# ONLINE APPENDIX TO: WAGE EQUALIZATION AND REGIONAL MISALLOCATION: EVIDENCE FROM ITALIAN AND GERMAN PROVINCES

**Tito Boeri** Bocconi University

**Enrico Moretti** University of Berkeley Andrea Ichino European University Institute

Johanna Posch Analysis Group

#### Appendix A: Description of the data

Employment rates are obtained from the national statistical offices and are for individuals aged 15-64. The wage data for Italy are from the National Italian Statistical Office (ISTAT) quarterly labor force statistics<sup>1</sup> and, for Germany, from the Institute for Employment Research (IAB)<sup>2</sup>. They include all private and public employees for Italy (individual-level wage data for 2009-2013); and all private and public employees who are subject to social security contributions for Germany (individual-level wages for 1992-2014). In both countries, wages are defined excluding all non recurrent additional components of pay and we focus on full-time workers.

For Italy, housing cost information comes from the Osservatorio Mobiliare Italiano and contains transaction-level data on residential real estate sales in Italy between 2004 and 2011. The dataset also contains information about the characteristics of the individual unit and the municipalities (or urban area) in which it is located. German housing data are obtained from the Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR), which is the regional planning authority in Germany. They provide aggregate data on rental prices observed in Germany for the years 2004-2014 as posted online or on newspaper advertisements, controlling for key characteristics such as flat size and whether or not it is furnished.

E-mail: tito.boeri@unibocconi.it (Boeri); andrea.ichino@eui.eu (Ichino); moretti@berkeley.edu (Moretti); johanna.posch@analysisgroup.com (Posch)

<sup>1.</sup> The surveys cover about 600,000 individuals. Wage information is provided only since 2009 by the Italian Labor Force Survey. Respondents provide the net monthly wage (excluding extraordinary items such as unusual overtime premia) and the average hours worked in the previous four weeks. This information is used to obtain hourly net wages. In this study we focus on the period 2009-13.

<sup>2.</sup> This study uses the factually anonymous Sample of Integrated Labour Market Biographies (version 1975-2014). Data access was provided via a Scientific Use File supplied by the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB).

Data on gross value added for all industries are from OECD (2018).

# A.1. Geographical Unit of Analysis.

Choosing a geographic definition of local labor markets constitutes an important assumption in our study. Ideally, we would like to use a geographic unit akin to US Metropolitan Statistical Areas or Commuting Zones, which are small enough to encompass economically-meaningful units but large enough that most residents both live and work within a single region. Administrative boundaries of municipalities are likely to be too small; Italian and German workers easily commute across municipalities. For Italy, our definition of local labor markets is based on 103 provinces, with average working-age population of 495,104. The minimum and maximum working-age population are 76,884 to 3,418,941. For Germany, we base our definition on 96 "Spatial Planning Regions" (Raumordnungsregion) with an average working age population of 737,448 and a range of 187,990 to 3,030,240.

In Italy, we define North and South by including in the North the following regions: Emilia-Romagna, Friuli Venezia Giulia, Liguria, Lombardia, Marche, Piemonte, Toscana, Trentino-Alto Adige, Umbria, Valle d'Aosta, Veneto. The South is everything else. In Germany, we define West and East based on the historical Cold War division, with Berlin assigned to East.

## A.2. Wages, Employment and Informal Sector.

We seek to measure mean wages by province and year in Italy and Germany, controlling for differences in worker quality and industry mix across provinces. To estimate *conditional* average wages net of workers' characteristics and industry effects, we use worker-level data for Italy and Germany<sup>3</sup> to regress

$$w_i = \alpha + Z_i \beta + u_i$$

where  $w_i$  is the hourly (Italy) or daily (Germany) wage of worker *i* and  $Z_i$  is a vector that includes worker *i*'s gender, age, age squared, education, and industry. The regression is performed separately for Italy and Germany. We take the average residual  $\hat{u}_i$  for every province-year pair, which we interpret as the average wage in that

<sup>3.</sup> SIAB data in Germany are recorded in spells of employment providing the resulting daily wage of each spell. We construct yearly average daily wages for each person in the panel by calculating the average of the daily wages of each full-time spell in the particular year, weighted by its length (if it is only one spell, the wage is the daily wage of that spell). In the few cases where workers change their characteristics (education, industry) in the course of the year, we would use the state they were in for most of the time and the corresponding wage. The data are also top coded at the upper limit of social security contributions. We do not adjust the data any further as a result of this. We do note however, that any variability across local areas in Germany may in general be downward biased.

province and year holding constant worker observable characteristics and industry.<sup>4</sup> In the remainder of the paper, the term "wages" will refer to conditional mean wages.

