(0.0174)	(0.0123)
1000 ou mais	0.6764***	0.2959***	0.1018***	0.1921***
	(0.0092)	(0.0112)	(0.0203)	(0.0133)
N		68 547		
J		68 226		
T		8		
N·T		548 376		
n^0 *matches* diferentes		99 907		

Nota 1: erros-padrão abaixo dos coeficientes, entre parênteses.
Nota 2: * = p-valor < 0.1 ** = p-valor < 0.05 *** = p-valor < 0.01
Nota 3: controles presentes em todas as regressões: *dummies* para tamanho do estabelecimento, grau de instrução do trabalhador, faixa etária do trabalhador, tempo no emprego (*tenure*), unidade federativa do estabelecimento empregador, região metropolitana do estabelecimento empregador, indústria (seção CNAE 1.0), ocupação (grandes grupos CBO 1994), *dummies* para ano e *dummy* para emprego temporário.

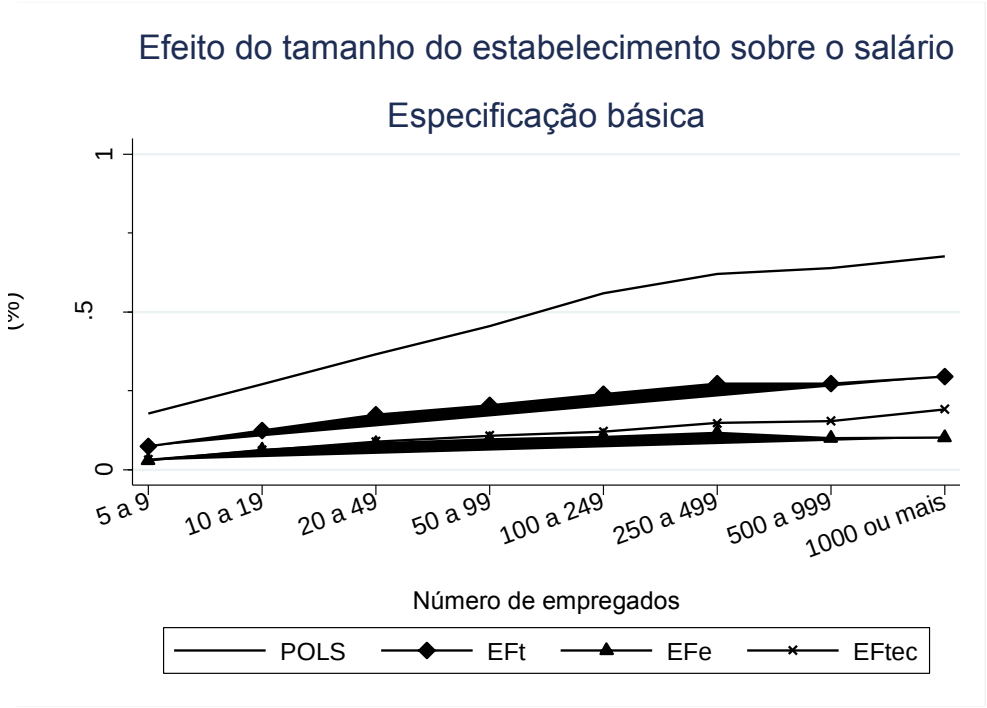

Gráfico 1 – Especificação básica

Os resultados obtidos para o efeito do tamanho do estabelecimento (ETE) seguem o seguinte padrão geral. Todos os coeficientes são significantes a 1%. Independentemente da especificação, o ETE é sempre crescente com o tamanho do estabelecimento, com exceção dos resultados para o *EFe*, em que a categoria "500 a 999 empregados" possui coeficiente levemente menor que o da categoria "250 a 499 empregados". Os ETE estimados por *POLS* são os maiores, qualquer que seja o tamanho de estabelecimento considerado. Do lado oposto, os ETE estimados por *EFe* são os menores, para todo tamanho de estabelecimento, exceto para a categoria "10 a 19 empregados", cujo coeficiente é ligeiramente maior que o seu correspondente no *EFtec*. Se ignorarmos esta ligeira diferença, podemos dizer que nenhuma das curvas no gráfico se cruza com outra, fato este que facilita as interpretações a seguir. Muito embora os coeficientes de interesse sejam um vetor menor contido em $\boldsymbol{\beta}$, nas interpretações usaremos álgebra matricial envolvendo todo o vetor $\boldsymbol{\beta}$ com o fim de facilitar a leitura, ao evitar detalhamentos desnecessários dos cálculos.

Comecemos comparando a queda nos coeficientes que ocorre quando passamos do *POLS* para o *EFt*. Subtraindo-se a equação (II*) da equação (I*) e usando a propriedade $\boldsymbol{M_A A} = \mathbf{0}$, podemos dizer que os resultados empíricos obtidos acima fornecem evidências de que

$$\boldsymbol{\beta}_{POLS} - \boldsymbol{\beta}_{EFt} > \mathbf{0} \Rightarrow$$

$$\Rightarrow (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{D}\boldsymbol{\theta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{F}\boldsymbol{\psi} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{G}\boldsymbol{\phi} - \\ -(\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{D}}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{D}}\boldsymbol{F}\boldsymbol{\psi} - (\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{D}}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{D}}\boldsymbol{G}\boldsymbol{\phi} > \mathbf{0} \Rightarrow$$

$$\Rightarrow (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'[\boldsymbol{D}\boldsymbol{\theta} + \boldsymbol{F}\boldsymbol{\psi} + \boldsymbol{G}\boldsymbol{\phi}] \ > \ (\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{D}}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{D}}[\boldsymbol{F}\boldsymbol{\psi} + \boldsymbol{G}\boldsymbol{\phi}] \Rightarrow$$

$$\Rightarrow (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'[\boldsymbol{D}\boldsymbol{\theta} + \boldsymbol{F}\boldsymbol{\psi} + \boldsymbol{G}\boldsymbol{\phi}] \ > \ (\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{D}}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{D}}[\boldsymbol{D}\boldsymbol{\theta} + \boldsymbol{F}\boldsymbol{\psi} + \boldsymbol{G}\boldsymbol{\phi}] \qquad (19)$$

ou seja, pensando nas variáveis de $\boldsymbol{X}$ que nos interessam – que são as categorias de tamanho do empregador –, temos evidências de que a soma das correlações não condicionais do tamanho do empregador com os três efeitos fixos é maior que a mesma soma de correlações, mas condicionais na identidade dos trabalhadores. Isto significa que a correlação entre o tamanho do empregador e os três efeitos fixos tende a ser maior na população do que na história de dado trabalhador da amostra, e quanto maior esta divergência, maior a diferença entre os ETE estimados por *POLS* e os estimados por *EFt*. Desta forma, ao controlarmos para o efeito fixo do trabalhador, isto é, ao restringirmos a estimação à variação sofrida pelos dados somente na história de cada trabalhador, evitamos parte do viés do efeito do tamanho do empregador, parte esta oriunda da correlação desta variável com as heterogeneidades não observáveis no nível da população.

Comparando os coeficientes do *EFt* com os do *EFe*, o mesmo raciocínio anterior pode ser feito partindo de (II*) e (III*) para dizermos que temos evidências de que, como fica claro no Gráfico 1,

$$\boldsymbol{\beta}_{EFt} - \boldsymbol{\beta}_{EFe} > \mathbf{0} \Rightarrow$$

$$\Rightarrow (\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{D}}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{D}}[\boldsymbol{F}\boldsymbol{\psi} + \boldsymbol{G}\boldsymbol{\phi}] \ > \ (\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{F}}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{F}}[\boldsymbol{D}\boldsymbol{\theta} + \boldsymbol{G}\boldsymbol{\phi}] \Rightarrow$$

$$\Rightarrow (\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{D}}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{D}}[\boldsymbol{D}\boldsymbol{\theta} + \boldsymbol{F}\boldsymbol{\psi} + \boldsymbol{G}\boldsymbol{\phi}] \ > \ (\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{F}}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{F}}[\boldsymbol{D}\boldsymbol{\theta} + \boldsymbol{F}\boldsymbol{\psi} + \boldsymbol{G}\boldsymbol{\phi}] \qquad (20)$$

isto é, temos evidências de que a soma das correlações, condicionais na identidade dos trabalhadores, do tamanho do empregador com os três efeitos fixos, é maior do que a mesma

soma de correlações, mas condicionais na identidade dos empregadores. Isto significa que a correlação do tamanho do empregador com os três efeitos fixos tende a ser maior na história de um dado trabalhador do que através da população de trabalhadores de um dado empregador – incluindo todos os empregados em todos os períodos de tempo. Em outras palavras, as mudanças sofridas pelo tamanho do empregador ao longo da história de dado trabalhador (seja no mesmo empregador, seja ao migrar entre empregadores) tendem a ser acompanhadas por mudanças na soma dos três efeitos fixos com maior intensidade do que as mudanças sofridas pelo tamanho do empregador ao longo da história de dado empregador.

Nas duas comparações acima (de *POLS* com *EFt* e de *EFt* com *EFe*), não sabemos qual das três correlações – com $\boldsymbol{\theta}$, com $\boldsymbol{\psi}$ ou com $\boldsymbol{\phi}$ – tende a ser maior em valores absolutos, nem qual tende a ser o sinal de cada uma delas. Com as duas próximas comparações, de *EFt* e de *EFe* com *EFtec*, exploraremos alguns destes detalhes.

Note no Gráfico 1 que os coeficientes do *EFtec* situam-se entre os do *EFt* e do *EFe*. Assim, usando (II*), (III*) e (IV*) podemos dizer que temos evidências de que as seguintes duas desigualdades são verdadeiras:

$$\boldsymbol{\beta}_{EFt} - \boldsymbol{\beta}_{EFtec} > \mathbf{0} \Longrightarrow (\boldsymbol{X}'\boldsymbol{M_D}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M_D}\boldsymbol{F}\boldsymbol{\psi} + (\boldsymbol{X}'\boldsymbol{M_D}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M_D}\boldsymbol{G}\boldsymbol{\phi} > \mathbf{0} \Longrightarrow$$

$$\Longrightarrow (\boldsymbol{X}'\boldsymbol{M_D}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M_D}[\boldsymbol{F}\boldsymbol{\psi} + \boldsymbol{G}\boldsymbol{\phi}] > \mathbf{0} \Longrightarrow \tag{21}$$

$$\Longrightarrow (\boldsymbol{X}'\boldsymbol{M_D}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M_D}[\boldsymbol{D}\boldsymbol{\theta} + \boldsymbol{F}\boldsymbol{\psi} + \boldsymbol{G}\boldsymbol{\phi}] > \mathbf{0}$$

$$\boldsymbol{\beta}_{EFe} - \boldsymbol{\beta}_{EFtec} < \mathbf{0} \Longrightarrow (\boldsymbol{X}'\boldsymbol{M_F}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M_F}\boldsymbol{D}\boldsymbol{\theta} + (\boldsymbol{X}'\boldsymbol{M_F}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M_F}\boldsymbol{G}\boldsymbol{\phi} < \mathbf{0} \Longrightarrow$$

$$\Longrightarrow (\boldsymbol{X}'\boldsymbol{M_F}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M_F}[\boldsymbol{D}\boldsymbol{\theta} + \boldsymbol{G}\boldsymbol{\phi}] < \mathbf{0} \Longrightarrow \tag{22}$$

$$\Longrightarrow (\boldsymbol{X}'\boldsymbol{M_F}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M_F}[\boldsymbol{D}\boldsymbol{\theta} + \boldsymbol{F}\boldsymbol{\psi} + \boldsymbol{G}\boldsymbol{\phi}] < \mathbf{0}$$

Agora podemos dizer que a soma das correlações do tamanho do empregador com os três efeitos fixos (lado esquerdo de (19)) tende a ser positiva. Ademais, podemos dizer que a

desigualdade em (20) tende a valer porque, na história de dado trabalhador, a soma das correlações do tamanho do empregador com os efeitos do empregador e da qualidade do *match* tende a ser positiva e, no nível do empregador, a soma das correlações do tamanho do empregador com os efeitos dos trabalhadores e da qualidade do *match* tende a ser negativa.

Se tivéssemos ignorado a existência de heterogeneidade não observável na qualidade dos *matches*, diríamos de (21) que ao longo da história de dado trabalhador, conforme ele migrasse de empregadores menores para maiores, o efeito fixo do empregador tenderia a crescer, explicando parte do ETE. Contudo, existindo o efeito da qualidade do *match*, devemos reconhecer que outras interpretações são possíveis e igualmente razoáveis. Por exemplo, é possível que ao longo de sua história de migração entre empregadores, o trabalhador que migra para empregadores maiores tenda a fazê-lo por esperar melhor aproveitamento (e com isso melhor remuneração) de suas habilidades específicas, e não necessariamente por esperar ter um poder de barganha maior no novo empregador.

Da mesma forma, ao ignorar a existência de heterogeneidade não observável na qualidade dos *matches*, diríamos de (22) que conforme um empregador crescesse, tenderia a contratar trabalhadores cada vez menos habilidosos. Esta tendência é o oposto do previsto por Barron *et al* (1987), que preveem maior seletividade nas contratações por parte dos grandes empregadores. Porém, se aceitarmos a existência do efeito da qualidade do *match*, teremos espaço para interpretações menos contra intuitivas. Por exemplo, poderíamos dizer que muito embora empregadores maiores tendessem a contratar trabalhadores menos habilidosos, tais empregadores tenderiam a conhecer e alocar sua mão-de-obra de maneira mais eficiente internamente, obtendo desta alocação melhor aproveitamento de habilidades específicas dos trabalhadores e, com isso, ganhos de produtividade com uma melhor relação custo-benefício do que se adotassem a política de contratar somente trabalhadores altamente habilidosos.

