Poster Session #2
Time: Thursday, August 9, 2012 PM

Paper Prepared for the 32nd General Conference of
The International Association for Research in Income and Wealth

Boston, USA, August 5-11, 2012

The Inequality of Real Wages in Germany

Christoph Lakner

For additional information please contact:

Name: Christoph Lakner
Affiliation: University of Oxford

Email Address: christoph.lakner@economics.ox.ac.uk

The inequality of real wages in Germany

Christoph Lakner*
University of Oxford

July 2012
Preliminary, do not cite

Abstract

This paper investigates the implications of regional price differences for earnings differentials and inequality in Germany. I combine a district-level price index with administrative earnings data from social security records. First, I investigate how earnings differentials across districts are affected by accounting for regional prices. Mean earnings differ substantially between East and West Germany and there is almost no overlap in the two regional distributions. I find support for an equalising effect of prices on district wage differentials in West Germany. This effect is weaker in East Germany and in the combined sample. I argue that these differences between East and West Germany can be explained by a more concentrated price distribution in the East. In the second part of the analysis, I investigate how inequality measures change after accounting for the price level of the place of residence. Accounting for regional price differences does not change aggregate inequality measures, which is robust to separate analyses for the East and West. However, I find a much stronger effect of the price adjustment on between-district inequality in West than in East Germany, which is consistent with the results for the district averages. Overall inequality measures are unaffected by accounting for prices because the overwhelming majority of inequality is within rather than between districts.

*Department of Economics, University of Oxford, christoph.lakner@economics.ox.ac.uk.

1 Introduction

This paper investigates how taking account of regional price variation alters regional earnings disparities and aggregate inequality in Germany. There is substantial inter-regional variation in both nominal earnings and prices, and I am interested in how these two variables interact to produce differences in real earnings. This analysis will also offer insights to the wage determination process, such as whether local wages compensate for higher living costs.

Germany is an interesting country to study regional earnings differentials. Germany is a large country and a relatively young nation state. It is less centralised than most other European countries, which would suggest that economic activity is more dispersed and thus regional differentials relatively small. Furthermore, twenty years after the fall of the Berlin Wall, it remains to be seen how real earnings compare between and within East and West Germany. Taking account of regional differences in the cost of living also has important policy implications in areas such as social security payouts and national wage bargaining. To illustrate the regional patterns of earnings, Figure 1 shows the regional distribution of labour market earnings per employee in 2007 using national accounts data. East Germany appears substantially poorer than West Germany. The urbanised districts in the West and in particular the South-West lie in the top earnings categories. In a later part of the paper, I also analyse the distribution of district-level averages obtained from the micro data used in this paper and the comparison with the macroeconomic data in Figure 1 serves as a robustness check.

I use a new district-level price index for 2007 published by the German Federal Ministry of Transport, Building and Urban Development (BBSR, 2009), which allows regional analyses at an unprecedented disaggregated level. I am not aware of any other research which uses the full regional information contained in this index to study individual earnings inequality in Germany. Most of the literature has focused only on the differential between East and West Germany, which ignores the substantial variation within these regions. In order to study real earnings inequality I combine the price data with earnings information from the Sample of Integrated Labour Market Biographies (SIAB) data published by the Institute for Employment Research (IAB) (SIAB, 2010). This data set has the necessary regional representativeness and sample size, as well as reasonable accuracy due to the administrative nature of the data.

The motivation for studying real wages is obviously that they reflect the material standard of living, which a given amount of money could buy in different locations. There are a couple of conceptual issues which I want to highlight. First, the price index holds weights constant across localities. Therefore, I am assuming that the basket of goods of the representative consumer is constant over space. There are clearly differences in tastes across Germany, but because food only constitutes such a small proportion of spending, I do not expect this to be a major issue. I am more concerned whether climatic conditions

change consumption baskets for example through housing costs and fuel use. For example, in the US, we would expect that fuel for heating makes up a larger proportion of spending in Northern states than in the South.[1] There is surely a significant amount of weather variation within Germany; the South and East generally having colder winters than the North. However, these differences are relatively small. Second, I need to argue that prices do not compensate for local amenities, such as areas of natural beauty, cultural institutions and access to public services. For example, some of the districts in the Bavarian Alps, a very attractice area, have a relatively high price level. It could also be argued that higher prices in cities compensate for better access to cultural institutions and other public services, such as education or public transport. As a result, the price level in the urban areas would be overstated relative to the rural areas, since city-dwellers receive additional benefits.[2] This is a difficult conceptual issue and there is generally no way around it. Third, there is a potential confounding issue pertaining to regional variation in tax rates. In contrast to the US, sales tax is set at a federal level in Germany. Municipalities, however, have discretion over some property taxes and trade taxes.[3] These constituted respectively 2.1% and 6.7% of tax revenues in 2010 (BMF, 2011b), i.e. a significant proportion of revenues. Their variation across municipalities, however, is limited and they are mostly levied on businesses and not consumers.

This paper is structured in six main parts. First, I give a brief overview of the literature which motivated my analysis. This is not yet a full literature review. Second, I describe the two data sets I use and their potential alternatives. The distribution of prices is strongly right-skewed and there are strong differences between East and West Germany. Controlling for states and the degree of urbanisation explains approximately 60% of the price variation across districts. In the third part, I describe how I constructed my sample and how I dealt with the right-censoring of the earnings variable. The fourth part analyses the distribution of mean wages across districts. I find that controlling for prices has strong effects in West but not East Germany. I argue that this might be explained by the fact that the price distribution in the East is more concentrated than in the West. In section five I present the results of the inequality analysis. I investigate how inequality measures change after accounting for the price level of the place of residence. I repeat the analysis for a number of earnings variables, which differ in how they account for the censoring issue. I consider the effects separately by East and West Germany. Furthermore, I decompose inequality in its between and within district components.

[1] Here I am abstracting from expenses for air conditioning.

[2] Commuting costs would also result in the urban price level to be overstated. However, the regional unit under consideration here is probably too large for considering commuting costs, for example between urban and suburban areas.

[3] Municipalities levy a number of other taxes, such as dog licenses. However, their revenues are relatively insignificant. In addition to the directly collected taxes, municipalities also receive a share of income tax, corporation tax and interest income tax (BMF, 2011a).

2 Literature review

The paper most closely related to my analysis is Goebel et al. (2009) who use the same regional price level data. Because their analysis is based on the German Socio-Economic Panel (SOEP), a nation-wide household survey which in the original sample contains only 6000 households, they cannot exploit the full information contained in the BBSR index. Their analysis is thus restricted to comparing East and West Germany. They find that significant differences in real incomes remain between the two regions. Inequality, as measured by the Gini, within the East and West and Germany as a whole, changes very little (less than 0.5 percentage points) after accounting for regional differences in the cost of living. However, they do find significant changes in the regional distribution of poverty; it declines in the East and increases in the West. Using more aggregated price data, Hillringhaus and Peichl (2010) find that accounting for prices increases poverty in Southern Germany.

With the SIAB (2010) data I cannot analyse poverty because these data do not contain information about non-labour incomes and household composition. However, I can extend their analysis in two major ways. First, by using a much larger data set, I will be able to exploit the full price information for the first time and examine earnings differentials at the district level. A comparison between East and West Germany hides a significant amount of variation which exists within these regions. In fact, while the average price level tends to be substantially lower in the East than in the West, the least expensive district according to the BBSR index is not in Eastern Germany but in Northeast Bavaria. A district-level analysis also allows me to analyse rural-urban earnings differentials by comparing urbanised and rural districts. Second, the Gini index is only one measure of inequality and considering other inequality measures, such as percentile ratios will offer a more complete picture of real wage inequality in Germany.

This research is also closely related to Moretti (2009), who finds that accounting for regional price differences decreases the college wage premium in the US. A large body of literature has shown an increase in the college wage premium during the last three decades. Moretti (2009) shows that the prices faced by college graduates have increased relatively faster because their jobs are located in metropolitan areas with high housing costs. The college wage premium can be seen as a measure of inequality and thus his analysis is closely related to this paper.

Aaberge et al. (2008) study the effects of accounting for geographical differences in the cost of living on poverty rates in Norway. They conclude, in agreement with Goebel et al. (2009), that aggregate poverty changes very little while the regional distribution of poverty does change significantly after accounting for geographic price differentials. Their study is interesting because they have income data at the municipal level, so they can draw very detailed conclusions about rural-urban differentials. An obvious limitation of their analysis is that they only account for regional variation in house prices and exclude

other goods.

The availability of new, reliable price data has replaced older literature which aimed to estimate regional price indices using econometric methods and a range of district-level characteristics which might explain price variation. For example, Roos (2006) has conducted an earlier study of regional income differentials in Germany at the state level. His analysis is based on price-level data for 50 cities (Ströhl, 1994) and econometric extrapolation to the state-level. Roos (2006) concludes that real disposable income has converged between East and West Germany by the mid-2000s.

3 Data

3.1 Price data

In 2009, the German Federal Ministry of Transport, Building and Urban Development published for the first time a comprehensive regional price index for the entire country (BBSR, 2009). The unit of observation is a district region ("Kreisregion" in German), which has on average approximately 200,000 residents. The price data uses 2006 district boundaries, before boundaries were changed in the Eastern states of Sachsen and Sachsen-Anhalt. According to the 2006 boundaries there were 439 administrative districts. The majority of them are rural districts ("Landkreis" in German); larger cities are districts on their own. The price index is constructed not for these administrative units, but instead small urban districts of less than 100,000 residents are included within the surrounding rural district, resulting in 393 district *regions*.

The location identifiers in the earnings data are different, because they refer to districts rather than district regions and because they use 2008 boundaries. My analysis is conducted at the district level, assigning the same price to districts in the same district region. Between 2006 and 2008, the number of districts in Sachsen and Sachsen-Anhalt was reduced. The analysis is conducted using the new territorial allocations and the prices collected for the old district regions are matched with the new district identifiers. I can match districts in Sachsen before and after the reform because old districts were simply combined into new units. The situation is more complicated in Sachsen-Anhalt, where some old districts were split up. As a result, I cannot match six new districts to prices in this state. In summary, my analysis is conducted at the level of the 413 administrative districts according to the 2008 boundaries. For these districts I have 407 non-missing price observations, stemming from the original price data for 393 district regions.

The index incorporates two thirds of the national consumer price index. The consumption weights are held constant across districts, so that any changes of this index over time reflect a pure price effect. I expect the regional variation in consumption baskets across Germany, for example due to different weather, to be relatively small. Consistent with evidence from other countries, most of the inter-regional price differentials are driven

by differences in housing costs. For ease of comparison with earlier efforts at creating a regional price index in Germany, Bonn is chosen as the base category.

Because the compilation of an accurate regional price index requires a large amount of pricing information, the price data were collected between 2004 and 2008, with most of the price information stemming from between 2006 and 2008. Prices were collected at different locations at the same time. For example, the price of fuel for all districts was collected in 2006 whereas the prices for personal medication is for 2008. This procedure assumes that prices changed uniformly across regions during this period. My initial analysis is for the year 2007, which is the average year, but I will repeat my analysis for all these years separately.[4]

Overall, there are relatively large differences in prices across Germany. The least expensive district, Tirschenreuth in Northeast Bavaria, is 16.6% cheaper than Bonn (the base category) and Munich is 14.4% more expensive, so there is a price dispersion of 31% overall. This is a remarkable amount of dispersion considering that these are prices for the same goods. For these price differentials to exist in equilibrium, they must be non-tradable goods, such as housing and non-tradable services (e.g. see Balassa-Samuelson hypothesis).

Figure 2 shows the regional distribution of prices for the entire country. Figure 3 shows the distribution of prices across districts. It is strongly asymmetrical and has a long right-hand tail. The coefficient of variation of the price index is 5.34%, the p90/p10 ratio is 1.13 and the p75/p25 ratio is 1.06. The distribution of prices appears approximately log normal but I need to test for this formally.

Figure 3, however, hides important differences between East and West Germany. As Figure 4 shows, prices are substantially more concentrated in the East than in the West. In fact, the coefficient of variation of the price distribution in the East is approximately half of the value in the West. Furthermore, the range of prices in the East lies within the prices in the West, with the three least expensive districts being in rural Bayern and Niedersachsen in the West. The ten most expensive districts are all in the West (and, in particularly, are all in the three Southern states Baden-Württemberg, Bavaria and Hesse). While the average for East Germany is below that of West Germany, the ranking of East and West Germany in terms of prices is not clear cut. As I will show in a later section, there is almost no overlap in terms of average district earnings between East and West Germany, which contrasts sharply with the price distributions. The level differences in prices between East and West can be partly explained by the fact that goods which are comparatively inexpensive in the East enter the consumer price index only with a low weight (BBSR, 2009). However, the arguably more interesting question concerns not the level but the dispersion of prices, which is substantially more concentrated in the East. There are no obvious differences in price setting mechanisms between the two parts of the

[4]Goebel et al. (2009) conduct their analysis for the years 2005 to 2008. I suspect they exclude 2004 because very few prices fall into this period.

country. The structure of the housing market deserves particular attention.

Most of the literature on regional real earnings disparities in Germany has focused on comparing the East and West of the country (see for example Goebel et al. (2009)). This is because authors are interested in whether the country has become more unified economically 20 years after the fall of the Berlin Wall. Furthermore, there is a strong prior that the geographical fault line in terms of regional differences in economic livelihoods runs between the East and West. The substantial overlap of the price distributions in Figure 4, question the validity of these claims and suggest that a more disaggregated analysis is necessary, at least in terms of prices.

3.1.1 Factors explaining price dispersion

In Table 1, I show a simple decomposition of the price variance into between and within state components. While there is a significant state component, most of the variation in prices is within states, which further supports my claim that an analysis at a relatively disaggregated level is necessary.

I conducted some initial investigations into factors which might explain the variation in prices across districts, shown in Table 2. A linear regression of prices on a rural/urban classification used by the BBSR and state-dummies explains 60% of the variation in prices, which is quite remarkable (column 5 of Table 2). Prices increase with urbanisation; core cities in agglomerations for example command a 10% price premium. After controlling for urbanisation, Southern states still command a price premium, whereas Eastern states are insignificant with the exception of Saxony which is significantly negative and Mecklenburg-Vorpommern which is significantly positive. A further interesting result is the interaction between state dummies and urbanisation. Controlling for urbanisation makes some state dummies significant and vice versa. For example, when only considering state dummies, there is no significant difference between Bavaria and North Rhine-Westphalia. On the other hand, once I control for urbanisation, districts located in Bavaria are approximately 5% more expensive ceteris paribus. Therefore, the price level not only differs with the level of urbanisation but it also matters in which state the district is located.