In measuring employment in Italy, a potentially important issue is the existence of a large informal sector, which includes workers paid under the table to avoid taxes and Social Security contributions. Since the informal sector is widely understood to be larger in Southern provinces than in Northern provinces, this has the potential to lead us to underestimate employment rates in the South. Two points are worth noting. First, the employment rates that we use are computed using data that come from Istat Quarterly Labor Force Survey which is anonymous. In contrast to employment rates obtained from Social Security records, which are based on earnings reported by employers to the government and therefore miss the informal sector by construction, workers in the sample that we use to compute our employment rates have limited incentive to misreport their employment status. Indeed, differences between LFS and Census (or social security) data are generally used as measures of the informal sector. Second, we will present additional estimates that are based on *corrected* employment rates, obtained by adding to our baseline employment rates estimates of informal employment by province produced by the Italian National Statistical Institute (Istat, 2014).

## A.3. Housing Costs and Local Cost of Living Indexes.

We seek to measure mean cost of housing by province and year, controlling for differences in housing quality across provinces. For Italy, we follow the same approach that we use for wages. We regress

$$p_i = \alpha + X_i \beta + u_i$$

where  $p_i$  is the price per square meter of housing unit *i* and  $X_i$  is a vector that includes unit size, presence of a balcony, terrace or cellar, brightness, views, orientation, measures of quality of construction, distance to retail stores, distance to public transport, distance to public park or garden, parking, indicators for neighborhood type, and whether the municipality is located by the sea or in the mountains (these last two variables come from ISTAT). As with the wage residuals above, we then average the residuals from this regression by year and province.

The German housing data are rent prices collected by the regional planning authority BBSR from online or newspaper advertisements. To compute prices per square meter, the BBSR uses non-furnished flats of a size between 40 and 130 sqm in announcements listed for less than half a year. They filter out implausible prices and

<sup>4.</sup> To scale the average wage properly, we add to each province-year mean residual the mean national wage in 2010. This re-scaling simply means that our measures of conditional wages for Italy and Germany are scaled so that their averages equal the 2010 average wage in Italy and Germany, respectively. Due to data availability, for Italy, we use hourly wages net of taxes, while for Germany we use daily wages gross of taxes. We will present robustness checks to assess whether using wages net or gross of taxes in the two countries matters for our results.

luxury flats. We compute weighted averages by Raumordnungsregion, using weights that reflect the stock of housing in the area.

The main source of geographical differences in local cost of living is represented by differences in cost of housing. An additional source is represented by differences in the price of non-tradable goods and services. The price of non-tradables tends to vary regionally with the price of housing. For example, a sandwich or haircut in Milan tend to cost more than a sandwich or haircut in Palermo. We build a local Consumer Price Index (Local CPI) following the methodology proposed by Moretti (2013) and using data on regional measures of CPI from the Italian and German statistical offices.<sup>5</sup>

#### A.4. Productivity.

To our knowledge, the only measure of firm productivity available at a fine geographic level in Italy and Germany is gross value added. Gross value added is firm output valued at basic prices less intermediate inputs valued at purchaser prices. The basic price is the amount receivable by the producer from the purchaser for a unit of a product or service minus any tax on the product plus any subsidy on the product. Gross value added per worker in each province is obtained by dividing this measure by employment in that province. We obtain data on gross value added per worker for all industries at the local level from OECD (2018).

$$CPI_{pt} = \omega HP_{pt} + (1 - \omega) \left[ \pi HP_{pt} + (1 - \pi)NHP_t \right]$$

(see Moretti (2013) for details). The housing weight in consumption  $\omega$  for Germany is obtained from the German Federal Statistical Office (Destatis). For Italy we use consumption weights from households consumption surveys for the years 2005-2011. See also Jappelli and Pistaferri (2000).

