

Persyn, Damiaan

**Working Paper**

## Spatial aggregation bias in wage curve and NAWRU estimation

JRC Working Papers on Territorial Modelling and Analysis, No. 02/2020

**Provided in Cooperation with:**

Joint Research Centre (JRC), European Commission

*Suggested Citation:* Persyn, Damiaan (2020) : Spatial aggregation bias in wage curve and NAWRU estimation, JRC Working Papers on Territorial Modelling and Analysis, No. 02/2020, European Commission, Joint Research Centre (JRC), Seville

This Version is available at:

<https://hdl.handle.net/10419/227682>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>



European  
Commission

JRC TECHNICAL REPORT

# Spatial aggregation bias in wage curve and NAWRU estimation

JRC Working Papers on  
Territorial Modelling and Analysis  
No 02/2020



Authors:  
Persyn, D.

2020

Joint  
Research  
Centre

This publication is a Technical report by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication. For information on the methodology and quality underlying the data used in this publication for which the source is neither Eurostat nor other Commission services, users should contact the referenced source. The designations employed and the presentation of material on the maps do not imply the expression of any opinion whatsoever on the part of the European Union concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

Contact information

Name: Simone Salotti

Address: Edificio Expo, C/Inca Garcilaso 3, 41092 Sevilla (Spain)

Email: [jrc-b3-secretariat@ec.europa.eu](mailto:jrc-b3-secretariat@ec.europa.eu)

Tel.: +34 954488250

EU Science Hub

<https://ec.europa.eu/jrc>

JRC120448

Seville: European Commission, 2020

© European Union, 2020



The reuse policy of the European Commission is implemented by the Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Except otherwise noted, the reuse of this document is authorised under the Creative Commons Attribution 4.0 International (CC BY 4.0) licence (<https://creativecommons.org/licenses/by/4.0/>). This means that reuse is allowed provided appropriate credit is given and any changes are indicated. For any use or reproduction of photos or other material that is not owned by the EU, permission must be sought directly from the copyright holders.

All content © European Union, 2020 (unless otherwise specified)

How to cite this report: Persyn, D. (2020). Spatial aggregation bias in wage curve and NAWRU estimation. JRC Working Papers on Territorial Modelling and Analysis No. 02/2020, European Commission, Seville, JRC120448.

# Spatial aggregation bias in wage curve and NAWRU estimation

Damiaan Persyn\*  
Joint Research Centre  
European Commission

*If one wishes to predict the rate of change of money wage rates, it is necessary to know not only the level of unemployment but also its distribution between the various markets of the economy.*

—Lipsey, G. (1960) *Economica* 27(1)

## Abstract

I argue in this paper that the estimation of wage curves and NAWRUs at the country level suffers from spatial aggregation bias. Using European data for the years 2000-2017, I find steeper country level wage curves and higher NAWRUs, compared to estimating at the underlying regional level. The distribution of regional unemployment rates within countries over time is not mean-scaled. Regions with low unemployment rates are the main drivers of changes in aggregate unemployment. The steepness of a log-linear wage curve in regions with low unemployment dominates at the aggregate (country) level, overestimating wage pressure. Lagged wages are important in explaining wage growth, together with unemployment. This suggests that a wage curve fits the data better than the assumption of a NAWRU or long run natural rate of unemployment. With regional wage curves, spatial aggregation bias can produce aggregate data that resembles such a natural rate of unemployment, however.

## 1 Introduction

The degree of wage rigidity in an economy plays a crucial role in determining how economic shocks affect employment and unemployment. This is the case both in the real world, and in the macro-economic models used to evaluate and steer fiscal and monetary

---

\*Contact: European Commission, Joint Research Centre, Calle Inca Garcilaso 3, 41092 Seville, Spain [damiaan.persyn@ec.europa.eu](mailto:damiaan.persyn@ec.europa.eu); +34633728923. I am greatly indebted to Ragnar Nymoen for many useful comments and insights that have significantly improved this paper. I also gratefully acknowledge helpful comments from Katja Heinisch, Javier Barbero, Enrique Lopez Bazo, Raul Ramos; participants at the Barcelona AQR group regional workshop, and the ERSA conference in Lyon. All remaining errors are mine. The views expressed are purely those of the author and may not in any circumstances be regarded as stating an official position of the European Commission.

policy, which rely on concepts such as wage-Phillips curves, NAWRUs or wage curves. The vast majority of empirical and theoretical macro-economic models considering wage rigidities operate on the national level. This seems intuitive since many fiscal and monetary policy questions are defined on the level of countries rather than at the regional level. Moreover, institutions such as labour unions or public unemployment insurance schemes operate at the national level or at least have an important national component. Another reason for performing analysis at the country-level is that the required data often is not available at the regional level.

This paper argues that estimation of wage curves and NAWRUs on the national level suffers from spatial aggregation bias. Using European regional data from 2000 to 2017, considering a host of different specifications, wage curves are consistently found to be steeper and, to a lesser extent, NAWRUs are higher when considering data aggregated at the country level. It was already noted by [Clar \*et al.\* \(2007\)](#) in their meta-study of 608 wage curve estimates that studies using national data on average find steeper wage curves compared to those using regional data. To the best of my knowledge, this is the first paper to investigate this effect of aggregation on wage curve and NAWRU estimation.

Starting with the seminal work of [Theil \(1954\)](#), several authors<sup>1</sup> have emphasised that heterogeneity in slope parameters across agents, or a shared but nonlinear relationship, implies that the slope parameters cannot in general be inferred from aggregate data without additional information or assumptions on the distribution of changes in at the micro-level. This type of *distributional aggregation bias* is a concept which is easy to understand and widely recognised in the context of, for example, demand estimation, but has received less attention in macro-econometric analysis. [Lewbel \(1992\)](#) considers the specific case of a log-linear relationship without parameter heterogeneity, and shows that estimation using aggregate data is biased if changes over time in the distribution of the explanatory variables at the micro-level are not proportional (mean-scaled). [van Garderen \*et al.\* \(2000\)](#) and [Albuquerque \(2003\)](#) consider log-linear aggregation in presence of slope heterogeneity.

In a related but separate strand of literature, [Pesaran and Smith \(1995\)](#), [Pesaran \*et al.\* \(1999\)](#), and numerous more recent contributions consider aggregation in the context of dynamic heterogeneity. Dynamic heterogeneity leads to residual autocorrelation in the aggregate series, and bias in the presence of lagged dependent variables. Bias through heterogeneous dynamics has received a lot of attention in the macro-econometric literature, but mostly on the question whether pooled estimation can be used on such data, rather than on the consequences of aggregating data (see for example [Canova, 2011](#), chapter 8). [Pesaran \*et al.\* \(1999\)](#) propose using the pooled mean group and mean group estimators on the micro-level data as strategies to avoid bias from pooling in the context of dynamic heterogeneity.