Para finalizar esta seção, podemos dizer que controlar para a heterogeneidade não observável nos três níveis citados não é suficiente para eliminar todo o efeito positivo do tamanho do estabelecimento sobre o salário. Apesar disso, a comparação do resultado do *POLS* com o do *EFtec* sugere que mais da metade do efeito do tamanho do empregador é explicada pela correlação desta variável com as três heterogeneidades.

5.2 Qualificação agregada da mão-de-obra no nível do estabelecimento

No modelo de Kremer (1993), empregadores maiores pagam salário maior para seus trabalhadores tanto porque estes são mais habilidosos quanto porque a habilidade agregada da mão-de-obra no nível do empregador é maior e esta tem, em si, um efeito positivo adicional sobre a produtividade. O aspecto relevante da "habilidade" no modelo é o quanto que ela determina a probabilidade de erro na tarefa desempenhada pelo trabalhador. Podemos dizer aqui que é razoável assumir que além da habilidade não observável, também o grau de instrução afeta esta probabilidade, sendo que trabalhadores com mais instrução errariam, em média, menos ao executar suas tarefas. Assim, se controlássemos para a qualificação – no sentido de grau de instrução – agregada da mão-de-obra no nível do estabelecimento na equação salarial, deveríamos encontrar ETEs menores que os encontrados na especificação básica.

O caminho que seguimos não foi o de inserir um controle na equação salarial, mas sim o de restringir a amostra a uma sub-amostra na qual a variação da qualificação agregada é menor. Mais especificamente, restringimos a amostra a estabelecimentos nos quais os trabalhadores qualificados são maioria. Se a previsão da teoria estiver certa, os ETEs estimados nesta sub-amostra devem ser menores que os estimados sobre a amostra total, pois os ETEs medem diferença salarial relativamente ao grupo de referência (tamanho do estabelecimento igual a "até 4 empregados"), o qual, nesta sub-amostra, será mais semelhante às outras categorias de tamanho do empregador quanto à qualificação agregada de sua mão-de-obra.

Para construir a sub-amostra, o primeiro passo foi criar oito variáveis contínuas, cada uma igual ao percentual de trabalhadores pertencente a cada uma das oito categorias de grau de instrução, no nível do estabelecimento em cada ano. Estes percentuais foram calculados a partir dos trabalhadores disponíveis na amostra. Por exemplo, se a amostra possui 5 empregados em 1997 para o estabelecimento A, dos quais 2 têm curso superior completo e 3 têm o ensino médio completo, então o valor da variável "percentual com curso superior completo" será 0,4 para A em 1997, e o valor da variável "percentual com ensino médio completo" será 0,6 para A em 1997; as outras seis variáveis que medem percentual serão todas iguais a zero para A em 1997. Em anos diferentes, todas as oito variáveis podem mudar de valor para A, conforme ocorrem contratações e demissões.

O segundo passo da construção da sub-amostra foi descartar todas as observações correspondentes a trabalhadores empregados, em qualquer ano, em estabelecimentos com mais de 50% dos empregados que tivessem qualificação menor ou igual a "até 8ª série completo". O grupo de observações que restou foi daqueles trabalhadores empregados, em cada ano do período 1995 a 2002, em estabelecimentos que continham 50% ou mais de seus trabalhadores com qualificação igual ou maior que "ensino médio incompleto".

Como em nossa amostra 90% dos estabelecimentos têm no máximo oito empregados observados em cada ano, os percentuais calculados medem o percentual verdadeiro com imprecisão, pois levam em conta somente os trabalhadores presentes na amostra, e não todos os trabalhadores que o estabelecimento de fato tinha em cada ano. Por isso, a sub-amostra que construímos pode conter, não intencionalmente, observações correspondentes a empregados de estabelecimentos que não têm maioria de trabalhadores qualificados, como pretendido.

O componente da habilidade do modelo de Kremer (1993) que se poderia atribuir a habilidades não observáveis é controlado através do efeito fixo do trabalhador nas especificações *EFt* e *EFtec*. Infelizmente, nenhuma de nossas especificações controla para algum tipo de medida de "habilidade não observável agregada" no nível do estabelecimento – a não ser aquela eventualmente captada pelo efeito fixo do empregador, se constante no tempo. Abaixo, a Tabela 3 e o Gráfico 2 com os resultados para a sub-amostra construída segundo o procedimento explicado acima.

Os ETEs estimados por *EFtec*, que é a especificação sem vieses de variável omitida, são, de forma geral, muito próximos dos obtidos para a amostra na especificação básica da seção anterior, e são maiores que os ETEs da Tabela 2. Assim, nossa evidência empírica nos leva a rejeitar a teoria de Kremer (1993).

Se a análise tivesse sido feita usando-se apenas a especificação que não controla para nenhum dos três efeitos fixos, o *POLS*, teríamos encontrado ETEs claramente maiores para a sub-amostra desta seção, em clara contradição com a previsão da teoria de Kremer (1993). Desta forma, teríamos rejeitado a teoria do autor, assim como fizemos com base nos resultados do *EFtec*. Se a análise tivesse usado apenas o *EFt*, também teríamos chegado à mesma conclusão a que chegamos com o *EFtec*, dada a ausência de um padrão claro entre os ETEs das Tabelas

2 e 3 para o *EFt*, em contradição com o que se esperaria se a teoria fosse válida. Se a análise tivesse usado somente o *EFe*, novamente diríamos que os resultados contradisseram a teoria, ainda que com menos clareza que no *POLS*, dada a maior proximidade entre os ETEs estimados nas Tabelas 2 e 3 para o *EFe*, especialmente para as categorias de estabelecimentos menores. Em suma, o modelo de Kremer (1993) é rejeitado em qualquer uma das especificações estimadas.

Tabela 3 – Estabelecimentos com mais instruídos

Variável dependente: logaritmo do salário

	POLS	*EFt*	*EFe*	*EFtec*
Número de empregados				
5 a 9	0.2888***	0.1011***	0.0486***	0.0445***
	(0.01875)	(0.01369)	(0.01165)	(0.01093)
10 a 19	0.5330***	0.1553***	0.0880***	0.0809***
	(0.01854)	(0.01719)	(0.01535)	(0.01436)
20 a 49	0.6657***	0.1986***	0.1014***	0.0996***
	(0.01804)	(0.01912)	(0.01816)	(0.01674)
50 a 99	0.7657***	0.2126***	0.1111***	0.1090***
	(0.01986)	(0.02070)	(0.02199)	(0.01947)
100 a 249	0.8893***	0.2411***	0.1366***	0.1304***
	(0.01873)	(0.02083)	(0.02458)	(0.02158)
250 a 499	0.9199***	0.2672***	0.1455***	0.1601***
	(0.01961)	(0.02156)	(0.02914)	(0.02325)
500 a 999	0.9481***	0.2630***	0.1602***	0.1586***
	(0.01929)	(0.02179)	(0.03296)	(0.02476)
1000 ou mais	0.9142***	0.2858***	0.1765***	0.2022***
	(0.01820)	(0.02207)	(0.03712)	(0.02603)
N		22 671		
J		20 306		
T		8		
N·T		181 368		
nº *matches* diferentes		33 666		

Nota 1: erros-padrão abaixo dos coeficientes, entre parênteses.
Nota 2: * = p-valor < 0.1 ** = p-valor < 0.05 *** = p-valor < 0.01
Nota 3: controles presentes em todas as regressões: *dummies* para tamanho do estabelecimento, grau de instrução do trabalhador, faixa etária do trabalhador, tempo no emprego (*tenure*), unidade federativa do estabelecimento empregador, região metropolitana do estabelecimento empregador, indústria (seção CNAE 1.0), ocupação (grandes grupos CBO 1994), *dummies* para ano e *dummy* para emprego temporário.

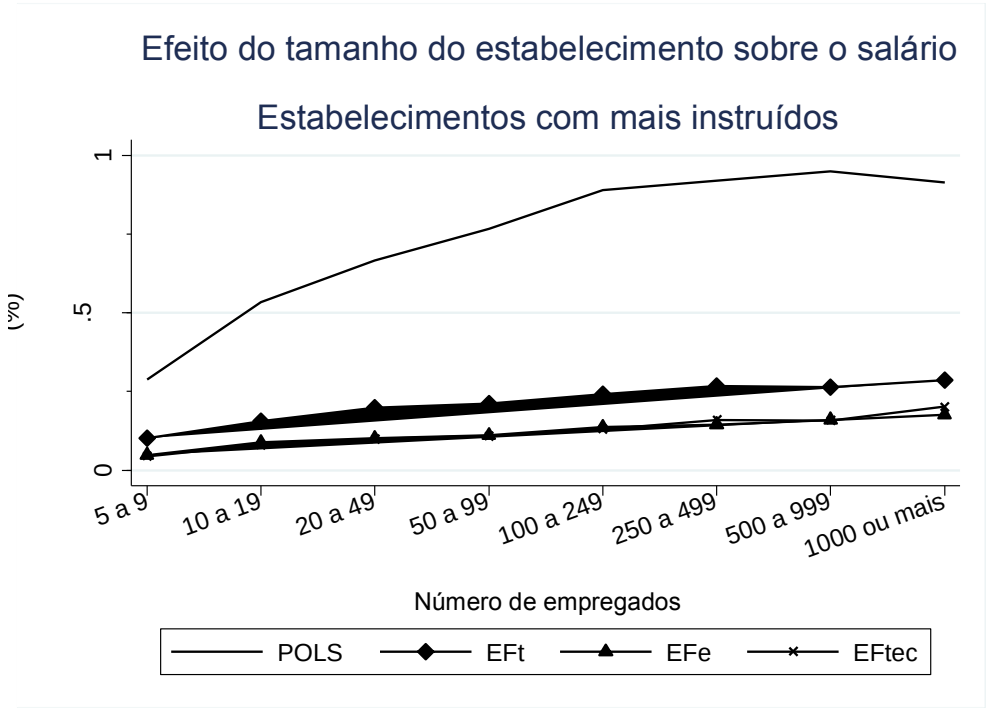

Gráfico 2 – Estabelecimentos com mais instruídos

Note a proximidade entre os ETEs estimados na Tabela 3 para o *EFe* e o *EFtec*, visível no Gráfico 2. Com base em (22), o significado desta proximidade é que temos evidências de que a correlação do tamanho do empregador com a heterogeneidade dos trabalhadores é o oposto da correlação do tamanho com a qualidade dos *matches*, correlações estas no nível do empregador. Em outras palavras, no grupo dos estabelecimentos com maioria qualificada, o crescimento de um estabelecimento tende a ser acompanhado, na recomposição de sua mão-de-obra, de movimentos opostos – mas de intensidade aproximada – na habilidade dos trabalhadores e na qualidade dos *matches* formados. Como vimos anteriormente no caso da especificação básica, na amostra total, que contém estabelecimentos com maioria qualificada ou não, o crescimento dos estabelecimentos tende a ser acompanhado por contratações cujo saldo em termos de habilidade e qualidade do *match* é negativo. Aqui, este saldo é aproximadamente zero, de forma que os vieses sobre o ETE anulam um ao outro na especificação *EFe*. Se nesta interpretação tivéssemos ignorado a existência de efeitos da qualidade do *match*, teríamos dito erroneamente que a correlação do tamanho do empregador com a habilidade dos trabalhadores tenderia a ser zero no nível do empregador no grupo dos empregadores com maioria qualificada, e que portanto não seria necessário controlar para o

efeito fixo do trabalhador nas estimações deste grupo, bastando controlar apenas para o efeito fixo do empregador.

5.2.1 Avaliação mais rigorosa

As comparações feitas acima entre os coeficientes da Tabela 3 e os da Tabela 2 não foram rigorosas, isto é, não se avaliou a significância estatística da diferença entre eles. Todavia, as conclusões anteriores se mostram novamente válidas após uma avaliação estatisticamente mais rigorosa, como vemos a seguir. A Tabela 4 apresenta os coeficientes das categorias de tamanho do estabelecimento, bem como os coeficientes das interações entre cada tamanho de estabelecimento e uma variável *dummy* igual a 1 se o trabalhador está empregado em um estabelecimento com alta qualificação agregada (alta no sentido usado para as regressões da Tabela 3) e 0 caso contrário. Tal estimação foi feita sobre a amostra total (a mesma usada na estimação da Tabela 2).

Podemos ver que cinco dos coeficientes de interação são estatisticamente significantes. Contudo, enquanto os três correspondentes às menores categorias de tamanho do empregador são positivos, os dois correspondentes às maiores categorias são negativos. Temos, com isso, evidência empírica novamente contraditória à previsão teórica de Kremer (1993).