<sup>5.</sup> Both countries' statistical offices provide regional measures of CPI that cannot be used for geographic comparisons as they are normalized to 1 in a given year. Nevertheless, they can be used in quantifying the relationship between prices of housing in an area and the price of other goods and services in that area. Specifically, we regress changes over time in CPI for a province on changes of its housing component. We run these regressions separately for Italy and Germany. Using these estimates, we then construct a Local CPI for Italy and one for Germany. The two Local CPI's measure differences in cost of living across provinces within each of the two countries. Local  $CPI_{pt}$  is defined as  $CPI_{pt} = \omega HP_{pt} + (1 - \omega)NHP_{pt}$  where  $HP_{pt}$  is housing price in province p and year t,  $NHP_{pt}$  is the price of non-housing or non-tradables, and ( $\omega$ ) is the housing weight. Some part of NHP varies with the housing price so that  $NHP = \pi HP + \nu$ . Therefore when we regress  $\Delta CPI_{pt}$  on  $\Delta HP_{pt}$ ,  $\beta = (\omega + (1 - \omega)\pi)$ . Then, we use  $\omega$  to compute:  $\pi = ((\beta - \omega_n)/(1 - \omega_n))$ . We then use the province specific housing prices obtained through our own calculations and construct the local CPI as:

### **Appendix B: Tables and Figures**

Italy	Mean	SD	Ν
Value added per worker	54837	6227	103
Local CPI	100.0	16.19	103
Local housing price	100.0	25.17	103
Nominal wage - hourly	8.5	0.44	103
Real wage - hourly	8.8	1.47	103
Non-empl rate	42.5	9.58	103
Non-empl rate corrected	34.0	7.55	103
Germany			
Value added per worker	52901	8134	96
Local CPI	100.0	14.18	96
Local housing price	100.0	18.53	96
Nominal wage - daily	91.6	11.28	96
Real wage - daily	92.3	10.77	96
Non-empl rate	27.9	3.59	96

TABLE B.1. Summary statistics - 2010

Note: Value-added is computed across all industries in each geographic area, as calculated by the OECD (Germany) and ISTAT (Italy), and it is divided by employment in the corresponding area. Housing prices are average prices for a square meter with similar characteristics in each area. The Local CPI is constructed using those housing prices according to the method describes in Section 4. Nominal wages are obtained after controlling for individual characteristics such as age, education, gender, and industry, as explained in Section 4. Note that for Italy we have "hourly" wages net of taxes, while for Germany we have "daily" wages gross of taxes. Real wages are deflated using Local CPI. Non-employment refers to the number of people age 15-64 out of employment over the total population of that age group. For Italy, we also report non-employment corrected for the presence of informal work.

TABLE B.2.  $R^2$  from a regression of individual wages on worker characteristics, industry and year fixed effects

	(1)	(2)
	Italy	Germany
Without province FE	.352	.389
With province FE	.36	.463
Difference	.008	.074

Note: the first row of this Table reports the  $R^2$  of regressions of individual wages on worker characteristics (gender, age, age squared and education), industry and year fixed effects, for Italy and Germany respectively. The second row reports the  $R^2$  of the same regressions when province fixed effects are added to the specification. The third row reports the change in the  $R^2$  deriving from the inclusion of province fixed effects. Data refer to 2009-2013 for Italy and to 1992-2014 for Germany.

	Ita	Italy		Germany		
	(1)	(2)	(3)	(4)		
Log value added	0.365	0.263	1.015	0.542		
	(0.020)	(0.029)	(0.018)	(0.013)		
Region FE:	No	Yes	No	Yes		
Provinces:	103	103	96	96		

TABLE B.3. Regression of mean nominal wages on mean value added - manufacturing only

Note: Entries are the coefficients of log mean value added in a regression of log mean nominal wage for workers in manufacturing only on log value added of each province pooling years 2000-2014 for Germany and 2009-2013 for Italy. All regressions include year fixed effects. Regressions in columns 2 and 4 include fixed effects for the North in Italy and the West in Germany. Standard errors in parentheses.