Although the parameters estimated using aggregate data correctly summarize the *observed* relationship between the macro-aggregates, there will be atypical changes in the micro-variables, for example policy-induced, leading to changes in macro-aggregates

---

<sup>1</sup>See for example [Stoker \(1986\)](#) for an overview.

that deviate from the estimated relationship. A relevant example of distributional aggregation bias which matches the key characteristics of the European regional data considered in this paper, is that of a log-linear wage curve which is operational at the regional level, and changes in aggregate unemployment which are driven mainly by underlying changes in regions with low unemployment rates. The variation in unemployment in the regions with low unemployment rates causes large changes in wage pressure, both locally and at the country level, because the log-linear wage curve in levels is steep in these regions. This leads to steep estimates of the wage curve and high NAWRU estimates at the country level. Policies, however, typically do not target low unemployment regions. A reduction in unemployment in regions with an average or high unemployment rate will lead to less aggregate wage pressure than a researcher or policy maker would have been led to believe from the country-level analysis, because the wage curve in levels is relatively flat at higher levels of unemployment. Higher variability of unemployment in low-unemployment regions leads to over-estimating wage curve elasticities when using aggregate data; even if the underlying wage curve is log-linear, with an identical elasticity in all regions. This bias can even occur if the wage curve expressed in logs is concave, such that the wage curve elasticity is smaller in regions with low unemployment, as long as the regional wage curve in levels is a convex function of the local unemployment rate.

Conditions for the distributional aggregation bias described by [Stoker \(1986\)](#) and [Lewbel \(1992\)](#) to occur are that (1) conditions at the regional level matter for local wage setting, and not just national variables; (2) that the underlying relationship between regional wages and unemployment is non-linear; and (3) that the regional unemployment rates are sufficiently different and that changes therein are not 'mean scaled'. There is ample empirical evidence that these conditions hold to a large degree in Europe.

First, the importance of local factors for wage determination has been attested by a vast empirical literature on the wage curve. Also the early Phillips curve literature related regional wage changes and regional unemployment levels. The still limited response of migration and labour mobility to labour demand shocks in Europe has been well documented (see for example [Beyer and Smets, 2015](#); [Arpaia \*et al.\*, 2016](#); [Basso \*et al.\*, 2019](#)) and is a contributing factor to the long lasting effect of local shocks on local wages and unemployment. In the formal derivation of the aggregation bias and empirical specifications, both regional and national unemployment rates are explicitly allowed to affect regional wages and wage inflation. An early study considering nonlinearity, distributional effects and the relationship between regional and national NAWRU estimates is [Lipsey \(1960\)](#). Other papers that consider the relation between regional and country level variables in the context of wage setting are for example [Roberts \(1997\)](#), [Jimeno and Bentolila \(1998\)](#) and [Campbell \(2008\)](#). Also related are studies such as [Kosfeld and Dreger \(2018\)](#) that consider regional unemployment, but also unemployment in neighbouring regions, through spatial spillovers and spatial autocorrelation in wage setting. Controlling for variables at the aggregate level implies controlling for spatial autocorrelation and a common correlated effect such as described in [Pesaran \(2006\)](#) and [Chudik and Pesaran \(2015\)](#).

Second, the nonlinearity of the relationship between unemployment rates and wages is well attested. It received attention in the early literature on the Phillips curve where log-linear relationships between wage inflation and unemployment rates were considered, but also higher order polynomials describing more convex relationships (see for example [Lipsev, 1960](#)). Also the wage curve literature ([Blanchflower and Oswald, 1994](#); [Card, 1994](#)) has typically considered a log-linear relationship rather than a linear relation between the unemployment rate and wages. Also in the data on European regions, the relationship between regional wages and unemployment rates is convex and approximately log-linear.

Third, regional unemployment differences in the EU are large, both between and within countries. In 2013, for example, the average unemployment rate in Germany was 5.2 percent, ranging from 2.5 percent in Oberbayern to 10.4 percent in Berlin. In Italy the average was 12.2, ranging from 4.4 percent in Bolzano to 22.3 in Calabria. I show that changes over time in the distribution of unemployment rates within countries are not mean-scaled. Increases (decreases) in country level unemployment rates are on average accompanied by lower (higher) dispersion of unemployment rates. The ratio of lower (higher) quantiles of regional unemployment rates to the country level unemployment rate are on average positively (negatively) correlated with the country level unemployment rate. This is only possible if the distribution of regional unemployment rates is compressed (extended) mostly on the left, e.g. if changes in aggregate unemployment rates are mainly driven by regions with relatively low levels of unemployment.

Throughout the analysis, the short time series of the available data forces us to consider a single result for the whole of the EU, either by pooling time series and imposing homogeneity of slope parameters, or by considering a single mean or median of the parameters in case of mean group estimation. To consider results at the level of individual countries would require longer time series, and is left for future research. As a robustness check, the mean group estimator and pooled mean group estimators of [Pesaran and Smith \(1995\)](#) and [Pesaran \*et al.\* \(1999\)](#) are considered. The conclusions also hold when using these methods which are robust to pooling under dynamic heterogeneity.

The policy implications from overestimating wage pressure and NAWRUs are significant. The European Commission judges whether a EU member state has an excessive fiscal deficit based on, among other things, an estimate of the output gap. The size of the NAWRU directly affects the output gap estimates and therefore the allowed deficit. While there has been much discussion on whether the NAWRU estimation techniques which are used lead to overly pro-cyclical NAWRU estimates (and therefore NAWRU estimates that are too high, and an underestimation of the fiscal space after the recent economic crisis), this paper points to the spatial level of analysis as another channel through which NAWRU estimates may be inflated.

## 2 Data

The data used in the empirical analysis is of annual frequency, at the NUTS2 level of regional disaggregation, and freely available online from the Eurostat website. The



sample consists of 246 regions in 18 countries. For most countries the available data runs from 2000 to 2017 resulting in 4138 region-year observations with an average of 16.82 yearly observations per region. Table 1 gives an overview. The NUTS2 regions vary greatly in size, and therefore the relative size of the region in the national aggregate hours worked is used as weights in the regressions.<sup>2</sup> For completeness, I also report the estimated wage curve elasticity and NAWRU from unweighted OLS.

When regional regressions consider country-level explanatory variables, these are calculated while excluding the region under consideration and therefore refer to ‘the rest of the country’. This is similar to a spatial lag of the variable with a weight of 1 for the other regions belonging to the same country, and 0 otherwise. However, larger regions have a larger influence on this country-level variable, which is not usually the case in spatial econometric analysis. The sum of the elasticity on the own-region and country level unemployment rates then reflects the effect on regional wages of a homogeneous increase of all regional unemployment rates in a country. The sum of these elasticities can be compared to the single elasticity obtained when using aggregate national data.

I take great care to ensure that the data at the country level perfectly matches the regional data. Only years in which all regions in a country have data are considered, and the country-level data is aggregated using these strictly balanced regional series. Small countries that consist of only one NUTS2 region cannot be considered<sup>3</sup> and these are therefore excluded from the analysis.