In column 1 of Table 2, I use a classification system which divides districts into four types. These categories are ordered with decreasing urbanisation and together they explain 22% of the variation in prices (using the adjusted R-squared measure). The signs of the coefficients are what we expect with cities being more expensive than rural areas. The base category is rural, so rural districts cost on average 88.55% of Bonn's price level. Core cities cost on average 94.79 of the price level of Bonn. High-density suburban areas are cheaper than core cities but there is still a 3 percentage point premium compared to rural districts. There is no significant difference in the price level between rural surroundings and rural districts. In column 2, I use a more detailed system, which divides districts into nine categories. These categories are not ordered in the same way

as the four category system in column 1. The results are consistent with what we find for the less detailed categories. Prices rise with increasing urbanisation. There is no significant difference between rural districts located in agglomerations, urban environments or rural places. Column 3, I uses only state dummy variables. I have used Nordrhein-Westfalen, the most populous state in Germany, as the base category. It is also the state where Bonn is located, the district base category of the index. Therefore, the unweighted average price level of Nordrhein-Westfalen is 92.12% of Bonn's price level. Niedersachsen and all Eastern states, except Berlin, are less expensive than Nordrhein-Westfalen. Baden-Württemberg, Hessen and Hamburg are more expensive. The Bavarian case is interesting. At a state-level average, there is no significant difference between Bavaria and Nordrhein-Westfalen. However, Munich is the most expensive district overall. This means that the price premium of Munich is washed out by some of the cheapest districts also being located in Bayern. Column 4 combines state dummies and the four-category system. More urbanised districts are still more expensive and the size of the premium is unchanged by controlling for state fixed-effects. But controlling for urbanisation has some interesting effects for the coefficients on the state dummies. Conditional on urbanisation, Schleswig-Holstein and Bayern now have a price premium, whereas they had insignificant price differentials before. The reverse happens for Niedersachsen and the Eastern states of Brandenburg and Mecklenburg-Vorpommern, which have no significant price effect after controlling for urbanisation. Column 5 combines the nine-category system with state dummies. There is not much action for the population density dummies, except for the category "Agglomerations: rural" which becomes highly significant after the inclusion of state dummies. Regarding the differentials across states we observe a similar pattern to column 4; Schleswig-Holstein, Bayern, Rheinland-Pfalz and Saarland become significant after controlling for population density. In contrast to the previous regression, Mecklenburg-Vorpommern remains significant after including population density controls whereas Sachsen-Anhalt and Thüringen become insignificant.

3.2 Earnings data

I use the SIAB (2010) data, which is based on a 2% sample of social security records. The data base from which the observations are drawn includes employees, benefit recipients and individuals looking for work. In this paper, I use only the data on employees subject to social security contributions. The advantages of the SIAB data are its large sample size and its long panel structure starting in 1975. Furthermore, because the data set is based on administrative records it is relatively reliable. However, it only includes individuals covered by the social insurance system, i.e. approximately 80% of the workforce, and excludes particularly civil servants, the self-employed and individuals with very low wages (Dustmann et al., 2009). In addition, it contains no detailed information on hours worked, but only whether an individual is working full-time or part-time. Furthermore, we have no

information on other sources of income, so the analysis is limited to earnings from work. The by far most important drawback of the SIAB data, however, is the right-censoring of earnings. Earnings are only recorded up to an upper earnings threshold in accordance with social insurance legislation. Details of how I have dealt with the right-censoring are included in the next chapter.

There are a number of alternative data sources I initially considered. The most preferable would have surely been the income tax data by the Federal Statistics Office, used for example by Bach et al. (2009). However, due to data security protection, the data do not contain a district identifier. The cross-sectional "Mikrozensus" survey contains only limited information on income and wealth. Surely the most widely used German data set to study questions of inequality is the SOEP data. Although it has many attractive features, in particular the rich income data and household composition information, it is unsuitable for my purpose because its sample size is too small. Furthermore, the observations are very unevenly distributed geographically and some districts are not represented at all. In fact, only seven district have more than 100 observations. Therefore, analyses with the SOEP, such as Goebel et al. (2009), can only be conducted at the level of East and West Germany but not at a more disaggregated regional level. In addition, the SOEP data have severe problems with attrition and non-response, especially in the most recent waves.

4 Description of the data

This paper investigates real earnings, which are defined as nominal earnings divided by the appropriate price index. I calculate real earnings of individual i as

$$Realearn_i = \frac{Nomearn_i}{cspricelevel_{D(i)}} \tag{1}$$

where $D(i)$ is the district person i lives in. The cross-section price index of district d relative to district r is computed as

$$cspricelevel_d = \frac{\sum_{k=1}^{n} p_{k,d} \cdot q_k}{\sum_{k=1}^{n} p_{k,r} \cdot q_k} \tag{2}$$

where $q_{k,d} = q_{k,r} = q_k$ since the CPI weights are the same across locations. The variables $p_{k,d}$ and $p_{k,r}$ are the prices of good k in districts d and r respectively. I have divided the price index by 100, so for individuals living in Bonn, the base category, real and nominal earnings are the same.

4.1 Sample construction

The population under investigation consists of all individuals in a job covered by social security legislation working full time, who are living in Germany and have a non-missing

place of residence.[5] The SIAB (2010) data set combines a number of different data sources, such as data concerning benefit recipients and job seekers. In the present analysis, I only use the observations from the so-called Employee History file[6], which consists of employees subject to social security or in marginal part-time employment. I further exclude individuals in part-time employment or for whom that identifier is recorded as missing. The data rely on notifications submitted by employers to the social security agency either at the end of the year for ongoing employment or at the termination of the employment relationship.

The unit of observation is the individual, as I do not have any information on households. I restrict my attention to full time workers because the data contain no information on hours worked. The data are a random sample from the population so I do not need to use sampling weights.

The earnings variable is the mean daily wage, weighted by the duration of a certain employment spell. The data record all employment spells for a given individual. Individuals might be working with one employer for the entire year, so only one spell is recorded, or they might switch employers within the year, resulting in numerous spells. I am interested in analysing the distribution of the daily wage for the year 2007. The literature has approached this issue in two ways. Blien et al. (2009) use the wage paid on 30th June of any given year as the daily wage.[7] Thus they do not consider any employees who do not have a valid employment spell at this point in time. Furthermore, they ignore for how long a given individual has been receiving this wage. I instead follow the approach used by Dustmann et al. (2009), which computes an average daily wage, which is weighted by the duration of employment. I simply compute total annual earnings as the sum of total earnings in every spell within a given year, where the latter is the daily wage of a spell multiplied by its duration. My wage measure is then defined as total annual earnings divided by the total duration of all spells within that year. Therefore, for every individual the daily wage recorded in my data set refers to the average wage they receive during the time they have spent working. Alternatively, I could use a daily wage measure which is averaged over the entire year. This measure, however, would understate the amount of economic resources individuals have at their disposal, since during periods of unemployment they would receive some transfer income.

In addition, I weight across individuals using their individual duration of employment. This means, that someone who has only worked for half a year receives half the weight of someone who worked the entire year.

When reducing the spell data to yearly observations per individual for variables other than the wage, I select the non-missing information recorded in the longest spell. This

[5] I drop all observations for which the district of residence is missing. I further exclude all individuals living abroad, as I have no price information for these individuals.

[6] Beschäftigten-Historik or BeH

[7] 30th June is the date for which the employer information is recorded, which can be merged with the employee data.

includes variables such as nationality, gender, year of birth and education, which should be time-invariant.[8] The district of residence and occupational characteristics are potentially time-varying. I argue that the information for the longest spell is both the most accurate and the most relevant.

There is a slight issue in matching the earnings data with the price information using the district identifiers, which I have already mentioned in the description of the price data. The price data use identifiers referring to the year 2006. In 2007 and 2008, the number of districts was reduced in Sachsen-Anhalt and Sachsen respectively. The earnings data have identifiers which are correct for the new territorial allocation in 2008. For Sachsen and a number of districts in Sachsen-Anhalt, entire pre-existing districts were merged together to form new districts. However, there are six districts in Sachsen-Anhalt, where pre-existing districts were split up and merged to form new districts. At this stage, I have thus not been able to match the price information for these districts.

In addition, the price data are recorded for a district region, whereas the earnings data rely on districts. District regions combine small urban districts which have less than 100,000 residents with the surrounding rural district. Districts which do not fall into this category keep their identifier. I use a matching between districts and district regions provided by the BBSR, and I need to make one manual replacement where the matching is not successful.[9]

4.2 Construction of the earnings variable

4.2.1 Censored earnings

Earnings are only recorded up to the upper earnings limit for the pension insurance. Below that limit, contributions to the pension insurance system are calculated as a percentage of gross earnings. Earnings above the limit are not subject to contributions, thus providing a cap on the amount paid to the pension insurance. The upper earnings thresholds differ by year and differ between East and West Germany, accounting for different wage levels between the two parts of the country. Because Berlin was divided between East and West Germany, the law applies the threshold for West Germany to the former western part of the city and vice versa. Unfortunately, there is no way to distinguish between the two parts of the city in the earnings data, since Berlin now forms only one district and state. I chose to apply the threshold for West Germany to Berlin, because Berlin is substantially richer than the rest of East Germany.

[8] Prior to this modification of the data, I use an algorithm provided by Bernd Fitzenberger which imputes missing time-invariant observations using information from previous and later spells.

[9] Specifically, this concerns Zwickau in Sachsen. The matching between the districts and the district regions lists two districts, an urban district ("Zwickau-Stadt") with ID 14167 and a rural district ("Zwickauer Land") with ID 14193. The price data, however, record only once district region called "Zwickauer Land/Zwickau" with ID 14193, i.e. combining the two districts into a district region. I kept the two districts as recorded in the classification of district regions, and thus also the earnings data, and replaced the price recorded for the combined district region in the price file.

A further complication arises from the fact that within the social security system, there are two different pension insurance schemes with different upper earnings thresholds. The limit of the insurance scheme for wage and salary earners is generally below that of the pension insurance for miners. In the present analysis, I have only applied the threshold for the wage and salary earners which is by far the bigger group in the sample. But I will need to investigate this issue in more detail, assigning individuals their correct censoring threshold.

The daily upper earnings limits in 2007 were EUR 149.59 for East and EUR 172.60 for West Germany. In Figure 5 I plot the distribution of nominal earnings weighted by the employment duration. There are clear spikes in the distribution at the various earnings limits. In the distribution for Berlin we see spikes at the East and West earnings limits, as would be expected given that both limits apply in different parts of the city. It is surprising, however, that there appears to be a mass point in the East German wage distribution at the Western earnings limit. This might be caused by commuting between the two parts of the country. My regional classification relies on the place of residence, which I perceive as the relevant category for prices. The social security legislation, however, depends on the place of work. It is conceivable that there is a significant number of people living in East Germany but commuting to work in West Germany, or in former West Berlin. Note that the kernel distribution also clearly displays some spikes at the lower end of the wage spectrum, which are caused by a lower earnings limit. I have not yet taken care of the censoring from below, but this clearly deserves further attention.

Figure 5 further shows that while there is a spike in the distribution at the earnings limits, these spikes are not discontinuous jumps, as we would expect from the legislation. An explanation for this phenomenon is that people are only employed at a censored wage for some part of the year. The averaged yearly wage is then less than the censoring limit. Obviously their earnings are still censored in the sense that I would observe a higher wage in the absence of the censoring limit. To further investigate this point, Table 3 shows the number and proportion of observations above different earnings limits. The bold numbers are the official upper earnings limits. While these figures clearly seem inappropriate from Figure 5, the lowest thresholds are probably too low. I have worked with different earnings thresholds, which I do not all report here. Overall the thresholds of EUR169 in the West and Berlin and EUR146 in the East seem to be the most appropriate.

There are a number of cases for which reported earnings exceed the upper thresholds, as shown by the long right-hand tail in Figure 5, where I have already recoded earnings exceeding EUR 300.00 per day to missing. This might be due to legitimate reasons, in particular the treatment of bonus payments. However, it might also be due to measurement error either in the duration of the employment spell or the wage paid. However, given that the earnings information directly influences social security payments, I would expect the information to be quite reliable.

4.2.2 Different approaches to the censoring problem

I construct a range of different earnings variables. In this analysis I report the results for three earnings measures. The first is the raw wage with no adjustments for censoring. The second earnings variable simply drops all the censored wages from the analysis. The last measure uses a censored normal regression model to impute the censored wages, following the approach by Dustmann et al. (2009). I account for heteroscedasticity between different age/education groups. An alternative might be to allow for heteroscedasticity across districts and genders, which might be more appropriate given that I am interested in regional wage inequality.

The appropriateness of the different earnings variables obviously depends on the question of interest. In the case when I drop the censored observations, the population of interest becomes employees covered by social security with earnings in the middle part of the distribution. My analysis would then investigate what happens to "middle" earnings inequality after accounting for price differences. This might be a less interesting question of interest, but this has to weighed up against the credibility of the imputation methods.

Table 4 shows summary statistics for the different wage variables, both nominal and real. The average wages I obtain from my data are rather close to results from the SOEP. For individuals working full-time, the mean gross earnings per month in 2007 was EUR2818 in West and EUR1974 in East Germany, as reported in Frick (2008). This translates into a daily wage of EUR92.58 in West and EUR64.85 in East Germany. Using the population shares for East and West Germany in 2007, this corresponds to an average of EUR87.96, which is relatively close to the mean wage I find in my data. Differences between results from the SOEP and the SIAB might arise from the fact that the SIAB only includes employment subject to social security legislation, the censoring of wages in the SIAB or from measurement error in the SOEP.

Earnings measure 1: Truncate at EUR169.00 and EUR146.00 I recode all earnings above EUR169.00 in West Germany and Berlin, and above EUR146.00 in East Germany to missing. Therefore, I exclude all censored observations from the analysis and concentrate on the middle part of the wage distribution. I choose a cut-off limit slightly below the legislative upper earnings limit because of the fact that earnings might only be censored for parts of the year, as discussed above.

Earnings measure 2: Imputed earnings, censoring at EUR169.00 and EUR146.00 I use a censored normal regression model to impute the censored earnings, similar to Dustmann et al. (2009). The dependent variables are three education categories, eight age categories and their interactions. Again, I use the duration of employment as a weight.

5 Results at the district level

In the first part of the analysis, I consider the effect of accounting for prices on district-level averages. A district-level price index leaves the position of an individual within its district unchanged. In other words, a person's position in the aggregate wage distribution can only change if accounting for prices shifts around districts. This means that a substantial part of the analysis should consider what happens to the position of districts when we account for prices. In most of the analysis I concentrate on the mean rather than the median. I have carried out some initial robustness checks with the median and the results are broadly similar. Furthermore, in the interest of space, I restrict myself to results using the wage measure which is imputed above a daily wage of EUR169.00 in West and a daily wage of EUR146.00 in East Germany.

The SIAB data have a very good regional representativeness compared with other data sets available. In total there are 440,399 observations on the truncated wage measure. Given that there are 413 districts according to the after-reform boundaries, this means on average there are about 1000 observations per district. There are 75 districts with less than 500 observations (58 in West and 17 in East Germany) and 21 with less than 300 observations (18 in West and 3 in East Germany).

Figure 7 shows the distribution of district averages by region. It is important to note that whenever I analyse regions separately and only consider East and West Germany, I effectively exclude Berlin, which is not part of either region. Because there is very little overlap between the distributions for East and West Germany, it is interesting to consider them separately. There is almost no overlap in the distributions of mean wages between the two regions. If anything, the real distributions appear more disjointed than the nominal distributions. I would not have expected such a stark contrast at the district average level, given that the individual wage distributions overlap quite substantially. Accounting for regional prices appears to reduce the dispersion of average earnings across districts, at least in the West (for more detail see Table 7 below). Table 5 shows the ten richest and poorest districts according to average nominal wages together with their price level and their rank in the price distribution. The stark contrast between the East and West is evident even from this very coarse illustration; the ten richest districts are all in the West, whereas the ten poorest districts are all in the East. I will consider the interaction between nominal wages and prices in more detail below, but the price ranks in the last column already suggest that the correlation between price ranks and nominal wage ranks is not perfect.[10]

[10]The third poorest district Rügen has a relatively high price level which might be explained by having a large tourism sector.

5.1 Relationship between mean wages and prices

Figures 8 and 9 analyse the association between earnings and prices in more detail by plotting average nominal (Figure 8) and real (Figure 9) wages against the price level. I have also included a 45° line and linear fits through the nominal and real series. In general, there is a strong positive relationship between average earnings and prices across districts. The fitted line for the real earnings measure is flatter than the fitted line for the nominal variable. This suggests that accounting for prices reduces the variation in mean earnings across districts. Table 6 summarises correlation measures between both nominal and real average earnings and the price. This clarifies the results from the graphs namely that the correlation between nominal wages and prices is much stronger than the association between real wages and prices.