TABLE B.4. Average wage differences between macro regions - manufacturing only

	North -	- South	West - East		
	in It	taly	in Germany		
	(1)	(2)	(3)	(4)	
	Nominal	Real	Nominal	Real	
% Difference	0.0789	-0.0546	0.414	0.275	
	(0.005)	(0.018)	(0.004)	(0.007)	
Provinces:	103	103	96	96	

Note: This Table reports the coefficients of regressions of log mean conditional wages of manufacturing workers in all local areas on an indicator of North/West. We pool all available years. For nominal wages: 2000-2014 (Germany) and 2009-2013 (Italy). For real wages: 2004-2014 (Germany) and 2009-2011 (Italy). All specifications control for year fixed effects. Standard errors in parentheses.

		North - South			
TT	. 1		0	. 1	

TABLE B.5. Nominal and real wages corrected for taxes

		North - South				
	Uncor	rected	Corrected	Corrected		
	(1)	(2)	(3)	(4)		
	nominal	real	nominal - corr	real - corr		
% Difference	0.0425	-0.0921	0.0717	-0.0629		
	(0.003)	(0.017)	(0.004)	(0.017)		
Year FE:	Yes	Yes	Yes	Yes		
Provinces:	103	103	103	103		

Note: This Table reports North-South differences in mean conditional wages obtained from regressions of log mean conditional wages on a dummy for the North. Columns 1 and 2 report, for convenience, the same estimates of Table 7 for Italy, based on wages net of taxes. Columns 3 and 4 are based instead on estimated wages gross of taxes. Wages used for Italy in the first two columns are from Istat and are net of taxes. To generate the corrected wages used in the remaining columns, we take the mean gross and net wages of all full-time workers from the Italian social security agency (INPS) to generate a net/gross ratio for every province. We then correct the Istat wages dividing the net wage of each province by the corresponding net/gross ratio derived from INPS. For nominal wages the data are for years 2009-2013 while for real wages they are for years 2009-2011. All specifications control for year fixed effects. Standard errors are in parentheses

	So	outh	No	North		Italy	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Level	Change	Level	Change	Level	Change	Change %
Average hourly	wage: in E	Euros					
Status quo	8.36		8.68		8.54		
Counterfactual 1	7.39	-0.97	8.39	-0.29	7.94	-0.59	-6.88
Counterfactual 2	6.92	-1.45	8.23	-0.46	7.64	-0.90	-10.50
Employment ra	te: in % co	orrected for	informal w	ork			
Status quo	57.32		71.00		64.86		
Counterfactual 1	70.17	12.85	71.80	0.79	71.07	6.21	11.80
Counterfactual 2	71.95	14.63	72.41	1.41	72.20	7.34	13.74
Aggregate labor	r income p	er capita: i	n Euros per	r month			
Status quo	766.63		986.68		887.89		
Counterfactual 1	830.10	63.48	964.51	-22.17	904.17	16.28	3.48
Counterfactual 2	795.03	28.40	953.35	-33.33	882.28	-5.61	0.69

TABLE B.6. Counterfactual scenarios - Variant 2 - top 5 provinces

Note: This Table is like Table 9, but the reference group used to calculate  $A_n$ ,  $E_n$  and  $W_n$  is the median of the top five provinces in terms of value added, rather than the median of all Northern provinces.

	South		No	North		Italy	
	(1) Level	(2) Change	(3) Level	(4) Change	(5) Level	(6) Change	(7) Change %
Average hourly	wage: in E	Euros					
Status quo	8.36		8.68		8.54		
Counterfactual 1	7.38	-0.98	8.39	-0.30	7.94	-0.60	-7.00
Counterfactual 2	6.93	-1.43	8.25	-0.43	7.66	-0.88	-10.28
Employment ra	te: in % co	orrected for	informal w	ork			
Status quo	57.32		71.00		64.86		
Counterfactual 1	69.66	12.35	71.32	0.32	70.58	5.72	11.01
Counterfactual 2	71.34	14.03	71.82	0.82	71.61	6.75	12.78
Aggregate labor	r income p	er capita: i	n Euros per	r month			
Status quo	766.63		986.68		887.89		
Counterfactual 1	822.80	56.17	957.39	-29.29	896.97	9.07	2.62
Counterfactual 2	790.23	23.60	948.41	-38.27	877.40	-10.50	0.11

TABLE B.7. Counterfactual scenarios - Variant 3: top 10 provinces

Note: This Table is like Table 9, but the reference group used to calculate  $A_n$ ,  $E_n$  and  $W_n$  is the median of the top ten provinces in terms of value added, rather than the median of all Northern provinces.

