IAB Discussion Paper
Beiträge zum wissenschaftlichen Dialog aus dem Institut für Arbeitsmarkt- und Berufsforschung

No. 15/2007

Expensive and low-price places to live

Regional price levels and the agglomeration wage differential in Western Germany

Uwe Blien, Hermann Gartner, Heiko Stüber, Katja Wolf

Bundesagentur für Arbeit

Expensive and low-price places to live

Regional price levels and the agglomeration wage differential in Western Germany

Uwe Blien (IAB, Technical University of Kaiserslautern, IZA),
Hermann Gartner (IAB),
Heiko Stüber (IAB, Leibniz University Hannover),
Katja Wolf (IAB, University of Erlangen-Nürnberg)

Auch mit seiner neuen Reihe „IAB-Discussion Paper“ will das Forschungsinstitut der Bundesagentur für Arbeit den Dialog mit der externen Wissenschaft intensivieren. Durch die rasche Verbreitung von Forschungsergebnissen über das Internet soll noch vor Drucklegung Kritik angeregt und Qualität gesichert werden.

Also with its new series "IAB Discussion Paper" the research institute of the German Federal Employment Agency wants to intensify dialogue with external science. By the rapid spreading of research results via Internet still before printing criticism shall be stimulated and quality shall be ensured.

Contents

Abstract

Recently, there has been a renewed interest in the agglomeration wage differential. One of the open research question is whether wage differences between large cities and the rural country are due to unobserved differences in regional price levels. In this paper information on regional price levels for western German regions is used to assess this wage differential. Since for many regions price information is not available Multiple Imputation is used to generate completed data sets. It can be shown that this strategy is more reliable than simply using predictions of regression analysis.

The results obtained show that the agglomeration wage differential in Germany is smaller than in the US, and that only a minor part of it can be explained by differences in prices.

JEL classification: R12, J31, C11

Keywords: Agglomeration wage differential, regional prices, Multiple Imputation, regional disparities

The authors would like to thank Elke Amend and Karen Scott-Leuteritz for very valuable advice.

1 Introduction

Information on regional price levels is fundamental for many research questions in regional science. For example, empirical analyses show a marked agglomeration wage premium in many countries. But this premium might be an illusion if it only reflects price differences.

Though the value of information on regional prices is obvious, there is a lack of empirical data in many countries. In some cases official statistics provide information about relatively large regional units, but this is not of much help since large regions include agglomerations and stretches of rural areas as well. Furthermore, price indexes for regions are often not comparable (see Appendix 1); they show the development in the time dimension (regional inflation) but not in the spatial dimension.

In this paper one of the very few surveys showing regional price levels for 32 small regional units is used to estimate regional price differences for the whole area of western Germany. This is done by Multiple Imputation, which is a method that takes care of the uncertainty included in this estimation by dealing correctly with the variance of estimates accomplished using imputed data.

The completed data is used to estimate the agglomeration wage differential, which has received considerable attention recently. Most of the explanations are based on the idea that the concentration of production increases productivity, which is at least partly reflected by wages. However, there is an older debate concerning whether regional wage differences simply balance out observed or unobserved (dis-)amenities of regions (early representatives are Goldfarb, Yezer 1976, Roback 1982), for example disadvantages of congestion in cities, or whether they are due to unobserved price differences. In more recent research the relation of the wage differential to cost-of-living differences is also discussed (Glaeser, Maré 2001: 319f., Yankow 2006).

The outline of the paper is as follows: In section 2 we give a brief overview over the current state of research. Section 3 provides background regarding Multiple Imputation. The application of the imputation procedure is shown in section 4. In section 5 we present the estimation results on the agglomeration wage differential. Section 6 provides some concluding remarks.

2 Current state of research

From many empirical studies it is known that prices and wages are higher in metropolitan areas. It should be noted, however, that the existence of high prices in city centres is contrary to the predictions of most New Economic Geography (NEG) approaches. An interesting theoretical model that avoids the inconsistency is given by Suedekum (2006). He adds a home goods sector to Krugman's (1991) standard approach. It is then possible for a price index to be higher in the core than in the periphery.

As regards empirical research on the agglomeration wage differential, Glaeser and Maré (2001) showed in their seminal article that the raw differential is 33 % in the US. If the heterogeneity of the labour force is controlled for and experience, education and race of workers are taken into account it falls by 6.5 %. The authors (2001: 340) concluded "that there is a substantial urban wage premium, which is not simply the result of omitted ability factors". In the case of Germany, Möller and Haas (2003) presented evidence for the existence of such a premium using quantile regression. In Germany the raw differential is smaller, it comes up to 20 % (see also Möller, Lehmer 2006). Generally, empirical studies show that individuals with identical characteristics receive higher wages in agglomerations than in rural areas.

To see to what extent the wage differential is due to a variation of price levels we will have to extend the scope of the analyses. However, for many countries information on regional price levels is scarce. In the case of the US, cost-of-living information is available for metropolitan areas but not for the rural country. Consequently, Yankow (2006) includes the price level when he studies the wage premium for large compared to smaller cities.

In general, empirical information is scarce because collecting data on prices is a very expensive process. For a span of over thirty-five years only four surveys on regional price levels have been published by the German Federal Statistical Office. Each of them covers only a subset of regions (see Appendix 1). The survey conducted in 1993 for fifty towns and cities of reunified Germany is the most recent one (see Ströhl 1994). The data is representative for goods and services consumed by private households, but does not include rents. This survey could be used to estimate the regional price variation in the whole country. Even though some time has passed since this survey was conducted, it would be valuable to assess the structure and extent of regional price differences. The results could be used to derive information about the situation of more recent years.

In fact, Roos (2006) has done an estimation for regional price levels of all districts in Germany (NUTS-3) on the basis of the data provided by Ströhl (1994). He used a standard regression approach with the population, the GDP per capita, the average annual wage, the rental rate of retail outlets, the population density, a tourism dummy and a dummy for the eastern part of Germany as independent variables (Roos 2006: 1557f.). Applying this model to the fifty regional units for which price information is available, he obtained regression coefficients, which he used in a second step to calculate expected values for all of the 390 remaining districts. Furthermore, he calculated price levels of federal states by population weighted averages. Using the latter and the inflation rates at state level, the price levels of federal states (Bundesländer) are extrapolated up to the year 2002.

