Slippery Slopes...and Intercepts

Pedro Portugal

Spring 2025, Carcavelos

Navigation

- The importance of high-dimensional fixed effects
  - The sources of wage variation
  - The sources of gender discrimination
  - The sources of wages dispersion
- The notion of high-dimensional slopes
  - Random trends
  - CAPM models
  - Heterogenous returns to education
  - Insider outsider wages
- Extensions

Worker heterogeneity

WORKER-SIDE explanations for the wage differentials:

- Perfect competition: wage differentials reflect differences in workers' productivity, which depends on:
  - Workers' skills (observed or not)
  - Motivation
  - Ability
  -
- Role of education and other human capital variables (seniority, experience, age, etc.): Becker (1962), Spence (1973),
- Assessment by means of earnings functions: Mincer (1974)

Firm heterogeneity

FIRM-SIDE explanations for the wage differentials:

- Theories that explain why firms find it profitable to pay non-competitive wages
- Firms design incentive schemes to retain their workers, attract better workers, and enhance their productivity (compensation and retention policies)
  - implicit contracts
  - principal-agent
  - efficiency-wages
  - rent-sharing
  - insider-outsider
  - Labor market frictions explanations for the wage differentials: job search and matching literature

Job-title heterogeneity

- Third important dimension of wage formation: JOB TITLE HETEROGENEITY
- There are compensating differentials for certain occupations involving:
  - Risks of accidents/injuries
  - Stressful working conditions
  - Complexity of tasks (requiring specific training or unusual skills)
  - Overcrowding of some occupations
  - Possibility to inflict losses on employers/society (unions; industrial action)
  - Entry barriers
  - Technology that can foster unionization

The sources of wage variation: data requirements

- Rich set of information available in the longitudinal matched employer-employee dataset for Portugal, on:
  - The collective agreement that regulates the employment contract applicable to each worker (300 negotiated per year, on average)
  - Detailed occupational categories defined for each collective agreement (100 categories defined by each collective agreement, on average)
- Job title: combination of collective agreement and professional category (around 30,000 per year)

A Reseacher's Dream Dataset

- Quadros de Pessoal
- Years: 1986 to 2017 (1990 and 2001 not available)
- Variables: Characteristics of firms, characteristics of workers, earnings (several components), hours of work (normal and overtime) and unique identifiers for workers (Social Security codes), firms and job titles
- Final dataset (after restrictions):
  - 27,020,044 observations
  - 567,739 firms; 5,492,332 workers; 95,9 thousand job titles
  - 19 years
- Hourly wage = (sum of 5 comp. of wages)/(sum of 2 types of hours)

Institutional wage setting

Wage bargaining system in Portugal prevailing over the sample period:

- Mandatory minimum wage
- Collective bargaining takes place at a sectoral level, but mandatory extensions are commonplace
- Around 30 000 job-title wage floors are settled each year
- Despite very low unionization rates (less than 10 percent)

A wage regression equation with worker, firm, and job title fixed effects

$$y_{ijft} = \lambda_i + \theta_f + \gamma_j + \alpha_t + \beta X_{ift} + \epsilon_{ijft} \quad (1)$$

- y_{ijft} represents the logarithm of the hourly wage for each individual i, in job j, working for firm f in year t
- X_{ift} are observed time-varying characteristics of individual i and firm j in year t
  - Workersime-varying characteristics (age, age squared, seniority, and seniority squared)
  - Firmsime-varying characteristics (log of size)
- λ_i is a worker fixed effect
- θ_f is a firm fixed effect
- γ_j is a job title fixed effect
- α_t are 18 year dummies
- ϵ_{ijft} is assumed to follow the conventional assumptions

Estimation algorithm, Guimarães and Portugal (2010)

C. *Estimation Strategy*

Controlling simultaneously for worker, firm, and job title-specific effects requires the introduction of three high-dimensional fixed effects in the linear regression model. To illustrate our estimation strategy, consider the following linear regression model in matrix form:

$$\mathbf{Y} = \mathbf{Z}\beta + \mathbf{D}_1\lambda + \mathbf{D}_2\theta + \mathbf{D}_3\gamma + u, \tag{2}$$

where $\mathbf{Z}$ is a matrix of time-varying explanatory variables and $\mathbf{D}_1$, $\mathbf{D}_2$, and $\mathbf{D}_3$ are high-dimensional matrices for the fixed effects. The normal equations may be rewritten as

$$\begin{bmatrix} \beta = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{Y} - \mathbf{D}_1\lambda - \mathbf{D}_2\gamma - \mathbf{D}_3\theta) \\ \lambda = (\mathbf{D}_1'\mathbf{D}_1)^{-1}\mathbf{D}_1'(\mathbf{Y} - \mathbf{Z}\beta - \mathbf{D}_2\gamma - \mathbf{D}_3\theta) \\ \theta = (\mathbf{D}_3'\mathbf{D}_3)^{-1}\mathbf{D}_3'(\mathbf{Y} - \mathbf{Z}\beta - \mathbf{D}_1\lambda - \mathbf{D}_2\gamma) \\ \gamma = (\mathbf{D}_2'\mathbf{D}_2)^{-1}\mathbf{D}_2'(\mathbf{Y} - \mathbf{Z}\beta - \mathbf{D}_1\lambda - \mathbf{D}_3\theta) \end{bmatrix},$$

suggesting an iterative solution that alternates between estimation of β, λ, θ, and γ.

The reghdfe stata procedure

reghdfe - Estimates a linear regression model with two or three high dimensional fixed effects.
Syntax
reghdfe depvar [indepvar] [if] [in] , id1(varname) id2(varname) [id3(varname)] [options]
or
reghdfe depvar [indepvar], indata(name) [options]
or
reghdfe var, improve(name) [options]

Description
This command implements the algorithm of Guimarães and Portugal for estimation of a linear regression model with two or three high dimensional fixed effects. The command is particularly suited for use with large data sets because in a first step you can remove the high dimensional fixed effects from the data and then use the transformed regression variables to try alternative specifications of the model.

Author
Paulo Guimaraes, Universidade do Porto, Portugal

Decomposition of wage variability (Torres, Portugal, Addison and Guimarães)

Ten components estimated:

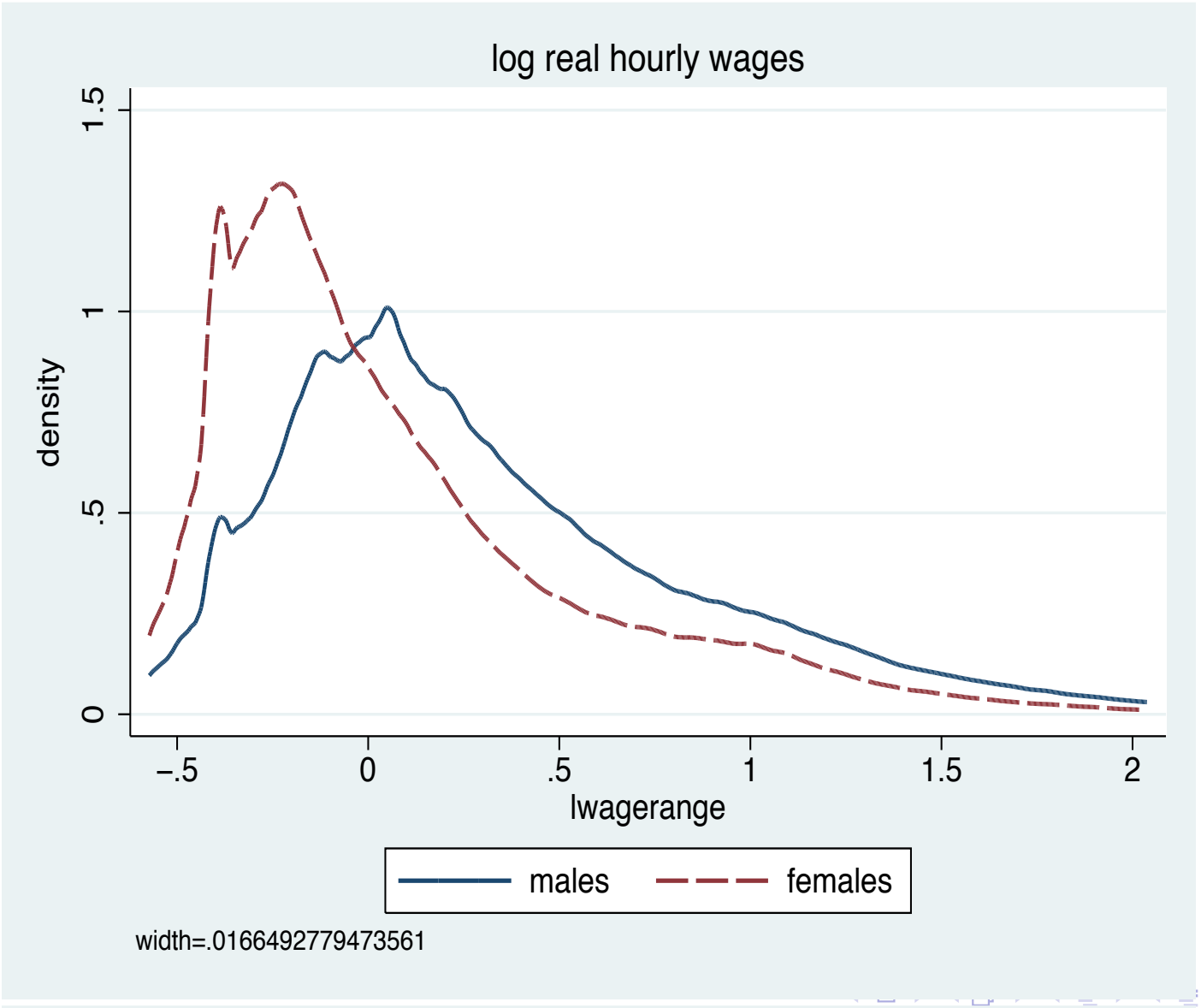
- Observed worker and firm time-varying characteristics (3 components)
- Worker effects, of which:
  - Observed constant (non-time-varying) worker characteristics
  - Unobserved constant worker characteristics
- Firm effects, of which:
  - Observed constant firm characteristics
  - Unobserved constant firm characteristics
- Job title effects, of which:
  - Observed constant job title characteristics
  - Unobserved constant job title characteristics
- Residual component