The variables used are the following<sup>4</sup>:

- *w*: nominal hourly cost of employees. Calculated as the total compensation of employees divided by the total hours supplied by employees, on the region or country level
- *gvap*: gross value added deflator, on the country level
- $rw = w/gvap$ : real hourly wage cost
- *prod*: real value added per hour supplied by employees, on the region or country level
- *u*: unemployment rate, on the region or country level

<sup>2</sup>As a robustness check, I considered the higher-level NUTS-1 region in all NUTS-1 regions which contain a NUTS-2 region with less than 150,000 employees in any year, resulting in a sample of 216 regions; I also considered a cutoff of 500,000 employees resulting in 146 regions; and also repeated the entire analysis on the level of 86 NUTS-1 regions. In these analysis the results are qualitatively similar, with changes that are expected from using larger regions: e.g. larger own-region effects and smaller spatial lagged effects for higher levels of aggregation, and a smaller difference between national and regional estimates

<sup>3</sup>This excluded Lithuania, Latvia, Estland, Luxemburg, Malta, Cyprus and Slovenia from the analysis. For Croatia, the youngest EU member state, there are no time series available at the regional level. For Poland the most recent years were dropped due to clear coding errors in the data.

<sup>4</sup>The eurostat datasets used are *nama\_10r\_2coe* for the compensation of employees; *nama\_10r\_2emhrw* for total hours supplied by employees; *nama\_10r\_3gva* for real value added; *nama\_10\_a10* for the gross value added deflator; and lastly *lfst\_r\_lfu3rt* and *lfst\_r\_lfu3pers* for the number of unemployed, the size of the labour force and the unemployment rate.



**Table 1:** The countries, the first and last year contained in the sample, the number of regions, and some summary statistics. All regional series within a country are strictly balanced. The reported smallest and largest unemployment rate and nominal wage are over all years and regions.

	min(year)	max(year)	#regions	min(urate)	max(urate)	min(wage)	max(wage)
Austria	2000	2016	9	2	11.3	12.2	28.3
Belgium	2000	2016	11	1.9	19.2	15.3	37.5
Bulgaria	2000	2017	6	2.9	24.6	0.7	5.4
Czech Republic	1999	2016	8	1.9	15.2	2.2	11.3
Denmark	2000	2017	5	3.2	8.2	20.1	40.4
France	2003	2015	22	5.2	15	16.6	36.4
Germany	2000	2017	37	2	22.4	13.5	32.4
Greece	2000	2016	13	4.7	31.6	3.5	10.8
Hungary	2000	2015	7	3.7	16.4	2.1	7.3
Italy	2000	2016	21	1.8	27.3	8.5	18.4
Netherlands	2000	2016	12	1.2	11	14.7	29.7
Poland	2000	2012	16	5.5	27.3	1.6	6.3
Portugal	1999	2016	5	1.9	18.5	4.8	12.3
Romania	2000	2016	8	3	10.8	0.5	7.9
Slovakia	2000	2016	4	3.4	25	2.2	11.6
Spain	2000	2017	17	4.1	36.2	8.1	20.2
Sweden	2000	2016	8	3.2	10.3	15.3	32.8
United Kingdom	2000	2017	37	1.8	13	12.9	35.8

A Harris-Tzavalis panel unit root test<sup>5</sup> does not reject the  $H_0$  of unit roots in the series for any of these variables (even for the comparatively stable unemployment rate the p-value is 0.85). The same test strongly rejects the presence of a unit root in the variables in first differences. All variables are therefore assumed to be  $I(1)$ .

Nominal wages are quite erratic compared to unemployment rates. A simple regression of wages on unemployment rates may lead to spurious inference although such regressions are frequently used on micro-data with a limited number of yearly observations. With productivity defined as  $prod = Y/H$  with  $Y$  aggregate real value added and  $H$  aggregate hours worked, the wage share in aggregate income is  $wsh \equiv \frac{w \times H}{gvap \times Y} = \frac{w}{gvap \times prod}$ . A stationary wage share in national income therefore implies that wages are co-integrated and homogeneous in prices and productivity. I therefore mostly consider prices and productivity as explanatory variables alongside unemployment to explain wages, and test for cointegration, or directly use the log of the wage share in value added as the dependent variable which imposes homogeneity of wages in prices and productivity.

- $\ln(wsh) = \ln(w) - \ln(prod) - \ln(gvap)$ .

Using the unemployment rate to explain changes in the wage share amounts to a model where labour market tightness influences the division of national income between capital and labour. This could happen when actors in a collective bargaining process take into

<sup>5</sup>This panel unit root test is appropriate because its asymptotic results are derived assuming  $T$  fixed, contrary to most panel unit root tests. This matches well with this dataset which has only up to 18 yearly observations. The HT test requires strongly balanced series, and therefore only the 15 countries and 185 regions which have 17 yearly observations are used for the calculation of this test.

account the unemployment situation, or when the unemployment affects the outside option of a labour union.

### 3 Exploring and explaining distributional aggregation bias

#### 3.1 The dynamic wage curve

The wage curve elasticity and the NAWRU are useful measures of wage pressure in a region. Both are estimated using a dynamic wage curve based on [Nymoen and Rødseth \(2003\)](#). Write  $w_{rt}$  for the nominal hourly cost of employees in region  $r$  (belonging to country  $c$ ) and year  $t$ ,  $prod$  for the real output per hour worked and  $gvap$  for the value added deflator. Different versions of the following error-correction model will be estimated:

$$\begin{aligned} \Delta \ln(w_{rt}) = & \sum_{i=1}^j \gamma_{ri} \Delta \ln(w_{r,t-i}) \\ & + \sum_{k=0}^l \left[ \beta_{0rk} \Delta \ln(prod_{r,t-k}) + \beta_{1rk} \Delta \ln(prod_{c,t-k}) + \beta_{2rk} \Delta \ln(gvap_{c,t-k}) \right] \\ & + \alpha_{0r} + \alpha_{1r} \ln(w_{r,t-1}) + \alpha_{2r} \ln(prod_{r,t-1}) + \alpha_{3r} \ln(prod_{c,t-1}) \\ & + \alpha_{4r} \ln(gvap_{c,t-1}) + \alpha_{5r} \ln(u_{r,t-1}) + \alpha_{6r} \ln(u_{c,t-1}) + v_{rt}. \end{aligned} \quad (1)$$

This is a general framework that embeds both the case of a wage curve in the tradition of [Blanchflower and Oswald \(1994\)](#), e.g. a relationship between the level of the unemployment rate and the level of wages, and a wage Phillips curve which posits a relationship between the level of the unemployment rate and wage growth. Start by assuming equal slope parameter vectors between regions  $\beta_r = \beta$  and  $\alpha_r = \alpha$  apart from the level effect  $\alpha_{0r}$ , and use pooled (fixed effects) estimation. Due to the limited number of annual observations per region on average  $l = 0$  and  $l = 1$  at most are considered, and often  $\gamma = 0$  and or other constraints are imposed on the parameters. A first constraint is one of dynamic homogeneity where  $\beta_{0rk} + \beta_{1rk} = \beta_{2rk} = 1$ , such that changes in prices and productivity fully translate into the nominal wage  $w$  and do not affect the share of labour in national income.