Tabela 4 – Interação com qualificação agregada

Variável dependente: logaritmo do salário

Número de empregados		Interação de tamanho com qualificação agregada
5 a 9	0.0224***	0.0274***
	(0.0058)	(0.0081)
10 a 19	0.0505***	0.0319***
	(0.0071)	(0.0087)
20 a 49	0.0845***	0.0178**
	(0.0083)	(0.0085)
50 a 99	0.1087***	0.0026
	(0.0097)	(0.0081)
100 a 249	0.1234***	-0.0011
	(0.0106)	(0.0062)
250 a 499	0.1532***	-0.0070
	(0.0116)	(0.0057)
500 a 999	0.1683***	-0.0241***
	(0.0126)	(0.0057)
1000 ou mais	0.2099***	-0.0299***
	(0.0137)	(0.0055)
N	68 547	
J	68 226	
T	8	
N·T	548 376	
n^0 *matches* diferentes	99 907	

Nota 1: erros-padrão abaixo dos coeficientes, entre parênteses.
Nota 2: * = p-valor < 0.1 ** = p-valor < 0.05 *** = p-valor < 0.01
Nota 3: os coeficientes são de uma especificação *EFtec* única. A primeira coluna da tabela contém os coeficientes das *dummies* de tamanho do empregador, e a segunda coluna os coeficientes de interações de *dummies* do tamanho do empregador com uma *dummy* indicando se o trabalhador pertence a um estabelecimento com maioria qualificada (valor 1) ou não (valor 0).
Nota 4: controles presentes na regressão: *dummies* para tamanho do estabelecimento, *dummies* de interação (ver Nota 3), grau de instrução do trabalhador, faixa etária do trabalhador, tempo no emprego (*tenure*), unidade federativa do estabelecimento empregador, região metropolitana do estabelecimento empregador, indústria (seção CNAE 1.0), ocupação (grandes grupos CBO 1994), *dummies* para ano e *dummy* para emprego temporário.

5.3 População de trabalhadores

No modelo de Bulow e Summers (1986), a relação positiva entre tamanho do empregador e salário pago é resultado de uma otimização sob restrição por parte dos empregadores, cuja equação de restrição, uma "condição de garantia de esforço" (*no-shirk condition*), podemos

representar aqui em nosso texto por $w - \overline{w} = f(N, q, d_2)$, onde w é o salário pago, $\overline{w}$ é a produtividade marginal do trabalho, N é a população de trabalhadores do mercado de trabalho, q é a taxa exógena de desligamentos entre empregador e empregado em cada período e d_2 é a probabilidade de detecção do esforço baixo (nulo) pelo empregador. Tal condição é necessária para o resultado dos autores, e indica qual o diferencial de salário que será pago em relação à produtividade marginal do trabalho (produtividade esta que é igual para todo trabalhador da população), em função de N, q e d_2, para garantir que todos os empregados realizem esforço pleno. A condição de *no-shirk* possui a seguinte propriedade: $f_N(N, q, d_2) < 0$. Ou seja, em mercados de trabalho com mais trabalhadores o efeito do tamanho do empregador é menor.

No mesmo sentido, tanto Weiss e Landau (1984) quanto Strand (1987) e Burdett e Mortensen (1998) fornecem a relação entre tamanho do empregador e salário, seja de forma explícita algebricamente ou apenas qualitativamente. Em todos os três casos, a população de trabalhadores é um parâmetro do modelo, que se aumentado, leva à redução do efeito do tamanho do empregador sobre o salário.

Usando como base esta previsão comum dos modelos, verificamos se a magnitude do ETE é menor em mercados de trabalho com mais trabalhadores. Para isso, estimamos a especificação *EFtec* separadamente para nove sub-amostras, correspondentes às nove regiões metropolitanas (RM) discriminadas na amostra total: Belo Horizonte, Fortaleza, São Paulo, Rio de Janeiro, Belém, Recife, Porto Alegre, Curitiba e Salvador. As observações que não pertencem a nenhuma das nove RMs foram descartadas na análise desta seção. Também nos restringimos apenas aos trabalhadores que permaneceram os oito anos empregados numa mesma RM, de forma que temos uma sub-amostra balanceada para cada uma delas. Abaixo, na Tabela 5, estão os resultados das estimações. O Gráfico 3 apresenta os ETEs de quatro RMs: São Paulo, Rio de Janeiro, Porto Alegre e Curitiba.

O tamanho das sub-amostras difere bastante entre as RMs, sendo proporcional às suas populações, o que talvez em parte explique por que os coeficientes não se mostraram significantes nas RMs menores como nas RMs de Belém e de Fortaleza. Assim, podemos nos focar na comparação entre as RMs do Rio de Janeiro (RMRJ) e de São Paulo (RMSP), que são as RMs com maior número de observações e possuem mercados de trabalho com números

diferentes de trabalhadores, como sugere o fato de que a RMRJ teve população total[9] de cerca de 60% daquela da RMSP no período 1995 a 2002. Os resultados da comparação da RMRJ com a RMSP na Tabela 6 são consistentes com a previsão dos modelos teóricos citados acima de que o ETE é menor em mercados de trabalho com mais trabalhadores. Note que o ETE é menor na RMSP para todo tamanho de estabelecimento.

Nem toda comparação dois a dois entre as RMs nos leva à mesma conclusão sobre a validade da teoria. Por exemplo, a região metropolitana de Curitiba (RMC) teve cerca de 73% da população total da região metropolitana de Porto Alegre (RMPA) no período 1995 a 2002, de forma que o mercado de trabalho da RMC provavelmente contou com menos trabalhadores que a RMPA no mesmo período. Porém, comparando os ETEs das duas RMs na Tabela 5, vemos que os ETEs são maiores na RMPA do que na RMC, em especial nas categorias intermediárias de tamanho do estabelecimento, o que contradiz a teoria.

[9] Esta e outras proporções foram calculadas a partir de estimativas da população residente em 1º de julho para os anos de 1995 a 2002, estimativas estas do IBGE (Instituto Brasileiro de Geografia e Estatística) obtidas do Ipeadata. Ver citação detalhada da fonte Ipeadata nas Referências.

Tabela 5 – Especificação *EFtec* para sub-amostras de 9 regiões metropolitanas

Variável dependente: logaritmo do salário

	Belo Horizonte	*Porto Alegre*	*Rio de Janeiro*	*Belém*	*Curitiba*	*Fortaleza*	*Recife*	*Salvador*	*São Paulo*
Número de empregados									
5 a 9	0.0078	0.0519	0.0494***	-0.0472	0.0169	-0.0269	0.0056	0.0452	0.0252*
	(0.0240)	(0.0410)	(0.0164)	(0.0853)	(0.0378)	(0.0279)	(0.0257)	(0.0471)	(0.0143)
10 a 19	0.0463	0.0936	0.1005***	-0.0286	0.0914*	0.0269	0.0161	0.0272	0.0385**
	(0.0327)	(0.0580)	(0.0222)	(0.0987)	(0.0501)	(0.0403)	(0.0370)	(0.0670)	(0.0179)
20 a 49	0.0951***	0.1908***	0.1409***	0.0412	0.0655	0.0118	0.0956**	-0.0035	0.0652***
	(0.0344)	(0.0693)	(0.0265)	(0.1136)	(0.0588)	(0.0461)	(0.0487)	(0.0843)	(0.0201)
50 a 99	0.0953	0.1999**	0.1552***	0.2630	0.0724	0.0372	0.1121*	-0.0293	0.0837***
	(0.0590)	(0.0798)	(0.0304)	(0.1975)	(0.0642)	(0.0627)	(0.0580)	(0.0962)	(0.0225)
100 a 249	0.0674	0.2188**	0.1573***	0.2701	0.1142*	0.1028	0.0874	0.0490	0.1024***
	(0.0591)	(0.0880)	(0.0333)	(0.2543)	(0.0687)	(0.0734)	(0.0621)	(0.1045)	(0.0243)
250 a 499	0.0099	0.2732***	0.2149***	0.3715	0.1575**	0.0632	0.1497**	0.0221	0.1350***
	(0.0647)	(0.0892)	(0.0360)	(0.2497)	(0.0754)	(0.0844)	(0.0656)	(0.1050)	(0.0265)
500 a 999	0.0246	0.2881***	0.2222***	0.2295	0.1912**	0.1064	0.1595**	0.1513	0.1578***
	(0.0692)	(0.0916)	(0.0393)	(0.2607)	(0.0821)	(0.0899)	(0.0701)	(0.1062)	(0.0289)
1000 ou mais	0.0612	0.3049***	0.2821***	0.3059	0.2459***	-0.0073	0.1980***	0.3759***	0.1784***
	(0.0753)	(0.0937)	(0.0451)	(0.2834)	(0.0864)	(0.0975)	(0.0765)	(0.1173)	(0.0306)
N	2 219	2 125	6 186	664	1 283	949	1 571	1 493	13 498
J	2 277	2 090	6 105	612	1 382	949	1 486	1 403	13 092
T	8	8	8	8	8	8	8	8	8
N·T	17 752	17 000	49 488	5 312	10 264	7 592	12 568	11 944	107 984
nº *matches* diferentes	3 305	2 991	8 514	794	1 894	1 269	2 059	1 943	19 228

Nota: aplicam-se as notas 1 e 2 da Tabela 2. Controles presentes em todas as regressões: *dummies* para tamanho do estabelecimento, grau de instrução do trabalhador, faixa etária do trabalhador, tempo no emprego (*tenure*), indústria (seção CNAE 1.0), ocupação (grandes grupos CBO 1994), *dummies* para ano e *dummy* para emprego temporário.

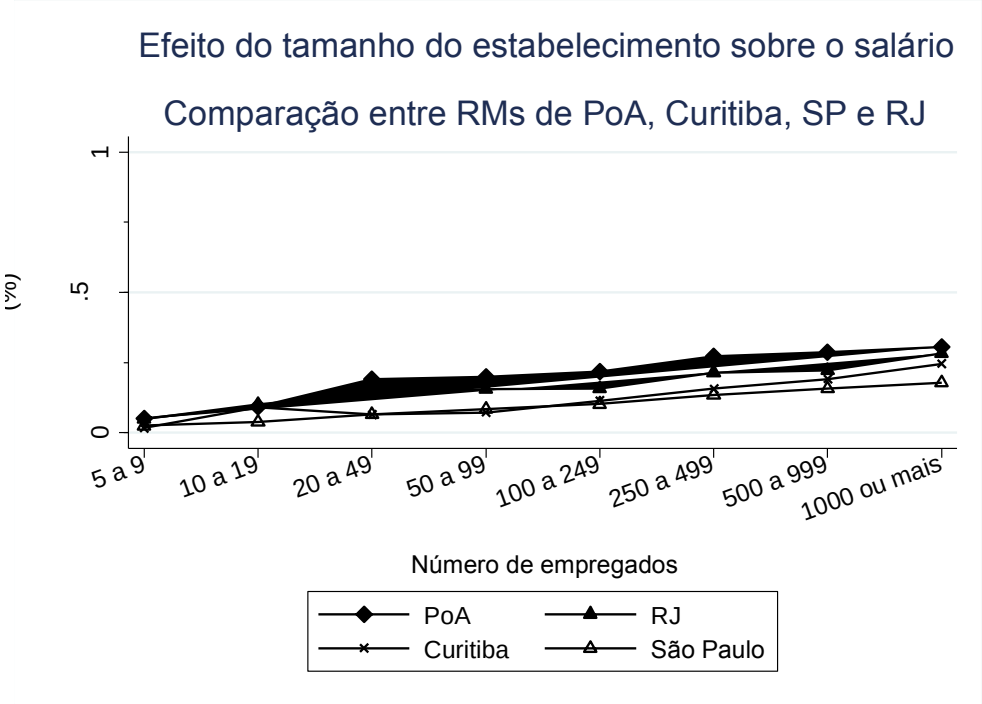

Gráfico 3 – Comparação entre RMs de PoA, Curitiba, SP e RJ
Nota ao Gráfico 3: PoA = Porto Alegre.

As Tabelas 6 e 7 abaixo mostram os resultados para as especificações *POLS*, *EFt*, *EFe* e *EFtec* para sabermos o que aconteceria se tivéssemos comparado os dois pares de RMs (RMRJ com RMSP e RMC com RMPA) através de especificações que não controlam para os três efeitos fixos simultaneamente, e que portanto teriam estimadores potencialmente viesados. Se tivéssemos feito a análise por *POLS*, teríamos rejeitado a previsão dos modelos teóricos tanto no caso da comparação da RMRJ com a RMSP, quanto da comparação da RMC com a RMPA. Isto é, nossa conclusão no caso da comparação da RMRJ com a RMSP seria o oposto da análise com *EFtec*, e na comparação da RMC com a RMPA, não teríamos, como temos na Tabela 5, uma diferença com um padrão tão persistente e claro entre os ETEs, mas ainda assim rejeitaríamos a previsão teórica.

Se tivéssemos feito a análise por *EFt*, teríamos encontrado resultados consistentes com a teoria na comparação da RMRJ com a RMSP, como ocorreu com o *EFtec*, apesar de que a diferença dos ETEs entre as duas RMs seria subestimada numa análise que usasse apenas o *EFt*, a ponto desta diferença ser tão pequena que se poderia colocar em questão a conclusão que fosse elaborada. O uso do *EFt* para a comparação da RMC com a RMPA nos levaria a dizer que os resultados são consistentes com a teoria, dada a clareza da diferença entre os ETEs nesta especificação, especialmente nas maiores categorias de tamanho do estabelecimento. Numa análise que usasse somente o *EFe*, teríamos resultados razoavelmente consistentes com a teoria na comparação da RMRJ com a RMSP, mas claramente contraditórios com a teoria na comparação da RMC com a RMPA, até porque os ETEs da RMC não são estatisticamente significantes, ao contrário dos ETEs da RMPA. Vemos que, assim como na seção anterior, as conclusões podem variar muito dependendo da especificação usada para a análise, uma vez que cada especificação controla para efeitos fixos diferentes e está sujeita, portanto, a vieses diferentes de variável omitida, aos quais a especificação *EFtec* não está.