Decomposition of wage variability

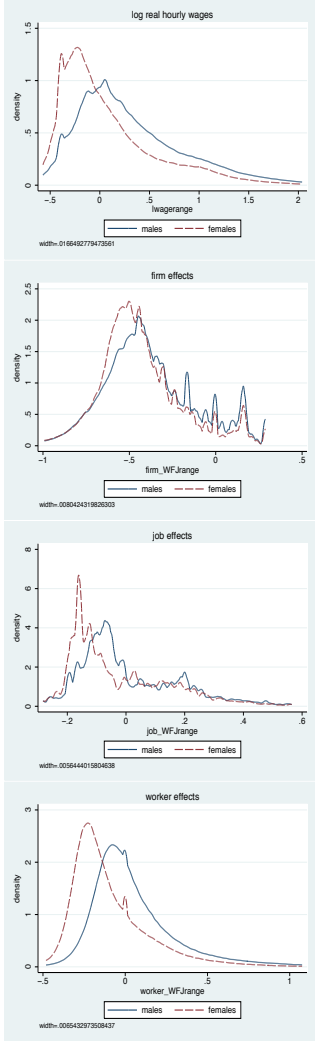
Contribution of the ten components to the real hourly wages:

- Worker fixed effects: 36.0%
  - Unobserved component: 21.0%
  - Observed component (gender and education): 15.0%
- Firm fixed effects: 28.7%
  - Unobserved component: 14.6%
  - Observed component (region, capital ownership, and industry): 14.0%
- Job title fixed effects: 9.7%
  - Unobserved component: 1.9%
  - Observed component (occupation and collective agreement): 7.9%
- Individual time-varying characteristics: 17.4%
  - Time: 6.2%
  - Time-varying observable characteristics of workers (age and seniority): 2.9%
  - Time-varying observable characteristics of firms (size): 5.3%

Cardoso, Guimarães and Portugal (2012)



All you always wanted to know about sex discrimination, Cardoso, Guimarães and Portugal (2012)



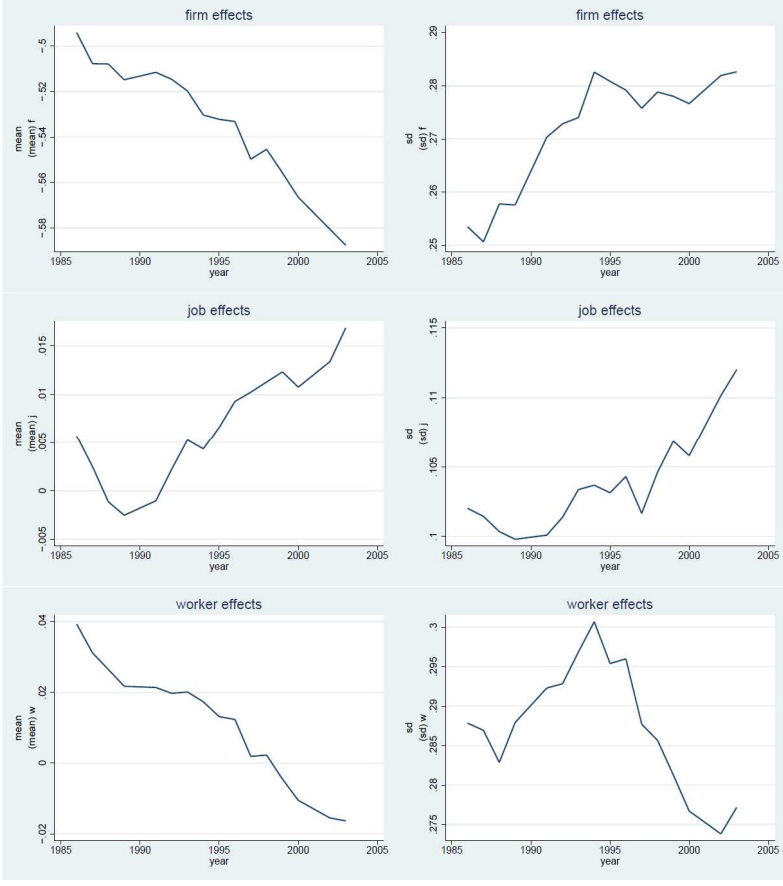
Cardoso,Guimarães and Portugal (2012)

Table: Conditional Decomposition of the Gender Wage Gap

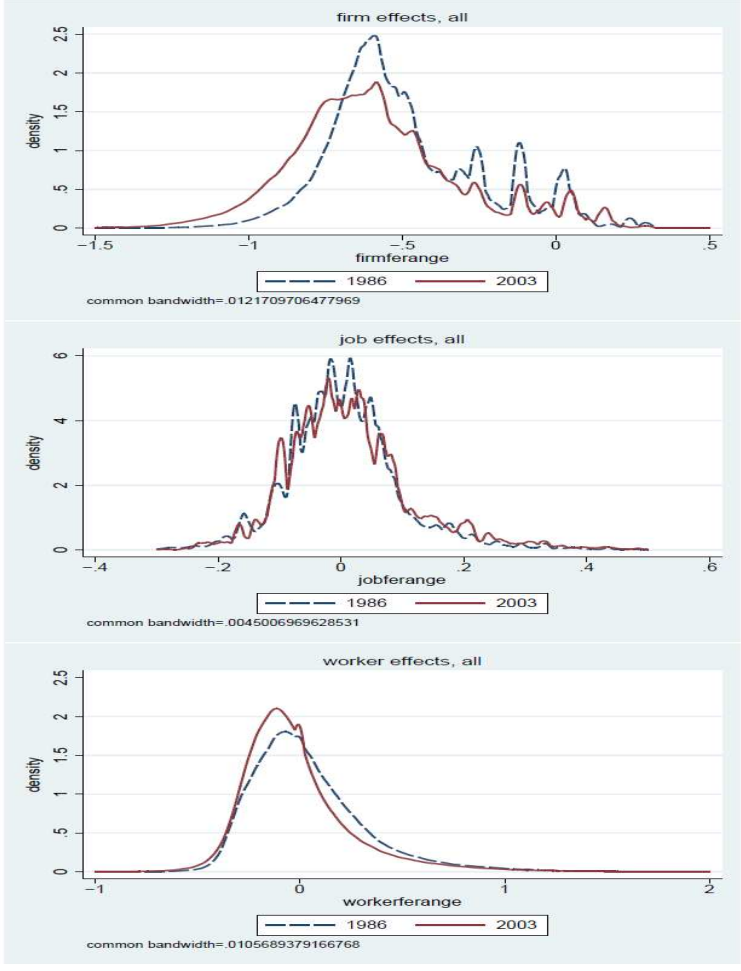
Worker FE	Firm FE	Job FE	Job & Firm FE
0.1987	0.0567		
(0.0001)	(0.0001)		
0.1793		0.0761	
(0.0001)		(0.0001)	
0.1573	0.0514	0.0467	
(0.0001)	(0.0001)	(0.00004)	
0.1332			0.1219
(0.0001)			(0.0001)

Note: Decompositions based on Gelbach (2009).

The sources of wage dispersion, Cardoso, Guimarães and Portugal (2012)



The sources of wage dispersion, Cardoso, Guimarães and Portugal (2012)



A random trend wage regression equation with worker fixed effecst

$$y_{ijft} = \lambda_i + \alpha_t + \phi_i t + \beta X_{ift} + +\epsilon_{ijft} \tag{2}$$

- y_{ijft} represents the logarithm of the hourly wage
- X_{ift} are observed time-varying characteristics of individual i and firm j in year t
- λ_i is a worker fixed effect
- ϕ_t is a random worker trend
- α_t are 18 year dummies
- ϵ_{ijft} is assumed to follow the conventional assumptions

A way to account for the Ashenfelter dip.