Of key interest is the long-run equilibrium, which is found by setting all terms in first differences to 0 (or a constant  $c$ ) in equation (1). If  $\alpha_1 \neq 0$  the following log-linear relationship between wages and the unemployment rate, e.g. a wage curve, is obtained:

$$\begin{aligned} \ln(w_{rt}) = & -\frac{\alpha_0}{\alpha_1} - \frac{\alpha_2}{\alpha_1} \ln(prod_{rt}) - \frac{\alpha_3}{\alpha_1} \ln(prod_{ct}) \\ & - \frac{\alpha_4}{\alpha_1} \ln(gvap_{ct}) - \frac{\alpha_5}{\alpha_1} \ln(u_{r,t-1}) - \frac{\alpha_6}{\alpha_1} \ln(u_{c,t-1}). \end{aligned} \quad (2)$$

For given national and regional unemployment rates, and assuming regional and national productivity are moving in line, a constant long-run labour share requires another

homogeneity restriction, e.g.  $-\alpha_1 = \alpha_2 + \alpha_3 = \alpha_4$ . Variations of this long run equilibrium relationship are frequently estimated in the empirical literature on the wage curve. Typically spatial lags and productivity are ignored. Regressions considering real wages as the dependent variable amount to assuming that the coefficient on prices equals 1. Estimating the ECM from equation (1), a larger value for the ratio of the estimated coefficient on the lagged level of unemployment  $\alpha_5$  to the estimate of the coefficient on the lagged level of wages  $\alpha_1$ , implies a steeper long-run wage curve, with wage pressure building up faster as the unemployment rate decreases.

Considering equation (1) with  $\alpha_1 = \alpha_2 + \alpha_3 = \alpha_4 = 0$  and excluding the spatial lag of unemployment ( $\alpha_6 = 0$ ), the long-run equilibrium rather corresponds to a vertical line in  $w$ - $u$  space, the long run vertical wage Phillips curve defined by

$$\ln(u^*) = -\frac{\alpha_0}{\alpha_5}. \quad (3)$$

Also here one can set some parameters to 1 and bring variables to the left hand side to consider changes in real wages, or changes in the wage share, rather than nominal wages, and I will do so in the empirical analysis. The unemployment rate at which wages are constant is called the non increasing wage rate of unemployment, or NIWRU. It is a basic estimate of the natural rate of unemployment in the economy. With  $\gamma = 1$  acceleration in nominal wages, real wages or wage shares is considered, rather than increases. The level of unemployment for which wage growth is constant is the NAWRU.

I will investigate both the case of stable wage inflation being defined by either a stable log-linear relationship between wages and the unemployment rate (a wage curve) as in equation (2), or rather a fixed level of unemployment (the NIWRU or NAWRU) as in equation (3).

### 3.2 Aggregation bias in wage curve estimation: basic regressions

To explore the basic properties of the dataset and possible presence of aggregation bias, some basic wage curve estimations are compared between the region and country level. The results are reported in Table 2. All specifications include a full set of cross-sectional and time dummies. Regions with a larger workforce will naturally have a larger weight in the calculation of the country level average wage. The regional share in the country level total hours worked is therefore used as weights in the reported regional regression results. This assures an apple-to-apples comparison between the regional and country level. The table separately reports the long run wage curve elasticity using weights in the row 'LR-elast. (weigh.)'. This correspond to the coefficient on the unemployment rate for regressions in levels; is calculated as the sum of the coefficients on regional and country level unemployment for specifications including a spatial lag; and as in equation (2) for the specifications with differenced wages (or wage shares) as the dependent variables.

Giving more weight to larger regions changes the focus of the regional analysis, however. Since fiscal policy and especially EU policy have an important regional dimension and do not necessarily target large regions, also the estimated unweighted

**Table 2: Aggregation bias: The estimated long run wage curve elasticities (LR-elast.) are larger (more negative) when estimating on aggregated data (columns 2, 4 and 6) compared to regional data (columns 1, 3, 5, 7 and 9). This is the case both when using unweighted regressions (unw.) or weighting by the regional share in aggregate hours worked (weigh.).**

	Specification 1		Specification 2		Specification 3		Specification 4		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\ln(u_{rt})$	-0.126*** (-13.00)		-0.0186*** (-4.68)		-0.00554 (-1.41)				
$\ln(u_{ct})$		-0.161*** (-6.94)		-0.0197** (-2.28)		-0.0137* (-1.66)			
$\ln(gvap_{ct})$			0.884*** (65.10)	0.904*** (33.45)					
$\ln(prod_{rt})$			0.955*** (79.69)						
$\ln(prod_{ct})$				1.002*** (36.49)					
$\ln(wsh_{r,t-1})$							-0.246*** (-10.94)		-0.249*** (-11.09)
$\ln(u_{r,t-1})$							-0.0184*** (-7.88)		-0.00521 (-1.02)
$\ln(wsh_{c,t-1})$								-0.192*** (-3.62)	
$\ln(u_{c,t-1})$								-0.0212*** (-4.71)	-0.0164*** (-3.12)
Constant	-14.18*** (-186.23)	-14.09*** (-86.19)	-0.243 (-1.37)	0.184 (0.48)	-0.662*** (-36.69)	-0.689*** (-21.02)	-0.218*** (-12.19)	-0.192*** (-4.66)	-0.229*** (-12.70)
LR-elast. (weigh.)	-0.126 (-13.00)		-0.0190 (-4.683)		-0.00600 (-1.407)		-0.0750 (-6.869)		-0.0870 (-7.674)
LR-elast. (unw.)	-0.0860 (-17.39)	-0.161 (-6.938)	-0.00500 (-2.130)	-0.0200 (-2.284)	0.00600 (2.579)	-0.0140 (-1.659)	-0.0510 (-8.517)	-0.111 (-3.390)	-0.0620 (-9.404)
N.Obs.	4138	304	4138	304	4138	304	3892	286	3892
Level	region	country	region	country	region	country	region	country	region
R-sq	0.971	0.970	0.998	0.999	0.928	0.944	0.260	0.316	0.264
Q AR(1) p	0	0	0	0	0	0	0	0.0290	0
Q AR(2) p	0	0	0	0	0	0	0.0100	0.220	0.0100
HT I(1) z	-1.522	-0.129	-9.210	-1.100	-8.669	-0.897	-17.93	-6.374	-17.73

Robust t-statistics in parentheses. \*\*\*: p<0.01; \*\*: p<0.05; \*: p<0.1. Cross-sectional and year dummies are included. HT-test: reject H0 of no-cointegration for z < -1.65.

long run wage curve elasticity is reported in the row ‘LR-elast. (unw.)’. The coefficients underlying the calculation of this elasticity are not reported to preserve space.

The first regression considers the real hourly wage cost at the regional level as the dependent variable, with the regional unemployment rate as the sole explanatory variable (column 1). This amounts to estimating the long-run wage curve of equation (2) without spatial lags, imposing a coefficient of prices equal to 1, and ignoring productivity. The long run wage elasticity is -0.126, the coefficient on the log of the regional unemployment rate. For the unweighted regression it is -0.086. Both are close to the value of -0.1 typically found in the literature. Taking the same specification, with the same data aggregated at the country level (column 2), the estimated elasticity is -0.161. Weighting by employment (hours) closes about half of the gap between the unweighted regional and national estimates. This tends to be the case in many specifications which are considered below. This very basic specification is used frequently in the literature, especially in studies based on detailed micro-data with short time series. The bottom of the table shows a set of diagnostic tests which should make us sceptical of the results, however. The z-value of the Harris-Tzavalis (HT) test shows a lack of evidence<sup>6</sup> against

<sup>6</sup>The HT tests rejects the H0 of presence of I(1) in the residuals which would imply no cointegration for z

the H0 of unit roots in the regression residuals, suggesting that the observed relationship may be spurious. The p-values of the Ljung-Box portmanteau (Q) tests reported in the table strongly reject the absence of first and second order autocorrelation in the residuals. This suggests that a more elaborate dynamic specification is required.