Conforme já salientado antes, a interpretação destes vieses pode ser feita erroneamente se ignorarmos a existência do efeito da qualidade do *match*. Por exemplo, note na Tabela 6 que a queda nos ETEs quando passamos do *EFt* para o *EFtec* é maior na RMSP que na RMRJ. Tal fato significa, tendo (21) em mente, que temos evidências de que conforme um trabalhador migra de empregadores menores para maiores, a soma dos efeitos fixos do empregador e da qualidade do *match* tende a crescer mais intensamente na RMSP do que na RMRJ. Tal diferença entre as RMs pode se dever tanto a diferenças sistemáticas nos empregadores das duas RMs quanto a diferenças sistemáticas na capacidade dos trabalhadores das duas RMs de se empregarem em estabelecimentos onde formem bons *matches* – para citar apenas duas possibilidades. Se tivéssemos ignorado a existência da qualidade do *match*, poderíamos acabar dizendo que a diferença entre os ETEs por *EFt* e por *EFtec* é maior para a RMSP porque nesta RM empregadores maiores tenderiam a apresentar efeitos fixos do empregador maiores que os da RMRJ, não havendo espaço para considerações sobre possíveis diferenças no comportamento dos trabalhadores na geração do viés de variável omitida presente no *EFt*.

Desta seção concluímos, com base em nossos resultados da especificação sem vieses (*EFtec*), que a previsão teórica de que o ETE é menor em mercados de trabalho com mais trabalhadores deve ser válida para determinados mercados, mas não para outros.

Tabela 6 – Especificações *POLS, EFt, EFe* e *EFtec* para RMs do Rio de Janeiro e São Paulo

Variável dependente: logaritmo do salário

	RM do Rio de Janeiro				*RM de São Paulo*			
	POLS	*EFt*	*EFe*	*EFtec*	*POLS*	*EFt*	*EFe*	*EFtec*
Número de empregados								
5 a 9	0.1556***	0.0778***	0.0412**	0.0494***	0.2933***	0.0703***	0.0277*	0.0252*
	(0.0267)	(0.0215)	(0.0183)	(0.0164)	(0.0217)	(0.0172)	(0.0162)	(0.0143)
10 a 19	0.1894***	0.1596***	0.1053***	0.1005***	0.3194***	0.1189***	0.0465**	0.0385**
	(0.0280)	(0.0272)	(0.0271)	(0.0222)	(0.0204)	(0.0213)	(0.0203)	(0.0179)
20 a 49	0.2434***	0.2169***	0.1608***	0.1409***	0.4081***	0.1728***	0.0622***	0.0652***
	(0.0266)	(0.0314)	(0.0341)	(0.0265)	(0.0197)	(0.0236)	(0.0229)	(0.0201)
50 a 99	0.3563***	0.2402***	0.1593***	0.1552***	0.4777***	0.1872***	0.0636**	0.0837***
	(0.0303)	(0.0360)	(0.0410)	(0.0304)	(0.0212)	(0.0247)	(0.0272)	(0.0225)
100 a 249	0.4741***	0.2371***	0.1392***	0.1573***	0.5766***	0.2216***	0.0947***	0.1024***
	(0.0307)	(0.0385)	(0.0474)	(0.0333)	(0.0207)	(0.0259)	(0.0314)	(0.0243)
250 a 499	0.5460***	0.2977***	0.1249**	0.2149***	0.6683***	0.2585***	0.1273***	0.1350***
	(0.0344)	(0.0406)	(0.0556)	(0.0360)	(0.0222)	(0.0269)	(0.0360)	(0.0265)
500 a 999	0.5163***	0.2883***	0.1088*	0.2222***	0.6676***	0.2709***	0.0959**	0.1578***
	(0.0345)	(0.0430)	(0.0603)	(0.0393)	(0.0239)	(0.0279)	(0.0434)	(0.0289)
1000 ou mais	0.4753***	0.2942***	0.1708**	0.2821***	0.6744***	0.2615***	0.1003**	0.1784***
	(0.0338)	(0.0460)	(0.0705)	(0.0451)	(0.0230)	(0.0291)	(0.0478)	(0.0306)
N	6 186				13 498			
J	6 105				13 092			
T	8				8			
N·T	49 488				107 984			
nº *matches* diferentes	8 514				19 228			

Nota: idem à Tabela 5.

Tabela 7 – Especificações *POLS, EFt, EFe* e *EFtec* para RMs de Porto Alegre e Curitiba

Variável dependente: logaritmo do salário

	Porto Alegre				*Curitiba*			
	POLS	*EFt*	*EFe*	*EFtec*	*POLS*	*EFt*	*EFe*	*EFtec*
Número de empregados								
5 a 9	0.2719***	0.0839*	0.0732*	0.0519	0.1385***	0.1079**	0.0539	0.0169
	(0.0540)	(0.0432)	(0.0415)	(0.0410)	(0.0526)	(0.0459)	(0.0420)	(0.0378)
10 a 19	0.3353***	0.1063*	0.1295**	0.0936	0.2655***	0.2147***	0.1024*	0.0914*
	(0.0526)	(0.0563)	(0.0577)	(0.0580)	(0.0563)	(0.0598)	(0.0529)	(0.0501)
20 a 49	0.4184***	0.1934***	0.2218***	0.1908***	0.4357***	0.2189***	0.0334	0.0655
	(0.0493)	(0.0632)	(0.0699)	(0.0693)	(0.0537)	(0.0677)	(0.0640)	(0.0588)
50 a 99	0.5163***	0.1903***	0.2977***	0.1999**	0.4969***	0.2785***	0.0554	0.0724
	(0.0556)	(0.0639)	(0.0900)	(0.0798)	(0.0587)	(0.0650)	(0.0689)	(0.0642)
100 a 249	0.5359***	0.2266***	0.3002***	0.2188**	0.5439***	0.3247***	0.0331	0.1142*
	(0.0526)	(0.0648)	(0.0932)	(0.0880)	(0.0575)	(0.0656)	(0.0758)	(0.0687)
250 a 499	0.6296***	0.2434***	0.3123***	0.2732***	0.7377***	0.4127***	0.1023	0.1575**
	(0.0561)	(0.0659)	(0.1020)	(0.0892)	(0.0606)	(0.0696)	(0.0820)	(0.0754)
500 a 999	0.7016***	0.2642***	0.3222***	0.2881***	0.5928***	0.4063***	-0.1286	0.1912**
	(0.0566)	(0.0671)	(0.1096)	(0.0916)	(0.0652)	(0.0795)	(0.1007)	(0.0821)
1000 ou mais	0.6407***	0.2565***	0.4412***	0.3049***	0.8505***	0.5609***	0.1253	0.2459***
	(0.0581)	(0.0709)	(0.1145)	(0.0937)	(0.0621)	(0.0781)	(0.1072)	(0.0864)
N	2 125				1 283			
J	2 090				1 382			
T	8				8			
N·T	17 000				10 264			
nº *matches* diferentes	2 991				1 894			

Nota: idem à Tabela 5.

5.3.1 Avaliação mais rigorosa

Nesta subseção fazemos uma avaliação semelhante à feita na subseção 5.2.1. Neste sentido, estimamos uma especificação *EFtec* sobre uma sub-amostra composta apenas dos trabalhadores presentes na RMRJ ou na RMSP e na qual, além dos controles para o tamanho do empregador, controlamos para a interação do tamanho do empregador com uma *dummy* igual a 1 se o trabalhador estivesse na RMRJ, e 0 caso contrário. O mesmo foi feito para a comparação da RMPA com a RMC, em que a *dummy* vale 1 para o trabalhador de Curitiba, e 0 caso contrário. Note que a *dummy* corresponde ao menor dos mercados de trabalho em cada comparação. A Tabela 8 abaixo contém os resultados.

Na comparação da RMSP com a RMRJ, os ETEs sem interação são muito parecidos entre a Tabela 8 e o *EFtec* da RMSP da Tabela 6, especialmente os ETEs das menores categorias de tamanho. Nas maiores categorias, o ETE é ligeiramente maior na Tabela 8. Ao somarmos os ETEs sem interação com os respectivos coeficientes de interação, assim como feito na seção 5.2.1, teremos valores menores que suas contrapartidas no *EFtec* da RMRJ da Tabela 6 (mesmo que considerássemos os coeficientes não significantes como significantes, o que não fazemos aqui). No entanto, mesmo assim estas somas apontam para um efeito marginal maior para a RMRJ, muito embora apenas para duas categorias de tamanho: “10 a 19” e “20 a 49”. Na comparação da RMPA com a RMC, os ETEs sem interação são novamente muito parecidos entre a Tabela 8 e o *EFtec* da RMPA da Tabela 7, sendo que os da Tabela 7 são ligeiramente maiores. A soma dos ETEs sem interação com os coeficientes de interação resulta nos mesmos valores, já que todos os coeficientes de interação não têm significância estatística.

Tabela 8 – Interação com RM

Variável dependente: logaritmo do salário

	RMSP x RMRJ		*RMPA x RMC*	
Número de empregados		Interação de tamanho com RM		Interação de tamanho com RM
5 a 9	0.0255*	0.0227	0.0509	-0.0278
	(0.01424)	(0.02160)	(0.04138)	(0.05584)
10 a 19	0.0390**	0.0591**	0.0925	0.0067
	(0.01781)	(0.02829)	(0.05905)	(0.07737)
20 a 49	0.0673***	0.0645**	0.1898***	-0.1170
	(0.01997)	(0.03285)	(0.07014)	(0.09134)
50 a 99	0.0869***	0.0575	0.1940**	-0.1065
	(0.02230)	(0.03715)	(0.08061)	(0.10273)
100 a 249	0.1056***	0.0350	0.2138**	-0.0850
	(0.02409)	(0.04012)	(0.08861)	(0.11161)
250 a 499	0.1411***	0.0460	0.2652***	-0.0884
	(0.02621)	(0.04336)	(0.08948)	(0.11623)
500 a 999	0.1644***	0.0270	0.2774***	-0.0692
	(0.02855)	(0.04706)	(0.09164)	(0.12159)
1000 ou mais	0.1881***	0.0494	0.3022***	-0.0300
	(0.03018)	(0.05277)	(0.09379)	(0.12636)
N	21 105		3 420	
J	20 635		3 503	
T	8		8	
N·T	168 840		27 360	
nº *matches* diferentes	29 503		4 923	

Nota 1: erros-padrão abaixo dos coeficientes, entre parênteses.
Nota 2: * = p-valor < 0.1 ** = p-valor < 0.05 *** = p-valor < 0.01
Nota 3: para cada comparação de RMs, os coeficientes são de uma especificação *EFtec* única. A primeira coluna da tabela contém os coeficientes das *dummies* de tamanho do empregador, e a segunda coluna os coeficientes de interações de *dummies* do tamanho do empregador com uma *dummy* indicando se o trabalhador pertence à RMRJ (valor 1) ou à RMSP (valor 0). A terceira e a quarta colunas são análogas, mas para a comparação da RMC com a RMPA.
Nota 4: controles presentes em todas as regressões: *dummies* para tamanho do estabelecimento, *dummies* de interação (ver Nota 3), grau de instrução do trabalhador, faixa etária do trabalhador, tempo no emprego (*tenure*), unidade federativa do estabelecimento empregador, indústria (seção CNAE 1.0), ocupação (grandes grupos CBO 1994), *dummies* para ano e *dummy* para emprego temporário.

A conclusão a que tínhamos chegado na subseção anterior era de que a teoria era consistente com a evidência na comparação da RMSP com a RMRJ, mas contraditória na

comparação da RMPA com a RMC. Os resultados da Tabela 8 modificam esta conclusão para a comparação da RMRJ com a RMSP: os ETEs na RMRJ agora superam os da RMSP em apenas duas categorias de tamanho do empregador, enquanto que na especificação *EFtec* da Tabela 6 os ETEs da RMRJ são maiores que os ETEs da RMSP para toda categoria de tamanho. Assim, agora a evidência contradiz a previsão teórica também para estas duas RMs.

6 CONCLUSÃO

Neste trabalho avaliamos o poder explicativo de modelos teóricos que pretendem explicar a origem do efeito positivo do tamanho do empregador sobre o salário, efeito este persistentemente encontrado na literatura empírica. Os modelos teóricos avaliados foram reunidos em quatro grupos, de acordo com a explicação que propõem para a existência do efeito: aqueles que o atribuem à heterogeneidade não observável dos trabalhadores, dos empregadores e/ou da qualidade dos *matches*; aqueles que o atribuem a diferenças na qualificação agregada da mão-de-obra no nível do empregador; os que atribuem o efeito a diferenças na dificuldade de monitoramento do esforço entre empregadores maiores e menores; e os que atribuem o efeito a problemas de informação sobre a distribuição de salários na economia.