The regintfe stata procedure

reghintfe - Estimates a linear regression model with one interacted high dimensional fixed effect.

Syntax

reghintfe depvar [indepvar] [if] [in] , id1(varname) intvar(varname) [options]

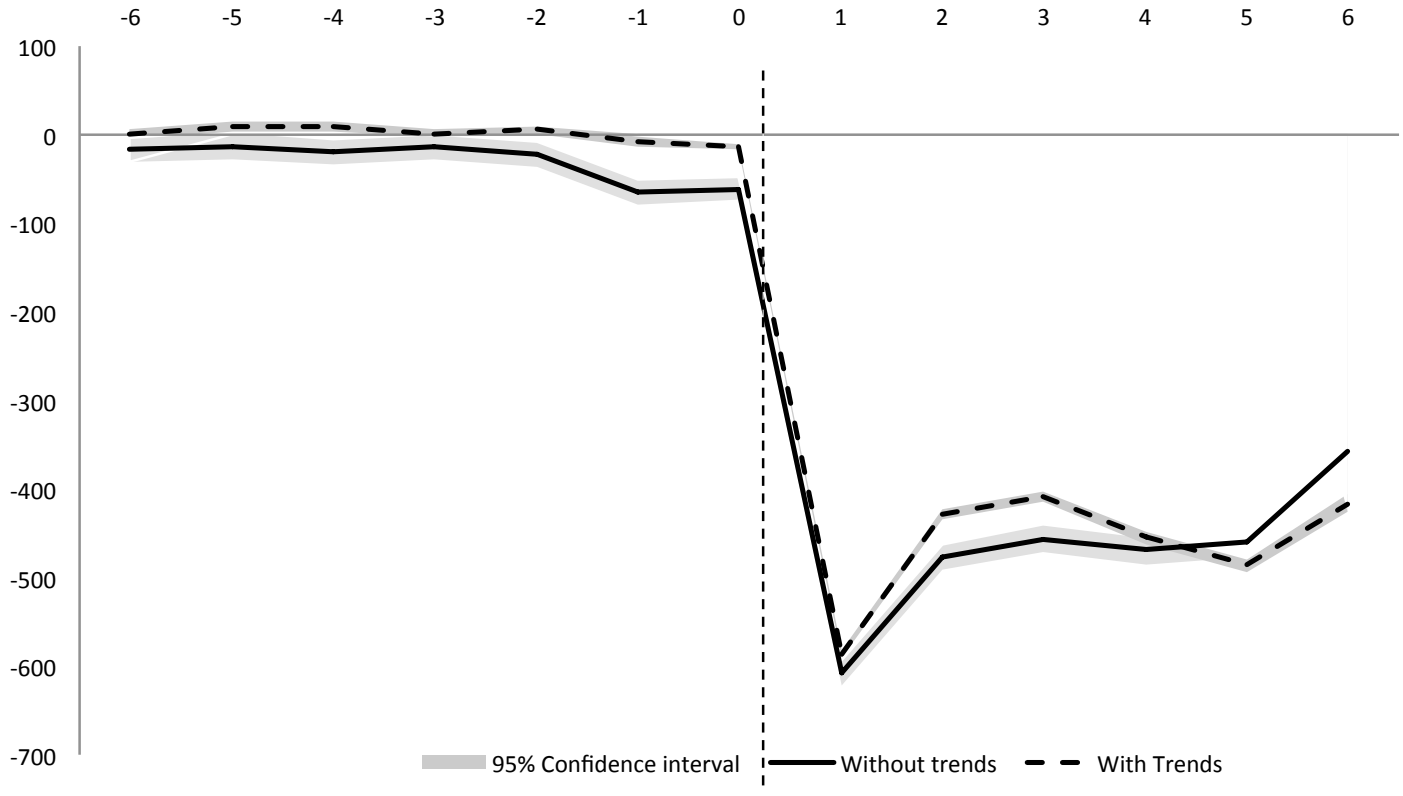
Description

This command estimates a linear regression model with one high-dimensional interacted fixed effect. The command makes use of the Frisch-Waugh-Lovell to avoid computing the dummy variables and interactions for the fixed effects.

Author
Paulo Guimaraes, Universidade do Porto, Portugal

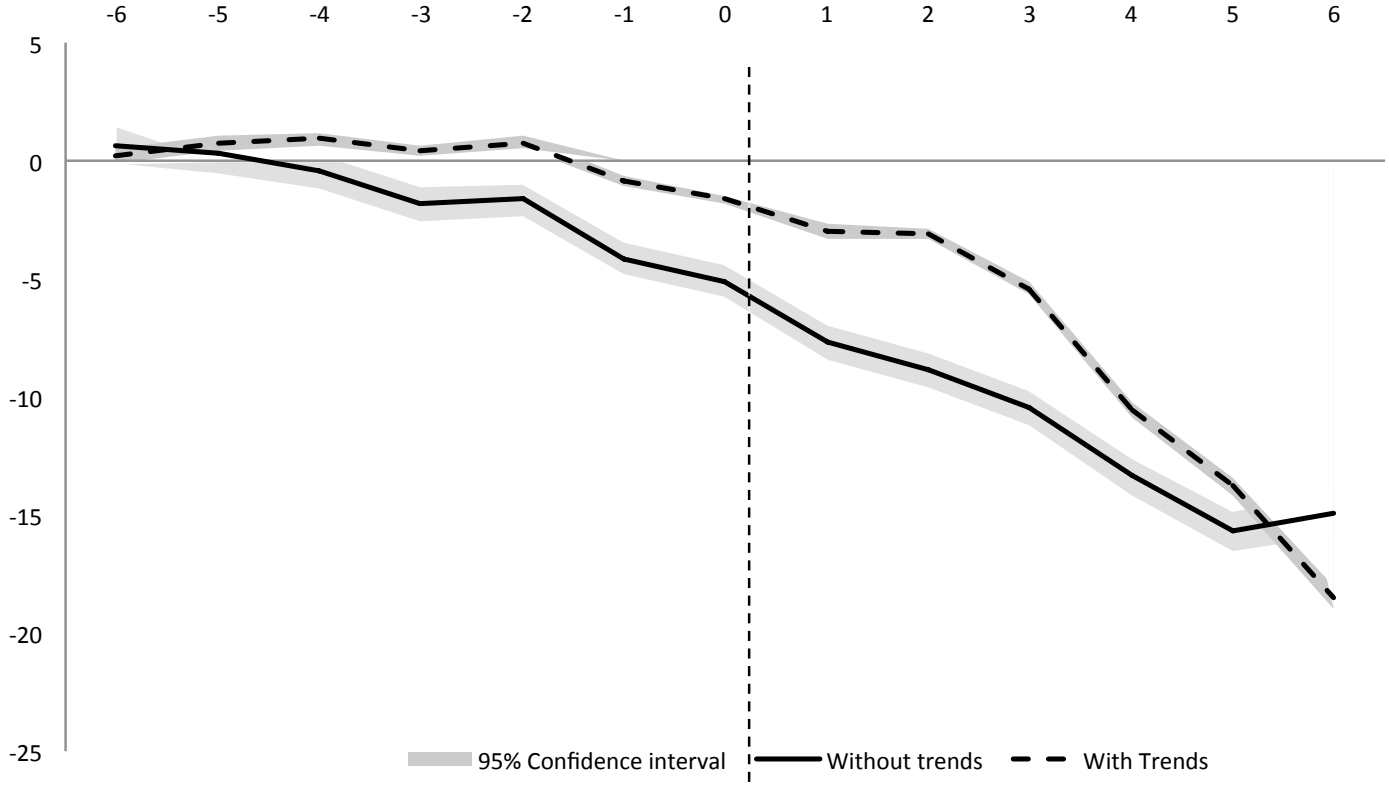
The Sources of the wage losses of the displaced workers (Raposo, Portugal and Carneiro, 2012)