One limitation of Roos' (2006) approach concerns the method he used. The imputation of missing values by a standard regression model ignores the fact that the estimated values are known only with some uncertainty. If the estimated values are used in a second step, e. g. for estimating the agglomeration wage differential, the standard formula for calculating the standard errors leads to biased results. Therefore, we use Multiple Imputation to obtain estimates of the missing regional price levels.

3 Multiple Imputation

The Multiple Imputation (MI) approach is standard for missing data problems in statistics and is being used more and more in applied research (see Rubin 1996). Following Little and Rubin (1987) there are three categories of missing data – missing completely at random, missing at random and missing not at random. MI is based on data missing at random (MAR). In this case the probability for a missing value may depend on the observed values but does not depend on the missing values. Schafer (1997) shows that even when the MAR assumption does not hold, estimations by MI are generally less biased than estimations using the complete cases only.

The principles of Multiple Imputation were developed by Rubin (1978). To obtain a complete data set, the missing values are drawn from a conditional random distribution. The correlation of the imputed values with the covariates is based on the correlation between the observed values. The completed data can be used for econometric analyses with standard techniques. But standard tests assume that the true data is used – whereas we are actually using imputed data. This causes additional uncertainty. By running the imputation procedure several times we can take this additional uncertainty into account.

Therefore Multiple Imputation consists of three steps: in the first one, we generate m imputed data sets. In the second step, in following our research question we run estimations separately for each of the m completed data sets using standard econometric methods. In the third one, we combine the results by the rules for MI estimates.

The origin of Multiple Imputation lies in Bayesian Statistics. To obtain proper MI results, the imputations must be drawn randomly from a posterior predictive distribution of the missing data given the observed data. This can be done by a Bayesian Markov-Chain-Monte-Carlo (MCMC) procedure. For variables that are jointly normally distributed Josef Schafer's program NORM can be used (see http://www.stat.psu.edu/~jls/misoftwa.html). Details on the algorithm are discussed in Schafer (1997).

To start a formal description, we need some definitions: the vector of the regional price levels Y (in logs) includes observed values Y_{obs} and missing values Y_{mis}, so that $Y = (Y_{obs}, Y_{mis})$. The matrix X contains further completely observed variables. We assume that the log of the regional price levels $Y = (Y_{obs}, Y_{mis})$ can be described by a normal distribution with the given parameters ξ and covariates X. In contrast to the parameters of an econometric model, the parameters ξ of the imputation model are of no further interest for the researcher. They are only needed for the imputation.

From a Bayesian point of view the parameters ξ are random variables. To impute the missing price levels given the observed levels and the covariates X, we need the distribution

$$P(Y_{mis},\xi \mid Y_{obs},X).$$

This distribution can be obtained by a Markov chain with an imputation step and a probability step. The imputation step is:

$$Y_{mis}^{(t+1)} \sim P(Y_{mis} \mid Y_{obs},X,\xi^{(t)}).$$

In this step the missing values are drawn from a normal distribution, given the observed values and $\xi^{(t)}$. The starting value $\xi^{(0)}$ is adopted from an expectation-maximizing algorithm. The probability step simulates new parameters ξ, given the observed values and the imputed values:

$$\xi^{(t+1)} \sim P(\xi \mid Y_{mis}^{(t+1)},Y_{obs},X),$$

where ξ is drawn from an inverted Wishard distribution. These new values of ξ are used in the next imputation step where we draw new Y_{mis}. The iteration between the imputation and probability steps results in a Markov chain. The distribution of the Markov chain $(Y_{mis}^{(t)},\xi^{(t)})$ converges to the posterior distribution $P(Y_{mis},\xi \mid Y_{obs},X)$.

We store Y_{mis} after $z>>1$ rounds to obtain the first imputed data set. In our case we run $z=1000$ rounds. After z further iterations, we store the next data set, and so on, until we have gained $m>1$ imputed data sets. This completes the first step of our analysis, the imputation of missing data and the generation of more than one completed data sets.

This enables us in the second step to calculate the imputed data estimate $\hat{\theta}^{(n)}=\hat{\theta}(Y^{(n)},X)$ and its estimated variance $\hat{var}(\hat{\theta}^{(n)})$, $n=1,2,\ldots,m$. $\hat{\theta}^{(n)}$ might be an estimator for the mean of a variable or the estimated coefficient from an econometric model in our case a wage analysis to check the existence of the agglomeration differential. The econometric analysis is carried out m times, once for each imputed data set.

In the third step these m estimates are used to compute the MI estimates and the MI variance. The MI estimate for θ is the average estimate

$$\hat{\theta}_{MI} = \frac{1}{m}\sum_{n=1}^{m}\hat{\theta}^{(n)} .$$

To obtain the variance for the MI estimate $\hat{\theta}_{MI}$, we first calculate the 'between-imputation' variance

$$B = \frac{1}{m-1}\sum_{n=1}^{m}\left(\hat{\theta}^{(n)} - \hat{\theta}_{MI}\right)^2 ,$$

and then the 'within-imputation' variance

$$W = \frac{1}{m}\sum_{n=1}^{m}\hat{var}\left(\hat{\theta}^{(n)}\right).$$

Finally, the estimated total variance T for the MI estimate is defined by

$$T = W + \frac{m+1}{m}B .$$

For large sample sizes, tests and two-sided $(1-\alpha)$-interval estimates can be based on Student's t-distribution

$$\frac{\hat{\theta}_{MI} - \theta}{\sqrt{T}} \sim t_v \text{ and } \hat{\theta}_{MI} \pm t_{v,1-\alpha/2}\sqrt{T}$$

with the degrees of freedom

$$v = (m-1)\left(1 + \frac{W}{(1+m^{-1})B}\right)^2 .$$

Note that the degrees of freedom depend on the number of imputations and the ratio of W to B. It can be shown that 3 to 5 imputed data sets are already sufficient if the share of missing values is not too high. More imputations would scarcely improve the efficiency of an estimate.