I repeat the same analysis separately for East and West Germany. The differences across the two regions are quite striking. Figures 12 and 13 show the results for East Germany. The estimated slopes are not very different from the estimation on the pooled sample. Figures 10 and 11 repeat the results for West Germany. The fitted line for real earnings is essentially flat. It has a slope of 0.07 with a standard error of 0.054, giving it a p-value of 0.22. The slope of the line fitted through the mean nominal earnings is not significantly different from one. This suggests that in West Germany prices and mean nominal earnings have a strong positive relationship, which suggests that wages compensate for high price locations. As a result the differentials in real terms are much smaller than the nominal differences.

As a robustness check, I have also repeated the analysis weighting districts by their population. Population figures were obtained from the Federal Statistics Office. This makes the real line fitted on the pooled sample flatter, i.e. more similar to the result estimated on the West German sample only. This effect of including the weights can be explained by the fact that districts in the West are on average 25% larger than in the East. In addition, there is a small number of very large districts in the West, which might influence the fitted line very strongly. Given that the distribution of district sizes is relatively compressed in the East, weighting does not make much of a difference to the estimated relationships in this part of Germany. In West Germany, the fitted line for the real earnings has even a negative slope after weighting by district population sizes, although again it is not significantly different from zero at a 5% significance level.

5.1.1 Summary of the relationship between mean wages and prices

In summary, I find that accounting for prices has a much bigger effect in West than in East Germany. This is because the distribution of prices is much more compressed in the East than in the West compared with the distribution of nominal wages. Figures 14 and 15 compare the distribution of prices with the distribution of average nominal and real wages for West and East. Because prices and wages are measured on different scales,

I scaled the district means and the price level by the appropriate regional (East and West) means. Table 7 summarises the coefficient of variation as a measure of dispersion for the three distributions in Figures 14 and 15. The dispersion of prices in the West is twice the Eastern level. Within the West, the dispersion of prices is comparable to the dispersion in district real mean wages.

5.2 Association between nominal and real average earnings

I also consider the relationship between nominal and real mean earnings across districts. In Table 8, I analyse the association between mean real and nominal earnings at the district level in more detail. It is just a simple linear regression of real earnings on nominal earnings. The first three columns use the truncated wage measures for nominal and real earnings; the last three columns use the imputed wage measures. Approximately 80% of the variation of mean real earnings across districts is explained by nominal earnings. Including dummy variable for the state and for the type of district, I can explain close to 90% of the variation. Furthermore, the Pearson product-moment correlation coefficients are 0.8922 for the truncated earnings and 0.9124 for the imputed earnings.

Figure 16 shows the strong association between mean real and nominal wages as a scatter plot. Confirming the regression results in Table 8, it is remarkable how well the linear prediction fits the data. Districts above the line, are richer in terms of real earnings than predicted on average. I have labelled a number of points. Wolfsburg in Niedersachsen, a state in North-West Germany, does very well both in terms of real and nominal earnings.[11] Both Munich and the Main-Taunus-Kreis in the vicinity of Frankfurt/Main have higher nominal wages than Wolfsburg, but their higher price level depresses real wages. The opposite happens to Helmstedt, a rural district close to Wolfsburg and the former border between the Federal Republic and the Democratic Republic. Its nominal earnings are substantially lower than those of Wolfsburg, but there is little difference in terms of real earnings due to the lower price level in Helmstedt. Munich, which has the highest price level amongst all districts, falls far below the linear prediction, because its price level is so far above the rest of the distribution. I further labelled the district with the lowest prices, Tirschenreuth in North Eastern Bavaria. As a robustness check, I again repeat the analysis separately for the West and East of the country in Figures 17 and 18 and. There is almost a one-to-one correspondence between real and nominal earnings for East Germany. On the other hand, the fitted line is flatter for the West, suggesting that the prices serve to equalise average earnings across districts.

Figures 19 and 20 plot the regional distribution of the imputed wage measure for nominal and real earnings respectively. The categorisation used in the maps corresponds to the percentiles of the nominal and real mean wage distributions respectively. Note that in the real earnings map (Figure 20), I cannot match five districts in Sachsen-Anhalt, an

[11] Wolfsburg is home to the headquarters of Volkswagen.

Eastern state, because of a reform of the district boundaries, giving rise to the large white area towards the North East. The East-West divide is very strong both in terms of nominal and real earnings, which I have also illustrated above in Figure 7. Comparing nominal and real earnings for the truncated wage measure, it appears that the Eastern states cannot gain from having lower prices on average.[12] Even under real earnings they are still mostly in the bottom two categories. However, within West Germany moving from nominal to real earnings changes the ranking of districts quite substantially. In terms of nominal wages, the top regions appear strongly clustered around the major urban centres especially in Southern Germany. The distribution in terms of real earnings is much more dispersed and, as expected, peripheral and relatively cheap districts gain relative to urban and expensive districts.

6 Results at the individual level

6.1 The effect of prices on the inequality of wages

Table 9 shows the first set of results. I investigate the effect of accounting for regional price levels on a number of different inequality measures. I repeat the analysis for the raw wage, the truncated and the imputed wage variables. For every wage measure, the first column shows inequality under nominal earnings, the second column accounts for regional prices and the last column computes the percentage difference between the inequality of real and nominal wages.

In general, I find a very small effect on wage inequality of accounting for regional price differences. The Gini coefficient changes by approximately 1.3% or 0.4 percentage points which is of a similar order of magnitude for the different wage variables. The percentage changes in some of the Generalised Entropy measures and Atkinson indices are somewhat bigger. My results confirm the conclusions reached by Goebel et al. (2009) with the SOEP, that accounting for regional price differentials has very little effect on aggregate inequality measures. I have also computed bootstrapped standard errors. Because of the large sample size, I obtain very small standard errors and as a result the change of the Gini is statistically significant, although numerically it is very small.

The inequality measures I obtain are of the same order of magnitude as in the SOEP. Frick (2008) report a Gini coefficient of 0.284 for West and 0.241 for East Germany. These figures refer to monthly personal income, i.e. they include other sources of revenue in addition to labour income. Bach et al. (2009), however, find a substantially higher Gini coefficient of 0.6522 for gross market income in 2003, the most recent year in their data. Their data set combines the SOEP with income tax data. These discrepancies might be due to the under-representation of top incomes in both the SOEP and the SIAB. But the

[12] Some coastal Eastern districts even appear to lose out, because they are amongst the ten poorest districts in nominal terms, but face a relatively high price level.

size of the differences seems implausibly large.

In Figure 6 I plot real earnings by region. Essentially it is the same plot as Figure 5, except that I divide by the price of the district of residence. It is remarkable that the spike at the censoring limit is smoothed out by accounting for prices in this way, in particular for West Germany. However, given that the distributions in Figures 5 and 6 look quite different, at least for West Germany, I would expect some effect of accounting for prices on measures of dispersion. Maybe the reason for not finding any large changes in Table 9 is due to analysing all regions jointly and not differentiating by East and West Germany and Berlin. This claim is supported by the fact that the price distributions are very different between the East and West, both in terms of the level and the shape of the distribution. Furthermore, the analysis of district-level averages below produces very different results for East and West Germany.

Measures of correlation In Table 10 I report the correlation (at an individual level) between nominal earnings and the price level. The Pearson and the Spearman correlation measures are very similar, which seems to suggest that the relationship between nominal earnings and the price level is not too far from being linear. The rank correlation measures indicate how an individual's ranking under nominal earnings compares with his ranking according to the price level he faces. Overall the correlation between nominal earnings and prices is rather low. This seems to suggest that prices do not fully adjust to nominal earnings.

A closely related comparison is conducted in Table 11, where I compare the correlation between nominal and real incomes. Consistent with the results in Table 10, I find a very strong association between nominal and real incomes. Again, the results for the rank correlation measures indicate that accounting for prices does not switch individuals' ranks by very much.

6.2 Separate analyses for East and West Germany

The district-level analysis above suggests that there are strong differences between East and West Germany. In order to investigate this relationship in more detail, I repeat the above analysis separately for East and West Germany. Note that this analysis effectively excludes Berlin. At the moment, I only consider the wage measure which imputes wages above EUR146.00 in East and EUR169.00 in West Germany. Table 12 computes a number of inequality measures for nominal and real earnings and the percentage difference between them separately for East and West Germany. The first three columns replicate the results from Table 9 for the entire country. Accounting for prices has a weaker effect on inequality in the two regions compared with the aggregate sample. The effect is substantially weaker in the East than the West, which is probably due to the price distribution being more compressed in the East. Furthermore, inequality in the East is generally higher than in

the West.

Table 13 decomposes a number of inequality measures into the between and within components by district. I use the Generalised Entropy and Atkinson indices because these measures are, in contrast to the Gini index, additively decomposable. I repeat the analysis for East and West Germany separately and for the pooled sample. Given the district-level analysis above, I would expect that accounting for prices has a bigger effect in the West than in the East. Between-group inequality is the component of inequality which is affected by the price adjustment, since prices are constant within a district. Indeed, I find that the price adjustment has a bigger effect on between-group inequality in West than in East Germany. The results also show that most of the inequality in Germany is within rather than between districts. Because the price adjustment only affects between-district inequality, this also explains why I do not find a large effect of accounting for prices on aggregate inequality.

7 Conclusion

This paper has exploited a new price index which for the first time allows an accurate price-level comparison at the district level. I have combined the index with wage data from social-security records, which provide a sufficient number of observations to perform a very disaggregated analysis. I find that accounting for prices has a strong equalising effect on average district wages in West, but not in East Germany. This can be explained by the fact that in East Germany, the distribution of prices is more concentrated than the distribution of mean wages.

Accounting for prices reduces between-group inequality by about 40% across Germany and by 30% in East and 50% in West Germany. Aggregate inequality in real wages is lower than nominal wage inequality but the difference, although statistically significant, is very small. The Gini index decreases by approximately 0.4 percentage points. The small effect on overall inequality is explained by the fact that most of the inequality in wages is within, rather than between districts. Accounting for cross-district price differences only affects the between-group inequality.

The present analysis considers cross-sectional price dispersion. A natural extension to the analysis would involve a dynamic dimension, i.e. considering how real wage disparities have changed over time. The evolution of real wage differences between East and West Germany is of particular policy interest, given the large transfer payments between the different parts of the country. In order to replicate the analysis for other time periods I would have to obtain time-series information on prices which is not included in the price data used here. One option would be to collapse the analysis to the state-level where inflation figures are available from the Federal Statistics Office. For analysing real earnings convergence between East and West Germany, this would be sufficient. However, there is not much price variation left at that level of aggregation because most of the variation in prices is within rather than between states. Furthermore, the strength of the price data is clearly the more disaggregated analysis. The alternative approach involves estimating my own inflation figures using for example district-level rent indices. While there exists some survey data, the sample size is not sufficient for estimating district-level inflation figures and there are real problems related to data access.[13] In sum, while a dynamic analysis would clearly be very interesting, I am constrained by the availability of data.

I plan to extend this paper in a number of ways. First, I will test for the robustness of my results to different imputation methods. I will consider imputations using the Pareto distribution and multiple imputation methods. Second, I want to consider factors which might explain the observed spatial dispersion of wages and prices. I plan to estimate

[13] The Income and Expenditure survey (Einkommens- und Verbrauchsstichprobe, EVS) collects detailed information on housing costs for approximately 60,000 households since the 1960s. The "Mikrozensus" also includes detailed rent data but it is not representative at the district level. Neither of these surveys is available as a scientific use file for researchers outside of Germany. A further alternative data source might be private providers, such as associations of estate agents, but this data might be of questionable quality.

a real wage curve, which gives particular attention to the local unemployment rate for explaining wage variation. Third, I plan to investigate the effect of accounting for regional price differences on the college wage premium, a particular measure of wage inequality.