Figure: **Monthly earnings loss of displaced workers**



The sources of the wage losses of the displaced workers

Figure: **Hourly wage loss of displaced workers (%)**



The Sources of the wage losses of the displaced workers

Table: **Gelbach 3 fixed effects decomposition of the wage loss**

Period relative to displacement	Base OLS hourly wage	Full OLS hourly wage	$\delta_k^{base} - \delta_k^{full}$	Worker fixed effect	Firm fixed effect	Job title fixed effect	checksum
D_{-6}	-21.9	0.9	-22.8	-6.6	-14.5	-1.7	0.0
D_{-5}	-23.0	0.8	-23.8	-7.0	-15.1	-1.8	0.0
D_{-4}	-23.8	-0.2	-23.6	-7.9	-13.5	-2.2	0.0
D_{-3}	-24.4	-1.4	-23.0	-7.5	-13.5	-2.0	0.0
D_{-2}	-28.1	-0.6	-27.5	-8.3	-16.2	-2.9	0.0
D_{-1}	-30.8	-1.6	-29.2	-8.8	-15.8	-4.6	0.0
D_0	-33.8	-1.8	-32.0	-9.7	-16.8	-5.4	0.0
D_1	-33.0	-4.0	-28.9	-7.5	-15.7	-5.7	0.0
D_2	-36.9	-4.2	-32.8	-9.5	-16.0	-7.4	0.0
D_3	-39.5	-4.3	-35.2	-10.3	-16.9	-8.0	0.0
D_4	-45.5	-4.6	-41.0	-11.9	-20.7	-8.4	0.0
D_5	-43.9	-5.3	-38.6	-10.5	-20.0	-8.1	0.0
D_6	-39.0	-4.3	-34.7	-7.7	-20.0	-6.9	0.0
$D_{-6} - D_0$	-26.6	-0.6	-26.0	-8.0	-15.1	-2.9	0.0
$D_1 - D_6$	-39.6	-4.4	-35.2	-9.6	-18.2	-7.4	0.0
Δ	-13.1	-3.9	-9.2	-1.6	-3.2	-4.5	0.0

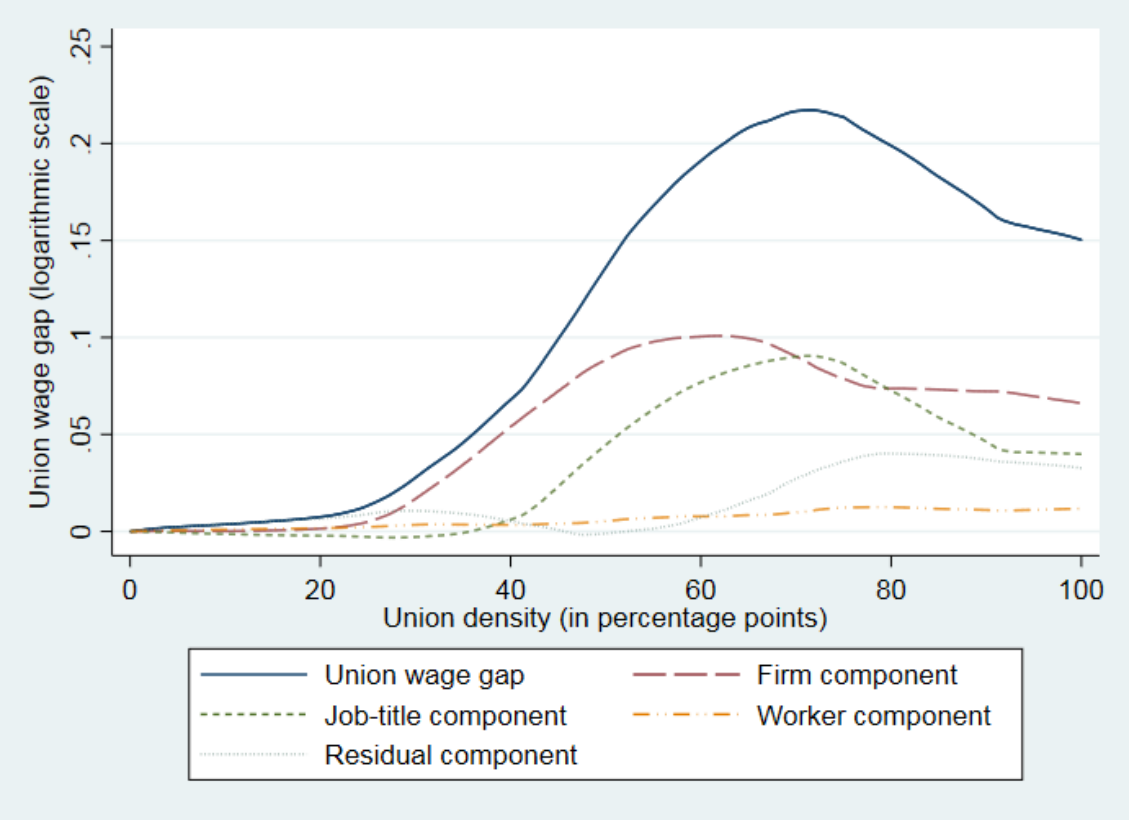
Decomposition of a curve

Decomposition of the union wage gap

Worker heterogeneity

Firm sorting

Job-title sorting.



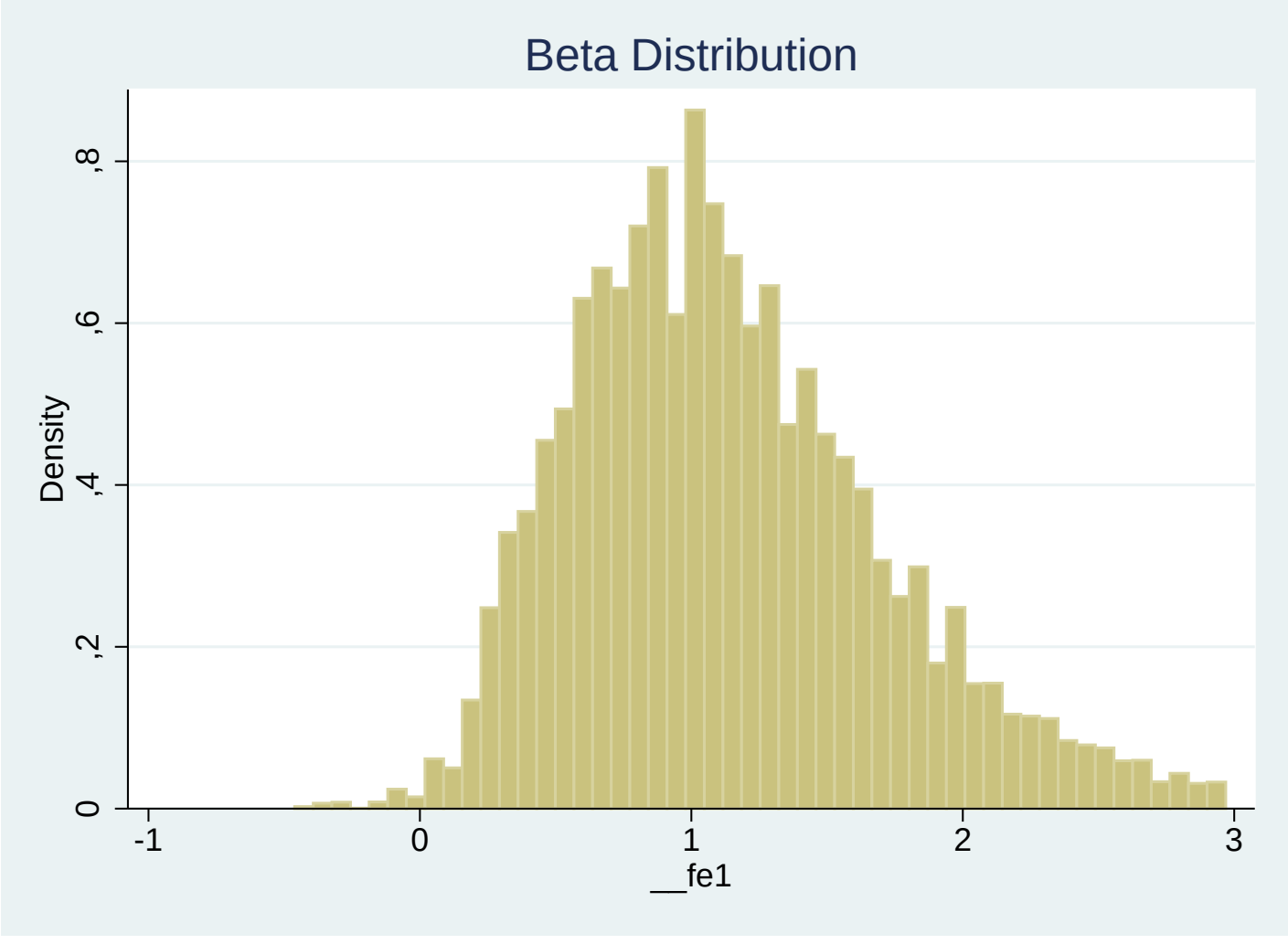
The Capital Asset Princing Model

$$(R_i - R_f)_{it} = \alpha_i + \beta_i(R_m - R_f)_{it} + \epsilon_{it} \tag{3}$$

- R_i represents return on equity i
- R_f is a risk free interest rate
- R_m represents the average market return
- ϵ_{it} is assumed to follow the conventional assumptions