Na avaliação, utilizamos uma especificação da equação salarial ainda não utilizada, segundo nosso conhecimento, nesta literatura empírica, e que controla, através de efeitos fixos, para a heterogeneidade não observável dos trabalhadores, dos estabelecimentos e da qualidade dos *matches* entre eles. Nossos resultados mostraram que mais da metade do efeito do tamanho do empregador sobre o salário é explicado por vieses de variável omitida atribuíveis às três heterogeneidades não observáveis citadas. Ademais, o modelo que atribui o efeito a diferenças na qualificação agregada, bem como os modelos que atribuem o efeito a problemas de monitoramento e de informação, todos eles tiveram sua validade empírica rejeitada pelos resultados das estimações.

Finalmente, cabe notar que apesar dos resultados terem apontado para a rejeição das previsões teóricas dos modelos testados, tais previsões mostraram-se válidas para algumas faixas restritas de variação do tamanho do empregador em alguns casos. Por exemplo, a teoria das diferenças na qualificação agregada foi corroborada, após avaliação mais rigorosa, para as categorias maiores de tamanho do empregador. Neste sentido, é possível que diferentes fatores determinem o efeito do tamanho do empregador em diferentes faixas de tamanho.

REFERÊNCIAS

BARRON, John *et al*. *Employer Size: The Implications for Search, Training, Capital Investment, Starting Wages, and Wage Growth*. **Journal of Labor Economics**. v. 5, n. 1, p. 76-89, January 1987.

BROWN, Charles; MEDOFF, James. *The Employer Size-Wage Effect*. **Journal of Political Economy**. v. 97, n. 5, p. 1027-1059, November 1989.

BULOW, Jeremy; SUMMERS, Lawrence. *A Theory of Dual Labor Markets, with Applications to Industrial Policy, Discrimination, and Keynesian Unemployment*. **Journal of Labor Economics**. 4, p. 376-414, July 1986.

BURDETT, Kenneth; MORTENSEN, Dale T. *Wage differentials, employer size, and unemployment*. **International Economic Review**. v. 32, n. 2, p. 257-273, May 1998.

DUNN, L. F. *Work disutility and compensating differentials: estimation of factors in the link between wages and firm size*. **Review of Economics and Statistics**. v. 69, n. 1, p. 67-73, February 1986.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Disponível em: <http://www.cnae.ibge.gov.br>. Acesso em: 25/09/2011.

IPEADATA. Disponível em: <http://www.ipeadata.gov.br/>. Acesso em: 29/09/2011.

KREMER, Michael. *The O-Ring Theory of Economic Development*. **The Quarterly Journal of Economics**. v. 108, n. 3, p. 551-575, 1993.

MELLOW, Wesley. *Employer Size and Wages*. **Review of Economics and Statistics**. v. 64, n. 3, p. 495-501, August 1982.

MINISTÉRIO DO TRABALHO E EMPREGO. Disponível em: <http://www.mte.gov.br/Empregador/CBO/procuracbo/conteudo/tabela.asp>. Acesso em: 25/09/2011.

STRAND, Jon. *The relationship between wages and firm size: an information theoretic analysis*. **International Economic Review**. v. 28, n. 1, p. 51-68, 1987.

TROSKE, Kenneth R. *Evidence on the employer size-wage premium from worker-establishment matched data*. **Review of Economics and Statistics**. v. 81, n. 1, p. 15-26, 1999.

WEISS, Andrew; LANDAU, Henry J. *Wages, Hiring Standards, and Firm Size*. **Journal of Labor Economics**. v. 2, n. 4, p. 477-499, October 1984.

WEISS, Leonard. *Concentration and Labor Earnings*. **American Economic Review**. v. 56, n. 1/2, p. 96-117, March 1966.

WOODCOCK, Simon D. *Match Effects*. Working Paper, Simon Fraser University, Canada, August 2008.

APÊNDICES

APÊNDICE A – DESCRIÇÃO DOS DADOS
APÊNDICE B – DERIVAÇÕES DE VIESES
APÊNDICE C – DERIVAÇÕES DAS EQUAÇÕES PARA ESTIMAÇÃO *WITHIN*
APÊNDICE D – ESTIMAÇÕES INCLUINDO TODAS AS SEÇÕES CNAE 1.0

APÊNDICE A – Descrição dos dados

Tabela 1A – Descrição das variáveis

Variável	Contínua/ Categórica	Subtipos (se contínua)/ Categorias (se categórica)	Média (se contínua)/ Percentual da amostra (se categórica)		Desvio-padrão	
			1995	2002	1995	2002
Salário real (remuneração real em dezembro)	Contínua		603.96	707.27	868.51	982.85
Tenure (meses)	Contínua		73.73	137.55	70.41	85.58
Tipo de vínculo	Categórica	Temporário	0.0055	0.0039		
		Ignorado*	0.0109	0.0088		
Unidade federativa do estabelecimento	Categórica	AC	0.15	0.15		
		AL	1.01	1.00		
		AM	1.15	1.15		
		AP	0.14	0.14		
		BA	4.11	4.13		
		CE	1.85	1.87		
		DF	1.82	1.93		
		ES	1.52	1.53		
		GO	1.91	1.87		
		MA	0.99	0.98		
		MG	8.84	8.82		
		MS	0.98	1.00		
		MT	0.95	0.93		
		PA	1.69	1.73		

Variável	Contínua/ Categórica	Subtipos (se contínua)/ Categorias (se categórica)	Média (se contínua)/ Percentual da amostra (se categórica)		Desvio-padrão	
			1995	2002	1995	2002
		PB	0.90	0.90		
		PE	3.69	3.65		
		PI	0.42	0.42		
		PR	5.30	5.24		
		RJ	11.67	11.60		
		RN	0.78	0.80		
		RO	0.42	0.42		
		RR	0.09	0.08		
		RS	7.09	7.05		
		SC	4.16	4.20		
		SE	0.61	0.60		
		SP	35.63	35.85		
		TO	0.25	0.26		
		Ignorado*	1.84	1.71		
Região metropolitana do estabelecimento	Categórica	Belém	1.02	1.03		
		Belo Horizonte	3.49	3.48		
		Curitiba	2.21	2.12		
		Fortaleza	1.56	1.55		
		Porto Alegre	3.32	3.38		
		Rio de Janeiro	9.60	9.49		
		Recife	2.50	2.45		
		Salvador	2.33	2.43		
		São Paulo	21.15	21.10		
		Outros	50.87	51.16		
		Ignorado*	1.95	1.81		

Variável	Contínua/ Categórica	Subtipos (se contínua)/ Categorias (se categórica)	Média (se contínua)/ Percentual da amostra (se categórica)		Desvio-padrão	
			1995	2002	1995	2002
Grau de instrução	Categórica	Analfabeto	3.64	3.17		
		4ª série incompleta	12.93	11.21		
		4ª série completa	18.66	16.54		
		8ª série incompleta	16.40	14.83		
		8ª série completa	14.38	15.67		
		2º grau incompleto	7.54	7.18		
		2º grau completo	15.14	18.16		
		Superior incompleto	3.21	3.50		
		Superior completo	6.49	8.71		
		Ignorado*	1.62	1.03		
Percentual de grau de instrução no estabelecimento	Contínua	Analfabeto	0.0363	0.0317	0.1692	0.1585
		4ª série incompleta	0.1288	0.1120	0.3033	0.2872
		4ª série completa	0.1860	0.1651	0.3469	0.3355
		8ª série incompleta	0.1634	0.1481	0.3251	0.3168
		8ª série completa	0.1438	0.1567	0.3070	0.3219
		2º grau incompleto	0.0753	0.0719	0.2259	0.2260
		2º grau completo	0.1521	0.1816	0.3070	0.3311
		Superior incompleto	0.0323	0.0350	0.1454	0.1522
		Superior completo	0.0657	0.0877	0.2039	0.2389
		Ignorado*	0.0162	0.0103	0.1213	0.0960
Tamanho do estabelecimento (nº de empregados)	Categórica	Até 4	14.43	14.23		
		de 5 a 9	10.17	10.45		
		de 10 a 19	10.46	10.54		
		de 20 a 49	12.67	12.97		
		de 50 a 99	9.11	9.21		

Variável	Contínua/ Categórica	Subtipos (se contínua)/ Categorias (se categórica)	Média (se contínua)/ Percentual da amostra (se categórica)		Desvio-padrão	
			1995	2002	1995	2002
		de 100 a 249	12.17	12.01		
		de 250 a 499	9.37	9.19		
		de 500 a 999	7.68	8.12		
		1000 ou mais	13.94	13.28		
Faixa etária (anos)	Categórica	18 a 24	19.45	7.97		
		25 a 29	19.44	13.74		
		30 a 39	32.13	35.14		
		40 a 49	16.72	26.07		
		50 a 64	8.26	13.27		
		Ignorado*	4.01	3.81		
Ocupação (Grande Grupo CBO 1994)**	Categórica	Grande grupo 0	5.22	5.63		
		Grande grupo 1	1.56	1.72		
		Grande grupo 2	2.74	3.88		
		Grande grupo 3	19.29	18.19		
		Grande grupo 4	10.27	10.45		
		Grande grupo 5	13.87	14.19		
		Grande grupo 6	6.33	6.08		
		Grande grupo 7	10.07	9.99		
		Grande grupo 8	10.52	10.52		
		Grande grupo 9	16.62	16.06		
		Ignorado*	3.52	3.28		
Indústria (seção CNAE 1.0)**	Categórica	Seção A	5.58	5.40		
		Seção B	0.07	0.07		
		Seção D	34.33	34.01		
		Seção F	5.56	5.30		

Variável	Contínua/ Categórica	Subtipos (se contínua)/ Categorias (se categórica)	Média (se contínua)/ Percentual da amostra (se categórica)		Desvio-padrão	
			1995	**2002**	**1995**	**2002**
		Seção G	20.57	20.85		
		Seção H	3.77	3.78		
		Seção J	5.03	5.08		
		Seção K	13.79	15.12		
		Não informado	7.53	7.17		
		Ignorado*	3.76	3.21		

*A categoria "ignorado" indica erros, omissões, e registros que não fazem sentido. Tal categoria faz parte da base original, não tendo sido criada por estes autores.
**As tabelas 3A e 4A descrevem o conteúdo dos grandes grupos da CBO 94 e das seções CNAE 1.0, respectivamente.

Tabela 2A – Comparação entre o painel não balanceado e o balanceado

Variável	Contínua/ Categórica	Subtipos (se contínua)/ Categorias (se categórica)	*Média (1995 a 2002)*		*Desvio-padrão (1995 a 2002)*	
			Amostra não balanceada	**Amostra balanceada**	**Amostra não balanceada**	**Amostra balanceada**
Salário real (remuneração real em dezembro)	Contínua		668.77	668.26	934.57	934.34
Tenure (meses)	Contínua		105 746	1 059 099	8 008 769	8 006 912
Tipo de vínculo	Categórica	Temporário	0.00469	0.00470	0.06831	0.06842
		Ignorado*	0.00924	0.00927	0.09566	0.09586
Unidade federativa do estabelecimento	Categórica	AC	0.00153	0.00153	0.03905	0.03913
		AL	0.00994	0.00993	0.09921	0.09917
		AM	0.01152	0.01155	0.10673	0.10683
		AP	0.00145	0.00145	0.03802	0.03802
		BA	0.04114	0.04118	0.19862	0.19871
		CE	0.01857	0.01859	0.13499	0.13508
		DF	0.01866	0.01871	0.13533	0.13550
		ES	0.01528	0.01531	0.12267	0.12280
		GO	0.01891	0.01889	0.13620	0.13615
		MA	0.00985	0.00982	0.09875	0.09861
		MG	0.08841	0.08833	0.28390	0.28377
		MS	0.00997	0.00989	0.09935	0.09893
		MT	0.00943	0.00944	0.09664	0.09672
		PA	0.01716	0.01717	0.12987	0.12989
		PB	0.00904	0.00904	0.09467	0.09466

Variável	Contínua/ Categórica	Subtipos (se contínua)/ Categorias (se categórica)	*Média (1995 a 2002)*		*Desvio-padrão (1995 a 2002)*	
			Amostra não balanceada	**Amostra balanceada**	**Amostra não balanceada**	**Amostra balanceada**
		PE	0.03668	0.03673	0.18798	0.18810
		PI	0.00421	0.00419	0.06476	0.06460
		PR	0.05297	0.05255	0.22398	0.22314
		RJ	0.11627	0.11644	0.32055	0.32075
		RN	0.00776	0.00776	0.08773	0.08776
		RO	0.00422	0.00421	0.06484	0.06478
		RR	0.00085	0.00085	0.02911	0.02917
		RS	0.07078	0.07063	0.25645	0.25621
		SC	0.04180	0.04182	0.20014	0.20017
		SE	0.00616	0.00616	0.07825	0.07826
		SP	0.35741	0.35773	0.47924	0.47933
		TO	0.00255	0.00255	0.05045	0.05046
		Ignorado*	0.01746	0.01753	0.13098	0.13123
		Belo Horizonte	0.03474	0.03478	0.18312	0.18322
		Porto Alegre	0.03367	0.03352	0.18038	0.18000
		Rio de Janeiro	0.09530	0.09543	0.29363	0.29381
		Belém	0.01036	0.01034	0.10126	0.10116
		Curitiba	0.02146	0.02135	0.14492	0.14455
Região metropolitana do estabelecimento	Categórica	Fortaleza	0.01547	0.01548	0.12342	0.12347
		Recife	0.02478	0.02479	0.15546	0.15548
		Salvador	0.02383	0.02387	0.15253	0.15263
		São Paulo	0.21100	0.21124	0.40802	0.40819
		Outros	0.51089	0.51063	0.49988	0.49989
		Ignorado*	0.01849	0.01857	0.13473	0.13499