8 Figures

Figure 1: Regional distribution of earnings from National Accounts

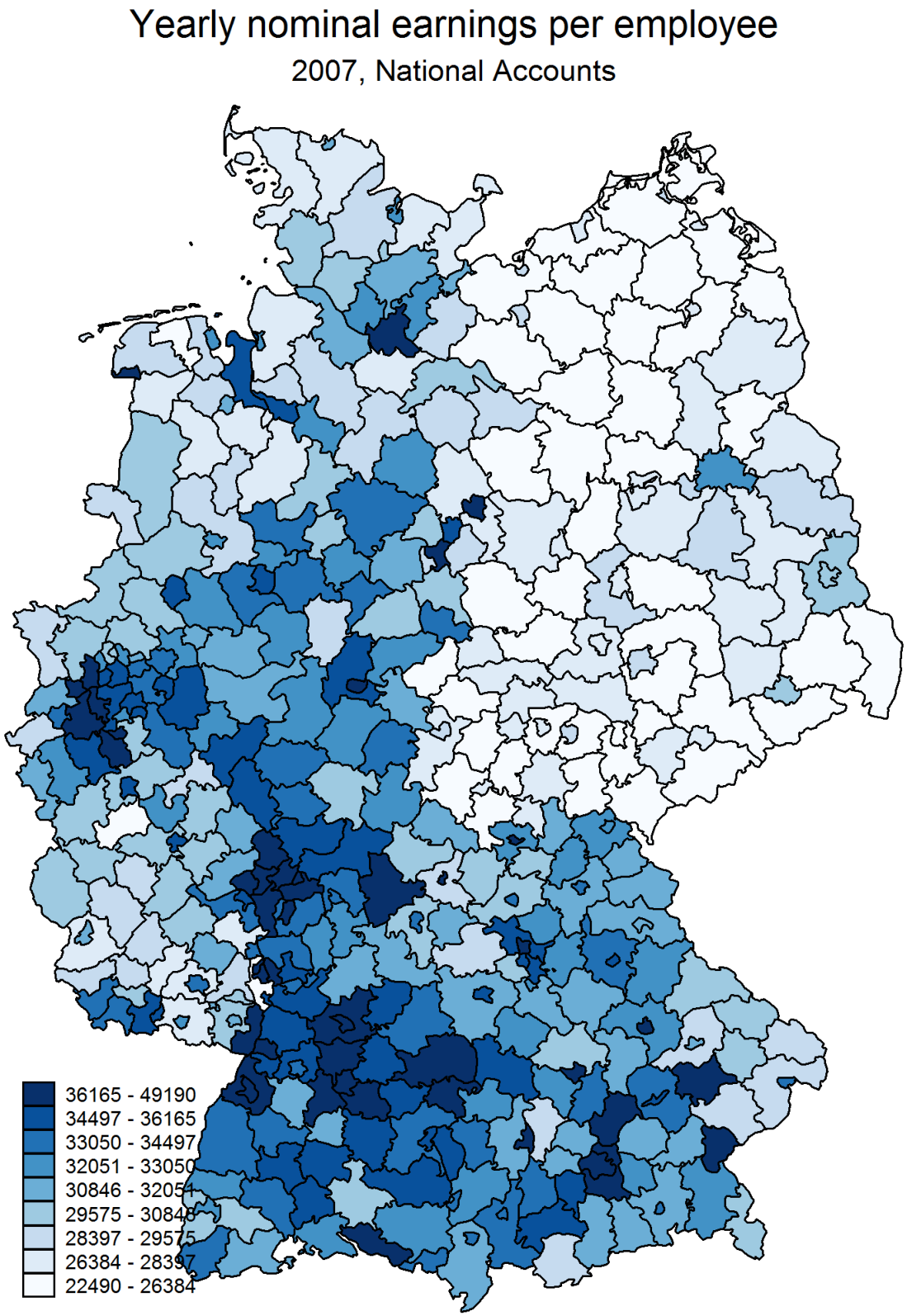


Figure 2: Regional price index

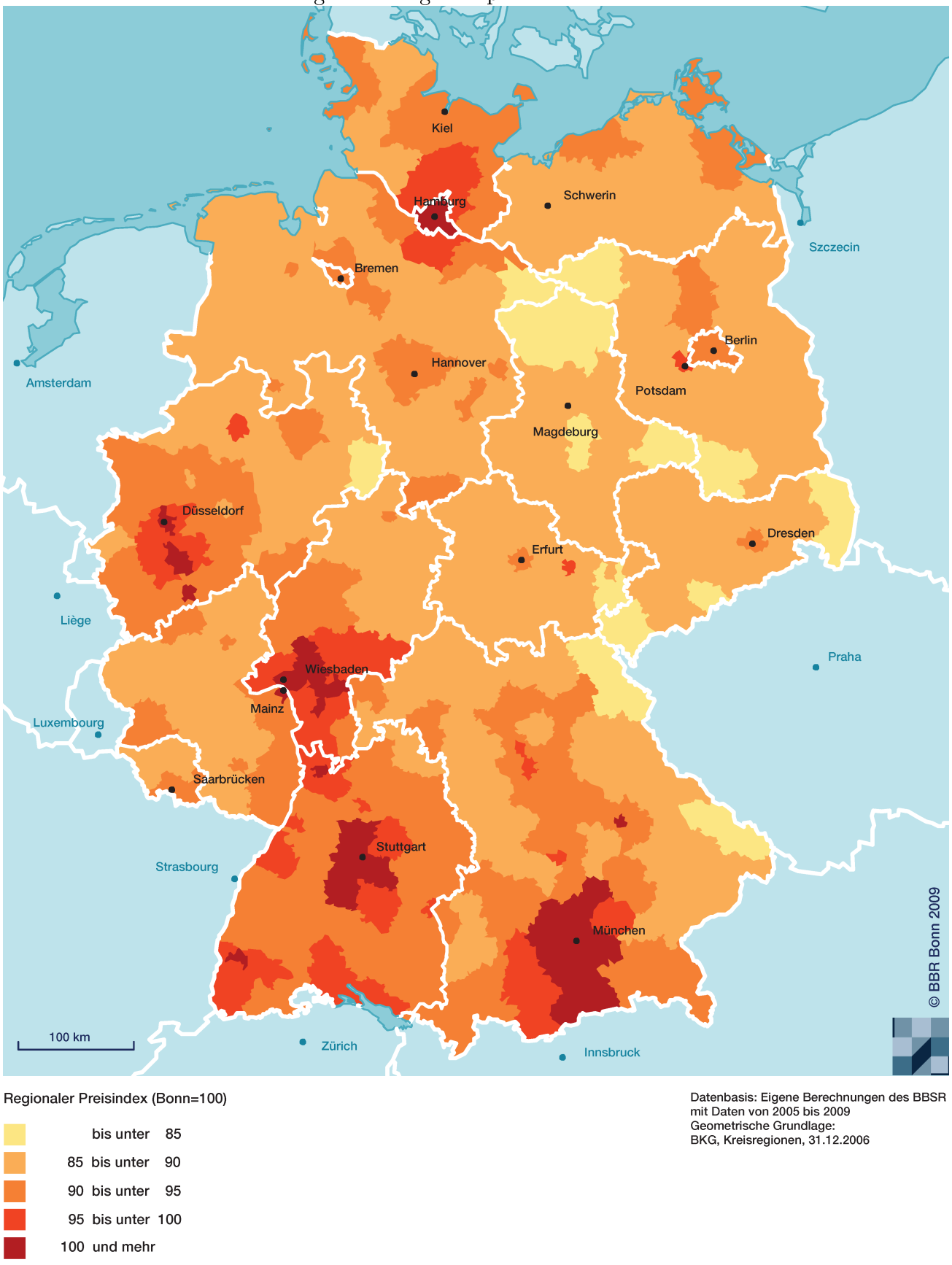


Figure 3:

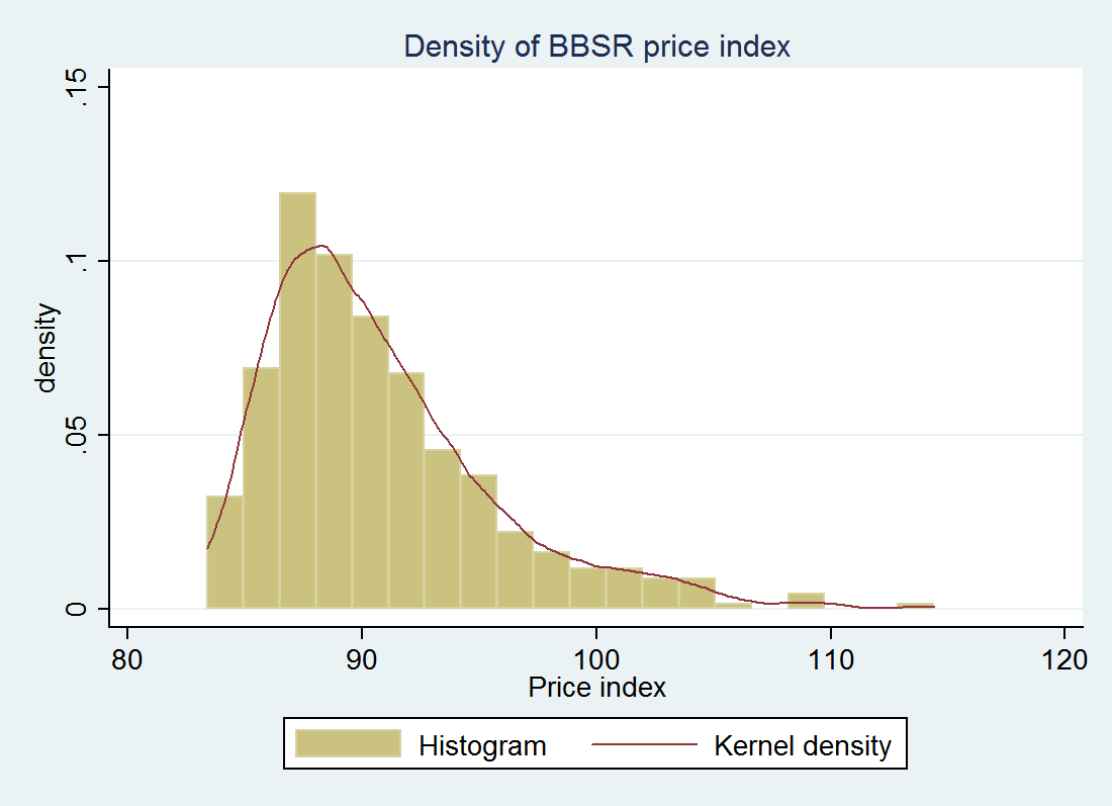


Figure 4:

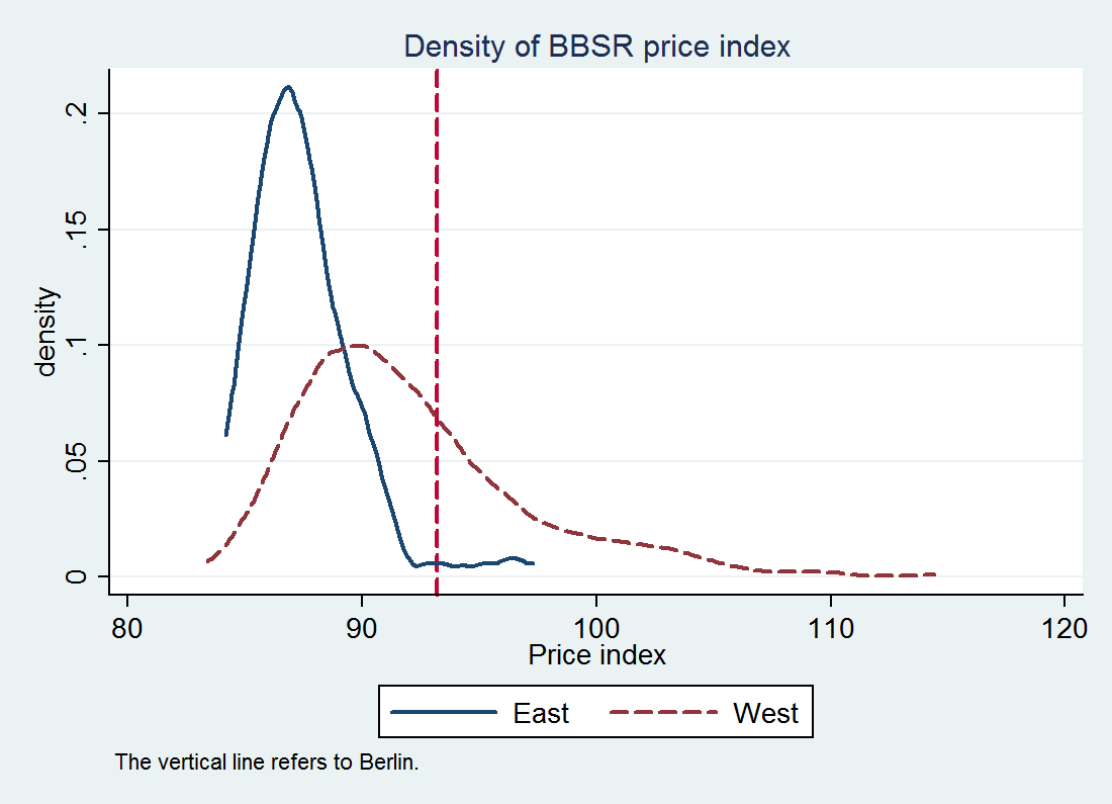


Figure 5: Nominal earnings by Berlin, East and West: *Raw wages*

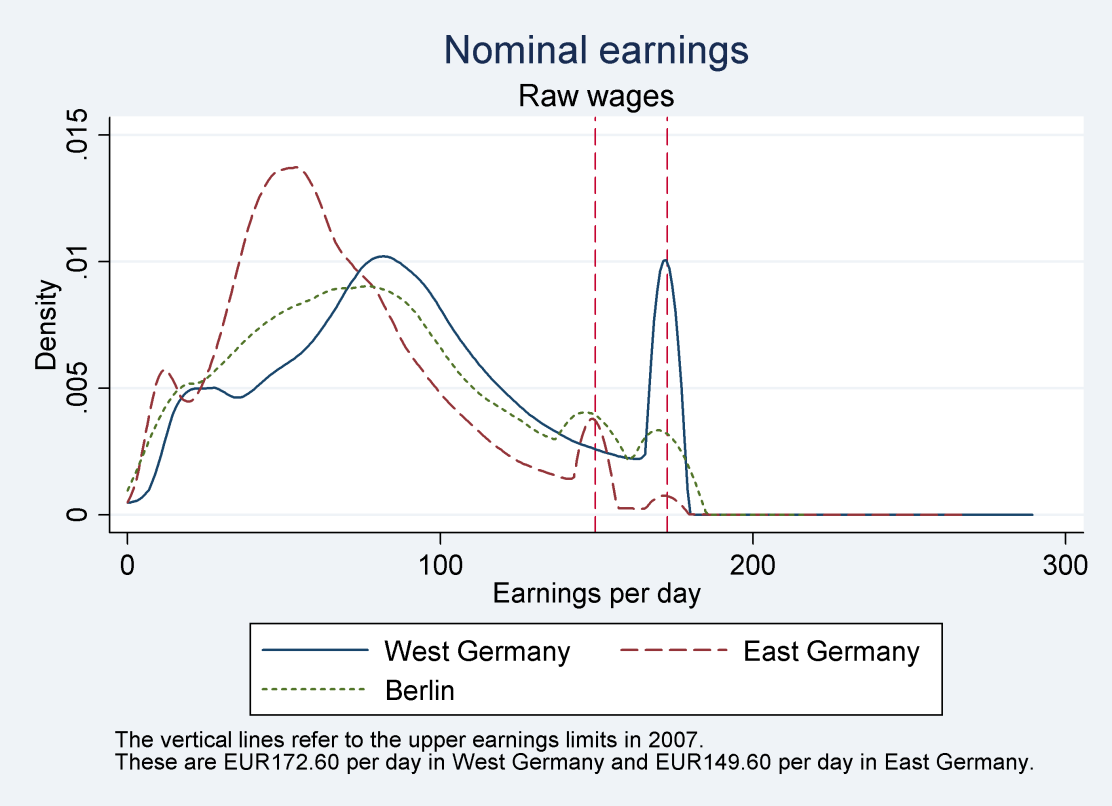


Figure 6: Real earnings by Berlin, East and West: *Raw wages*

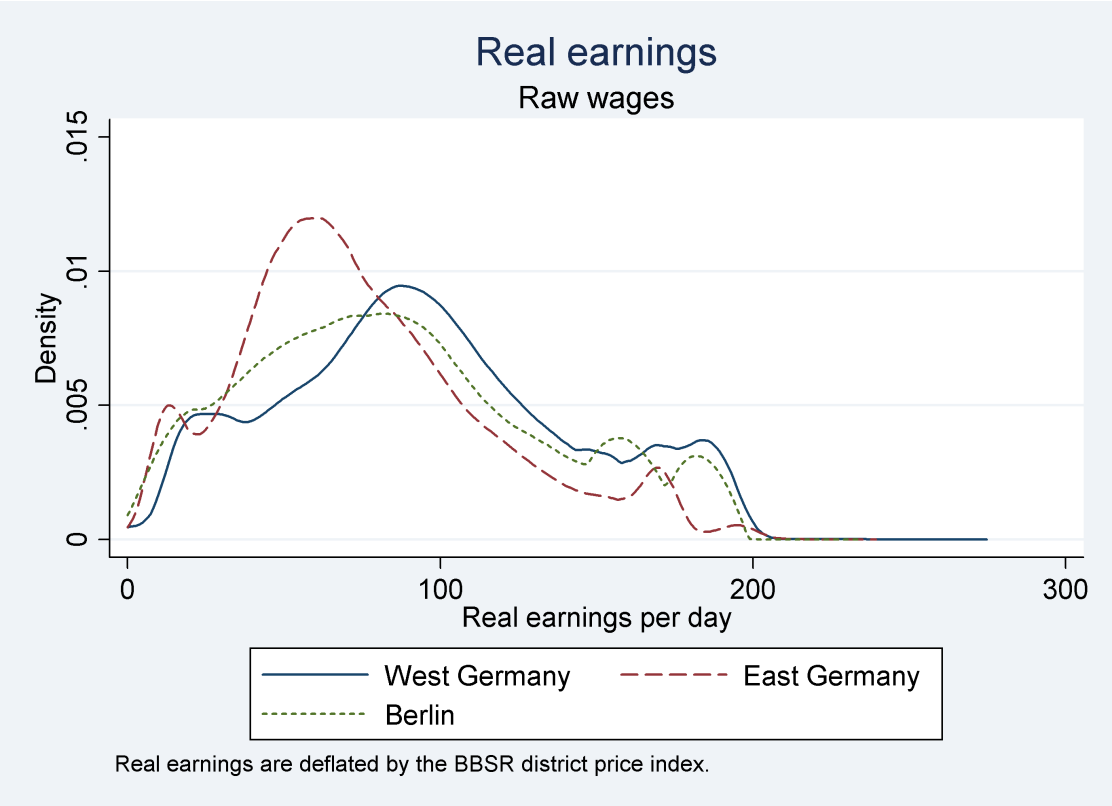


Figure 7:

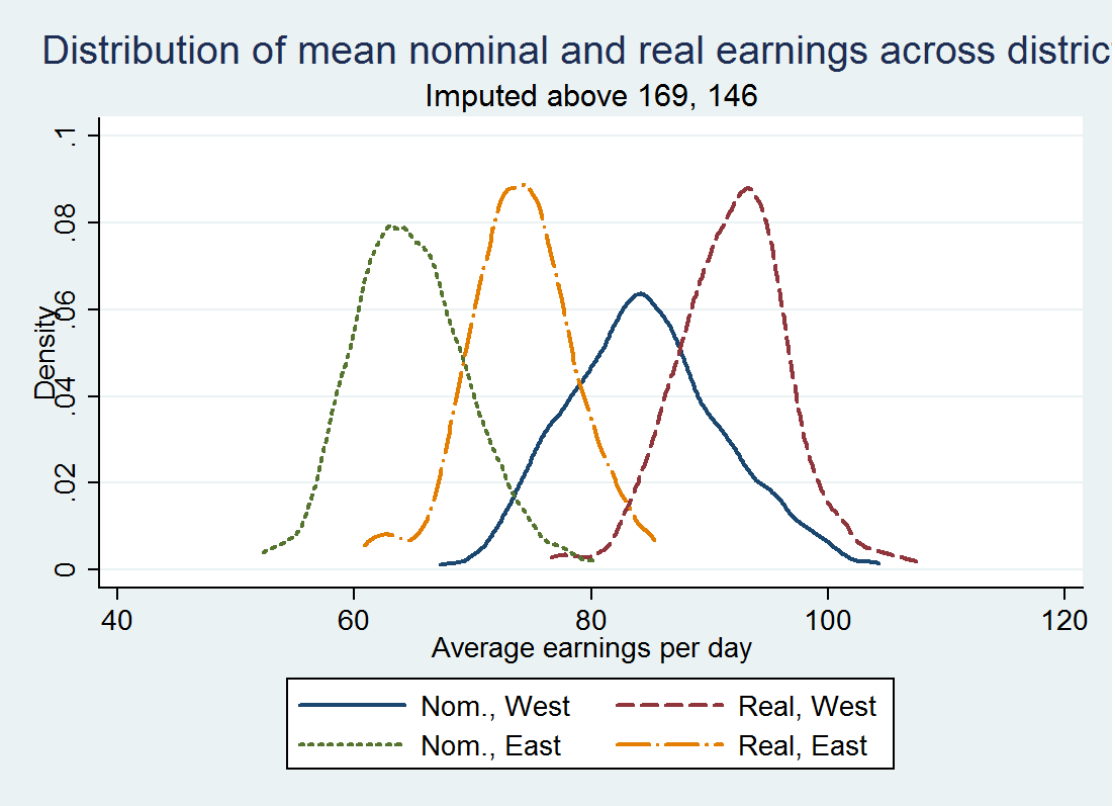


Figure 8: Mean nominal earnings and prices across districts

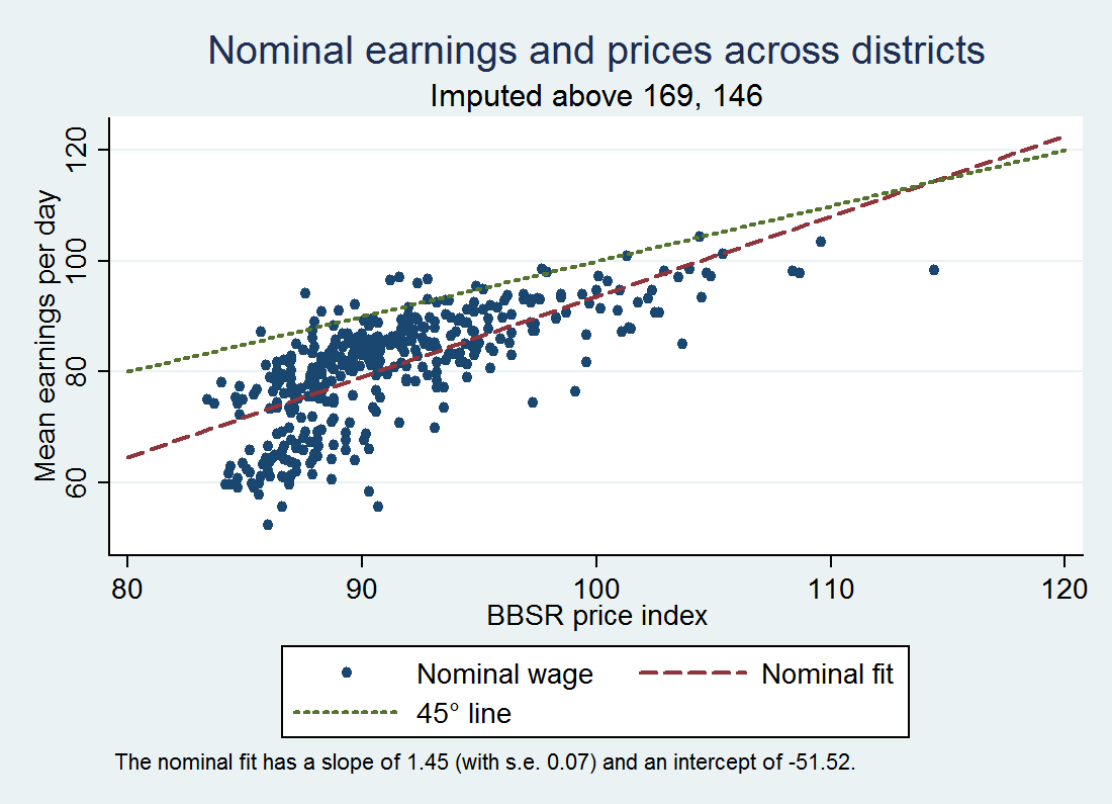


Figure 9: Mean real earnings and prices across districts

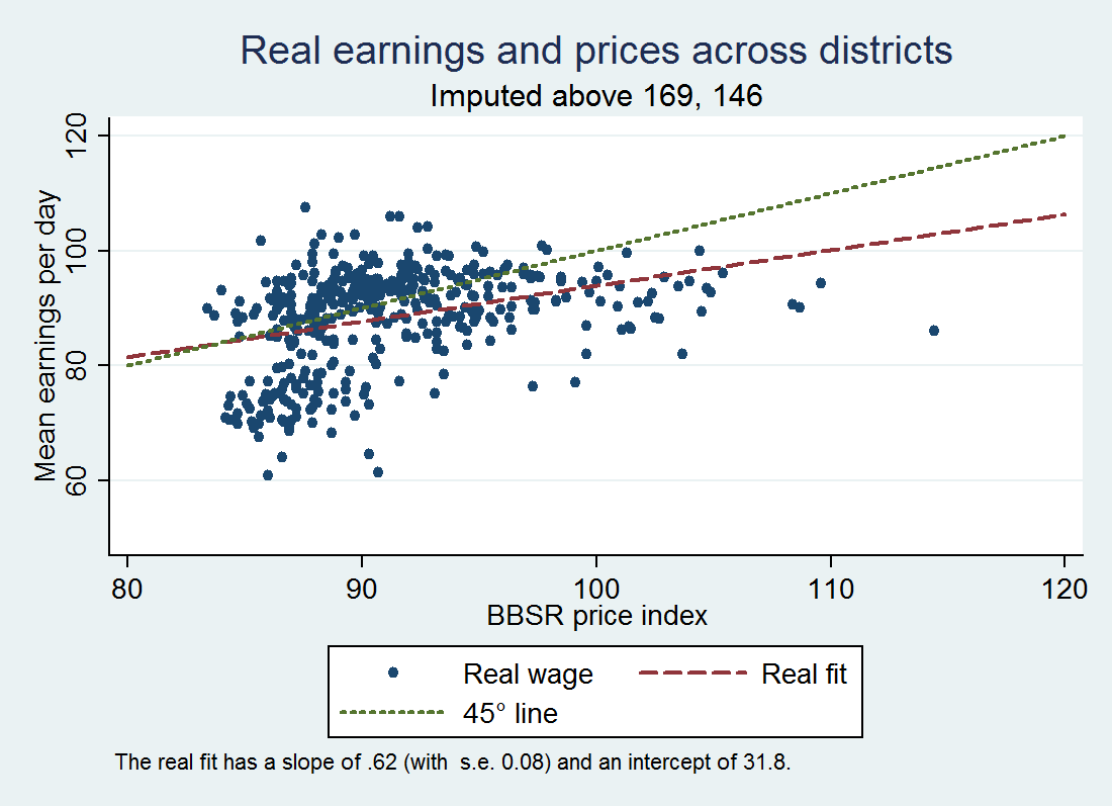


Figure 10: Mean nominal earnings and prices across districts; West Germany only

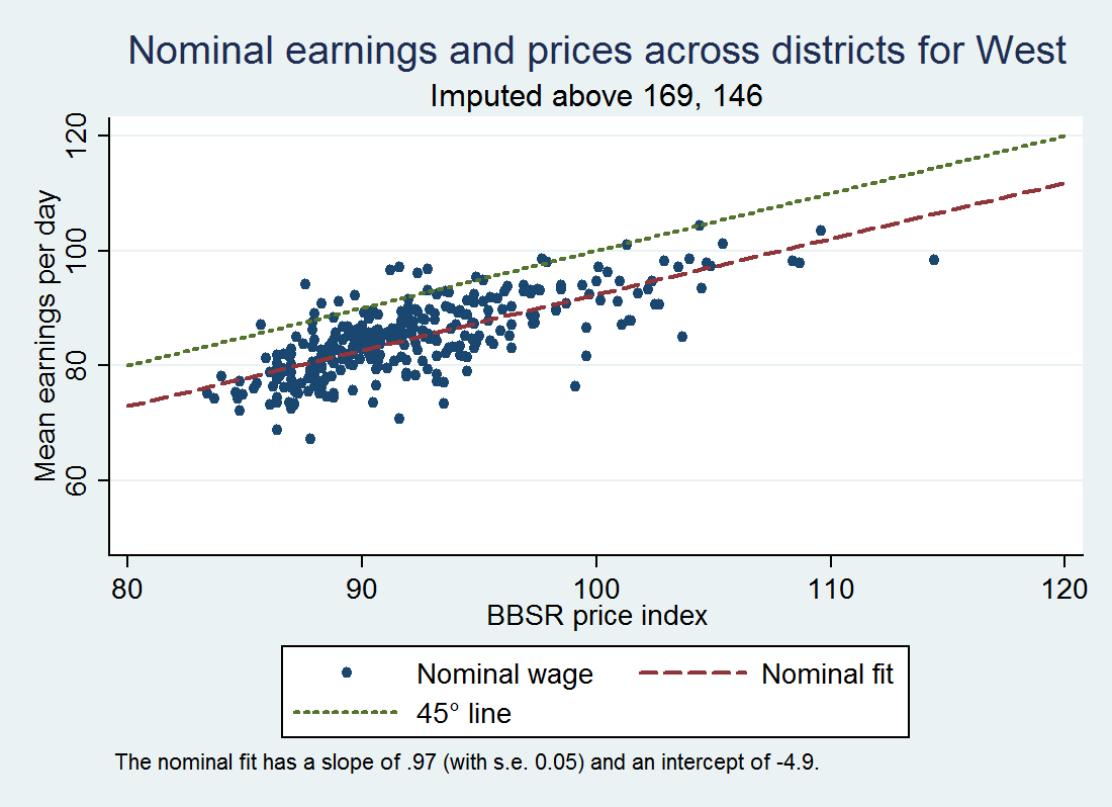


Figure 11: Mean real earnings and prices across districts; West Germany only

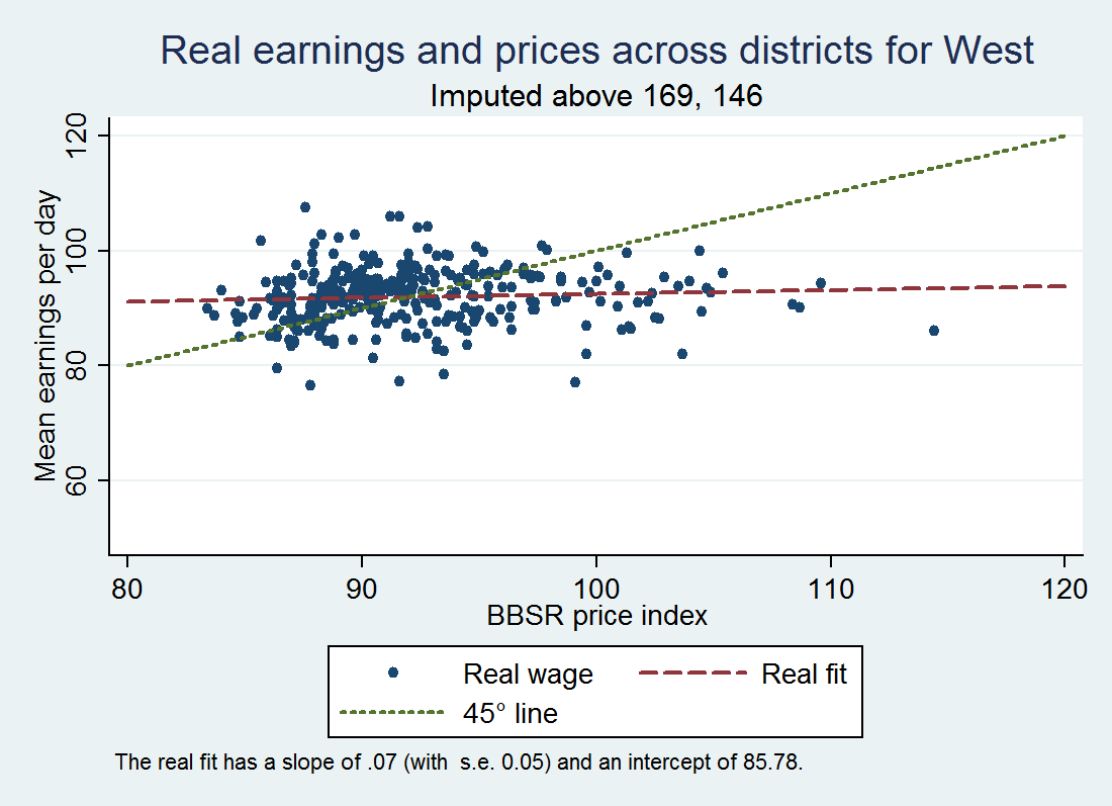


Figure 12: Mean nominal earnings and prices across districts; East Germany only

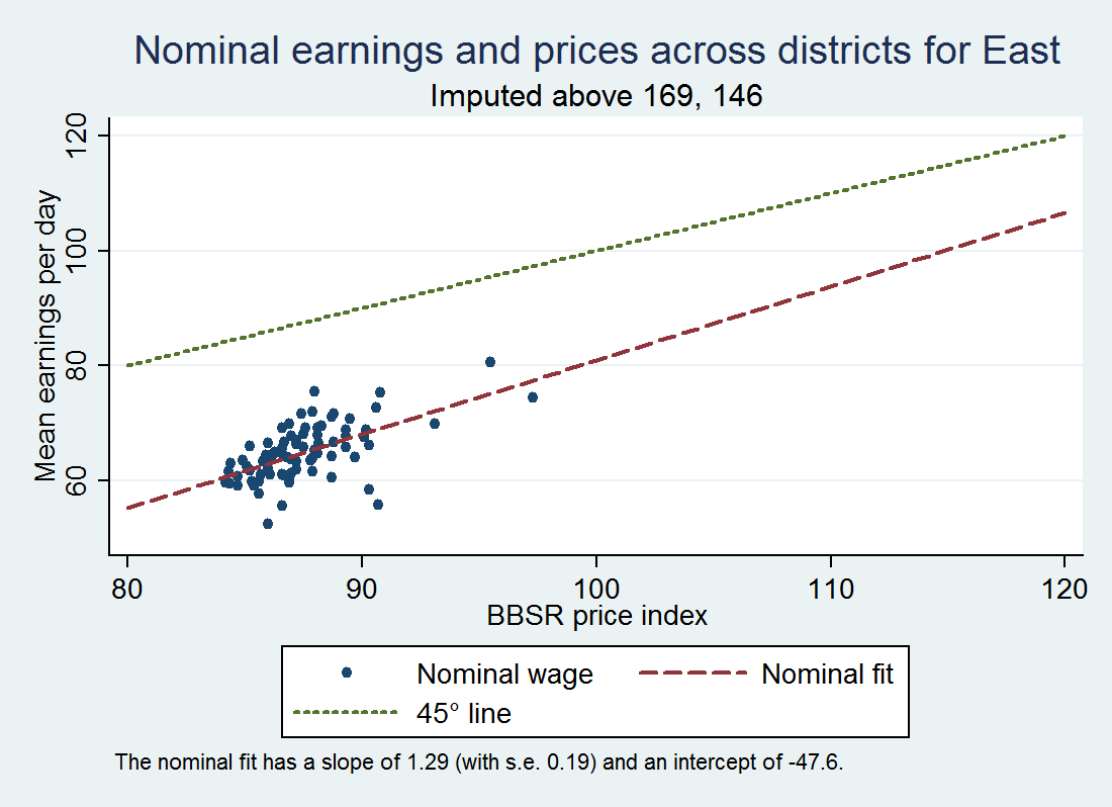


Figure 13: Mean real earnings and prices across districts; East Germany only

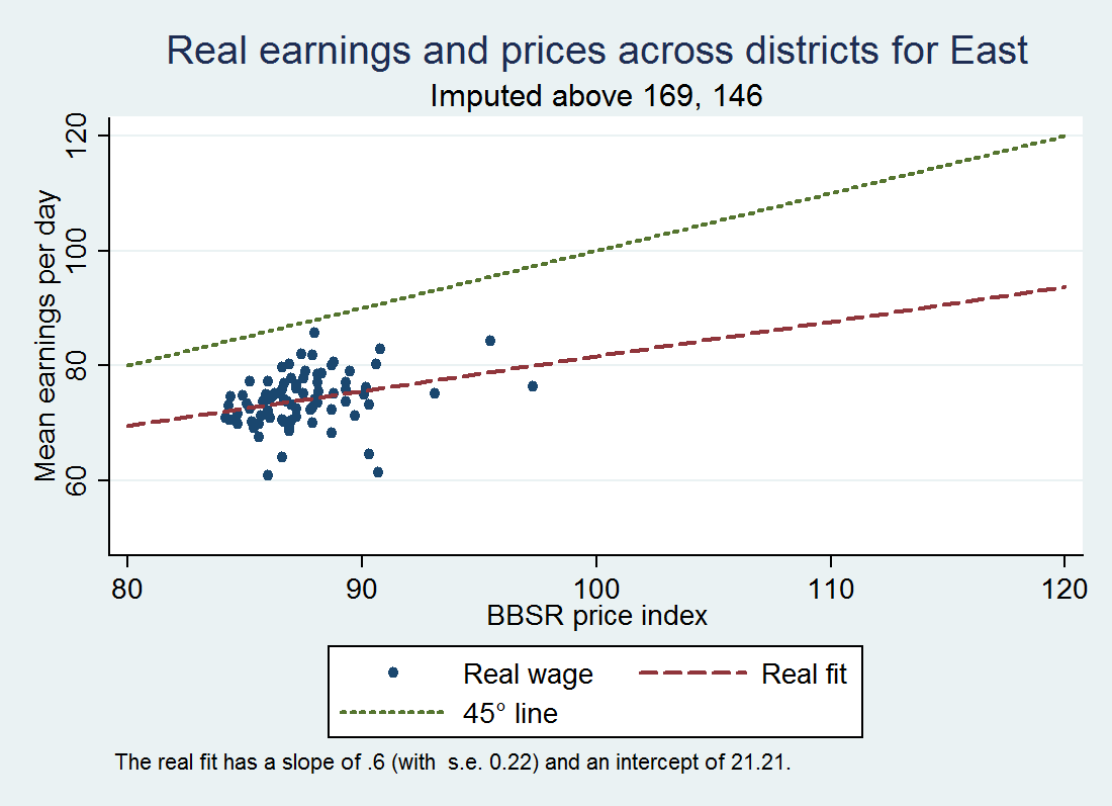


Figure 14: Comparing the distribution of mean earnings and prices; West Germany only

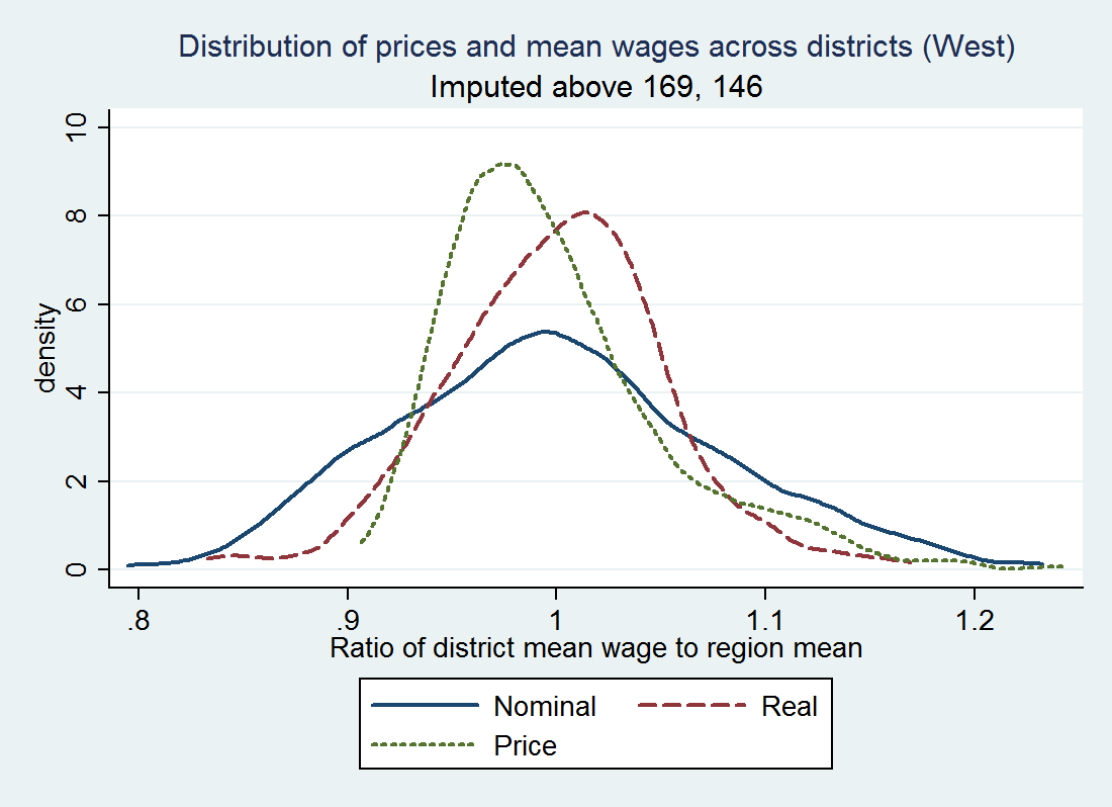


Figure 15: Comparing the distribution of mean earnings and prices; East Germany only