CAPM

Figure: **One thousand separate regressions**



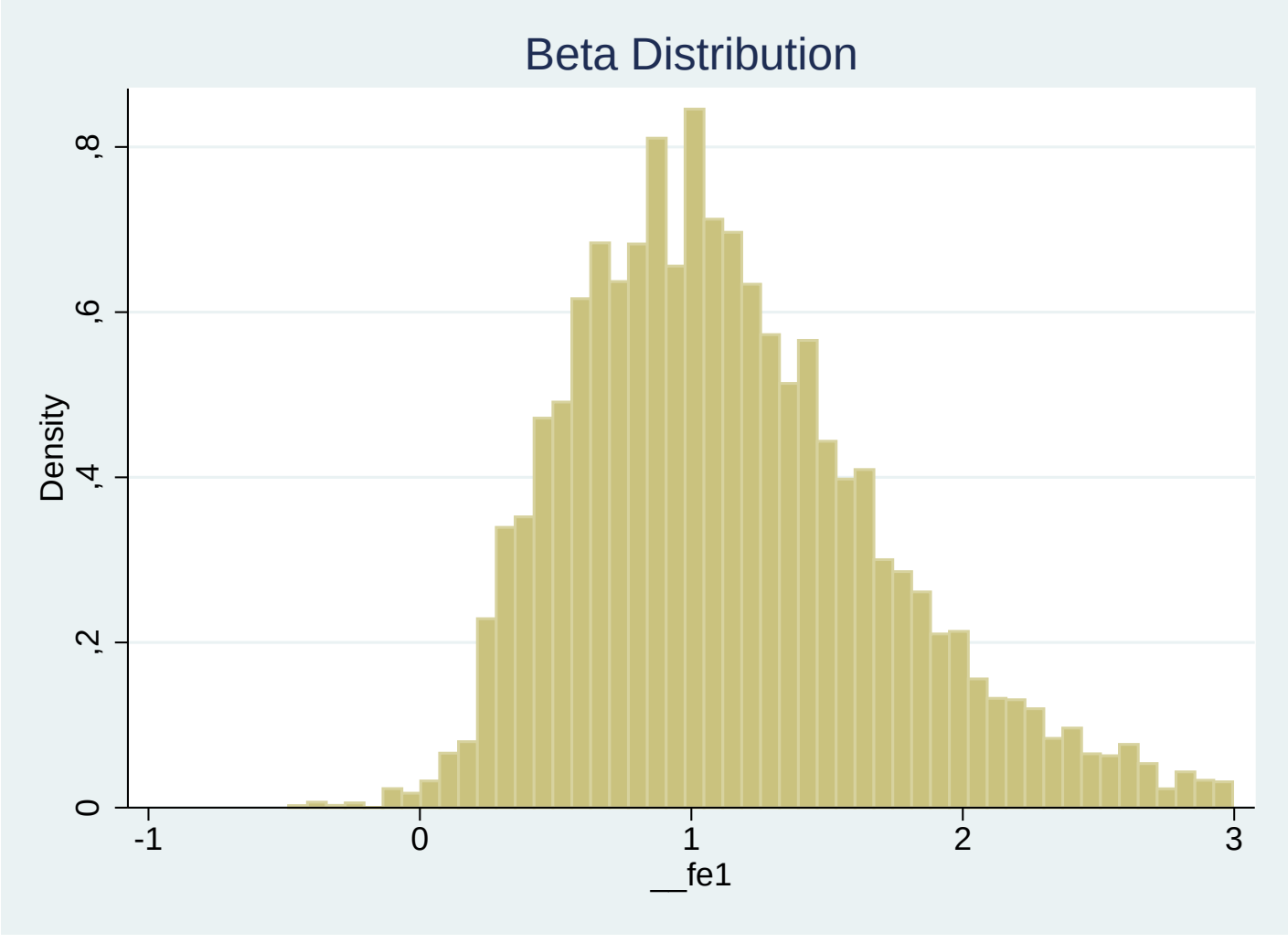
A extension on The Capital Asset Pricing Model

$$(R_i - R_f)_{it} = \gamma_t + \alpha_i + \beta_i(R_m - R_f)_{it} + \epsilon_{it} \quad (4)$$

- R_i represents return on equity i
- R_f is a risk free interest rate
- R_m represents the average market return
- ϵ_{it} is assumed to follow the conventional assumptions

CAPM

Figure: **One thousand separate regressions with time effects**



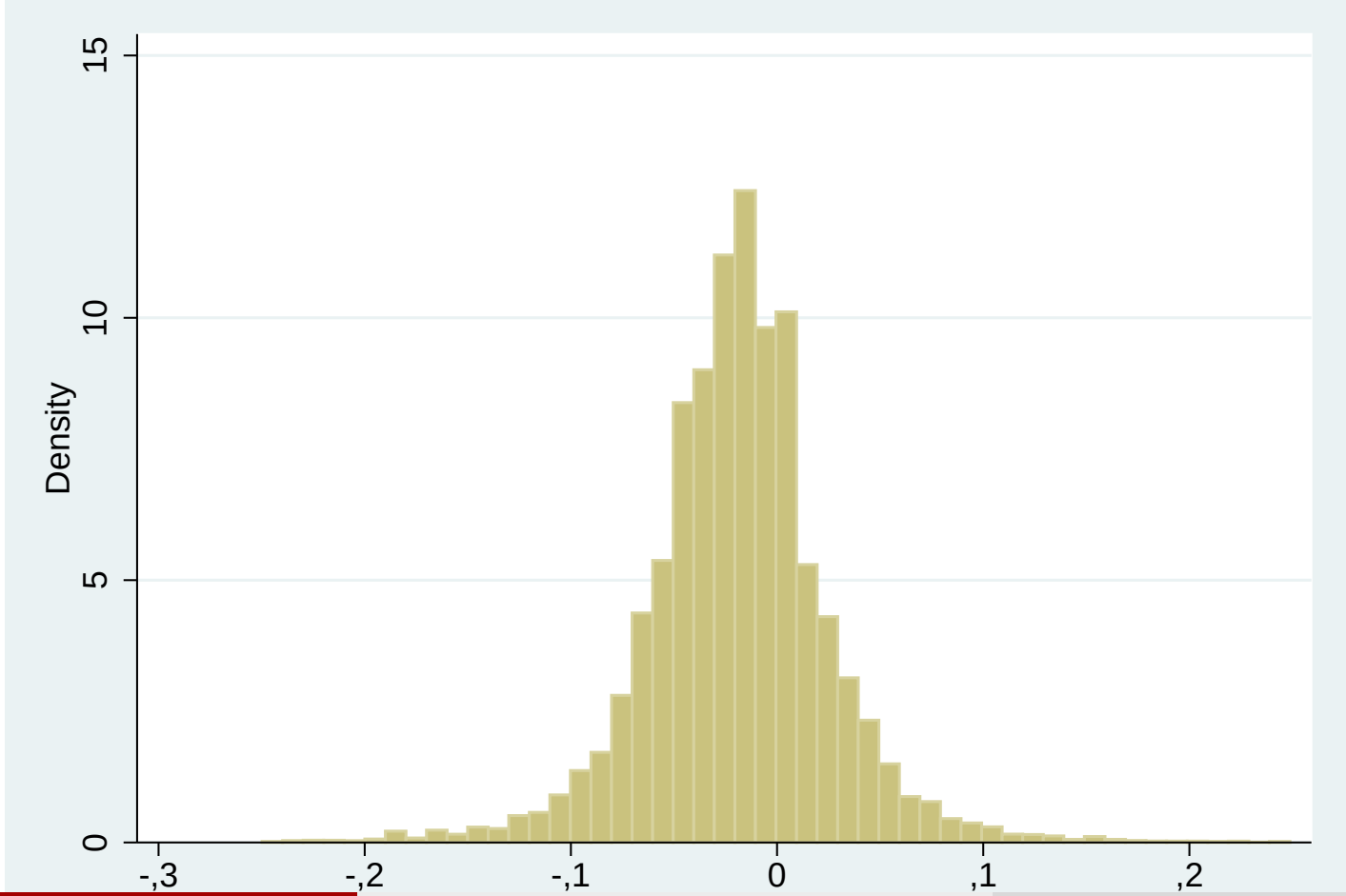
The cyclicality or real wages

$$y_{ijft} = \lambda_i + \theta_f + \gamma_j + \alpha_f cycle_t + \beta X_{ift} + \epsilon_{ijft} \quad (5)$$

- y_{ijft} represents the logarithm of the hourly wage for each individual i, in job j, working for firm f in year t
- X_{ift} are observed time-varying characteristics of individual i and firm j in year t
  - Workersime-varying characteristics (age, age squared, seniority, and seniority squared)
  - Firmsime-varying characteristics (log of size)
- λ_i is a worker fixed effect
- θ_f is a firm fixed effect
- γ_j is a job title fixed effect
- α_f is firm specific coefficient on the business cycle variable
- ϵ_{ijft} is assumed to follow the conventional assumptions