In the second specification considered in the table, column (3) considers nominal wages as the dependent variable, but adds both country level prices and regional productivity as explanatory variables to control for key stochastic drivers of wages. Also in this specification the wage curve is steeper (the elasticity estimate is more negative) using the aggregate data (column 4) compared to the regional data in column (3), but only marginally so when weighting the regional regression by the regional share in hours worked. The HT test now strongly rejects the H0 of a unit root in the residuals at the regional level, which suggests that wages, prices and productivity are co-integrated. At the country level there is only weak evidence of co-integration. The absence of AR(1) and AR(2) in the residuals is again strongly rejected.

Since the coefficients on prices and productivity are close to 1, the third specification in columns (5) and (6) brings these variables to the left hand side, replacing wages as the dependent variable by the labour share in value added. This amounts to estimating equation (2) without spatial lags, and with the restriction of long run homogeneity in prices and productivity imposed. Also here estimation on aggregated data leads to a higher elasticity. Still the HT test cannot reject unit roots in the residuals for the country level analysis. The absence of AR(1) and AR(2) in the residuals is strongly rejected.

In the last specification, columns (7) and (8) no longer start from the static long run equilibrium (2), but rather consider a simplified version of the dynamic wage curve of equation (1). The diagnostic test statistics for this dynamic specification are more promising. Although the Q test rejects the absence of AR(1) in the residuals, it does so less strongly, with the z-value (not reported) decreasing from about 21 to 5 in absolute value for the region-level regression. The HT test now strongly rejects unit roots in the residuals. This test result, together with the highly significant coefficient on lagged wages in column (7) and the highly significant coefficients on prices and productivity in explaining wages in column (3), suggests a cointegration relationship exists between wages, prices and productivity. The long run wage curve elasticity is calculated as the ratio of the coefficients on the lagged log unemployment rate and the lagged log wage share as in equation (2). Again, the unweighted regional elasticity is about half that obtained after aggregating data at the country level, and weighting by regional employment shares closes about half of that gap. Lastly, column (9) repeats the regional regression while including the country level unemployment rate. Although the coefficient on the regional level unemployment rate is not significant, again it is found that the sum of the long run elasticities on the regional and country level unemployment rate is smaller than what is obtained using the same data aggregated at the national level.

In conclusion, all these relatively simple wage curve estimations with specifications that are commonly found in the literature find a lower elasticity when estimation is per-  

---

< -1.65, at the one-sided 5 percent significance level; this critical value is reported at the bottom of the tables.

formed at the regional level, compared to the same data aggregated at the country level. The diagnostic tests suggest that an error correction term specification is preferred, and more elaborate dynamics are needed. section 4 considers more elaborate specifications in. First distributional aggregation bias is considered as a possible explanation of the observed difference between the regional and country level results, and whether the properties of the European regional data support this explanation.

### 3.3 Distributional aggregation bias in wage curve estimation: formal derivation

This section formally derives the conditions under which distributional aggregation bias can explain the observed difference between the regional and country level wage curve elasticity estimation. Consider estimating the long run equilibrium relationship of equation (2) without spatial spillovers in productivity  $\alpha_3 = 0$  and impose long run homogeneity in productivity and prices  $-\alpha_1 = \alpha_2 = \alpha_4$  such that

$$\ln(wsh_r) = \ln(w) - \ln(prod_r) - \ln(gvap_c) = b_0 + b_s \ln(u_c) + b \ln(u_r) + \nu, \quad (4)$$

where the constant is  $b_0 = -\alpha_0/\alpha_1$ , the coefficient on the spatial lag of the unemployment rate equals  $b_s = \alpha_6/\alpha_1$ , and the coefficient on the regional unemployment rate is  $b = \alpha_4/\alpha_1$ . I drop time indices for convenience. Assume that regions are approximately equally sized to abstract from weighting, such that the country level wage share and unemployment rate correspond to the average over regions  $wsh_c = E[wsh_r]$  and  $u_c = E[u_r]$ . To formally derive the bias that may occur when estimating equation (4) using national rather than regional data I follow [Lewbel \(1992\)](#) and [van Garderen et al. \(2000\)](#). Exponentiating and taking the expected value shows

$$E[wsh_r] = E \left[ \exp \left( b_0 + b_s \ln(u_c) + b \ln \left( \frac{u_r}{u_c} \right) + b \ln(u_c) + \nu \right) \right].$$

If the errors in the region-level wage equation are  $\nu \sim N(0, \sigma_\nu^2)$  then  $E[\exp(\nu)] = \sigma_\nu^2/2$  and

$$E[wsh_r] = \exp \left( b_0 + (b + b_s) \ln(u_c) + \frac{\sigma_\nu^2}{2} \right) E \left[ \exp \left( b \ln \left( \frac{u_r}{u_c} \right) \right) \right].$$

An estimation equation at the country level then would be

$$wsh_c = \exp \left( b_0 + (b + b_s) \ln(u_c) + \frac{\sigma_\nu}{2} \right) E \left[ \left( \frac{u_r}{u_c} \right)^b \right] \exp(\epsilon), \quad (5)$$

with  $E[\exp(\epsilon)] = 1$ . Note that then in general  $E[\epsilon] \neq 0$  (see for example [van Garderen et al., 2000](#); [Silva and Tenreyro, 2006](#)). If one is willing to assume that  $\epsilon \sim N(E[\epsilon], \sigma_\epsilon^2)$  it holds that  $E[\exp(\epsilon)] = \exp(E[\epsilon] + \frac{1}{2}\sigma_\epsilon^2) = 1$  or  $E[\epsilon] = -\frac{1}{2}\sigma_\epsilon^2$ . Defining  $\xi = \epsilon + \frac{1}{2}\sigma_\epsilon^2$  implies  $E[\xi] = 0$ , and the estimation equation becomes

$$\ln(wsh_c) = b_0 + (b + b_s) \ln(u_c) + \frac{\sigma_\nu}{2} - \frac{\sigma_\epsilon}{2} + \ln E \left[ \left( \frac{u_r}{u_c} \right)^b \right] + \xi. \quad (6)$$

Assuming that the variances  $\sigma_v$  and  $\sigma_e$  of the error terms in the micro and macro wage equations (4) and (5) are uncorrelated with the aggregate unemployment term (for example if these variances are time-invariant), shows the aggregation bias as an omitted variable bias: the expected value of the coefficient on  $\ln(u_c)$  when estimating equation (6) while omitting the term  $\ln E[(u_r/u_c)^b]$  equals

$$E[\widehat{b + b_s}] = b + b_s + \frac{\text{cov}\left(\ln E\left[\left(\frac{u_r}{u_c}\right)^b\right], \ln(u_c)\right)}{\text{var}(\ln(u_c))}. \quad (7)$$

The bias is increasing in the covariance between  $\ln E[(u_r/u_c)^b]$  and  $\ln(u_c)$ . If the  $u_r$  move proportionally with their mean, the covariance is 0, and the distribution of  $u_r$  is mean-scaled (Lewbel, 1992).