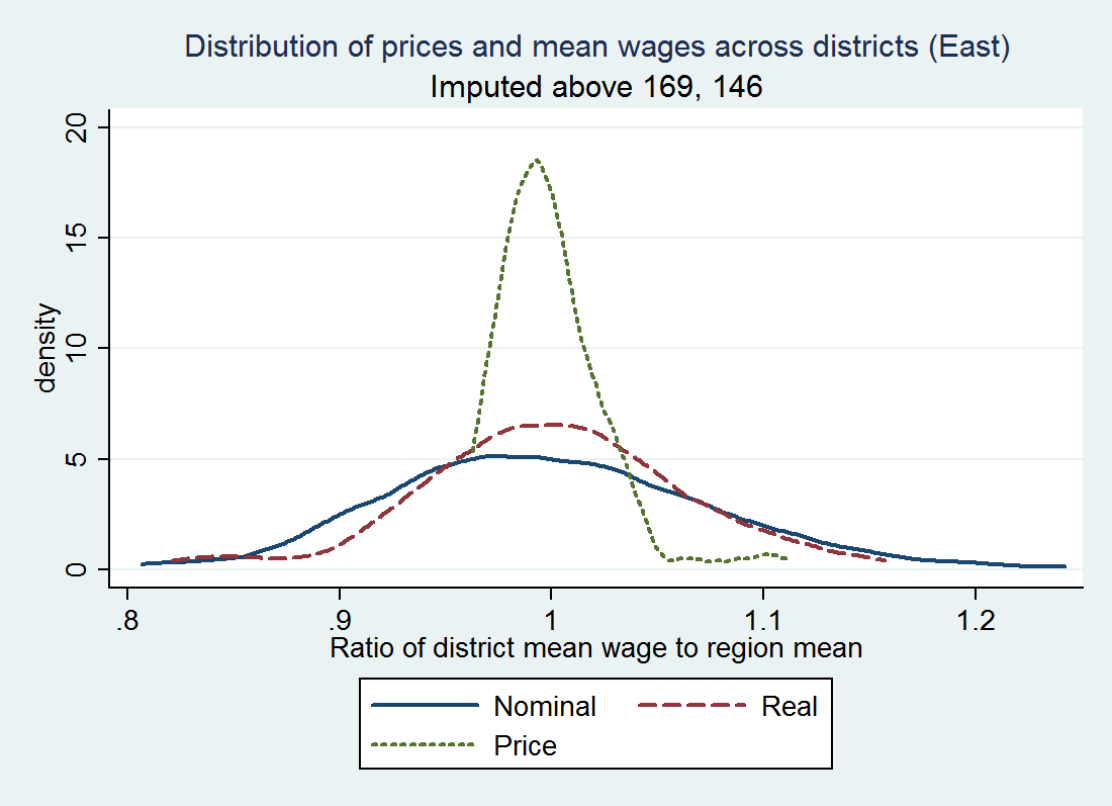


Figure 16: Relationship between nominal and real earnings across districts

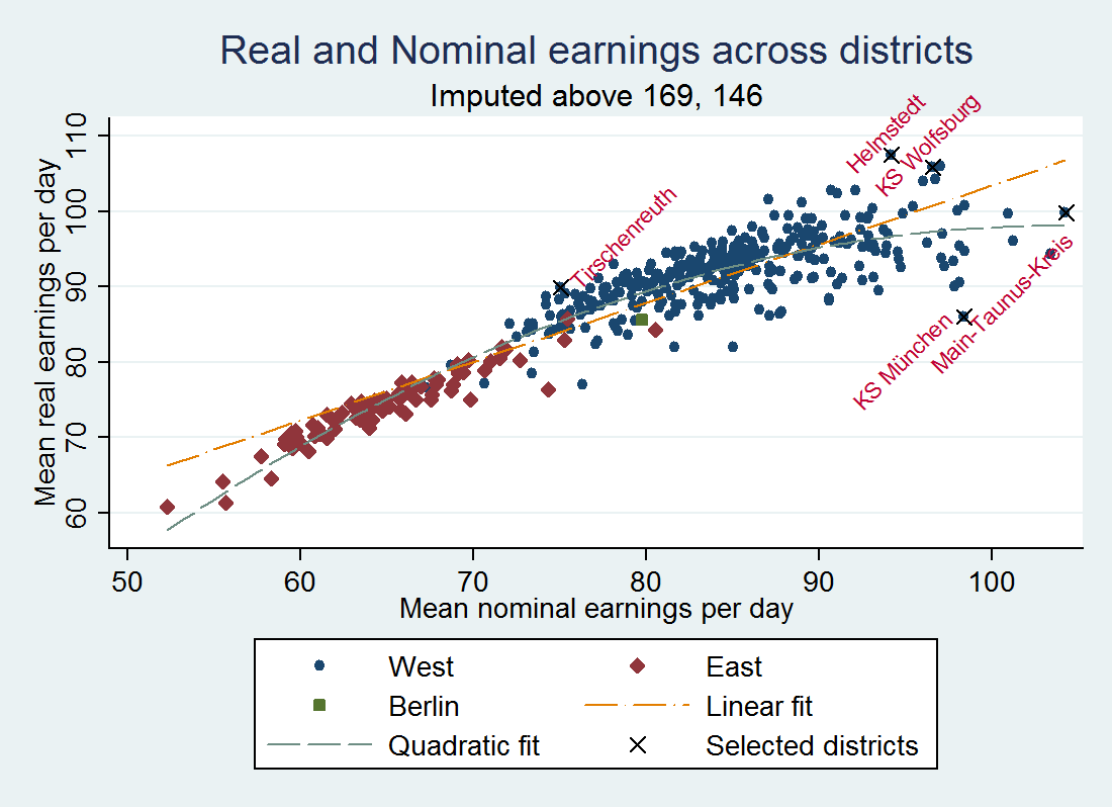


Figure 17: Relationship between nominal and real earnings across districts

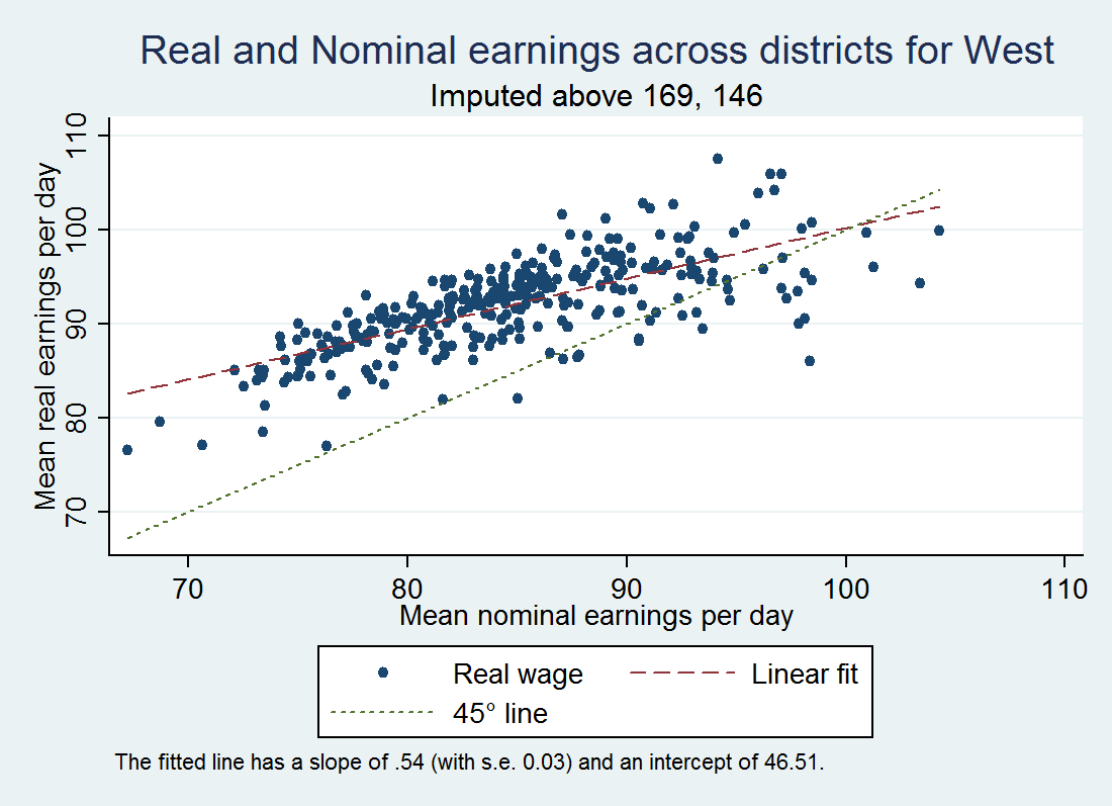


Figure 18: Relationship between nominal and real earnings across districts

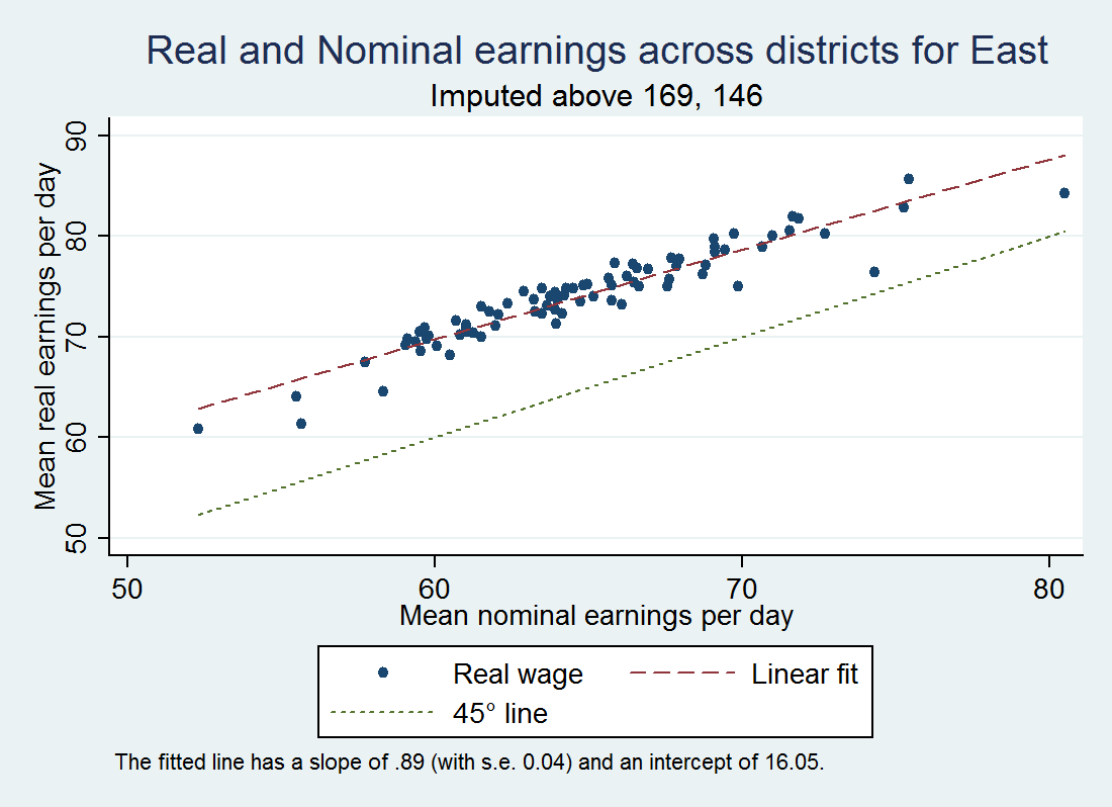


Figure 19: Regional distribution of nominal earnings (by percentile of distribution)

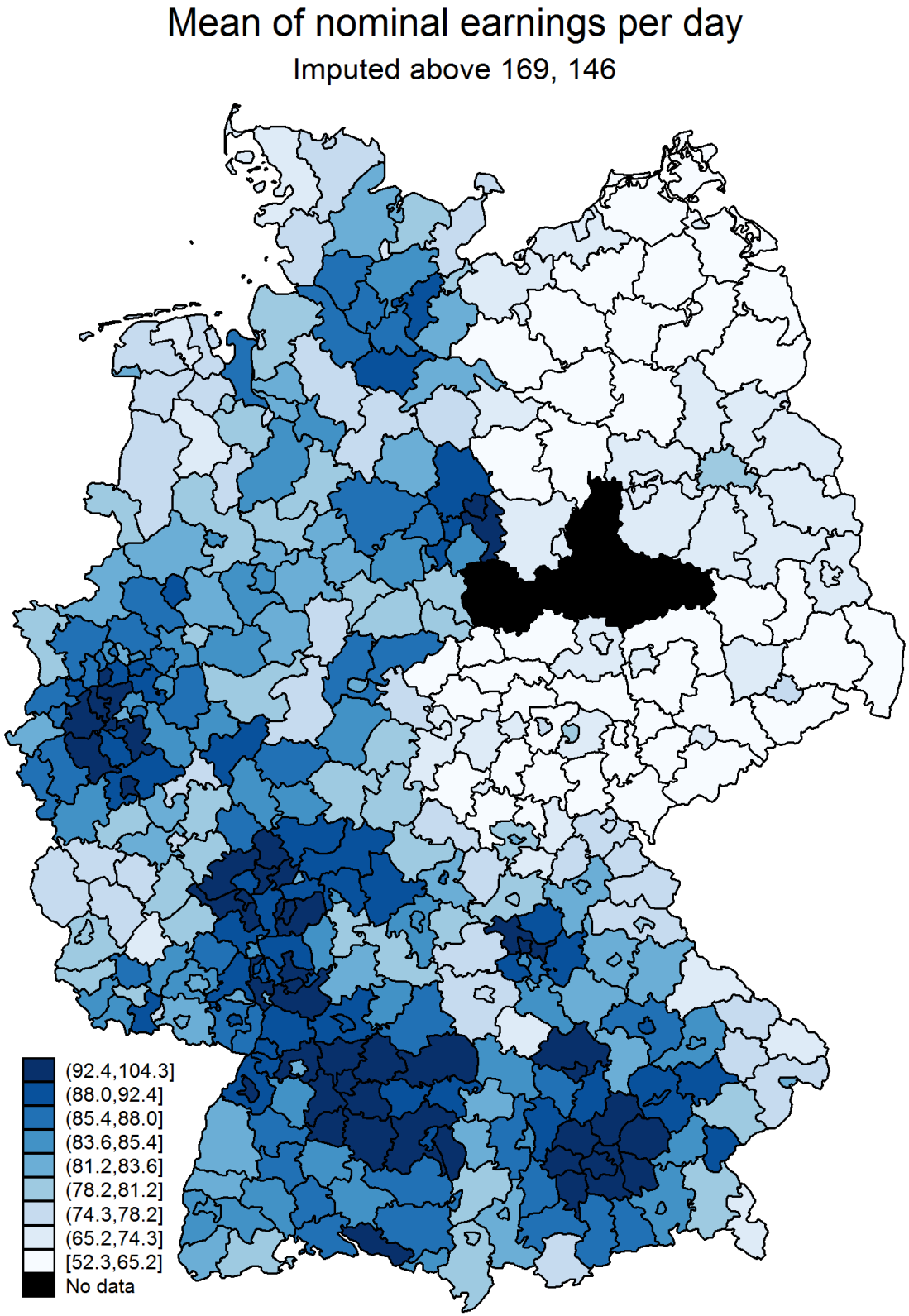
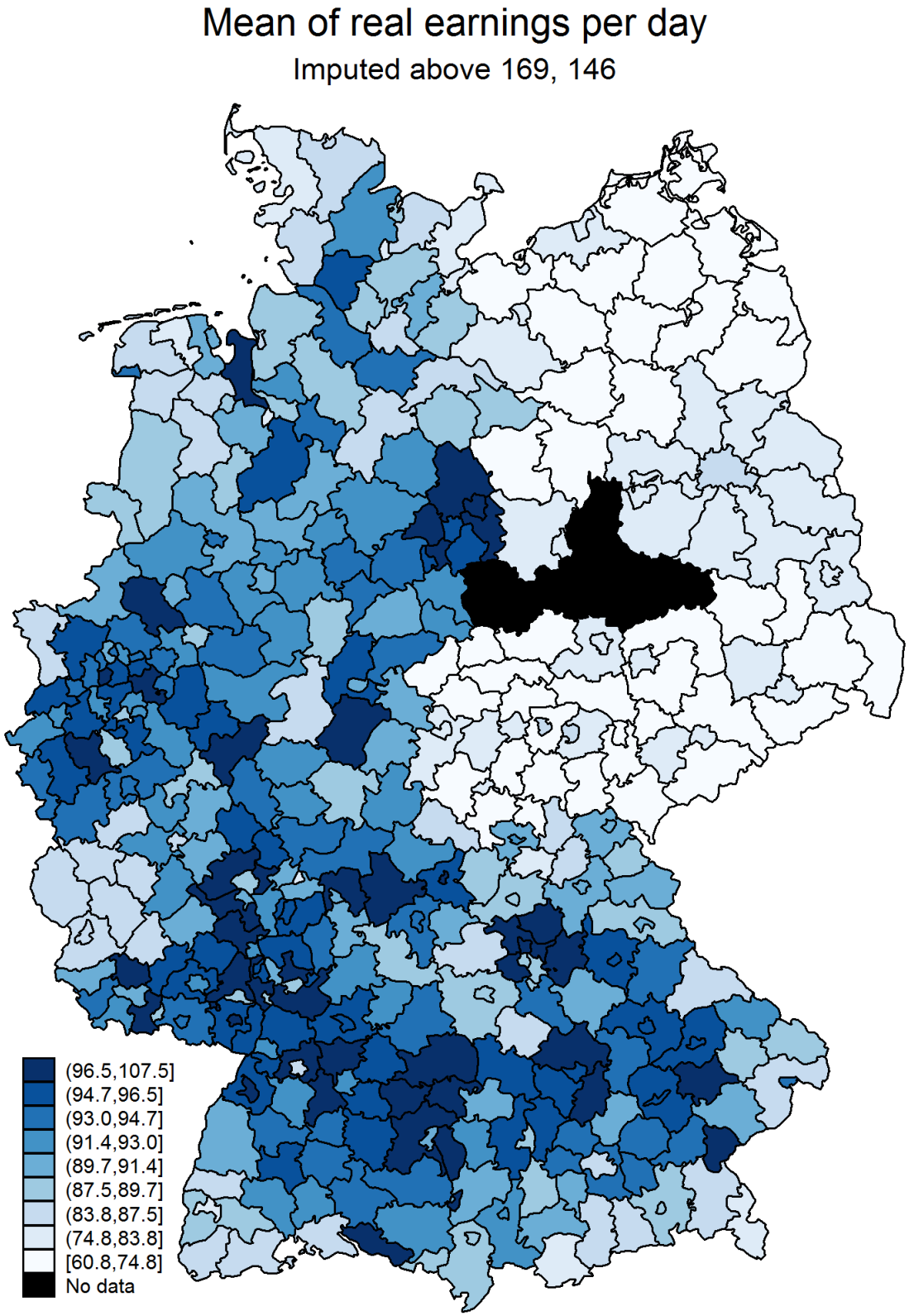


Figure 20: Regional distribution of real earnings (by percentile of distribution)



9 Tables

Table 1: Decomposition of price variation by state

Source	SS	df	MS	F	Prob > F
Between groups	2862.08	15	190.81	11.03	0.00
Within groups	6765.72	391	17.30		
Total	9627.79	406	23.71		

Bartlett's test for equal variances: chi2(13) = 118.11 Prob>chi2 = 0.000

Table 2: Regression of price index on rural/urban and state indicators

Price index	(1)	(2)	(3)	(4)	(5)
Core city	6.287***			6.986***	
High density periphery	3.777***			3.975***	
Rural periphery	0.0608			0.685	
Agglomerations: Core cities		8.679***			10.65***
Agglomerations: High density		8.181***			9.135***
Agglomerations: Medium density		5.737***			7.039***
Agglomerations: Rural		2.374*			4.204***
Urbanised regions: Core cities		5.845***			6.654***
Urbanised regions: Medium density		3.134***			3.423***
Urbanised regions: Rural		0.975			1.324
Rural: Major density		2.050*			1.760*
Schleswig-Holstein			0.333	3.075**	5.009***
Hamburg			9.280*	7.429*	7.008*
Niedersachsen			−3.433***	−0.994	1.407
Bremen			−1.170	−3.021	−1.445
Hessen			2.830**	4.223***	5.123***
Rheinland-Pfalz			−1.590	0.0587	2.440**
Baden-Württemberg			3.368***	4.654***	6.566***
Bayern			−0.369	3.035***	5.010***
Saarland			−2.804	−2.145	−3.116*
Berlin			1.080	−0.771	−1.192
Brandenburg			−4.070***	−0.206	0.788
Mecklenburg-Vorpommern			−3.581**	1.090	3.984***
Sachsen			−5.283***	−3.953***	−2.329*
Sachsen-Anhalt			−5.895***	−2.849*	0.156
Thüringen			−5.107***	−1.753	0.848
Constant	88.53***	87.30***	92.12***	86.99***	83.74***
R-squared	0.244	0.348	0.297	0.484	0.605
adjusted R-squared	0.239	0.334	0.270	0.460	0.582
Number of observations	407	407	407	407	407

The dependent variable is the BBSR priceindex

(1): Omitted category is Rural