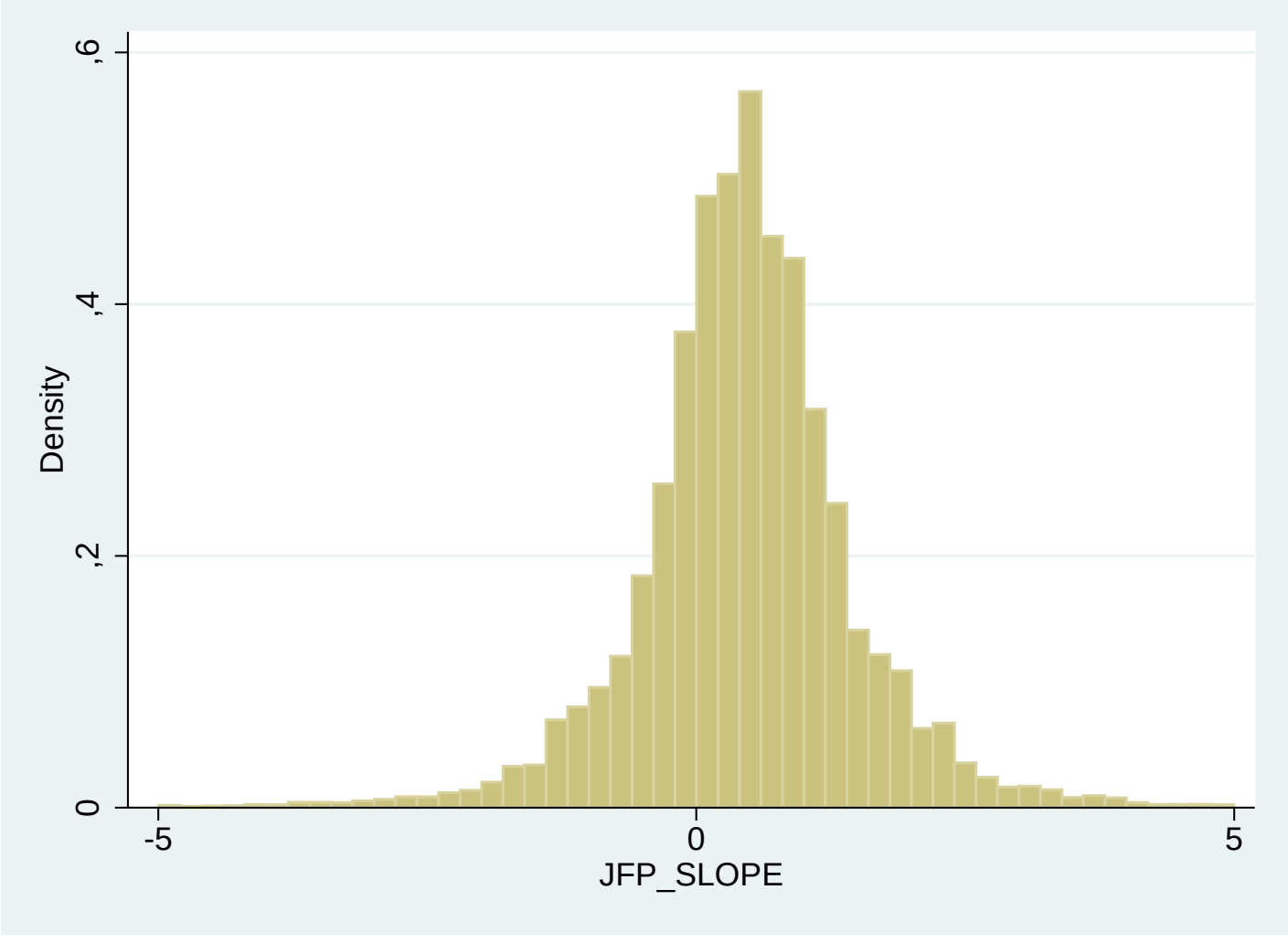
Cyclicality of real wages (Carneiro, Guimarães, and Portugal