4 Price imputation

The observed price levels Y_{obs} (included in logs) for a number of regions stem from a publication by Ströhl (1994). He reports results of a cross-section sample collected by the German Federal Statistical Office. The data contains price indexes for 32 western German towns and cities in 1993. The indexes are based on a representative basket of 367 goods and services. In total, nearly 44,000 prices were used. The price information is not the regional cost-of-living index for two reasons. First, housing prices are not included. Second, in constructing the index, national expenditure shares have been used as weights, so local expenditure patterns are not represented.

Table 1: Region classification and average imputed price levels

According to Görmar, Irmen (1991), price information following Ströhl (1994)

Region Type	Description	No. of regions … in West Germany	No. of regions … with price index information	Average imputed price levels (Reg.Type 1 = 100)
1	Core cities in regions with major agglomerations.	39	13	100
2	Very densely populated districts in regions with major agglomerations.	41	2	99.72
3	Densely populated districts in regions with major agglomerations.	29	0	99.71
4	Rurally structured districts in regions with major agglomerations.	10	0	99.83
5	Core cities in regions with conurbational features.	21	7	99.70
6	Densely populated districts in regions with conurbational features.	72	5	99.18
7	Rurally structured districts in regions with conurbational features.	48	1	98.76
8	Densely populated districts in rurally structured regions.	44	3	98.39
9	Rurally structured districts in rurally structured regions.	22	1	95.46
Total		326	32	

Of special interest is the part of the price differential that arises from agglomeration effects, since this should correspond with the wage differential analysed in the next step of our study. Compared to the US, for which the seminal studies (e. g. Glaeser, Maré 2001) were published, Germany is rather small and densely populated. No part of the country is really far away from an agglomeration and therefore the differences in economic conditions should not be as pronounced as they are in the US. To capture different degrees of agglomeration between central core cities and the rural country we introduce a widely used classification developed by the Federal Office for Building and Regional Planning (BBR). This typology for districts, i. e. NUTS-3 regions, uses centrality and population density as criteria as is shown in Table 1. In a strict sense, only the regions of Type 1 represent agglomerations and only those of Type 9 the rural country. The other types are in between.

The matrix of covariates X, which is used to simulate the missing price levels contains: size of regions, price of building land, squared price of building land, a variable for tourism, unemployment rate (all in logs), composition corrected regional wage, a dummy for cities which are a district of their own and nine region dummy variables indicating the type of district (Table 1). Because of a lack of observations in two regional groups, region Type 4 is subsumed to 7 and Type 3 to 6 only in the MI step. Detailed information on the other covariates is given in Appendix 2.

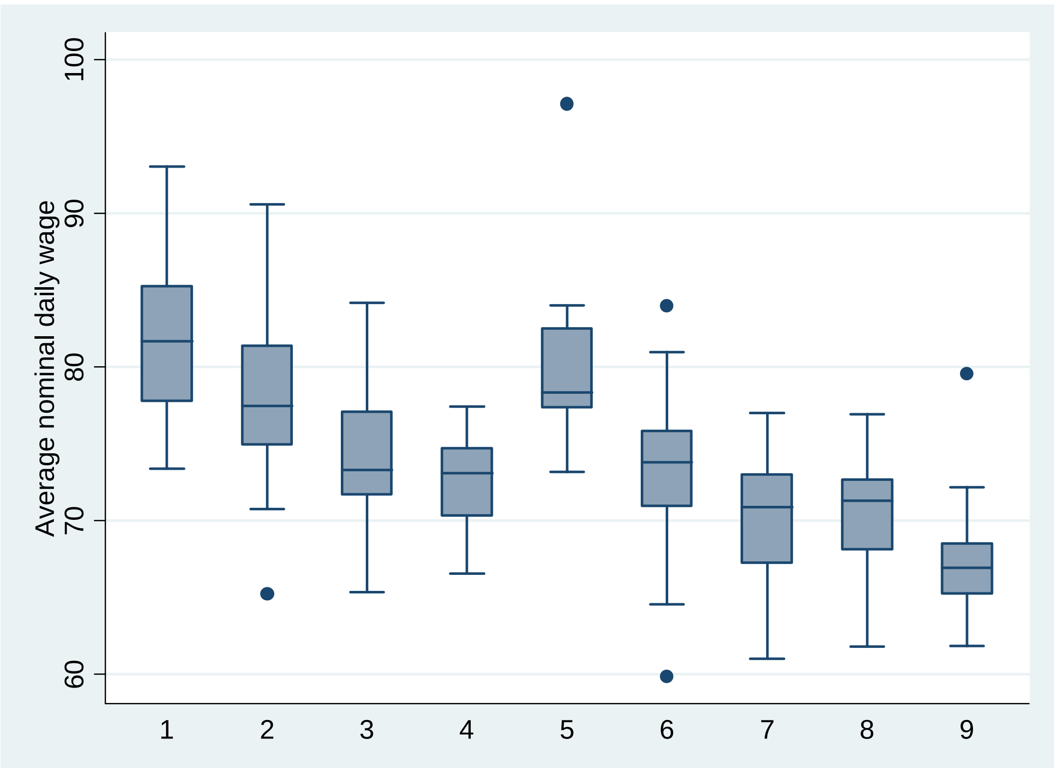
Following the method outlined in the previous section, these variables are used to generate regional levels for those districts where the information is missing. In our case we generated $m = 5$ data sets. This completes the first step of our analysis.

To obtain a rough indicator for the quality of the information used for imputation a regression is carried out for the 32 regions with complete information. The price level is the response variable and the variables employed for imputation are included as predictors. The regression gives an R^2 of 0.90. Therefore, we feel confident about the precision of the imputed data.

5 Estimation of agglomeration wage differentials

Before continuing with the wage regressions, we provide in Figure 1 a first impression of the regional disparities in nominal wages. As expected, the average wages are lower in rural areas (Types 8 and 9) than in core cities (Types 1 and 5). The data set underlying this figure and also the wage equations is the IAB Employment Sample (IABS-R01, for details see Hamann et al. 2004 and for an older version of the data Bender, Haas, Klose 2000). This is a two percent random sample of the Employment Statistics of the Federal Employment Services, which includes all employees covered by social security and represents about 80 percent of total employment in Germany. The Employment Statistics are generated from the administrative data used to calculate the contributions to the social security system. It is therefore highly reliable. The data includes information on employment spells, job and personal characteristics as well as earnings. The wage variable is measured for calendar days. Only workers, who were employed on 30th June, 1993 are included. Part-time workers and apprentices are not taken into account. The remaining sample contains 171,366 individuals covering the districts of western Germany.