Variável	Contínua/ Categórica	Subtipos (se contínua)/ Categorias (se categórica)	*Média (1995 a 2002)*		*Desvio-padrão (1995 a 2002)*	
			Amostra não balanceada	**Amostra balanceada**	**Amostra não balanceada**	**Amostra balanceada**
Grau de instrução	Categórica	Analfabeto	0.03434	0.03441	0.18209	0.18229
		4ª série incompleto	0.11904	0.11911	0.32384	0.32391
		4ª série completo	0.17609	0.17622	0.38090	0.38100
		8ª série incompleto	0.15701	0.15698	0.36381	0.36379
		8ª série completo	0.15189	0.15202	0.35891	0.35904
		2º grau incompleto	0.07354	0.07341	0.26102	0.26081
		2º grau completo	0.16694	0.16672	0.37292	0.37272
		Superior incompleto	0.03404	0.03402	0.18134	0.18129
		Superior completo	0.07528	0.07523	0.26384	0.26376
		Ignorado*	0.01184	0.01188	0.10818	0.10833
Tamanho do estabelecimento (nº de empregados)	Categórica	até 4	0.14303	0.14277	0.35010	0.34984
		de 5 a 9	0.10326	0.10325	0.30430	0.30429
		de 10 a 19	0.10500	0.10500	0.30655	0.30656
		de 20 a 49	0.12931	0.12940	0.33554	0.33564
		de 50 a 99	0.09225	0.09232	0.28938	0.28948
		de 100 a 249	0.12155	0.12153	0.32677	0.32675
		de 250 a 499	0.09352	0.09352	0.29116	0.29116
		de 500 a 999	0.07963	0.07972	0.27072	0.27086
		1000 ou mais	0.13245	0.13248	0.33898	0.33901
Faixa etária (anos)	Categórica	18 a 24	0.12864	0.12878	0.33480	0.33495
		25 a 29	0.17126	0.17121	0.37673	0.37669
		30 a 39	0.34387	0.34354	0.47500	0.47489
		40 a 49	0.21349	0.21341	0.40977	0.40971
		50 a 64	0.10418	0.10435	0.30550	0.30571

Variável	Contínua/ Categórica	Subtipos (se contínua)/ Categorias (se categórica)	*Média (1995 a 2002)* Amostra não balanceada	*Média (1995 a 2002)* Amostra balanceada	*Desvio-padrão (1995 a 2002)* Amostra não balanceada	*Desvio-padrão (1995 a 2002)* Amostra balanceada
		Ignorado*	0.03856	0.03871	0.19254	0.19291
Ocupação (Grande Grupo CBO 1994)	Categórica	grande grupo 0	0.05449	0.05437	0.22698	0.22674
		grande grupo 1	0.01635	0.01639	0.12681	0.12697
		grande grupo 2	0.03182	0.03187	0.17552	0.17565
		grande grupo 3	0.18730	0.18747	0.39015	0.39029
		grande grupo 4	0.10410	0.10406	0.30538	0.30534
		grande grupo 5	0.14048	0.14080	0.34749	0.34782
		grande grupo 6	0.06240	0.06207	0.24189	0.24127
		grande grupo 7	0.10064	0.10044	0.30085	0.30059
		grande grupo 8	0.10608	0.10589	0.30794	0.30770
		grande grupo 9	0.16262	0.16279	0.36902	0.36917
		Ignorado*	0.03372	0.03385	0.18051	0.18085
Indústria (seção CNAE 1.0)	Categórica	Seção A	0.05613	0.05573	0.23017	0.22940
		Seção B	0.00068	0.00067	0.02598	0.02579
		Seção D	0.34162	0.34102	0.47425	0.47405
		Seção F	0.05401	0.05407	0.22604	0.22616
		Seção G	0.20749	0.20745	0.40551	0.40548
		Seção H	0.03781	0.03784	0.19075	0.19080
		Seção J	0.05029	0.05047	0.21854	0.21892
		Seção K	0.14612	0.14646	0.35322	0.35357
		Ignorado*	0.03324	0.03337	0.17925	0.17960
		Não informado	0.07261	0.07292	0.25950	0.26001

*A categoria "ignorado" indica erros, omissões, e registros que não fazem sentido. Tal categoria faz parte da base original, não tendo sido criada por estes autores.
**As tabelas 3A e 4A descrevem o conteúdo dos grandes grupos da CBO 94 e das seções CNAE 1.0, respectivamente.

Tabela 3A – Descrição dos grandes grupos CBO 1994

Grande grupo	Conteúdo do grande grupo
0 e 1	Trabalhadores das profissões científicas, técnicas, artísticas e trabalhadores assemelhados
2	Membros dos poderes Legislativo, Executivo e Judiciário, funcionários públicos superiores, diretores de empresas e trabalhadores assemelhados
3	Trabalhadores de serviços administrativos e trabalhadores assemelhados
4	Trabalhadores de comércio e trabalhadores assemelhados
5	Trabalhadores de serviços de turismo, hospedagem, serventia, higiene e embelezamento, segurança, auxiliares de saúde e trabalhadores assemelhados
6	Trabalhadores agropecuários, florestais, da pesca e trabalhadores assemelhados
7, 8 e 9	Trabalhadores da produção industrial, operadores de máquinas, condutores de veículos e trabalhadores assemelhados

FONTE: elaborada pelos autores com base na "Tabela de Grande Grupo" em *site* do Ministério do Trabalho e Emprego (http://www.mte.gov.br/Empregador/CBO/procuracbo/conteudo/tabela.asp)

Tabela 4A – Descrição das seções CNAE 1.0

Seção	Conteúdo da seção
A	Agricultura, pecuária, silvicultura e exploração florestal
B	Pesca
C	Indústrias extrativas
D	Indústrias de transformação
E	Produção e distribuição de eletricidade, gás e água
F	Construção
G	Comércio; reparação de veículos automotores, objetos pessoais e domésticos
H	Alojamento e alimentação
I	Transporte, armazenagem e comunicações
J	Intermediação financeira, seguros, previdência complementar e serviços relacionados
K	Atividades imobiliárias, aluguéis e serviços prestados às empresas
L	Administração pública, defesa e seguridade social
M	Educação
N	Saúde e serviços sociais
O	Outros serviços coletivos, sociais e pessoais
P	Serviços domésticos
Q	Organismos internacionais e outras instituições extraterritoriais

FONTE: elaborada pelos autores com base na estrutura da CNAE 1.0 obtida em *site* do IBGE (http://www.cnae.ibge.gov.br)

APÊNDICE B – Derivações de vieses

B.1 Derivação de (4)

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'[\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{D}\boldsymbol{\theta} + \boldsymbol{F}\boldsymbol{\psi} + \boldsymbol{\varepsilon}] =$$

$$= \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{D}\boldsymbol{\theta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{F}\boldsymbol{\psi} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}$$

Sob a hipótese usual de que $E[\boldsymbol{\varepsilon}|\boldsymbol{X},\boldsymbol{D},\boldsymbol{F}] = \boldsymbol{0}$, temos

$$E(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta} + E_{\boldsymbol{X},\boldsymbol{D},\boldsymbol{F}}\big[E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{D}\boldsymbol{\theta}|\boldsymbol{X},\boldsymbol{D},\boldsymbol{F}]\big] + \\ +E_{\boldsymbol{X},\boldsymbol{D},\boldsymbol{F}}\big[E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{F}\boldsymbol{\psi}|\boldsymbol{X},\boldsymbol{D},\boldsymbol{F}]\big] + E_{\boldsymbol{X},\boldsymbol{D},\boldsymbol{F}}\big[E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}|\boldsymbol{X},\boldsymbol{D},\boldsymbol{F}]\big] =$$

$$= \boldsymbol{\beta} + E_{\boldsymbol{X},\boldsymbol{D},\boldsymbol{F}}\big[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{D}E[\boldsymbol{\theta}|\boldsymbol{X},\boldsymbol{D},\boldsymbol{F}]\big] + E_{\boldsymbol{X},\boldsymbol{D},\boldsymbol{F}}\big[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{F}E[\boldsymbol{\psi}|\boldsymbol{X},\boldsymbol{D},\boldsymbol{F}]\big] + \\ +E_{\boldsymbol{X},\boldsymbol{D},\boldsymbol{F}}\big[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'E[\boldsymbol{\varepsilon}|\boldsymbol{X},\boldsymbol{D},\boldsymbol{F}]\big] =$$

$$= \boldsymbol{\beta} + E_{\boldsymbol{X},\boldsymbol{D},\boldsymbol{F}}\big[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{D}E[\boldsymbol{\theta}|\boldsymbol{X},\boldsymbol{D},\boldsymbol{F}]\big] + E_{\boldsymbol{X},\boldsymbol{D},\boldsymbol{F}}\big[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{F}E[\boldsymbol{\psi}|\boldsymbol{X},\boldsymbol{D},\boldsymbol{F}]\big] =$$

$$= \boldsymbol{\beta} + E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{D}\boldsymbol{\theta}] + E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{F}\boldsymbol{\psi}]$$

B.2 Derivação de (9)

$$\widehat{\boldsymbol{\theta}} = (\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}'\boldsymbol{y} = (\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}'[\boldsymbol{D}\boldsymbol{\theta} + \boldsymbol{G}\boldsymbol{\phi} + \boldsymbol{\eta}] =$$

$$= \boldsymbol{\theta} + (\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}'\boldsymbol{G}\boldsymbol{\phi} + (\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}'\boldsymbol{\eta}$$

Sob a hipótese usual de que $E[\boldsymbol{\eta}|\boldsymbol{D},\boldsymbol{G}] = \boldsymbol{0}$, temos

$$E(\widehat{\boldsymbol{\theta}}) = \boldsymbol{\theta} + E_{\boldsymbol{D},\boldsymbol{G}}\big[E[(\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}'\boldsymbol{G}\boldsymbol{\phi}|\boldsymbol{D},\boldsymbol{G}]\big] + E_{\boldsymbol{D},\boldsymbol{G}}\big[E[(\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}'\boldsymbol{\eta}|\boldsymbol{D},\boldsymbol{G}]\big] =$$

$$= \boldsymbol{\theta} + E_{\boldsymbol{D},\boldsymbol{G}}\big[(\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}'\boldsymbol{G}E[\boldsymbol{\phi}|\boldsymbol{D},\boldsymbol{G}]\big] + E_{\boldsymbol{D},\boldsymbol{G}}\big[(\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}'E[\boldsymbol{\eta}|\boldsymbol{D},\boldsymbol{G}]\big] =$$

$$= \boldsymbol{\theta} + E_{\boldsymbol{D},\boldsymbol{G}}\big[(\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}'\boldsymbol{G}E[\boldsymbol{\phi}|\boldsymbol{D},\boldsymbol{G}]\big] =$$

$$= \boldsymbol{\theta} + E[(\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}'\boldsymbol{G}\boldsymbol{\phi}]$$

B.3 Derivação de (I*)

$$\widehat{\boldsymbol{\beta}}_{POLS} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'[\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{D}\boldsymbol{\theta} + \boldsymbol{F}\boldsymbol{\psi} + \boldsymbol{G}\boldsymbol{\phi} + \boldsymbol{\omega}] =$$

$$= \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{D}\boldsymbol{\theta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{F}\boldsymbol{\psi} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{G}\boldsymbol{\phi} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\omega}$$

Sob a hipótese usual de que $E[\boldsymbol{\omega}|\boldsymbol{X}, \boldsymbol{D}, \boldsymbol{F}, \boldsymbol{G}] = \boldsymbol{0}$, temos

$$E\left(\widehat{\boldsymbol{\beta}}_{POLS}\right) = \boldsymbol{\beta} + E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{D}\boldsymbol{\theta}] + E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{F}\boldsymbol{\psi}] + E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{G}\boldsymbol{\phi}] + \\ + E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\omega}] =$$

$$= \boldsymbol{\beta} + E_{\boldsymbol{X},\boldsymbol{D},\boldsymbol{F},\boldsymbol{G}}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{D}E[\boldsymbol{\theta}|\boldsymbol{X}, \boldsymbol{D}, \boldsymbol{F}, \boldsymbol{G}]\right] + E_{\boldsymbol{X},\boldsymbol{D},\boldsymbol{F},\boldsymbol{G}}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{F}E[\boldsymbol{\psi}|\boldsymbol{X}, \boldsymbol{D}, \boldsymbol{F}, \boldsymbol{G}]\right] + \\ + E_{\boldsymbol{X},\boldsymbol{D},\boldsymbol{F},\boldsymbol{G}}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{G}E[\boldsymbol{\phi}|\boldsymbol{X}, \boldsymbol{D}, \boldsymbol{F}, \boldsymbol{G}]\right] + E_{\boldsymbol{X},\boldsymbol{D},\boldsymbol{F},\boldsymbol{G}}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'E[\boldsymbol{\omega}|\boldsymbol{X}, \boldsymbol{D}, \boldsymbol{F}, \boldsymbol{G}]\right] =$$

$$= \boldsymbol{\beta} + E_{\boldsymbol{X},\boldsymbol{D},\boldsymbol{F},\boldsymbol{G}}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{D}E[\boldsymbol{\theta}|\boldsymbol{X}, \boldsymbol{D}, \boldsymbol{F}, \boldsymbol{G}]\right] + E_{\boldsymbol{X},\boldsymbol{D},\boldsymbol{F},\boldsymbol{G}}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{F}E[\boldsymbol{\psi}|\boldsymbol{X}, \boldsymbol{D}, \boldsymbol{F}, \boldsymbol{G}]\right] + \\ + E_{\boldsymbol{X},\boldsymbol{D},\boldsymbol{F},\boldsymbol{G}}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{G}E[\boldsymbol{\phi}|\boldsymbol{X}, \boldsymbol{D}, \boldsymbol{F}, \boldsymbol{G}]\right] =$$

$$= \boldsymbol{\beta} + E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{D}\boldsymbol{\theta}] + E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{F}\boldsymbol{\psi}] + E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{G}\boldsymbol{\phi}]$$