(2): Omitted category is Rural: Minor density

(3): Omitted category is Nordrhein-Westfalen

(4): Omitted categories are Rural and Nordrhein-Westfalen

(5): Omitted categories are Rural: Low population density and Nordrhein-Westfalen

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Number and proportion of censored observations for 2007

Censoring limit	East		West		Berlin	
(EURO)	No. obs.	Perc.	No. obs.	Perc.	No. obs.	Perc.
172.60			**24,990**	**6.54%**	**571**	**3.33%**
169.00			31,205	8.16%	725	4.22%
150.00			46,299	12.11%	1,214	7.07%
149.59	**1,014**	**1.24%**				
146.00	3,130	3.83%				
130.00	4,887	5.97%				
Total obs.	81,816		382,377		17,162	

Table 4: Summary statistics for different wage variables, nominal and real

Variable	Mean	Median	S.D.	Max	Min	N
Nom. raw wage (unweighted)	80.68	76.31	44.89	289.38	0.00	481,336
Nom. raw wage (weight)	85.36	81.08	44.11	289.38	0.00	481,336
Nom. wage 1 (trunc. at 169, 146) (weight)	78.10	77.30	37.69	169.00	0.00	440,399
Nom. wage 2 (imp. at 169, 146) (weight)	83.19	79.35	46.75	991.22	0.01	469,284
Real raw wage (unweighted)	86.87	82.81	47.53	325.33	0.00	475,262
Real raw wage (weight)	91.90	87.93	46.62	325.33	0.00	475,262
Real wage 1 (trunc. at 169, 146) (weight)	84.15	83.53	40.14	199.01	0.00	440,399
Real wage 2 (imp. at 169, 146) (weight)	89.37	85.64	49.49	1,106.83	0.01	469,284

Table 5: The 10 richest and poorest districts and their prices

District region	State	Region	Mean nom. daily wage	Price index	Price rank
Main-Taunus-Kreis	Hessen	West	104.26	104.4	9
München	Bayern	West	103.4	109.6	2
Hochtaunuskreis	Hessen	West	101.25	105.4	5
Böblingen	Baden-Württ.	West	100.96	101.3	21
Bodenseekreis	Baden-Württ.	West	98.47	97.7	40
Ebersberg	Bayern	West	98.46	104	10
KS München	Bayern	West	98.38	114.4	1
Starnberg	Bayern	West	98.15	108.4	4
KS Darmstadt	Hessen	West	98.14	102.9	13
KS Erlangen	Bayern	West	98.01	97.9	39
...	...	...	...	...	...
Güstrow	Meck.-Vorp.	East	59.54	86.9	339
Görlitz	Sachsen	East	59.5	84.4	402
Erzgebirgskreis	Sachsen	East	59.37	85.4	386
Elbe-Elster	Brandenburg	East	59.1	84.7	397
Mansfeld-Südharz	Sachsen-Anhalt	East	59.05	85.4	386
Ostvorpommern	Meck.-Vorp.	East	58.32	90.3	193
Demmin	Meck.-Vorp.	East	57.73	85.6	383
Rügen	Meck.-Vorp.	East	55.66	90.7	172
Müritz	Meck.-Vorp.	East	55.5	86.6	351
Uecker-Randow	Meck.-Vorp.	East	52.3	86	374

Nominal wages are imputed above EUR146 or EUR169 per day.