Figure 1: Boxplots of average nominal daily wages over region types according to Table 1



Source: Own calculations based on the IABS-R01

In the next step, the regressions with the 5 data sets completed by MI are carried out. To test for the existence of an agglomeration wage premium, we estimate equations with the log of wages as response variable. For the first one we use individual nominal wages, for the second one we calculate real wages using the imputed regional price levels. In order to quantify the agglomeration wage differentials we use again the classification scheme of the BBR (see Section 3). Because contributions to the social security system have to be paid only up to a contribution limit, the wages are right-censored. For this reason we estimate the wages by means of a Tobit regression. Because of the multilevel framework – employees are located within regions – we adjust the standard errors of each single regression for intragroup correlation (Moulton 1990).

Table 2 contains the results of the Tobit estimates for the regression model where only the region types are used as explanatory variables to quantify the raw urban wage premium. The reference category is Type 9 "Rurally structured districts in rurally structured regions". The estimations from the five imputed data sets are combined according to the principles of Multiple Imputation to obtain the results for the real wage model.

The nominal wages at individual level differ considerably between the region types. Employees working in rurally structured districts of Type 9 earn about 25 % less than those working in a core city of Type 1. This is less than the 33 % found by Glaeser and Maré (2001) for the US. The wage differentials decrease somewhat when using the real wage instead of the nominal wage. The value for Type 1 is now 19 %. All coefficients decrease in absolute values. Thus, about one quarter of the raw urban wage premium can be explained by regional price differences.

The results of Table 2 do not account for internal heterogeneity of regions. To control for variations in the distribution of jobs and employees, we estimate a Mincer wage equation at individual level. As is standard in the literature we assume a linear relationship between log wages and the explanatory variables. To obtain results comparable to those in the literature (say Glaeser, Maré 2001) we include education, gender, age and age squared.

Table 1: Agglomeration effects on nominal and real wages (Model 1)

Variable	log Nominal Wage				log Real Wage			
	Coef.	Std. Err.	95 % Conf. Interval		Coef.	Std. Err.	95 % Conf. Interval	
Constant	4.152	0.013	4.126	4.177	4.200	0.014	4.172	4.229
Region Type 1	0.250	0.019	0.214	0.286	0.194	0.017	0.160	0.228
Region Type 2	0.181	0.018	0.146	0.217	0.136	0.018	0.101	0.171
Region Type 3	0.107	0.017	0.074	0.139	0.065	0.018	0.029	0.102
Region Type 4	0.068	0.021	0.028	0.109	0.026	0.022	-0.018	0.070
Region Type 5	0.186	0.021	0.145	0.227	0.142	0.022	0.099	0.184
Region Type 6	0.099	0.014	0.071	0.127	0.060	0.016	0.028	0.092
Region Type 7	0.048	0.016	0.017	0.079	0.014	0.017	-0.021	0.049
Region Type 8	0.051	0.016	0.020	0.082	0.023	0.019	-0.017	0.062

Notes: Estimation method is Tobit; reference category is Region Type 9; all standard errors are robust; n = 171,366.

Table 2: Agglomeration effects on nominal and real wages with control variables (Model 2)

Variable	log Nominal Wage				log Real Wage			
	Coef.	Std. Err.	95 % Conf. Interval		Coef.	Std. Err.	95 % Conf. Interval	
Constant	3.240	0.026	3.190	3.290	3.286	0.026	3.235	3.337
Region Type 1	0.196	0.016	0.164	0.227	0.140	0.015	0.111	0.169
Region Type 2	0.146	0.015	0.117	0.175	0.101	0.015	0.072	0.130
Region Type 3	0.084	0.015	0.056	0.113	0.043	0.016	0.011	0.076
Region Type 4	0.049	0.017	0.015	0.083	0.007	0.019	-0.031	0.045
Region Type 5	0.140	0.018	0.105	0.176	0.096	0.019	0.059	0.133
Region Type 6	0.079	0.012	0.055	0.102	0.040	0.014	0.011	0.068
Region Type 7	0.039	0.014	0.011	0.066	0.005	0.016	-0.027	0.036
Region Type 8	0.037	0.013	0.012	0.062	0.009	0.017	-0.027	0.044
Vocational Qualification	0.205	0.004	0.198	0.212	0.205	0.004	0.198	0.212
University Degree	0.660	0.007	0.646	0.674	0.657	0.007	0.643	0.671
Women	-0.315	0.007	-0.329	-0.302	-0.316	0.007	-0.330	-0.302
Age	0.039	0.001	0.037	0.041	0.039	0.001	0.037	0.041
Age squared	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: Estimation method is Tobit; reference variables are Region Type 9, No Formal Education and Men; all standard errors are robust; n = 171,366.

By and large the estimated coefficients of the control variables show the magnitude and sign which is expected (see Table 3). For instance, women earn ceteris paribus less than men, highly qualified workers more than less qualified ones and age shows a positive but decreasing marginal effect. Concerning the region types, the estimated agglomeration wage differentials are smaller than in the model with no additional explanatory variables. The difference between region Types 1 and 9 is reduced to 20 percent in nominal terms. This is about the same result as in Möller, Haas (2003). In real terms the coefficients are reduced by about the same amount as before in Table 2. Therefore, the differential for Type 1 shrinks to 14 %.

Our data set provides us with the opportunity to control for even more variables. These include nine establishment-size categories, 30 industry categories and thirteen occupational categories. See Appendix 2 for a detailed description of the variables. These variables map attributes of employees (occupational category) and of establishments (industry category and establishment size). Glaeser, Maré (2001) and Yankow (2006) use models with fixed effects to control for unobserved worker heterogeneity within regions and to trace the origin of the agglomeration wage differential even further. This approach is not feasible for us, because our price level data is available for only one year. Models with an extended number of control variables, however, allow additional insights into the sources of the agglomeration wage differential. Results are shown in Table 4, but the coefficients for the additional variables are omitted for lack of space. The complete results can be obtained from the authors upon request.