B.4 Derivação de (II*)

$$\widehat{\boldsymbol{\beta}}_{EFt} = (\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{D}}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{D}}\boldsymbol{y} = (\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{D}}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{D}}[\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{D}\boldsymbol{\theta} + \boldsymbol{F}\boldsymbol{\psi} + \boldsymbol{G}\boldsymbol{\phi} + \boldsymbol{\omega}]$$

Usando a propriedade $\boldsymbol{M}_{\boldsymbol{A}}\boldsymbol{A} = \boldsymbol{0}$, temos

$$\widehat{\boldsymbol{\beta}}_{EFt} = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{D}}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{D}}\boldsymbol{F}\boldsymbol{\psi} + (\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{D}}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{D}}\boldsymbol{G}\boldsymbol{\phi} + (\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{D}}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{D}}\boldsymbol{\omega}$$

Sob a hipótese usual de que $E[\boldsymbol{\omega}|\boldsymbol{X}, \boldsymbol{D}, \boldsymbol{F}, \boldsymbol{G}] = \boldsymbol{0}$, temos

$$E\left(\widehat{\boldsymbol{\beta}}_{EFt}\right) = \boldsymbol{\beta} + E_{\boldsymbol{X},\boldsymbol{D},\boldsymbol{F},\boldsymbol{G}}\left[(\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{D}}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{D}}\boldsymbol{F}E[\boldsymbol{\psi}|\boldsymbol{X}, \boldsymbol{D}, \boldsymbol{F}, \boldsymbol{G}]\right] + \\ + E_{\boldsymbol{X},\boldsymbol{D},\boldsymbol{F},\boldsymbol{G}}\left[(\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{D}}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{D}}\boldsymbol{G}E[\boldsymbol{\phi}|\boldsymbol{X}, \boldsymbol{D}, \boldsymbol{F}, \boldsymbol{G}]\right] + \\ + E_{\boldsymbol{X},\boldsymbol{D},\boldsymbol{F},\boldsymbol{G}}\left[(\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{D}}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_{\boldsymbol{D}}E[\boldsymbol{\omega}|\boldsymbol{X}, \boldsymbol{D}, \boldsymbol{F}, \boldsymbol{G}]\right] =$$

$$= \boldsymbol{\beta} + E[(\boldsymbol{X}'\boldsymbol{M_D}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M_D}\boldsymbol{F}\boldsymbol{\psi}] + E[(\boldsymbol{X}'\boldsymbol{M_D}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M_D}\boldsymbol{G}\boldsymbol{\phi}]$$

B.5 Derivação de (III*)

$$\widehat{\boldsymbol{\beta}}_{EFe} = (\boldsymbol{X}'\boldsymbol{M_F}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M_F}\boldsymbol{y} = (\boldsymbol{X}'\boldsymbol{M_F}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M_F}[\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{D}\boldsymbol{\theta} + \boldsymbol{F}\boldsymbol{\psi} + \boldsymbol{G}\boldsymbol{\phi} + \boldsymbol{\omega}]$$

Usando a propriedade $\boldsymbol{M_A}\boldsymbol{A} = \boldsymbol{0}$, temos

$$\widehat{\boldsymbol{\beta}}_{EFe} = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{M_F}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M_F}\boldsymbol{D}\boldsymbol{\theta} + (\boldsymbol{X}'\boldsymbol{M_F}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M_F}\boldsymbol{G}\boldsymbol{\phi} + (\boldsymbol{X}'\boldsymbol{M_F}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M_F}\boldsymbol{\omega} =$$

Sob a hipótese usual de que $E[\boldsymbol{\omega}|\boldsymbol{X}, \boldsymbol{D}, \boldsymbol{F}, \boldsymbol{G}] = \boldsymbol{0}$, e por procedimento análogo ao da derivação número 4 acima, temos

$$E\left(\widehat{\boldsymbol{\beta}}_{EFe}\right) = \boldsymbol{\beta} + E[(\boldsymbol{X}'\boldsymbol{M_F}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M_F}\boldsymbol{D}\boldsymbol{\theta}] + E[(\boldsymbol{X}'\boldsymbol{M_F}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M_F}\boldsymbol{G}\boldsymbol{\phi}]$$

B.6 Derivação de (IV*)

$$\widehat{\boldsymbol{\beta}}_{EFte} = \left(\boldsymbol{X}'\boldsymbol{M}_{[\boldsymbol{D}\ \boldsymbol{F}]}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{M}_{[\boldsymbol{D}\ \boldsymbol{F}]}\boldsymbol{y} =$$

$$= \left(\boldsymbol{X}'\boldsymbol{M}_{[\boldsymbol{D}\ \boldsymbol{F}]}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{M}_{[\boldsymbol{D}\ \boldsymbol{F}]}[\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{D}\boldsymbol{\theta} + \boldsymbol{F}\boldsymbol{\psi} + \boldsymbol{G}\boldsymbol{\phi} + \boldsymbol{\omega}] =$$

$$= \left(\boldsymbol{X}'\boldsymbol{M}_{[\boldsymbol{D}\ \boldsymbol{F}]}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{M}_{[\boldsymbol{D}\ \boldsymbol{F}]}\left[\boldsymbol{X}\boldsymbol{\beta} + [\boldsymbol{D}\ \boldsymbol{F}]\begin{bmatrix}\boldsymbol{\theta}\\ \boldsymbol{\psi}\end{bmatrix} + \boldsymbol{G}\boldsymbol{\phi} + \boldsymbol{\omega}\right]$$

Usando a propriedade $\boldsymbol{M_A}\boldsymbol{A} = \boldsymbol{0}$, temos

$$\widehat{\boldsymbol{\beta}}_{EFte} = \boldsymbol{\beta} + \left(\boldsymbol{X}'\boldsymbol{M}_{[\boldsymbol{D}\ \boldsymbol{F}]}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{M}_{[\boldsymbol{D}\ \boldsymbol{F}]}\boldsymbol{G}\boldsymbol{\phi} + \left(\boldsymbol{X}'\boldsymbol{M}_{[\boldsymbol{D}\ \boldsymbol{F}]}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{M}_{[\boldsymbol{D}\ \boldsymbol{F}]}\boldsymbol{\omega}$$

Como o efeito fixo da qualidade do *match* é constante na história de um dado empregado em dado empregador, a segunda parcela do lado direito da equação acima é zero. Sabendo

disso, e sob a hipótese usual de que $E[\boldsymbol{\omega}|\boldsymbol{X},\boldsymbol{D},\boldsymbol{F},\boldsymbol{G}] = \mathbf{0}$, por procedimento análogo ao da derivação número 4 acima, temos

$$E\left(\widehat{\boldsymbol{\beta}}_{EFte}\right) = \boldsymbol{\beta}$$

APÊNDICE C – Derivações das equações para estimação *within*

C.1 Derivação de (16)

No caso do *EFt*, assume-se momentaneamente que o modelo verdadeiro é dado por

$$\ln w_{ijt} = \boldsymbol{x}'_{ijt}\boldsymbol{\beta} + \theta_i + \varepsilon_{ijt}$$

Partindo deste modelo e denotando por $J(i,t)$ a função que retorna o empregador j no qual o trabalhador i está empregado em t, podemos escrever a média da equação, calculada no nível do trabalhador i para todo empregador e instante do tempo:

$$\frac{\sum_t[\ln w_{iJ(i,t)t}]}{T} = \frac{\sum_t\left[\boldsymbol{x}'_{iJ(i,t)t}\right]}{T}\boldsymbol{\beta} + \frac{\sum_t[\theta_i]}{T} + \frac{\sum_t[\varepsilon_{iJ(i,t)t}]}{T} \Longrightarrow \overline{\ln w_i} = \overline{\boldsymbol{x}}'_i\beta + \theta_i + \bar{\varepsilon}_i$$

Agora, subtraindo esta equação da equação anterior, para cada trabalhador i, teremos o modelo em desvios da média no nível do trabalhador:

$$\ln w_{ijt} - \overline{\ln w_i} = \left(\boldsymbol{x}'_{ijt} - \overline{\boldsymbol{x}}'_i\right)\boldsymbol{\beta} + (\theta_i - \theta_i) + \left(\varepsilon_{ijt} - \bar{\varepsilon}_i\right) \Longrightarrow$$

$$\Longrightarrow \ln w_{ijt} - \overline{\ln w_i} = \left(\boldsymbol{x}'_{ijt} - \overline{\boldsymbol{x}}'_i\right)\boldsymbol{\beta} + \left(\varepsilon_{ijt} - \bar{\varepsilon}_i\right)$$

C.2 Derivação de (17)

No caso do *EFe*, assume-se momentaneamente que o modelo verdadeiro é dado por

$$\ln w_{ijt} = \boldsymbol{x}'_{ijt}\boldsymbol{\beta} + \psi_j + \varepsilon_{ijt}$$

Seja N_{it}^{j} o número de pares (i,t) presentes na amostra para o empregador j. Ele é dado por

$$N_{it}^{j} = \sum_i \sum_t 1(J(i,t) = j)$$

onde $1(\cdot)$ é a função indicadora, que assume valor 1 se seu conteúdo é verdadeiro, 0 se falso. Podemos escrever a média da equação salarial, média esta calculada no nível do empregador j para todo trabalhador e instante do tempo, assim:

$$\frac{\sum_i \sum_t [\ln w_{ijt}]}{N_{it}^{j}} = \frac{\sum_i \sum_t [\boldsymbol{x}'_{ijt}]}{N_{it}^{j}} \boldsymbol{\beta} + \frac{\sum_i \sum_t [\psi_j]}{N_{it}^{j}} + \frac{\sum_i \sum_t [\varepsilon_{ijt}]}{N_{it}^{j}} \Longrightarrow \overline{\ln w_j} = \overline{\boldsymbol{x}}'_j \boldsymbol{\beta} + \psi_j + \bar{\varepsilon}_j$$

Agora, subtraindo esta equação da equação anterior, para cada empregador j, teremos o modelo em desvios da média no nível do empregador:

$$\ln w_{ijt} - \overline{\ln w_j} = (\boldsymbol{x}'_{ijt} - \overline{\boldsymbol{x}}'_j)\boldsymbol{\beta} + (\psi_j - \psi_j) + (\varepsilon_{ijt} - \bar{\varepsilon}_j) \Longrightarrow$$

$$\Longrightarrow \ln w_{ijt} - \overline{\ln w_j} = (\boldsymbol{x}'_{ijt} - \overline{\boldsymbol{x}}'_j)\boldsymbol{\beta} + (\varepsilon_{ijt} - \bar{\varepsilon}_j)$$

C.3 Derivação de (18)

No caso do *EFtec*, assume-se que o modelo verdadeiro é dado por

$$\ln w_{ijt} = \boldsymbol{x}'_{ijt}\boldsymbol{\beta} + \theta_i + \psi_j + \phi_{ij} + \varepsilon_{ijt}$$

Seja T_{ij} o número de períodos que o *match* entre i e j aparece na amostra. Temos que

$$T_{ij} = \sum_t 1(J(i,t) = j) \ .$$

Podemos escrever a média da equação salarial, média esta calculada no nível do *match* entre i e j ao longo do tempo, da seguinte forma:

$$\frac{\sum_t[\ln w_{ijt}]}{T_{ij}} = \frac{\sum_t\left[\boldsymbol{x}'_{ijt}\right]}{T_{ij}}\boldsymbol{\beta} + \frac{\sum_t[\theta_i]}{T_{ij}} + \frac{\sum_t[\psi_j]}{T_{ij}} + \frac{\sum_t[\phi_{ij}]}{T_{ij}} + \frac{\sum_t[\varepsilon_{ijt}]}{T_{ij}} \Longrightarrow$$

$$\Longrightarrow \overline{\ln w_{ij}} = \overline{\boldsymbol{x}}'_{ij}\boldsymbol{\beta} + \theta_i + \psi_j + \phi_{ij} + \bar{\varepsilon}_{ij}$$

Agora, subtraindo esta equação da equação anterior para cada *match*, teremos o modelo em desvios da média no nível do *match*:

$$\ln w_{ijt} - \overline{\ln w_{ij}} = \left(\boldsymbol{x}'_{ijt} - \overline{\boldsymbol{x}}'_{ij}\right)\boldsymbol{\beta} + (\theta_i - \theta_i) + \left(\psi_j - \psi_j\right) + \left(\phi_{ij} - \phi_{ij}\right) + \\ + \left(\varepsilon_{ijt} - \bar{\varepsilon}_{ij}\right) \Longrightarrow$$

$$\Longrightarrow \ln w_{ijt} - \overline{\ln w_{ij}} = \left(\boldsymbol{x}'_{ijt} - \overline{\boldsymbol{x}}'_{ij}\right)\boldsymbol{\beta} + \left(\varepsilon_{ijt} - \bar{\varepsilon}_{ij}\right)$$