Figure: **Unemployment rate**



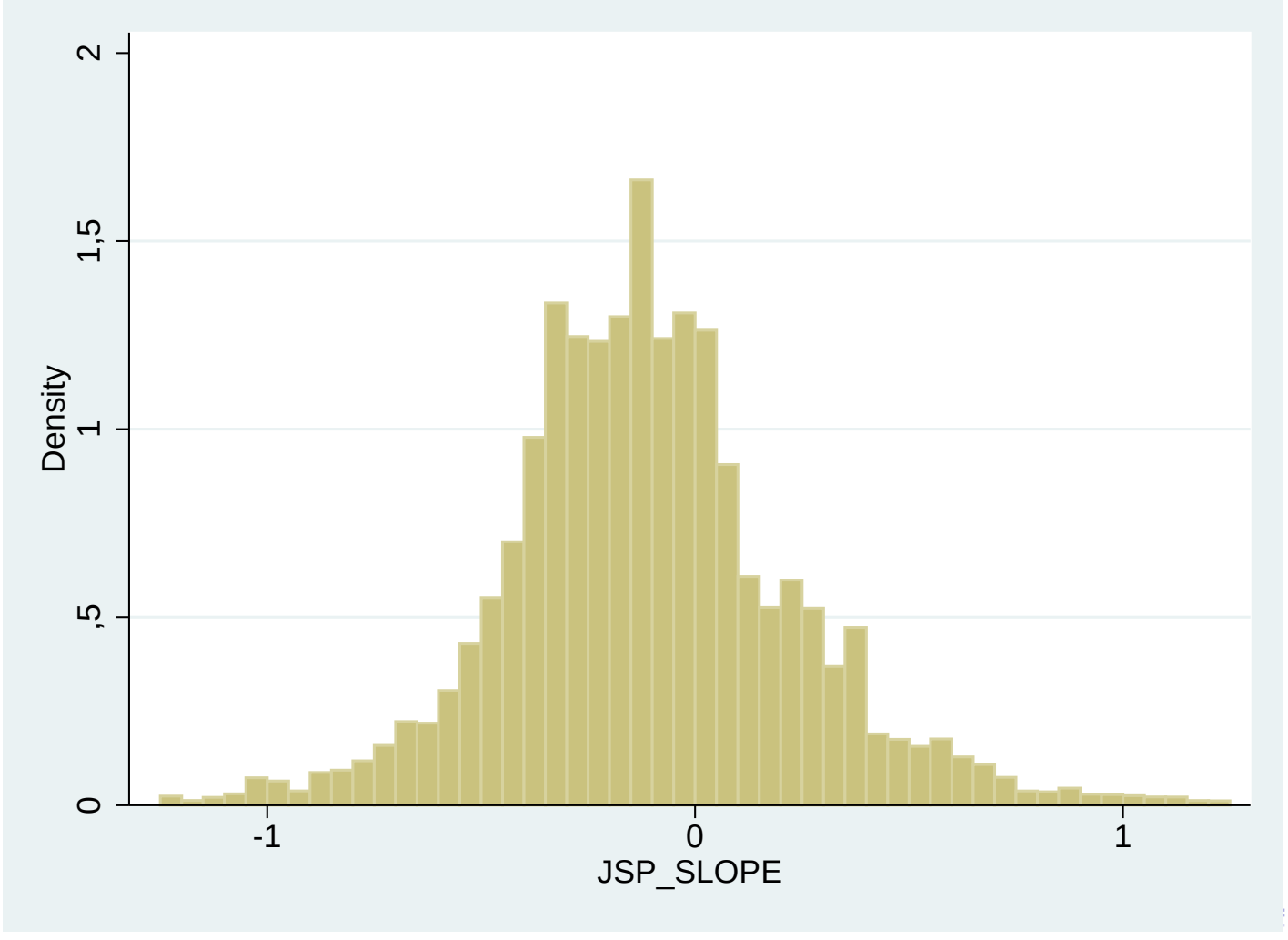
Cyclicality of real wages

Figure: **Job finding probability**



Cyclicality of real wages

Figure: **Job separation probability**



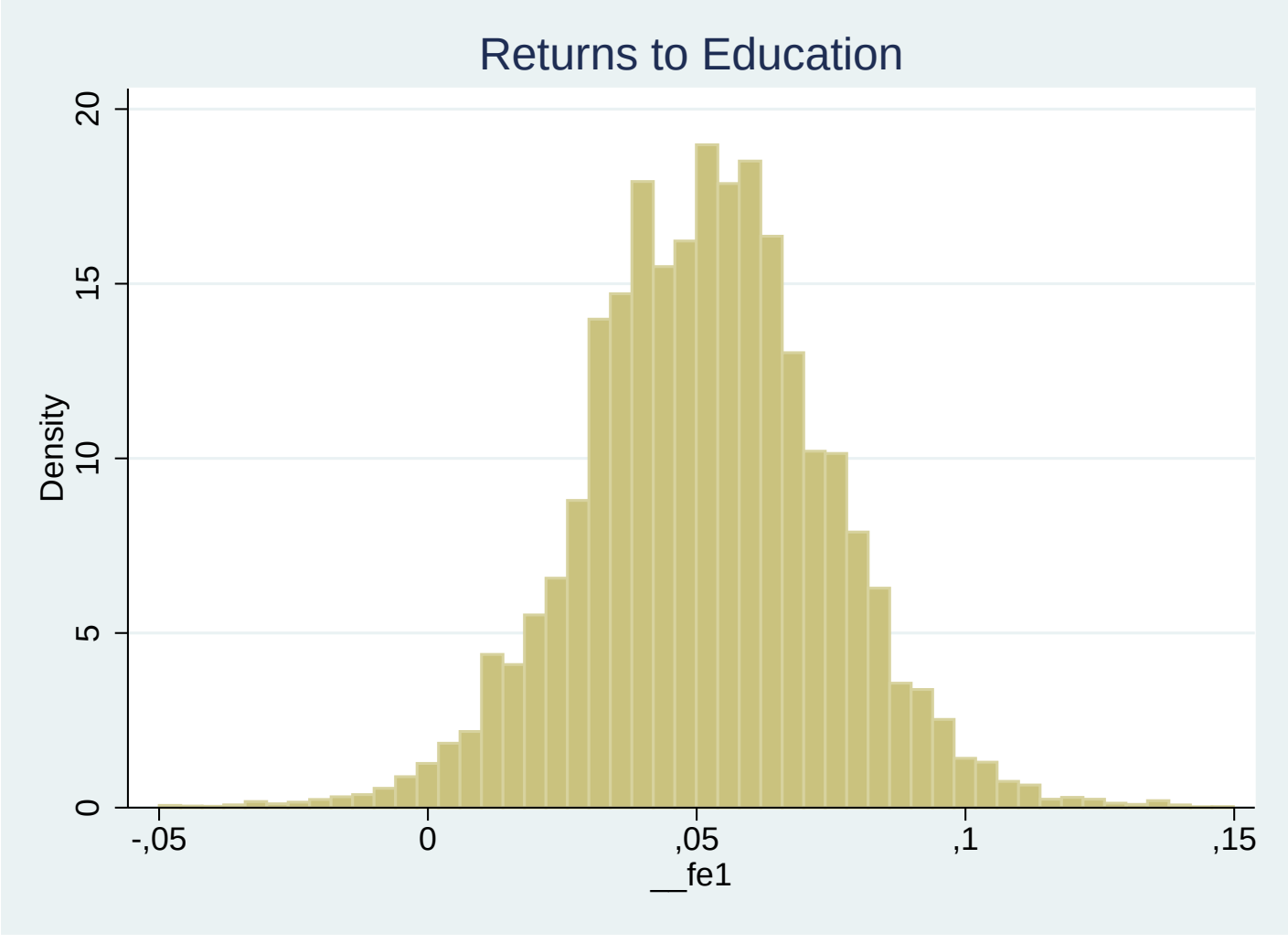
The Returns to Education

$$y_{ift} = \theta_f + \alpha_t + \beta X_{ift} + \phi_f Education_i + \epsilon_{ift} \quad (6)$$

- y_{ift} represents the logarithm of the hourly wage for each individual
- ϕ_f is a firm specific coefficient of education
- X_{ift} represents other observed time-varying characteristics of individual i and firm j in year t
- θ_f is a firm fixed effect
- ϵ_{ift} is assumed to follow the conventional assumptions

Heterogeneity on the returns to education

Figure: **The distribution of the returns to education**



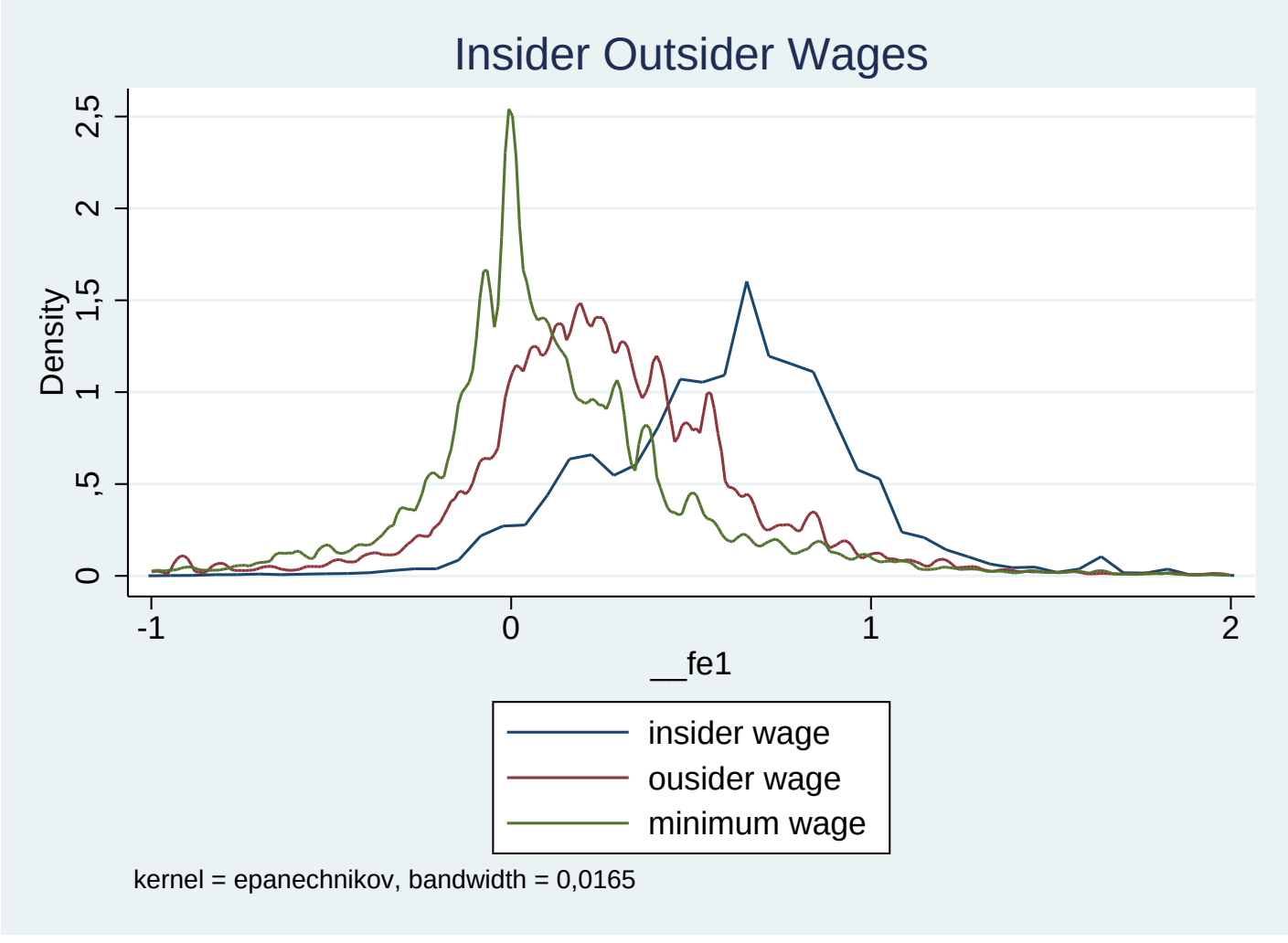
Insider-outsider wages (Guimarães, Martins and Portugal)

$$y_{ift} = \theta_f + \gamma_f \, wage_{insider} + \alpha_f \, wage_{outsider} + \phi_f \, wage_{minimum} + \epsilon_{ift} \qquad (7)$$

- y_{ift} represents the logarithm of the monthly wage
- γ_f is a firm specific coefficient on the firm's ruling wage
- α_f is a firm specific coefficient on the collective bargained wage
- θ_f is a firm fixed effect
- ϵ_{ift} is assumed to follow the conventional
- ϕ_f is a firm specific coefficient on the minimum wage