Now we obtain the surprising result that the agglomeration wage premium is only half of its previous size when nominal wages are chosen, and it is even smaller with real wages. The interpretation is simple: A large part of the agglomeration wage differential is due to the structure of jobs of regional economies. Better paying industries with larger establishments are concentrated in metropolitan areas.

Table 4: Agglomeration effects on nominal and real wages with extended number of control variables (Model 3)

Variable	log Nominal Wage				log Real Wage			
	Coef.	Std. Err.	95 % Conf. Interval		Coef.	Std. Err.	95 % Conf. Interval	
Constant	3.044	0.026	2.993	3.095	3.090	0.027	3.038	3.143
Region Type 1	0.093	0.010	0.072	0.113	0.039	0.011	0.017	0.060
Region Type 2	0.090	0.010	0.070	0.110	0.045	0.012	0.021	0.068
Region Type 3	0.056	0.011	0.034	0.077	0.015	0.014	-0.014	0.043
Region Type 4	0.038	0.012	0.014	0.062	-0.004	0.015	-0.035	0.027
Region Type 5	0.052	0.011	0.030	0.073	0.009	0.013	-0.018	0.035
Region Type 6	0.041	0.009	0.025	0.058	0.003	0.012	-0.022	0.028
Region Type 7	0.019	0.010	-0.001	0.039	-0.015	0.013	-0.042	0.012
Region Type 8	0.015	0.010	-0.004	0.034	-0.013	0.015	-0.047	0.020
Vocational Qualification	0.106	0.002	0.102	0.110	0.107	0.002	0.103	0.112
University Degree	0.300	0.006	0.288	0.312	0.299	0.006	0.287	0.311
Women	-0.317	0.006	-0.329	-0.305	-0.318	0.006	-0.330	-0.305
Age	0.030	0.001	0.029	0.032	0.031	0.001	0.029	0.032
Age squared	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: Estimation method is Tobit; reference variables are Region Type 9, No Formal Education and Men; all standard errors are robust; n = 171,366.
Variables for establishment sizes, industries and occupations are included but results not shown.

The differences between coefficients of the regression in real and in nominal terms are in most cases significant as they have been in the previous regressions. This is an important result since the analysis with real wages now leads to different conclusions. In models with fewer variables, the core cities of Region Type 5 are similar to the core cities of Region Type 1. Now, with real wages, the coefficient of Type 5 vanishes. The agglomeration differential revealed for this type can be completely explained by the composition of the work force. The differential for Type 1 is less than half of what it was in the analysis with nominal wages.

Finally, if only the control variables are included which are related to establishments (industries and size groups), the Type 1 dummy has coefficients of 0.160 (nominal) and 0.105 (real). If only variables relating to employees are included (Model 2 plus occupation groups) the coefficients are 0.141 and 0.086.

Our analyses show the important role of regional prices in explaining and understanding the agglomeration wage premium. In a descriptive dimension, regional prices influence the structure and magnitude of regional disparities in an economy. In an analytical dimension, regional prices are relevant for empirical tests of economic theories, because they are generally formulated in real and not in nominal terms. This is true not only for theories of regional science but for many other approaches. An empirical analysis of labour demand, for example, is affected to the extent that differences in wage levels are due to variations in regional price levels for which no information is available. As another example, a region's response to an external shock cannot be measured adequately if the price information is missing. Therefore, approaches describing regional business cycles have no direct empirical counterpart.

Though, obviously, regional prices are important, little has been done to supply the data required and to fill the gaps in empirical research (Hayes 2005). The analysis presented here is one step in this direction.

6 Conclusions

In this paper it has been demonstrated that Multiple Imputation is an important tool for generating completed data sets. One of the special properties of this approach is that it takes into account the additional uncertainty associated with the fact that some of the data are not observed. The estimation of a regional price index for western Germany has interesting consequences for the structure of regional wage disparities. It is obvious that the larger cities are places which are relatively expensive to live in. The rural country is cheaper. That is an expected result. It is important, however, that the "raw" wage differential does not vanish, it is "real". Without additional controls in the regression the wage differential is 25 % with nominal and 19 % with real wages.

When the composition of the labour force and the distribution and quality of jobs is taken into account, the wage differential is greatly reduced. In nominal terms it is now only 9 %. In real terms it is reduced to about 4 %.

Literature

Angermann, O. (1989): Vergleich des Verbraucherpreisniveaus in Bonn, Karlsruhe, München und Berlin (West); in: Wirtschaft und Statistik 4: 258-261.

Baltagi, B.H.; Blien, U.; Wolf, K. (2007): Phillips Curve or Wage Curve? Evidence from West Germany: 1980-2004. IAB Discussion Paper No. 14/2007. Nuremberg.

Bender, S.; Haas, A.; Klose, C. (2000): The IAB Employment Subsample 1975-1995; in: Schmollers Jahrbuch 120/4: 649-662.

Glaeser, E. L.; Maré, D. C. (2001): Cities and skills; in: Journal of Labor Economics 19/1: 316-342.

Goldfarb, R. S.; Yezer, A. M. J. (1976): Evaluating alternative theories of intercity and interregional wage differentials; in: Journal of Regional Science 16/3: 345-363.

Görmar, W.; Irmen, E. (1991): Nichtadministrative Gebietsgliederungen und -kategorien für die Regionalstatistik. Die siedlungsstrukturelle Gebietstypisierung der BfLR; in: Raumforschung und Raumordnung 49/6: 387-394.

Hamann, S. et al. (2004): Die IAB-Regionalstichprobe 1975-2001: IABS-R01; in: ZA-Information 55: 34-59.

Hayes, P. (2005): Estimating UK Regional Price Indices, 1974-96; in: Regional Studies 39/3: 333-344.

Krugman, P. (1991): Geography and Trade, Cambridge (Mass.) etc.: MIT Press.

Little, R.J.A.; Rubin, D.B. (1987): Statistical analysis with missing data, New York: John Wiley.

Möller, J.; Lehmer, F. (2006): The Rural/ Urban Wage Premium, Backwater- and Firm-Size Effects. Working Paper. University of Regensburg.