The bias can be further quantified by assuming log-normally distributed  $u_r$ . For  $\ln(u_r) \sim N(\mu, \sigma_c^2)$  with  $\mu$  and  $\sigma_c^2$  the cross-regional mean and variance of regional unemployment rates within country  $c$  within a time period, it holds that  $\ln(E[u_r^b]) = b\mu + b\sigma_c^2$  and therefore

$$\begin{aligned} \ln E\left[\left(\frac{u_r}{u_c}\right)^b\right] &= \ln E[u_r^b E[u_r]^{-b}] = -b \ln(E[u_r]) + \ln E[u_r^b] \\ &= -b\mu - b\sigma_c^2/2 + b\mu + b^2\sigma_c^2/2 = \frac{\sigma_c^2}{2}(b^2 - b). \end{aligned}$$

The bias expressed as the percentage difference between the expectation of the slope parameter in a country level wage curve, and the underlying parameter value  $b + b_s$  which shows how much wages would change assuming a homogeneous increase in unemployment rates in all regions of the country, then equals<sup>7</sup>

$$\frac{E[\widehat{b + b_s}] - (b + b_s)}{b + b_s} = \frac{b - 1}{2} \frac{\text{cov}(\ln(u_c), \sigma_c^2)}{\text{var}(\ln(u_c))} \frac{b}{b + b_s}. \quad (8)$$

The last factor adjusts for the relative importance of the regional versus the country level (spatially lagged) unemployment rate in the regional wage equation. If only national variables matter for regional wage setting ( $b = 0, b_s \neq 0$ ) there can be no distributional aggregation bias.

The small size of the own-region unemployment elasticity compared to the larger effect of changes in the other regions' unemployment rates within the same country found in the last column of Table 2 therefore argues against distributional aggregation bias as a major explanation for the difference between the regional and national results. The effect of the own-region unemployment rate will be larger in the specifications with more elaborate dynamics and a spatial lag for productivity, however, and also in those considering parameter heterogeneity between regions and countries in section 4. As may be expected, the effect of the own-region unemployment is also typically higher

<sup>7</sup>Lewbel (1992) reports  $\text{cov}(\ln(u_c), \sigma_c)$  and not  $\text{cov}(\ln(u_c), \sigma_c^2)$ .



in the weighted regressions, and when using regional data at a higher level of regional disaggregation (e.g. at the NUTS1 level).

For the EU as a whole, regional unemployment rates in any given year seem approximately log-normally distributed. Log-normal distributions appear quite naturally: if rates of change of regional unemployment rates are independent of the level of the unemployment rate, the distribution of regional unemployment rates will tend to a log-normal distribution as a result of the central limit theorem. It may be risky to put much faith in equation (8) in this application, however, given that many countries contain just a handful of NUTS2 regions.

### 3.4 Convexity of the wage curve and absence of mean-scaling in European regional unemployment rates

This section considers whether the data on European regional unemployment rates and wages supports distributional aggregation bias as an explanation of the higher wage curve elasticities which are observed at the country level compared to the regional level. I first show that the basic relationship between regional wages and unemployment rates is convex and approximately log-linear, before showing how changes over time in the distribution of regional unemployment rates deviate from mean scaling.

Table 3 shows that in Europe the relationship between regional wages and unemployment, in levels, is highly convex and reasonably approximated by a log-linear function. This is the case both at the regional and country level. Column (1) considers the unemployment rate in levels and the unemployment rate squared. Column (2) uses the log of the unemployment rate. Column (3) uses both the log and the squared log unemployment rate. Columns (4) to (6) repeat this on the country level. The regressions with the unemployment rate and its square (columns (1) and (4)) show that the relationship is highly convex. Although one can expect aggregation bias for any convex relationship in levels, the formal derivation of the bias in section 3.3 assumed log-linearity throughout. The data suggest that this assumption is reasonable: Comparing the unadjusted  $R^2$  reported at the bottom of the table between the specifications with a simple log (columns 2 and 5) and those including the squared log unemployment rate (columns 3 and 6), and the lack of significance of the squared log unemployment rates suggests that the specifications with a simple log of the unemployment rate adequately capture the non-linearity.

Also noteworthy is the fact that the squared terms (both in levels and logs) suggest that the relationship is more convex or less concave at the aggregate level. Finally, note that the main conclusion from before still holds when including quadratic terms: the estimated long run wage curve elasticities, now evaluated at the median unemployment rate of 7 percent, are higher when estimated using country level data.

Equation (7) shows that assuming a log-linear relationship between unemployment rates and wages at the regional level, correlation between  $u_r/u_c$  and  $u_c$  causes aggregation bias. Figure 1 illustrates this failure of mean-scaling graphically. The figure considers wage curves in two large regions with a labour force of equal size. Assume

**Table 3:** The relationship between regional wages and the unemployment rate is convex and approximately log-linear. The long run wage curve elasticity (LR-elast.), evaluated at the median level of unemployment of 7 percent, are higher at the country level.

