The Inequality of Real Wages in Germany

Christoph Lakner
University of Oxford

This paper investigates the implications of regional price differences for earnings differentials and inequality in Germany. There is substantial inter-regional variation in both nominal earnings and prices, and we are interested in how these two variables interact to produce differences in real earnings. We combine a new district-level price index for 2007 published by the German Ministry of Transport with earnings data from the SIAB data by the Institute for Employment Research. The richness of the price information allows a regional analysis at an unprecedented disaggregated level. Most of the literature has focused on the differential between East and West Germany, which ignores the substantial variation within these regions. Initial results suggest that inequality in all of Germany as measured by the Gini coefficient changes very little after accounting for regional differences in the cost of living. We have not yet finished the full analysis of the initial results although access to all the necessary data has been secured.

The first part of the paper compares inequality measures before and after accounting for regional price differences. We consider a number of different inequality measures, such as percentile ratios, Generalised Entropy indices, Atkinson indices and the Gini index. We also decompose inequality into the within- and between-state variation. These aggregate measures, however, do not measure if individual positions in the earnings distribution are changed when accounting for regional price differentials. In the second part we thus investigate positional changes by considering the correlation between real and nominal earnings. In the last part we account for spatial autocorrelations.

In 2009, the German Ministry of Transport published for the first time a comprehensive regional price index for the entire country. The unit of observation is a district region, which has on average approximately 200,000 residents. 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. Consistent with evidence from other countries, most of the inter-regional price differentials are driven by differences in housing costs.

There are relatively large differences in prices across Germany. Overall, there is a price differential of 31% between the most expensive and least expensive districts. 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. It is interesting that the ten least expensive districts are relatively evenly distributed between East and West Germany. On the other hand, the ten most expensive districts are all in the West. While the average for East Germany is below that of West Germany, the East-West divide in prices is relatively weak. This is because the goods which are comparatively inexpensive in the East enter the consumer price index only with a low weight. These facts alone question the validity of the literature focused on comparing East and West Germany and suggest that a more disaggregated analysis is necessary. A linear regression of prices on a rural/urban classification and state-dummies explains 60% of the variation in prices which is quite remarkable.

The earnings information comes from the SIAB data, which is based on a 2% sample of social security records. The most widely used German data to study questions of inequality is the German Socio-Economic Panel (SOEP), a nation-wide household survey which in the original sample contains only 6000 households. Although it has many attractive features, in particular the rich income data and household composition information, it is unsuitable for our 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 addition, the SOEP data has severe problems with attrition and non-response.

The advantages of the SIAB data are its large sample size and its long panel structure. The SIAB survey, previously called the IABS, started in 1975 and includes 200,000 people in West and 52,000 people in East Germany. Furthermore, because it is based on administrative records it is very 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, 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.

We compare the results using different approaches to the problem of right-censored earnings. One possibility is to drop all the censored observations and confine the analysis to the earnings distribution up to the 85th percentile as done in Dustmann et al. (2009). Second, we implement a number of different imputation methods. We estimate a standard interval regression model separately for each district. The independent variables include controls for human capital, gender, age, labour market experience and industrial sector of employment. We also use a multiple imputation method as described by Jenkins et al. (2011), which can be shown to outperform simple imputation methods.

The paper most closely related to our analysis is Goebel et al. (2009) who use the same regional price level data. Because their analysis is based on the SOEP data, they cannot exploit the full information contained in the BBSR index. Their analysis is thus restricted to comparing East and West Germany. A further closely related paper is Moretti (2009), who finds that accounting for regional price differences decreases the college wage premium in the US. He shows that the prices faced by college graduates have increased relatively faster because their jobs are located in metropolitan areas with high housing costs.