	South		No	North		Italy	
	(1) Level	(2) Change	(3) Level	(4) Change	(5) Level	(6) Change	(7) Change %
Average hourly	wage: in E	Euros					
Status quo	8.36		8.68		8.54		
Counterfactual 1	7.43	-0.93	8.43	-0.25	7.98	-0.56	-6.47
Counterfactual 2	7.02	-1.34	8.34	-0.35	7.75	-0.79	-9.27
Employment ra	t <b>e:</b> in % co	orrected for	informal w	ork			
Status quo	57.32		71.00		64.86		
Counterfactual 1	69.00	11.69	70.60	-0.40	69.89	5.03	9.89
Counterfactual 2	70.53	13.21	70.94	-0.06	70.76	5.90	11.41
Aggregate labor	r income p	er capita: i	n Euros per	r month			
Status quo	766.63		986.68		887.89		
Counterfactual 1	820.42	53.80	952.87	-33.80	893.41	5.52	2.20
Counterfactual 2	791.41	24.78	946.83	-39.84	877.06	-10.83	0.08

TABLE B.8. Counterfactual scenarios - Variant 4: top 20 provinces

Note: This Table is like Table 9, but the reference group used to calculate  $A_n$ ,  $E_n$  and  $W_n$  is the median of the top twenty provinces in terms of value added, rather than the median of all Northern provinces.



FIGURE B.1. Share of informal employment. This figure plots deviations from the country mean of the share of irregular employment out of total employment of each province in 2010. Means are not weighted by population. The share of irregular employment is provided by Istat, 2014



FIGURE B.2. Non-employment corrected for informal employment and value added. This figure shows the relationship between the log non-employment rate among 15-64-year-olds taking into account informal work, and log value added across provinces in 2010. Each province is represented by a dot.



FIGURE B.3. Non-employment rate corrected for informal employment. To account for employment in the informal sector in Italy we compute an "informal labor market - corrected" employment rate by adjusting our employment rate to the fact that we only observe a proportion  $1 - e_{inf}$  (1-rate of informal employment) of actual employment. We thus inflate the official employment rate by a factor  $((1)/(1 - e_{inf}))$ . This figure plots deviations from the country mean of the non-employment rate of each province in 2010, comparing the corrected and uncorrected figures. Means are not weighted by population. The share of irregular employment is provided by Istat, 2014.



FIGURE B.4. Housing costs. This figure plots deviations from the country mean of the housing price index of each province in 2010. Means are not weighted by population.



FIGURE B.5. Local cost of living index. This figure plots deviations from the country mean of the consumer price index of each province in 2010. Means are not weighted by population.



FIGURE B.6. Change in wages and employment by province in counterfactual 1. The maps show the percent change in mean wage and employment rate in each province under counterfactual 1. For nominal wages, a lighter color indicates a larger in size negative adjustment. For employment a darker color indicates a larger in size positive adjustment. There is no change in the North by assumption.



FIGURE B.7. Changes in wages and employment by province in counterfactual 2. The maps show the percent change in mean wage and employment rate in each province under counterfactual 2. For nominal wages, a lighter color indicates a larger in size negative adjustment. For employment a darker color indicates a larger in size positive adjustment. There is no change in the North by assumption.

Boeri, Ichino, Moretti & Posch Online Appendix: Wage Equalization and Regional Misallocation 15

### References

Istat (2014). "Noi Italia. 100 statistiche per capire il Paese in cui viviamo."

- Jappelli, Tullio and Luigi Pistaferri (2000). "The dynamics of household wealth accumulation in Italy." *Fiscal Studies*, 21(2), 269–295.
- Moretti, Enrico (2013). "Real wage inequality." *American Economic Journal: Applied Economics*, 5(1), 65–103.
- OECD (2018). "Regional Statistics." Date retrieved from https://stats.oecd.org on 17/10/2018.