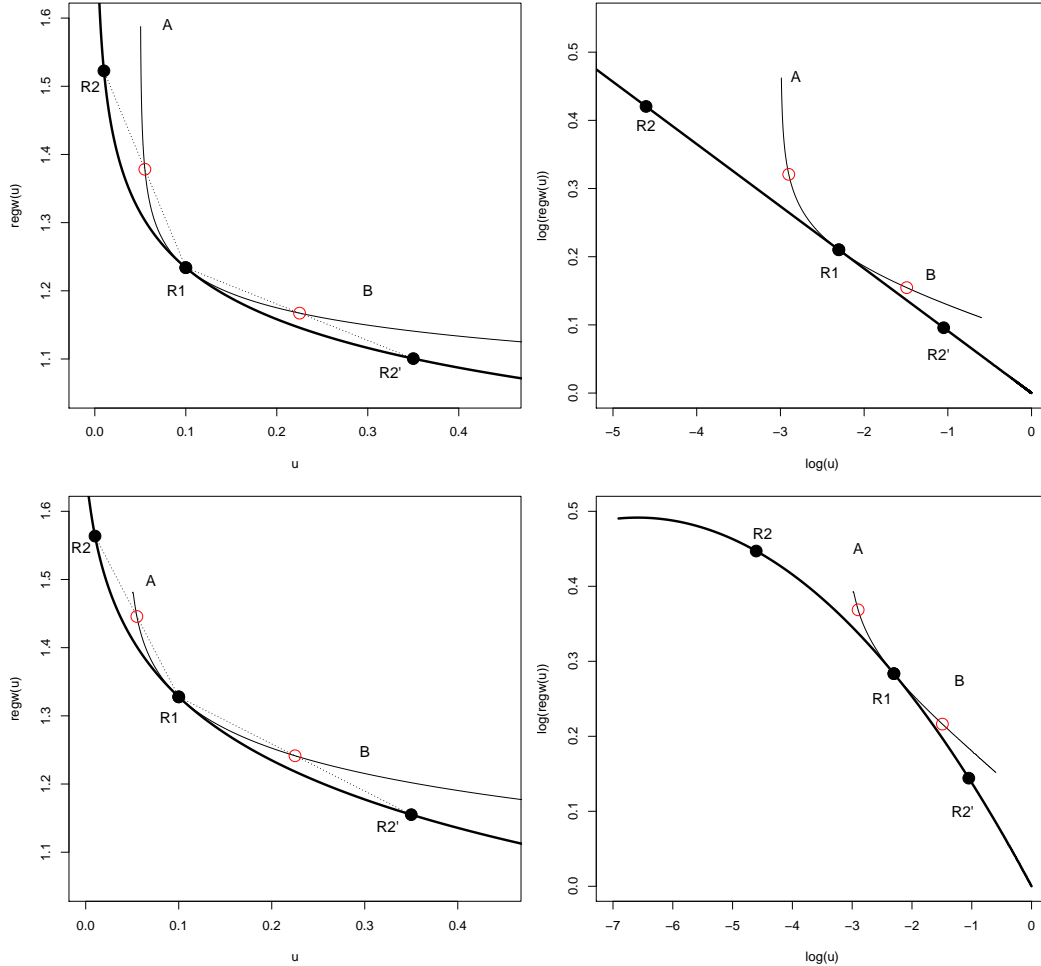
	(1)	(2)	(3)	(4)	(5)	(6)
	$\Delta \ln(wsh_{rt})$	$\Delta \ln(wsh_{rt})$	$\Delta \ln(wsh_{rt})$	$\Delta \ln(wsh_{ct})$	$\Delta \ln(wsh_{ct})$	$\Delta \ln(wsh_{ct})$
$\ln(wsh_{r,t-1})$	-0.246*** (-11.00)	-0.246*** (-10.94)	-0.246*** (-10.97)			
$u_{r,t-1}$	-0.325*** (-5.36)					
$u_{r,t-1}^2$	0.553*** (3.21)					
$\ln(u_{r,t-1})$		-0.0184*** (-7.88)	-0.0286*** (-3.18)			
$\ln(u_{r,t-1})^2$			-0.00205 (-1.20)			
$\ln(wsh_{c,t-1})$				-0.190*** (-3.61)	-0.192*** (-3.62)	-0.192*** (-3.61)
$u_{c,t-1}$				-0.383*** (-3.32)		
$u_{c,t-1}^2$				0.709* (1.87)		
$\ln(u_{c,t-1})$					-0.0212*** (-4.71)	-0.0280 (-1.28)
$\ln(u_{c,t-1})^2$						-0.00142 (-0.34)
Constant	-0.148*** (-9.74)	-0.218*** (-12.19)	-0.230*** (-11.05)	-0.109*** (-3.22)	-0.192*** (-4.66)	-0.199*** (-4.10)
LR-elast. $_{ u=0.07}$ (weigh.)	-0.0700 (-5.840)	-0.0750 (-6.869)	-0.0720 (-6.636)			
LR-elast. $_{ u=0.07}$ (unw.)	-0.0450 (-6.682)	-0.0510 (-8.517)	-0.0480 (-8.117)	-0.104 (-3.086)	-0.111 (-3.390)	-0.107 (-3.448)
N.Obs.	3892	3892	3892	286	286	286
Level	region	region	region	country	country	country
R-sq	0.260	0.260	0.261	0.316	0.316	0.316
Q AR(1) p	0	0	0	0.0280	0.0290	0.0280
Q AR(2) p	0.0100	0.0100	0.0100	0.230	0.220	0.230
HT I(1) z	-18.17	-17.93	-17.99	-6.639	-6.374	-6.429

Robust t-statistics in parentheses. \*\*\*:  $p < 0.01$ ; \*\*:  $p < 0.05$ ; \*:  $p < 0.1$ . Cross-sectional and year dummies are included  
HT-test: reject H0 of no-cointegration for  $z < -1.65$ .

that the wage curves are independent from variables in the other region for this example. In the top row, the stylized wage curve in both regions is log linear and equals  $\ln(w_{rt}) = -0.1\ln(u_{rt})$ . These regional wage curves overlap and are pictured in bold. The elasticity of the regional wage curve equals -0.1 everywhere. There is no heterogeneity in the slope parameter for the wage curve expressed in logs in this example. The right panel shows the log unemployment rate and log wages. The log-linear regional wage curve depicted in logs is obviously a straight line. The left panel shows the relationship in levels, and the wage curve therefore is a curve, with an asymptote at  $u = 0$ .

Assume that both regions initially have an unemployment rate of 10 percent. Con-

**Figure 1:** A graphical illustration of the aggregation bias. Bold: overlapping wage curves in two regions. Thin line: average (country level) values. Left panels: levels. Right panels: logs. The unemployment rate in region R1 is fixed at 10 percent. For values  $R2 < R1$  decreasing averages go with an increase in dispersion of unemployment rates and the country level wage curve is steeper than the regional one. For values of  $R2 > R1$  the opposite holds. Bottom: even if the regional relationship in logs is concave, the aggregate wage curve in logs is convex.



sider changes in the unemployment rate in only one of the regions,  $R2$ , keeping the unemployment rate in  $R1$  fixed. Such changes are clearly not mean-scaled. The thinner line shows how the average wage and unemployment rates change in response to the changes in region  $R2$ . Two specific values for the unemployment in  $R2$  are illustrated by a black dot, for unemployment rates of 0.01 and 0.35. For each, the average unemployment rate and wage levels are indicated by a red circle, which in the left panel lies at the midpoint of the line segment between the regions  $R1$  and  $R2$ .

Starting from the initial situation with equal unemployment rates, if the unemployment rate in  $R2$  is increasing, the relationship between region-averages is on the line

segment B and is flatter compared to the regional wage curves. If the unemployment in R2 rather decreases below R1, the observed aggregate relationship is steeper than the regional wage curves. The wage curve at the national level has an asymptote at an unemployment rate of 5 percent ( $u = 0.05$ ), which is the value of the national unemployment rate when the unemployment rate in R2 approaches 0. As shown in the top right panel, the relationship between the unemployment rate and wage observed at the country level is not log-linear, and more convex, as was observed in the regressions in Table 3.

One should not necessarily expect to observe the entire schedule [A, B] in the aggregate data. Depending on the nature of the changes in the distribution of regional unemployment rates over time, more of the flatter schedule B or rather the steeper part A would be observed, or a mix thereof, the composition of which may also change over time.

The steeper section A of the aggregate wage curve resembles a long run vertical Phillips curve near its asymptote at  $u = 0.05$ . Due to the distributional aggregation bias, there are conditions under which regional wage curves generate aggregate data that is observationally equivalent to a long run vertical Phillips curve. This extreme case is not supported by the data, however: the empirics find a significant effect of lagged wages (or wage shares) in the error correction models, suggesting that also at the country level wages are stable along a wage curve, rather than at a fixed level of unemployment. E.g. the aggregate wage curve is not estimated to be extremely steep.