Insider-outsider wages

Figure: **The determinants of entry wages**



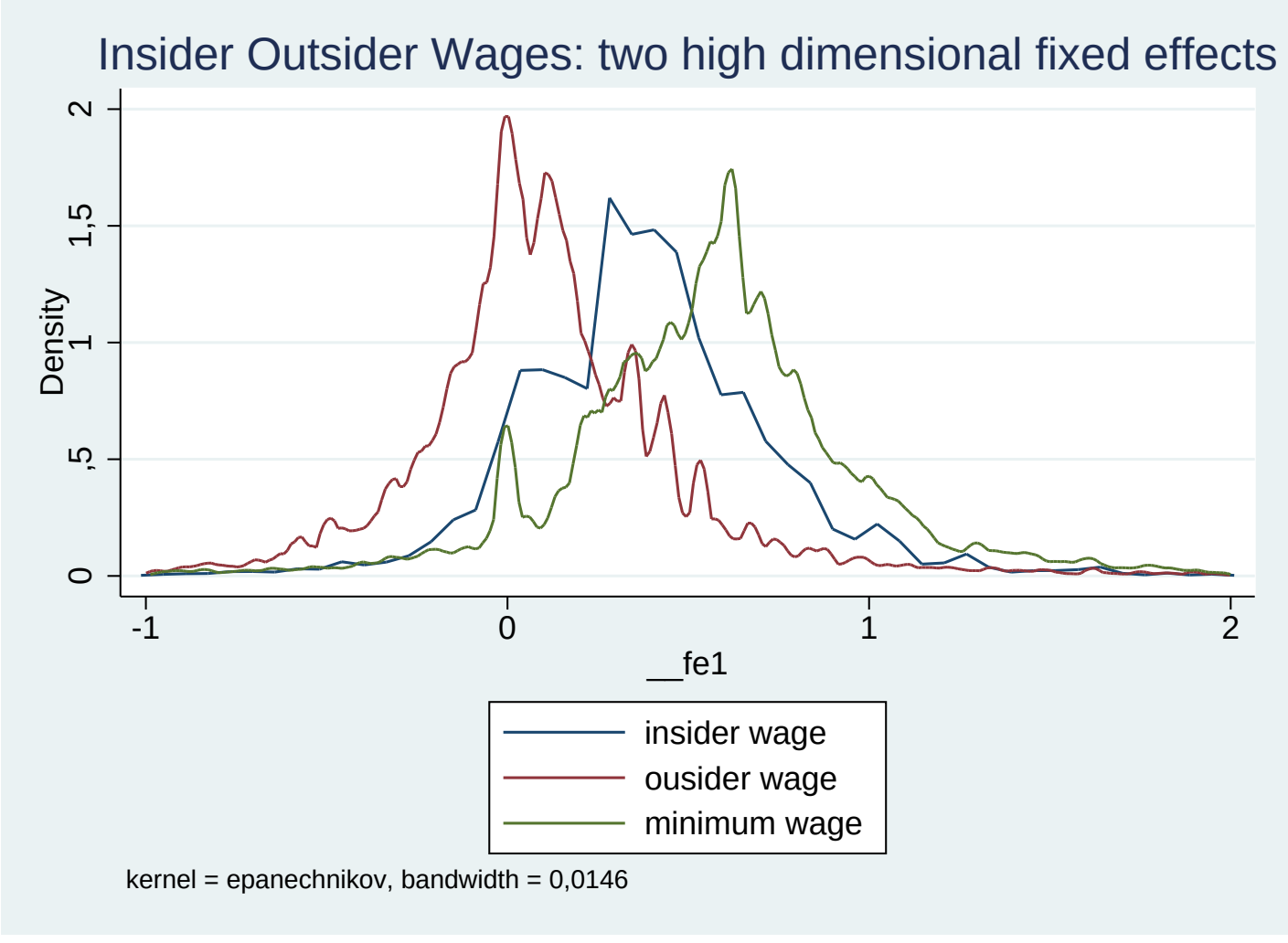
Insider-outsider wages

$$y_{iftj} = \phi_j \theta_f + \gamma_f \, wage_{insider} + \alpha_f \, wage_{outsider} + \phi_f \, wage_{minimum} + \epsilon_{iftj} \quad (8)$$

- y_{ift} represents the logarithm of the monthly wage
- γ_f is a firm specific coefficient on the firm ruling wage
- α_f is a firm specific coefficient on the collective bargained wage
- θ_f is a firm fixed effect
- ϕ_f is a firm fixed effect
- ϵ_{ift} is assumed to follow the conventional
- ϕ_f is a firm specific coefficient on the minimum wage

Insider-outsider wages

Figure: **The determinants of entry wages (two-fixed effects)**



A wage regression equation accounting for peer effects

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \alpha_i + \eta\overline{\alpha}_{-it} + \theta_{\mathbf{P}(i,t)} + \varepsilon_{it} \quad , \tag{9}$$

where

- $\overline{\alpha}_{-it}$ is the average of worker fixed effects over of worker i at time t
- δ is the associated coefficient.
- $\theta_{\mathbf{P}(i,t)}$ is the Establishment/Job/Year effect

- Estimation: Arcidiacono et al. (2012) empirical procedure to account for peers' unobservable attributes

The notion of peer

- Work in the same establishment
- Covered by the same collective agreement
- Work in the same job-title (categoria profossional)
- In the same year (in October)
-
- only Hugo is my peer
-
- 3.9 million peer groups
- 4.9 workers per peer group
- 14 peer groups by firm
- 47.8 peer groups by job title

A wage regression equation accounting for peer effects - Identification

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \alpha_i + \eta\overline{\alpha}_{-it} + \theta_{\mathbf{P}(i,t)} + \varepsilon_{it} \quad , \tag{10}$$

- identification of η would come strictly from changes on the size of the peer groups (N_p), eliminating any endogenous contamination from sorting into establishment and job-title

Table: Identification - Illustrative example

	Worker	α_i					N_p	$\overline{\alpha}_{it}$	$\overline{\alpha}_{-it}$	change in peer quality
		Ana	*Rute*	*Pedro*	*Paulo*	*Hugo*				
2010	*Ana*	2	0				1	1	0	
2011	*Ana*	2	0	1			2	1	1/2	0,5
2012	*Ana*	2	0	1	1		3	1	2/3	0,17
2013	*Ana*	2	0	1	1	1	4	1	3/4	0,08

Estimating a wage regression equation accounting for peer effects

The estimating equation can be written:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{D}\boldsymbol{\alpha} + \eta_0 \mathbf{W}\mathbf{D}\boldsymbol{\alpha} + \epsilon = \mathbf{X}\boldsymbol{\beta} + [\mathbf{I} + \eta_0 \mathbf{W}]\mathbf{D}\boldsymbol{\alpha} + \epsilon \qquad (11)$$

and from the first order conditions for minimization of *SSR* we get:

$$\begin{aligned}
\frac{\partial S(.)}{\partial \widehat{\boldsymbol{\beta}}} &= \mathbf{X}'\mathbf{e} = \mathbf{0} \\
\frac{\partial S(.)}{\partial \widehat{\eta}_0} &= \widehat{\boldsymbol{\alpha}}'\mathbf{D}'\mathbf{W}\mathbf{e} = \mathbf{0} \\
\frac{\partial S(.)}{\partial \widehat{\boldsymbol{\alpha}}} &= \left[\mathbf{D}' + \widehat{\eta}_0 \mathbf{D}'\mathbf{W}\right]\mathbf{e} = \mathbf{0}
\end{aligned}$$

Estimating a wage regression equation accounting for peer effects

- *Step 1* Given $\widehat{\boldsymbol{\alpha}}$ run an OLS regression on $\mathbf{X}$, $\mathbf{D}\widehat{\boldsymbol{\alpha}}$, and $\mathbf{WD}\widehat{\boldsymbol{\alpha}}$. The coefficients on $\mathbf{X}$ will provide an estimate of $\boldsymbol{\beta}$, while the coefficient on $\mathbf{WD}\widehat{\boldsymbol{\alpha}}$ is an estimate of η_0. $\mathbf{D}\widehat{\boldsymbol{\alpha}}$ should have a coefficient of 1.
- *Step 2* Given $\widehat{\boldsymbol{\beta}}$ and $\widehat{\eta}_0$ estimate $\boldsymbol{\alpha}$ using the updating equation

$$\widehat{\boldsymbol{\alpha}}_{[h]} = \mathbf{M_D}\left[\mathbf{I} + \widehat{\eta}_0\mathbf{W}\right]\left[\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right] - \widehat{\eta}_0\mathbf{M_D}\left[2\mathbf{I} + \widehat{\eta}_0\mathbf{W}\right]\mathbf{WD}\widehat{\boldsymbol{\alpha}}_{[h-1]} \quad (12)$$

There is, however, a faster approach to solve the f.o.c. $\left[\mathbf{D}' + \widehat{\eta}_0\mathbf{D}'\mathbf{W}\right]\mathbf{e} = \mathbf{0}$. We can then rewrite the equation as

$$\mathbf{D}'\widetilde{\mathbf{W}}\widetilde{\mathbf{W}}\mathbf{D}\widehat{\boldsymbol{\alpha}} = \mathbf{D}'\widetilde{\mathbf{W}}\left[\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right]$$

and since this is now written as a system of linear equations we apply the conjugate gradient method to obtain a solution for $\widehat{\boldsymbol{\alpha}}$.

The role of peers

Table: Wage Equation Accounting for Human Capital Spillovers

HC spillovers ($\overline{\alpha}_{-it}$)	0.2050
	(0.0006)
Worker effects (α_i)	✓
Establishment/Job-title/Year effects	✓

Note: Regression includes a quadratic term on age and tenure, gender, and firm size as covariates.

Decomposing the returns to education, a second pass

Table: Conditional Decomposition of the Return to Education - OLS

Base	Individual	Est./Job/Year	HC Spillovers
0.0791	0.0313	0.0422	0.0056

Note: Decompositions based on Gelbach (2016).

Table: Conditional Decomposition of the Return to Education - IV

Base	Individual	Est./Job/Year	HC Spillovers
0.0774	0.0236	0.0481	0.0057

Note: Decompositions based on Gelbach (2016).

Extensions

- Other dimensions of macronumerosity
  - Regions (countries, counties, etc)
  - Products
  - Interactions (e.g., firm and gender)
  - Heterogeneous treatment effects
- Estimation issues
  - incidental parameter problem
  - Identification
  - Statistical inference
  - Non-linear models
  - Four or higher dimensions
  - Speed of estimation