Möller, J.; Haas, A. (2003): The Agglomeration Wage Differential Reconsidered: An Investigation Using German Micro Data 1984-1997; in: Bröcker, J. et al. (ed.), Innovation Clusters and Interregional Competition, Berlin etc.: Springer: 182-217.

Moulton, B.R. (1990): An illustration of a pitfall in estimating the effects of aggregate variables on micro units; in: The Review of Economics and Statistics 72: 334-338.

Roback, J. (1982): Wages, Rents, and the Quality of Life; in: Journal of Political Economy 90/6: 1257-1277.

Roos, M. W. M. (2006): Regional price levels in Germany; in: Applied Economics 38/13: 1553-1566.

Rostin, W. (1979): Zwischenörtlicher Vergleich des Verbraucherpreisniveaus in 31 Städten; in: Wirtschaft und Statistik 6: 403-410.

Rubin, D. B. (1978): Multiple imputation in sample surveys – a phenomenological Bayesian approach to nonresponse; in: Proceedings of the Survey Resarch Methods Sections of the American Statistica: 20-40.

Rubin, D. B. (1996): Multiple imputation after 18+ years; in: Journal of the American Statistical Association 91: 473-489.

Schafer, J. L. (1997): Analysis of Incomplete Multivariate Data, Chapman & Hall, London.

Ströhl, G. (1994): Zwischenörtlicher Vergleich des Verbraucherpreisniveaus in 50 Städten; in: Wirtschaft und Statistik 6: 415-434.

Suedekum, J. (2006): Agglomeration and Regional Costs of Living; in: Journal of Regional Science 46/3: 529-543.

Yankow, J. J. (2006): Why do Cities Pay more? An Empirical Examination of some Competing Theories of the Urban Wage Premium; in: Journal of Urban Economics 60/2: 139-161.

Appendix 1: Regional price information

Surveys for regional price indexes carried out in Germany:

- 1971 for ten West German cities: The survey is mentioned by Rostin (1979: 403, 405). Neither the relevant division of the German Federal Statistical Office nor the European Office of Statistics (EUROSTAT) could provide the original source.
- 1978 for 31 West German cities (Rostin 1979),
- 1987 for the four German cities Bonn, Karlsruhe, Munich and Berlin (West) (Angermann 1989) and
- 1993 for 50 cities of reunified Germany (Ströhl 1994).

The Federal Statistical Office publishes inflation rates for the federal states (Bundesländer) of Germany. However, there is a lack of information in the cross-section. The price levels of federal states cannot be compared with each other since it could not be ruled out that commodities of different quality and therefore cost are used for the calculation of regional price indexes between states.

Appendix 2: Data description

The main data source is the IABS-R01, i. e. the regionalized version of the IAB Employment Sample of which a 50 % random subsample was used = a 1 % sample of the total employment covered by social security. Only full time employment in western Germany is included. The variables are:

- *Wage:* Gross calendar daily wage of the employee**.**
- *Age:* Age of the employee**.**
- *Establishment size:* Size of an establishment measured by the number of employees. This includes 9 categories. 1-4 employees, 5-9 employees, 10-49 employees, 50-99 employees, 100-199 employees, 200-499 employees, 500-999 employees, 1000-4999 employees, and more than 4999.
- *Composition corrected regional wage:* This variable is the estimated regional fixed effect of a Mincer-type equation for western Germany

in 1993. It gives information about the deviance of regional wages from their expected values. For further details see Baltagi, Blien and Wolf (2007).

- *Industry classification:* This variable defines the specific industry to which the employing establishment belongs. These include 30 categories: primary sector, energy & mining, chemistry, plastic products, stones & earth, glass products, quarrying & metals, metal construction, motor vehicles, computers & electronic equipment, jewellery & toys, wood, paper, textiles, food products, construction, trade, transport & telecommunication, banking & insurance, hotels & catering, health care, simple business-related services, qualified business-related services, security services, temporary help services, education, leisure-related services, household-related services, other social services, public administration, and not applicable.

- *Occupational group:* This variable describes the field of an employee's occupational specialization. These include 13 categories: agricultural, unskilled blue collar worker, skilled blue collar worker, technician, engineer, simple services, qualified services, semi-professional, professional, simple administrative, qualified administrative, managers and no classification applicable.

Other variables:

- *Cities and towns which are a district of their own:* 1 = cities that have district status, otherwise 0.

- *Population:* Population of the districts.

- *Price of building land:* Price per square meter of building land.

- *Price index for 1993:* Price index for German regions provided in Ströhl (1994).

- *Qualification level of an employee:* This variable includes 4 categories: No formal education, vocational qualification, university degree, no classification applicable. The qualification level "no classification applicable" is subsumed to "no formal education".

- *Regional unemployment:* The districts are the smallest regions for which unemployment figures can be obtained.
- *Region types:* The nine dummies indicate the type of the districts defined according to a classification of the Federal Office for Building and Regional Planning (BBR; Görmar, Irmen 1991), see Table 1.
- *Gender:* Female = 1, male = 0.
- *Size of the regions:* Size of the regions in square kilometres.
- *Tourism:* Tourism is composed of the number of overnight stays divided by the number of arrivals of visitors multiplied by the population density of the region.

The size of the regions, the population of the regions, the prices of building land and tourism are provided by the BBR. Some missing data was supplemented by the statistical offices of the states Brandenburg, Berlin, Lower Saxony, Baden-Württemberg, the towns Friedrichshafen and Nordhorn and the chamber of commerce and industry Osnabrück-Emsland. The unemployment rates of the districts are supplied by the Federal Employment Services.