APÊNDICE D – Estimações incluindo todas as seções CNAE 1.0

Tabela D1 – Especificação básica (todas as seções CNAE 1.0)

Variável dependente: logaritmo do salário

	POLS	*EFt*	*EFe*	*EFtec*
Número de empregados				
5 a 9	0.1888***	0.0853***	0.0272***	0.0321***
	(0.0074)	(0.0055)	(0.0052)	(0.0044)
10 a 19	0.2894***	0.1489***	0.0572***	0.0601***
	(0.0074)	(0.0067)	(0.0068)	(0.0057)
20 a 49	0.3929***	0.2090***	0.0820***	0.0901***
	(0.0071)	(0.0074)	(0.0084)	(0.0068)
50 a 99	0.4999***	0.2507***	0.0965***	0.1123***
	(0.0075)	(0.0079)	(0.0100)	(0.0078)
100 a 249	0.5883***	0.2860***	0.1040***	0.1313***
	(0.0071)	(0.0081)	(0.0113)	(0.0085)
250 a 499	0.6457***	0.3144***	0.1129***	0.1612***
	(0.0073)	(0.0083)	(0.0126)	(0.0090)
500 a 999	0.6939***	0.3280***	0.1155***	0.1869***
	(0.0074)	(0.0085)	(0.0139)	(0.0095)
1000 ou mais	0.6673***	0.3505***	0.1026***	0.2274***
	(0.0070)	(0.0088)	(0.0154)	(0.0100)
N	166 258			
J	100 348			
T	8			
N·T	1 330 064			
n^0 *matches* diferentes	236 497			

Nota 1: erros-padrão abaixo dos coeficientes, entre parênteses.
Nota 2: * = p-valor < 0.1 ** = p-valor < 0.05 *** = p-valor < 0.01
Nota 3: controles presentes em todas as regressões: *dummies* para tamanho do estabelecimento, grau de instrução do trabalhador, faixa etária do trabalhador, tempo no emprego (*tenure*), unidade federativa do estabelecimento empregador, região metropolitana do estabelecimento empregador, indústria (seção CNAE 1.0), ocupação (grandes grupos CBO 1994), *dummies* para ano e *dummy* para emprego temporário.

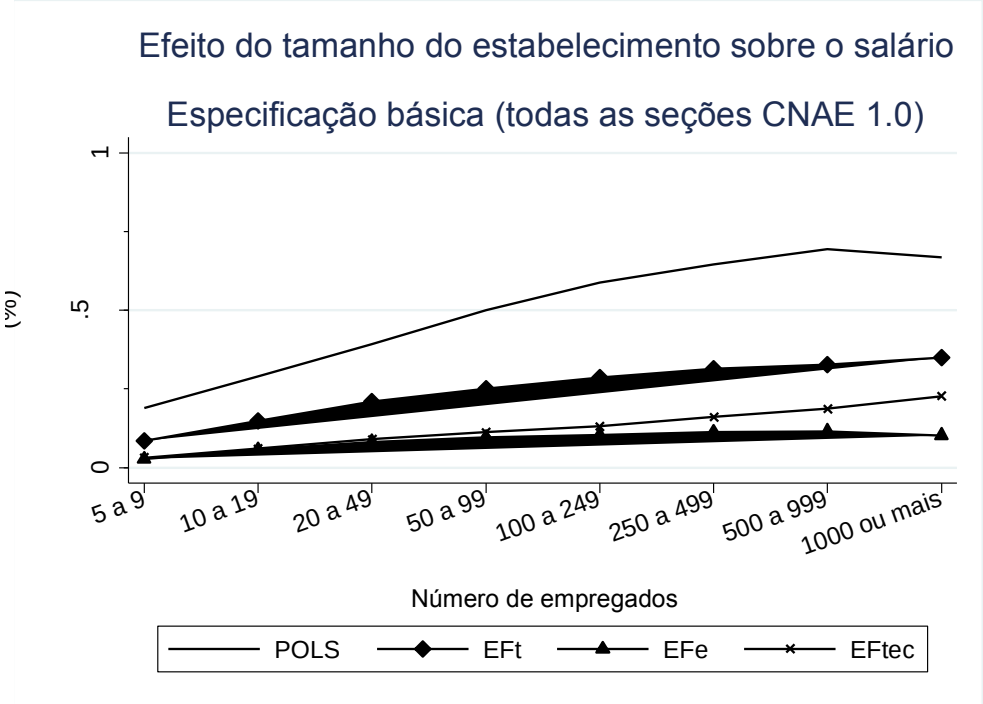

Gráfico D1 – Especificação básica (todas as seções CNAE 1.0)

Tabela D2 – Estabelecimentos com mais instruídos (todas as seções CNAE 1.0)

Variável dependente: logaritmo do salário

	POLS	*EFt*	*EFe*	*EFtec*
Número de empregados				
5 a 9	0.2852***	0.1216***	0.0380***	0.0399***
	(0.01566)	(0.01139)	(0.00917)	(0.00851)
10 a 19	0.5126***	0.2023***	0.0728***	0.0676***
	(0.01553)	(0.01384)	(0.01241)	(0.01120)
20 a 49	0.6323***	0.2582***	0.0888***	0.0882***
	(0.01458)	(0.01504)	(0.01460)	(0.01302)
50 a 99	0.7476***	0.2953***	0.0998***	0.1086***
	(0.01515)	(0.01589)	(0.01749)	(0.01476)
100 a 249	0.8622***	0.3360***	0.1216***	0.1407***
	(0.01405)	(0.01610)	(0.01985)	(0.01614)
250 a 499	0.8997***	0.3598***	0.1275***	0.1683***
	(0.01403)	(0.01628)	(0.02221)	(0.01684)
500 a 999	0.9554***	0.3796***	0.1602***	0.2035***
	(0.01384)	(0.01652)	(0.02384)	(0.01745)
1000 ou mais	0.8725***	0.4127***	0.1578***	0.2525***
	(0.01323)	(0.01670)	(0.02555)	(0.01793)
N		94 573		
J		35 849		
T		8		
N·T		756 584		
nº *matches* diferentes		126 809		

Nota 1: erros-padrão abaixo dos coeficientes, entre parênteses.
Nota 2: * = p-valor < 0.1 ** = p-valor < 0.05 *** = p-valor < 0.01
Nota 3: controles presentes em todas as regressões: *dummies* para tamanho do estabelecimento, grau de instrução do trabalhador, faixa etária do trabalhador, tempo no emprego (*tenure*), unidade federativa do estabelecimento empregador, região metropolitana do estabelecimento empregador, indústria (seção CNAE 1.0), ocupação (grandes grupos CBO 1994), *dummies* para ano e *dummy* para emprego temporário.

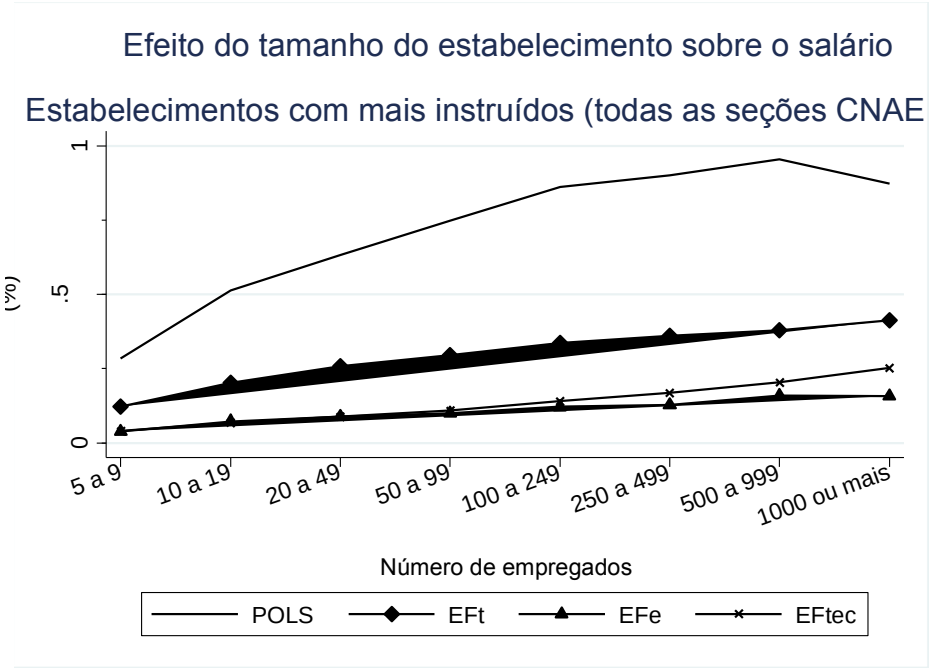

Gráfico D2 – Estabelecimentos com mais instruídos (todas as seções CNAE 1.0)

Tabela D3 – Especificação *EFtec* para sub-amostras de 4 RMs (todas as seções CNAE 1.0)

Variável dependente: logaritmo do salário

	São Paulo	*Rio de Janeiro*	*Porto Alegre*	*Curitiba*
Número de empregados				
5 a 9	0.0355***	0.0559***	0.0166	0.0126
	(0.0123)	(0.0145)	(0.0340)	(0.0358)
10 a 19	0.0491***	0.1069***	0.0568	0.0895*
	(0.0153)	(0.0193)	(0.0477)	(0.0463)
20 a 49	0.0853***	0.1417***	0.1307**	0.0606
	(0.0172)	(0.0231)	(0.0552)	(0.0540)
50 a 99	0.1045***	0.1634***	0.1469**	0.0742
	(0.0192)	(0.0264)	(0.0618)	(0.0576)
100 a 249	0.1288***	0.1781***	0.1630**	0.1168*
	(0.0207)	(0.0289)	(0.0678)	(0.0612)
250 a 499	0.1568***	0.2068***	0.2288***	0.1681**
	(0.0224)	(0.0305)	(0.0697)	(0.0666)
500 a 999	0.1899***	0.2421***	0.2544***	0.3310***
	(0.0241)	(0.0320)	(0.0722)	(0.0714)
1000 ou mais	0.2256***	0.2630***	0.3183***	0.3729***
	(0.0253)	(0.0341)	(0.0749)	(0.0754)
N	27 035	14 867	4 356	4 157
J	17 177	8 956	2 891	2 009
T	8	8	8	8
N·T	216 280	118 936	34 848	33 256
nº *matches* diferentes	37 781	21 614	5 905	5 752

Nota: aplicam-se as notas 1 e 2 da Tabela 2. Controles presentes em todas as regressões: *dummies* para tamanho do estabelecimento, grau de instrução do trabalhador, faixa etária do trabalhador, tempo no emprego (*tenure*), indústria (seção CNAE 1.0), ocupação (grandes grupos CBO 1994), *dummies* para ano e *dummy* para emprego temporário.

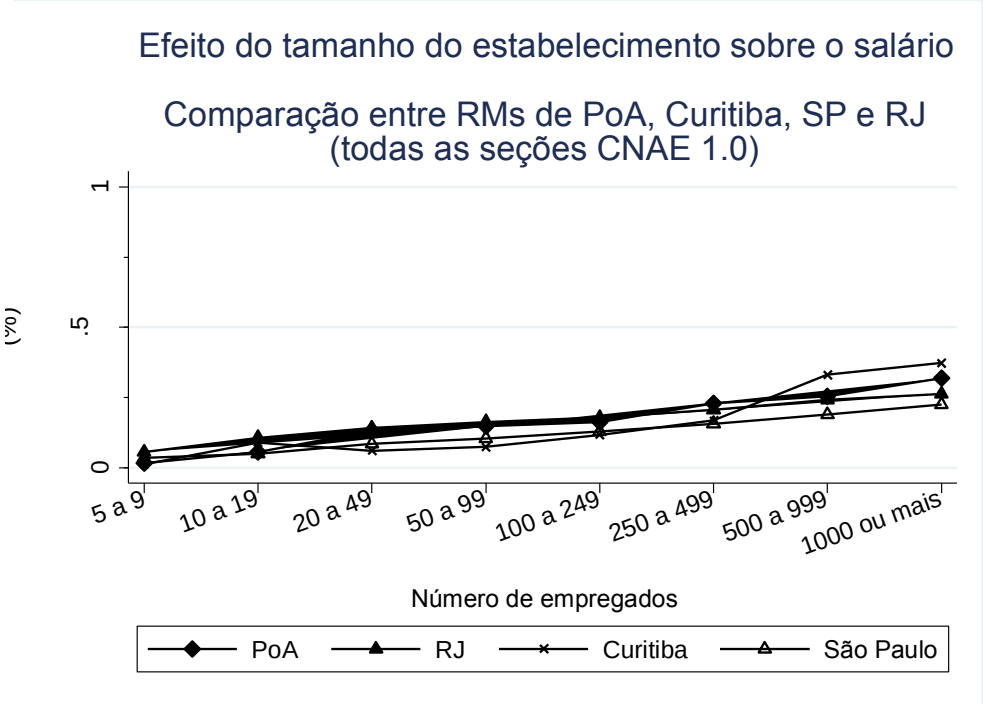

Gráfico D3 – Comparação entre RMs de PoA, Curitiba, SP e RJ (todas as seções CNAE 1.0)
Nota ao Gráfico D3: PoA = Porto Alegre.