Districts are ranked highest to lowest by their price level.

In total, there are 407 districts with valid price information.

Table 6: The correlation between average earnings and price *at district-level*

Variable	Pearson	Kendall tau-b	Spearman rho	N
Nominal raw	0.7241	0.5949	0.7807	407
Real raw	0.4235	0.3558	0.5119	407
Nominal 1: truncated at 169, 146	0.6309	0.5437	0.7270	407
Real 1: truncated at 169, 146	0.2138	0.1627	0.2462	407
Nominal 2: imputed above 169, 146	0.7038	0.5800	0.7662	407
Real 2: imputed above 169, 146	0.3529	0.2893	0.4229	407

Table 7: Coefficient of variation by region

	East	West
Price index	2.62%	5.35%
Nominal median wage, imputed	7.94%	7.09%
Real median wage, imputed	6.67%	5.51%

Table 8: The association between average real and nominal wages across districts

Real earnings	Wage 1: Trunc. above 169 or 146 (1)	(2)	(3)	Wage 3: Impute above 169 or 146 (4)	(5)	(6)
Mean nom. earn. 1	0.796***	0.554***	0.653***			
Schleswig-Holstein		−3.199***	−4.061***		−3.314***	−4.013***
Hamburg		−7.690*	−6.464*		−7.903**	−6.713*
Niedersachsen		1.237*	0.245		1.138	0.316
Bremen		−1.713	0.127		−2.238	−0.528
Hessen		−2.146**	−3.049***		−1.973**	−2.784***
Rheinland-Pfalz		−0.104	−0.858		−0.240	−0.941
Baden-Württemberg		−1.998**	−3.031***		−2.051***	−2.938***
Bayern		−0.876	−2.308***		−0.730	−1.856***
Saarland		2.300	1.870		2.025	1.610
Berlin		−3.628	−1.812		−4.105	−2.421
Brandenburg		−6.651***	−6.906***		−5.809***	−6.237***
Mecklenburg-Vorpommern		−8.961***	−8.950***		−8.943***	−9.001***
Sachsen		−7.001***	−5.679***		−6.176***	−5.453***
Sachsen-Anhalt		−5.751***	−5.351***		−5.475***	−5.275***
Thüringen		−6.804***	−6.469***		−6.361***	−6.275***
Core city			−3.639***			−2.941***
High density periphery			−1.438**			−0.762
Rural periphery			0.229			0.361
Mean nom. earn. 3				0.779***	0.598***	0.660***
Constant	22.94***	43.39***	37.63***	25.54***	41.96***	38.33***
R-squared	0.796	0.858	0.877	0.833	0.888	0.899
adjusted R-squared	0.796	0.852	0.871	0.832	0.883	0.894
Number of observations	407	407	407	407	407	407

The dependent variable is the mean real wage per district.

(2) and (5): Omitted category is Nordrhein-Westfalen

(3) and (6): Omitted categories are Rural and Nordrhein-Westfalen

(1) to (3): Refer to earnings measure 1: earnings truncated above 169 or 146.

(4) to (6): Refer to earnings measure 3: earnings imputed above 169 or 146.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 9: The effect of accounting for prices on inequality

Inequality measure	Raw wage			Wage 1 (trunc. 169, 146)			Wage 2 (imput. 169,146)		
	nominal	real	% diff	nominal	real	% diff	nominal	real	% diff
Coeff. of Var.	0.5167	0.5073	-1.83%	0.4826	0.4771	-1.14%	0.5620	0.5537	-1.47%
Log Variance	0.4194	0.4094	-2.38%	0.3847	0.3789	-1.52%	0.4049	0.3975	-1.81%
Poverty rate	0.2207	0.2178	-1.32%	0.2218	0.2183	-1.58%	0.2152	0.2123	-1.33%
p90/p10	5.3189	5.1785	-2.64%	4.6982	4.6271	-1.51%	4.7298	4.6543	-1.60%
p90/p50	1.9152	1.8655	-2.60%	1.6857	1.6639	-1.30%	1.7364	1.7111	-1.45%
p10/p50	0.3601	0.3602	0.04%	0.3588	0.3596	0.22%	0.3671	0.3676	0.15%
p75/p25	2.1292	2.0955	-1.58%	2.0459	2.0204	-1.25%	2.0474	2.0219	-1.24%
GE(-1)	0.3142	0.3060	-2.61%	0.2882	0.2827	-1.90%	0.3045	0.2976	-2.27%
GE(-1) SE	0.0187	0.0185		0.0189	0.0180		0.0177	0.0179	
GE(0)	0.1696	0.1649	-2.76%	0.1533	0.1506	-1.74%	0.1693	0.1658	-2.10%
GE(0) SE	0.0004	0.0004		0.0004	0.0004		0.0004	0.0004	
GE(1)	0.1376	0.1333	-3.14%	0.1225	0.1200	-1.98%	0.1458	0.1423	-2.40%
GE(1) SE	0.0003	0.0003		0.0003	0.0003		0.0004	0.0005	
GE(2)	0.1314	0.1265	-3.68%	0.1142	0.1115	-2.30%	0.1579	0.1533	-2.92%
GE(2) SE	0.0003	0.0002		0.0003	0.0002		0.0008	0.0009	
Gini	0.2910	0.2860	-1.70%	0.2725	0.2693	-1.16%	0.2913	0.2875	-1.31%
Gini SE	0.0003	0.0003		0.0003	0.0003		0.0004	0.0004	
A(0.5)	0.0731	0.0710	-2.87%	0.0658	0.0646	-1.81%	0.0747	0.0731	-2.17%
A(0.5) SE	0.0001	0.0001		0.0001	0.0001		0.0002	0.0002	
A(1)	0.1560	0.1520	-2.54%	0.1421	0.1398	-1.62%	0.1557	0.1527	-1.93%
A(1) SE	0.0003	0.0003		0.0003	0.0003		0.0004	0.0004	
A(2)	0.3859	0.3796	-1.62%	0.3656	0.3612	-1.21%	0.3785	0.3731	-1.42%
A(2) SE	0.0139	0.0141		0.0150	0.0146		0.0136	0.0139	
No. obs.	478,124	472,076		437,213	437,213		469,284	469,284	

All earnings figures are employment-duration weighted.

Rows (5)-(8): Percentile ratios

Rows (9)-(16): Generalized Entropy indices GE(a), a = income difference sensitivity

Rows (19)-(24): Atkinson indices A(e), e = inequality aversion

For the GE(a), A(e) and Gini coefficient, I use bootstrapped standard errors.

Table 10: The correlation between nominal earnings and prices

Earnings variable	Pearson (weighted)	Kendall tau-b	Spearman rho	N
Raw wage	0.1918	0.1202	0.1777	475,262
Wage 1: truncated at 169, 146	0.1508	0.0977	0.1450	440,399
Wage 2: imputed above 169, 146	0.1558	0.1081	0.1603	469,284

Table 11: The correlation between nominal and real earnings

Earnings variable	Pearson (weighted)	Kendall tau-b	Spearman rho	N
Raw wage	0.9907	0.9370	0.9938	475,262
Wage 1: truncated at 169, 146	0.9904	0.9344	0.9932	440,399
Wage 2: imputed above 169, 146	0.9919	0.9343	0.9932	469,284

Table 12: The effect of accounting for prices on inequality, separately by region

Inequality measure	Germany			West			East		
	nominal	real	% diff	nominal	real	% diff	nominal	real	% diff
Coeff. of Var.	0.5620	0.5537	-1.47%	0.5424	0.5364	-1.10%	0.6028	0.5998	-0.50%
Log Variance	0.4049	0.3975	-1.81%	0.3834	0.3806	-0.74%	0.4235	0.4220	-0.35%
Poverty rate	0.2152	0.2123	-1.33%	0.1908	0.1944	1.89%	0.3322	0.2944	-11.40%
p90/p10	4.7298	4.6543	-1.60%	4.6781	4.6215	-1.21%	4.4799	4.4596	-0.45%
p90/p50	1.7364	1.7111	-1.45%	1.7004	1.6771	-1.37%	1.8509	1.8406	-0.56%
p10/p50	0.3671	0.3676	0.15%	0.3635	0.3629	-0.16%	0.4132	0.4127	-0.10%
p75/p25	2.0474	2.0219	-1.24%	1.9523	1.9542	0.10%	2.0181	2.0150	-0.15%
GE(-1)	0.3045	0.2976	-2.27%	0.2948	0.2909	-1.32%	0.2967	0.2956	-0.37%
GE(-1) SE	0.0177	0.0179		0.0197	0.0190		0.0069	0.0069	
GE(0)	0.1693	0.1658	-2.10%	0.1602	0.1585	-1.03%	0.1782	0.1773	-0.54%
GE(0) SE	0.0004	0.0004		0.0005	0.0005		0.0013	0.0013	
GE(1)	0.1458	0.1423	-2.40%	0.1375	0.1356	-1.42%	0.1569	0.1557	-0.74%
GE(1) SE	0.0004	0.0005		0.0005	0.0005		0.0015	0.0015	
GE(2)	0.1579	0.1533	-2.92%	0.1471	0.1439	-2.18%	0.1817	0.1799	-1.00%
GE(2) SE	0.0008	0.0009		0.0009	0.0009		0.0034	0.0034	
Gini	0.2913	0.2875	-1.31%	0.2821	0.2802	-0.68%	0.2997	0.2985	-0.39%
Gini SE	0.0004	0.0004		0.0004	0.0004		0.0010	0.0010	
A(0.5)	0.0747	0.0731	-2.17%	0.0708	0.0699	-1.17%	0.0791	0.0786	-0.62%
A(0.5) SE	0.0002	0.0002		0.0002	0.0002		0.0006	0.0006	
A(1)	0.1557	0.1527	-1.93%	0.1480	0.1466	-0.95%	0.1632	0.1624	-0.50%
A(1) SE	0.0004	0.0004		0.0004	0.0004		0.0011	0.0011	
A(2)	0.3785	0.3731	-1.42%	0.3709	0.3678	-0.83%	0.3724	0.3715	-0.23%
A(2) SE	0.0136	0.0139		0.0136	0.0132		0.0048	0.0049	
No. obs.	469,284	469,284		377,129	377,129		75,169	75,169	

All earnings figures are employment-duration weighted.

Using the imputed wage measure.

Rows (5)-(8): Percentile ratios

Rows (9)-(16): Generalized Entropy indices GE(a), a = income difference sensitivity

Rows (19)-(24): Atkinson indices A(e), e = inequality aversion

For the GE(a), A(e) and Gini coefficient, I use bootstrapped standard errors.

Table 13: The effect of accounting for prices on inequality measures, decomposed by district, separately by region

			GE(-1)	GE(0)	GE(1)	GE(2)	A(0.5)	A(1)	A(2)	N
Germany	Nom.	With.	0.2966	0.1618	0.1386	0.1510	0.0712	0.1489	0.3443	469,284
		Betw.	0.0079	0.0075	0.0072	0.0070	0.0038	0.0080	0.0522	469,284
	Real	With.	0.2935	0.1618	0.1384	0.1496	0.0712	0.1489	0.3444	469,284
		Betw.	0.0041	0.0040	0.0038	0.0037	0.0021	0.0045	0.0438	469,284
	Diff	With.	-1.06%	0.00%	-0.07%	-0.88%	-0.06%	-0.03%	0.04%	
		Betw.	-47.55%	-47.23%	-47.08%	-47.10%	-45.29%	-43.34%	-16.05%	
West	Nom.	With.	0.3481	0.1809	0.1521	0.1618	0.0788	0.1656	0.3872	377,129
		Betw.	0.0028	0.0028	0.0028	0.0028	0.0014	0.0027	0.0411	377,129
	Real	With.	0.3465	0.1809	0.1518	0.1601	0.0787	0.1654	0.3868	377,129
		Betw.	0.0012	0.0012	0.0012	0.0012	0.0006	0.0014	0.0382	377,129
	Diff	With.	-0.45%	0.00%	-0.23%	-1.04%	-0.19%	-0.15%	-0.11%	
		Betw.	-56.36%	-56.45%	-56.60%	-56.81%	-52.68%	-47.68%	-7.08%	
East	Nom.	With.	0.3389	0.1950	0.1687	0.1942	0.0856	0.1775	0.4012	75,169
		Betw.	0.0025	0.0025	0.0025	0.0025	0.0011	0.0022	0.0075	75,169
	Real	With.	0.3387	0.1950	0.1685	0.1933	0.0855	0.1773	0.4009	75,169
		Betw.	0.0016	0.0016	0.0016	0.0016	0.0007	0.0015	0.0069	75,169
	Diff	With.	-0.03%	0.00%	-0.11%	-0.44%	-0.11%	-0.10%	-0.09%	
	Diff	Betw.	-36.01%	-36.75%	-37.51%	-38.30%	-36.44%	-32.39%	-7.95%	

All earnings figures are employment-duration weighted.

Using the imputed wage measure.

Rows (9)-(16): Generalized Entropy indices GE(a), a = income difference sensitivity

Rows (19)-(24): Atkinson indices A(e), e = inequality aversion

References

Aaberge, R., A. Langørgen, M. Mogstad, and M. Østensen: 2008, 'The Impact of Local Public Services and Geographical Cost of Living Differences on Poverty Estimates'. *Discussion Papers, Statistics Norway* **551**, 43.

Bach, S., G. Corneo, and V. Steiner: 2009, 'From Bottom To Top: The Entire Income Distribution In Germany, 1992-2003'. *Review of Income and Wealth* **55**(2), 303–330.

BBSR: 2009, 'Bundesinstitut für Bau-, Stadt- und Raumforschung (BBSR): Regionaler Preisindex'. *Berichte des BBR (Bundesamt für Bauwesen und Raumordnung)* **30**.

Blien, U., H. Gartner, H. Stüber, and K. Wolf: 2009, 'Regional price levels and the agglomeration wage differential in western Germany.'. *Annals of Regional Science* **43**(1), 71 – 88.

BMF: 2011a, 'Federal Ministry of Finance: Wer zahlt welche Steuern?'.

BMF: 2011b, 'Monatsbericht des Bundesministerium der Finanzen, Juni 2011'. *Federal Ministry of Finance* **06**.

Dustmann, C., J. Ludsteck, and U. Schönberg: 2009, 'Revisiting The German Wage Structure'. *Quarterly Journal of Economics* **124**(2), 843 – 881.

Frick, J. R.: 2008, 'SOEPmonitor 1984-2007: Time Series on selected Indicators about "Living in Germany"'. DIW Berlin.

Goebel, J., J. R. Frick, and M. M. Grabka: 2009, 'Preisunterschiede mildern Einkommensgefälle zwischen West und Ost'. *DIW Wochenbericht* **51**.

Hillringhaus, T. and A. Peichl: 2010, 'Die Messung von Armut unter Berücksichtigung regional divergierender Lebenshaltungskosten und öffentlicher Leistungen'. Technical Report 5344, IZA Discussion Paper.

Moretti, E.: 2009, 'Real Wage Inequality'. UC Berkeley Working Paper.

Roos, M.: 2006, 'Earnings Disparities in Unified Germany: Nominal versus Real'. *Jahrbuch für Regionalwissenschaft* **26**, 171–189. 10.1007/s10037-006-0003-2.

SIAB: 2010, 'Sample of Integrated Labour Market Biographies 1975-2008'. *Research Data Centre of the German Federal Employment Agency at the Institute for Employment Research, Nuremberg.*

Ströhl, G.: 1994, 'Zwischenörtlicher Vergleich des Verbraucherpreisniveaus in 50 Städten'. *Wirtschaft und Statistik* **6**, 415.