Recently published

No.	Author(s)	Title	Date
1/2004	Bauer, T. K. Bender, S. Bonin, H.	Dismissal protection and worker flows in small establishments published in: Economica	7/04
2/2004	Achatz, J. Gartner, H. Glück, T.	Bonus oder Bias? : Mechanismen geschlechtsspezifischer Entlohnung published in: Kölner Zeitschrift für Soziologie und Sozialpsychologie 57 (2005), S. 466-493 (revised)	7/04
3/2004	Andrews, M. Schank, T. Upward, R.	Practical estimation methods for linked employer-employee data	8/04
4/2004	Brixy, U. Kohaut, S. Schnabel, C.	Do newly founded firms pay lower wages? : first evidence from Germany published in: Small Business Economics, (2007)	9/04
5/2004	Kölling, A. Rässler, S.	Editing and multiply imputing German establishment panel data to estimate stochastic production frontier models published in: Zeitschrift für ArbeitsmarktForschung 37 (2004), S. 306-318	10/04
6/2004	Stephan, G. Gerlach, K.	Collective contracts, wages and wage dispersion in a multi-level model published as: Wage settlements and wage setting : results from a multi-level model. In: Applied Economics, Vol. 37, No. 20 (2005), S. 2297-2306	10/04
7/2004	Gartner, H. Stephan, G.	How collective contracts and works councils reduce the gender wage gap	12/04
1/2005	Blien, U. Suedekum, J.	Local economic structure and industry development in Germany, 1993-2001	1/05
2/2005	Brixy, U. Kohaut, S. Schnabel, C.	How fast do newly founded firms mature? : empirical analyses on job quality in start-ups published in: Michael Fritsch, Jürgen Schmude (Ed.): Entrepreneurship in the region, New York et al., 2006, S. 95-112	1/05
3/2005	Lechner, M. Miquel, R. Wunsch, C.	Long-run effects of public sector sponsored training in West Germany	1/05
4/2005	Hinz, T. Gartner, H.	Lohnunterschiede zwischen Frauen und Männern in Branchen, Berufen und Betrieben published in: Zeitschrift für Soziologie 34 (2005), S. 22-39, as: Geschlechtsspezifische Lohnunterschiede in Branchen, Berufen und Betrieben	2/05
5/2005	Gartner, H. Rässler, S.	Analyzing the changing gender wage gap based on multiply imputed right censored wages	2/05
6/2005	Alda, H. Bender, S. Gartner, H.	The linked employer-employee dataset of the IAB (LIAB) published as: The linked employer-employee dataset created from the IAB establishment panel and the process-produced data of the IAB (LIAB). In: Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften 125 (2005), S. 327-336 (shortened)	3/05
7/2005	Haas, A. Rothe, T.	Labour market dynamics from a regional perspective : the multi-account system	4/05
8/2005	Caliendo, M. Hujer, R.	Identifying effect heterogeneity to improve the efficiency of job creation schemes in Germany	4/05

	Thomsen, S. L.		
9/2005	Gerlach, K. Stephan, G.	Wage distributions by wage-setting regime published as: Bargaining regimes and wage dispersion. In: Jahrbücher für Nationalökonomie und Statistik, Bd. 226, H. 6 (2006)	4/05
10/2005	Gerlach, K. Stephan, G.	Individual tenure and collective contracts	4/05
11/2005	Blien, U. Hirschenauer, F.	Formula allocation : the regional allocation of budgetary funds for measures of active labour market policy in Germany published in: Economics Bulletin, Vol. 18, no. 7 (2006)	4/05
12/2005	Alda, H. Allaart, P. Bellmann, L.	Churning and institutions : Dutch and German establishments compared with micro-level data	5/05
13/2005	Caliendo, M. Hujer, R. Thomsen, S. L.	Individual employment effects of job creation schemes in Germany with respect to sectoral heterogeneity	5/05
14/2005	Lechner, M. Miquel, R. Wunsch, C.	The curse and blessing of training the unemployed in a changing economy : the case of East Germany after unification	6/05
15/2005	Jensen, U. Rässler, S.	Where have all the data gone? : stochastic production frontiers with multiply imputed German establishment data published in: Zeitschrift für ArbeitsmarktForschung, Jg. 39, H. 2, 2006, S. 277-295	7/05
16/2005	Schnabel, C. Zagelmeyer, S. Kohaut, S.	Collective bargaining structure and its determinants : an empirical analysis with British and German establishment data published in: European Journal of Industrial Relations, Vol. 12, No. 2. S. 165-188	8/05
17/2005	Koch, S. Stephan, G. Walwei, U.	Workfare: Möglichkeiten und Grenzen published in: Zeitschrift für ArbeitsmarktForschung 38 (2005), S. 419-440	8/05
18/2005	Alda, H. Bellmann, L. Gartner, H.	Wage structure and labour mobility in the West German private sector 1993-2000	8/05
19/2005	Eichhorst, W. Konle-Seidl, R.	The interaction of labor market regulation and labor market policies in welfare state reform	9/05
20/2005	Gerlach, K. Stephan, G.	Tarifverträge und betriebliche Entlohnungsstrukturen published in: C. Clemens, M. Heinemann & S. Soretz (Hg.): Auf allen Märkten zu Hause, Marburg 2006, S. 123-143	11/05
21/2005	Fitzenberger, B. Speckesser, S.	Employment effects of the provision of specific professional skills and techniques in Germany	11/05
22/2005	Ludsteck, J. Jacobebbinghaus, P.	Strike activity and centralisation in wage setting	12/05
1/2006	Gerlach, K. Levine, D. Stephan, G. Struck, O.	The acceptability of layoffs and pay cuts : comparing North America with Germany	1/06
2/2006	Ludsteck, J.	Employment effects of centralization in wage setting in a median voter model	2/06
3/2006	Gaggermeier, C.	Pension and children : Pareto improvement with heterogeneous preferences	2/06
4/2006	Binder, J. Schwengler, B.	Korrekturverfahren zur Berechnung der Einkommen über der Beitragsbemessungsgrenze	3/06
5/2006	Brixy, U. Grotz, R.	Regional patterns and determinants of new firm formation and survival in western Germany	4/06
6/2006	Blien, U.	Structural change and regional employment dynamics	4/06

	Sanner, H.		
7/2006	Stephan, G. Rässler, S. Schewe, T.	Wirkungsanalyse in der Bundesagentur für Arbeit : Konzeption, Datenbasis und ausgewählte Befunde published as: Das TrEffeR-Projekt der Bundesagentur für Arbeit : die Wirkung von Maßnahmen aktiver Arbeitsmarktpolitik. In: Zeitschrift für ArbeitsmarktForschung, Jg. 39, H. 3/4 (2006)	4/06
8/2006	Gash, V. Mertens, A. Romeu Gordo, L.	Are fixed-term jobs bad for your health? : a comparison of West-Germany and Spain published in: European Societies, 2007	5/06
9/2006	Romeu Gordo, L.	Compression of morbidity and the labor supply of older people	5/06
10/2006	Jahn, E. J. Wagner, T.	Base period, qualifying period and the equilibrium rate of unemployment	6/06
11/2006	Jensen, U. Gartner, H. Rässler, S.	Measuring overeducation with earnings frontiers and multiply imputed censored income data	6/06
12/2006	Meyer, B. Lutz, C. Schnur, P. Zika, G.	National economic policy simulations with global interdependencies : a sensitivity analysis for Germany	7/06
13/2006	Beblo, M. Bender, S. Wolf, E.	The wage effects of entering motherhood : a within-firm matching approach	8/06
14/2006	Niebuhr, A.	Migration and innovation : does cultural diversity matter for regional R&D activity?	8/06
15/2006	Kiesl, H. Rässler, S.	How valid can data fusion be? published in: Journal of Official Statistics, (2006)	8/06
16/2006	Hujer, R. Zeiss, C.	The effects of job creation schemes on the unemployment duration in East Germany	8/06
17/2006	Fitzenberger, B. Osikominu, A. Völter, R.	Get training or wait? : long-run employment effects of training programs for the unemployed in West Germany	9/06
18/2006	Antoni, M. Jahn, E. J.	Do changes in regulation affect employment duration in temporary work agencies?	9/06
19/2006	Fuchs, J. Söhnlein, D.	Effekte alternativer Annahmen auf die prognostizierte Erwerbsbevölkerung	10/06
20/2006	Lechner, M. Wunsch, C.	Active labour market policy in East Germany : waiting for the economy to take off	11/06
21/2006	Kruppe, T.	Die Förderung beruflicher Weiterbildung : eine mikroökonometrische Evaluation der Ergänzung durch das ESF-BA-Programm	11/06
22/2006	Feil, M. Klinger, S. Zika, G.	Sozialabgaben und Beschäftigung : Simulationen mit drei makroökonomischen Modellen	11/06
23/2006	Blien, U. Phan, t. H. V.	A pilot study on the Vietnamese labour market and its social and economic context	11/06
24/2006	Lutz, R.	Was spricht eigentlich gegen eine private Arbeitslosenversicherung?	11/06
25/2006	Jirjahn, U. Pfeifer, C. Tsertsvadze, G.	Mikroökonomische Beschäftigungseffekte des Hamburger Modells zur Beschäftigungsförderung	11/06
26/2006	Rudolph, H.	Indikator gesteuerte Verteilung von Eingliederungsmitteln im SGB II : Erfolgs- und Effizienzkriterien als Leistungsanreiz?	12/06
27/2006	Wolff, J.	How does experience and job mobility determine wage gain	12/06

		in a transition and a non-transition economy? : the case of east and west Germany	
28/2006	Blien, U. Kirchhof, K. Ludewig, O.	Agglomeration effects on labour demand	12/06
29/2006	Blien, U. Hirschenauer, F. Phan, t. H. V.	Model-based classification of regional labour markets : for purposes of labour market policy	12/06
30/2006	Krug, G.	Kombilohn und Reziprozität in Beschäftigungsverhältnissen : eine Analyse im Rahmen des Matching-Ansatzes	12/06
1/2007	Moritz, M. Gröger, M.	The German-Czech border region after the fall of the Iron Curtain: Effects on the labour market : an empirical study using the IAB Employment Sample (IABS)	1/07
2/2007	Hampel, K. Kunz, M. Schanne, N. Wapler, R. Weyh, A.	Regional employment forecasts with spatial interdependencies	1/07
3/2007	Eckey, H.- F. Schwengler, B. Türck, M.	Vergleich von deutschen Arbeitsmarktregionen	1/07
4/2007	Kristen, C. Granato, N.	The educational attainment of the second generation in Germany : social origins and ethnic inequality	1/07
5/2007	Jacob, M. Kleinert, C.	Does unemployment help or hinder becoming independent? : the role of employment status for leaving the parental home	1/07
6/2007	Konle-Seidl, R. Eichhorst, W. Grienberger-Zingerle, M.	Activation policies in Germany : from status protection to basic income support	1/07
7/2007	Lechner, M. Wunsch, C.	Are training programs more effective when unemployment is high?	2/07
8/2007	Hohendanner, C.	Verdrängen Ein-Euro-Jobs sozialversicherungspflichtige Beschäftigung in den Betrieben?	2/07
9/2007	Seibert, H.	Frühe Flexibilisierung? : regionale Mobilität nach der Lehrausbildung in Deutschland zwischen 1977 und 2004	2/07
10/2007	Bernhard, S. Kurz, K.	Familie und Arbeitsmarkt : eine Längsschnittstudie zum Einfluss beruflicher Unsicherheiten auf die Familienerweiterung	2/07
11/2007	Drechsler, J. Dundler, A. Bender, S. Rässler, S. Zwick, T.	A new approach for disclosure control in the IAB Establishment Panel : multiple imputation for a better data access	2/07
12/2007	Fuchs, J. Söhnlein, D.	Einflussfaktoren auf das Erwerbspersonenpotenzial : Demografie und Erwerbsverhalten in Ost- und Westdeutschland	3/07
13/2007	Hartmann, J. Krug, G.	Verknüpfung von Befragungs- und Prozessdaten: Selektivität durch fehlende Zustimmung der Befragten?	3/07
14/2007	Baltagi, B. H. Blien, U. Wolf, K.	Phillips Curve or wage curve? Evidence from West Germany: 1980-2004	4/07

Stand: 2.4.2007

Imprint

IAB*DiscussionPaper*
No. 15 / 2007

Editorial address
Institut für Arbeitsmarkt- und Berufsforschung
der Bundesagentur für Arbeit
Weddigenstr. 20-22
D-90478 Nürnberg

Editorial staff
Regina Stoll, Jutta Palm-Nowak

Technical completion
Jutta Sebald, Claudia Halbmeier

Download of this DiscussionPaper:
http://doku.iab.de/discussionpapers/2007/dp1507.pdf

Website
http://www.iab.de

For further inquiries contact the author:
Uwe Blien, Tel. 0911/179-3035,
or e-mail: uwe.blien@iab.de