Figure 1 may help to interpret the aggregation bias derived in section 3.3: start with the expression for log-normally distributed regional unemployment rates in equation (8). The term  $\text{cov}(\sigma_{ct}^2, \ln(u_{ct}))$  determines the sign of the bias. For the case  $b < 0$  the estimated slope at the aggregate level will be steeper than the slope at the regional level if increases in the aggregate unemployment rate  $\ln(u_{ct})$  are accompanied by decreases in the dispersion of regional unemployment rates as measured by  $\sigma_t^2$  and vice-versa. This is the case illustrated to the left of point R1 in Figure 1. The opposite holds to the right of point R1, where increases in the aggregate  $\ln(u_{ct})$  coincide with increases in  $\sigma_t^2$ .

Using the data on EU regional unemployment, I estimate  $\frac{\text{cov}(\sigma_{ct}^2, \ln(u_{ct}))}{\text{var}(\ln(u_{ct}))}$  as the coefficient on  $\ln(u_{ct})$  in a regression of country-year observations of a robust<sup>8</sup> estimate of the variance  $\sigma_{ct}^2$  of the regional unemployment rates within country  $c$  and year  $t$  on  $\ln(u_{ct})$ . The regressions include country dummies to ensure that only within-country variation is used, since the aggregation over regions takes place within countries. The resulting coefficient is -0.062 (t-value 3.2), which strongly rejects mean-scaling of the distribution of regional unemployment rates within EU countries but has a rather small predicted value of the aggregation bias under the assumption of log-normality of the regional unemployment rates: considering equation (6), ignoring a spatial lag for unemployment, and using the typical estimate of region-level wage curves of  $b = -0.1$ , a value of -0.062 for the covariance term, and an equal size of the regional and country elasticity, results

<sup>8</sup>Many countries have only a handful of regions. The interquartile range of regional unemployment rates divided by 1.349 therefore is useful as a robust estimator of the standard deviation. The idea follows from the fact that this ratio holds for the normal distribution but is clearly less sensitive to outliers.

in a predicted bias of only 1.7 percent. There are large differences between countries, however. For Italy, for example, a country with infamously large and persistent regional unemployment differences the coefficient is -0.44 (t-value 3.5), and the predicted bias under the assumption of log-normality is 12 percent. For 14 out of the 18 countries considered the estimated coefficient is negative (the predicted bias is for larger elasticities at the country level), and all of these estimates are significant at the 10 percent level. For none of the other 4 countries the coefficient is significantly positive at the 10 percent level.

For this application, assuming log-normality or another specific distribution of regional unemployment rates within countries may be risky, if only because some of the countries considered contain just a handful of regions. For any distribution, however, [Lewbel \(1992\)](#) shows that a necessary and sufficient condition for mean scaling and unbiased aggregation is that the ratio of the  $q$ 'th quantile  $s_{qct}$  of the distribution of the regional unemployment rates to the aggregate unemployment rate,  $rel_{qct} = \frac{s_{qct}}{u_{ct}}$  is independent from  $u_{ct}$ . [Table 4](#) shows the coefficients on  $u_{ct}$  in a regression of  $rel_{qct}$  on  $u_{ct}$  for several choices of  $s_{qct}$ . The first row shows the results when pooling all countries while including country dummies to ensure that the reported regression coefficient is reflecting only within-country variation. There is a clear pattern: the lower quantiles

**Table 4:** Regressing regional unemployment rate quantiles relative to the country level unemployment rate  $rel_{qct} = \frac{s_{qct}}{u_{ct}}$ , on the country level unemployment rate  $u_{ct}$ .

	(1) rel1	(2) rel5	(3) rel10	(4) rel25	(5) rel50	(6) rel75	(7) rel90	(8) rel95	(9) rel99
EU	0.484*** (3.78)	0.559*** (4.47)	0.416*** (3.45)	0.263* (2.30)	-0.318** (-3.04)	-0.812*** (-5.14)	-1.092*** (-4.47)	-1.743*** (-7.29)	-1.804*** (-6.65)
_cons	0.549*** (30.44)	0.545*** (30.98)	0.552*** (32.46)	0.698*** (43.31)	0.862*** (58.59)	0.993*** (44.61)	1.805*** (52.47)	1.838*** (54.56)	1.841*** (48.17)

expressed relative to the average tend to be positively correlated with the national average, and this is reversed for the higher quantiles. This violates mean scaling, and hints at the direction of the bias.

The pattern found in the first row of [Table 4](#) corresponds to section A of [Figure 1](#), where the decrease (increase) in the average unemployment rate was driven by a decrease (increase) in the regions with the lowest unemployment rate. The change in the lower quantiles of the regional unemployment rates is larger than the average, leading  $s_{qct}/u_{ct}$  to move in line with  $u_{ct}$ . The higher quantiles decrease less (or in section A of the figure, not at all), leading  $s_{qct}/u_{ct}$  to move in the opposite direction of  $u_{ct}$ , leading to a negative correlation. In short, the changes in the distribution of regional unemployment rates within countries over time as summarised in [Table 4](#) suggests that one should indeed expect wage curve estimates at the national level that are steeper compared to the regional level.

**Two more indications of distributional aggregation bias** This section aimed to illustrate the mechanisms of distributional aggregation bias and to verify that some of the

preconditions for it hold in European regional data. It was shown in turn that simple wage curve elasticities are higher at the aggregate country level compared to the regional level; that not only country level unemployment but also the regional unemployment rate matters for local wage settings (more elaborate wage curve specifications below will provide more evidence on this); that the relationship between regional and country level wages and unemployment rates is convex and approximately log-linear; and that changes in the distribution of regional unemployment rates are not mean-scaled. The observed deviation from mean-scaling is in line with the observation that wage curve elasticity estimates are higher (steeper) at the country level. Some of the findings run counter to the distributional aggregation bias argument: First, the effect of regional unemployment rates is considerably smaller than for the spatial lag (column 9 in Table 2). Second, the predicted size of the bias under the assumption of log-normally distributed regional unemployment rates is less than 2 percent.

There are two more indications that distributional aggregation bias is at work. Firstly, as was illustrated in Figure 1, a log-linear regional wage curve is expected to become more convex after aggregation. This is even the case for a regional relationship which is concave in logs (as in the bottom row of the figure). In that sense, the fact that the quadratic terms on the unemployment rate in Table 3 was found to be more positive (in levels) or less negative (in logs) at the country level is suggestive of distributional aggregation bias.

Second, another indication of distributional aggregation bias is obtained from adding the sample equivalent of the omitted term  $\ln E[(u_r/u_c)^b]$  in equation (6) to the country level regressions. Doing this for the specification reported in column (8) of table 2 results in a significant coefficient of -0.5 on this variable and lowers the estimated long run wage curve elasticity from -0.111 to -0.0996. This closes some but not all of the gap with the regional employment-weighted elasticities of -0.075 and -0.087 in specifications with and without a spatial lag of unemployment. Importantly, this result does not depend on any distributional assumption of the regional unemployment rates.

## 4 Empirical results considering dynamics and spatial autocorrelation

### 4.1 Dynamic wage curves

This section considers specifications based on the error correction equation (1), with elaborate dynamics and spatial lags. I continue assuming homogeneity of parameters other than the intercept across regions, such that (in absence of other complications) pooling and estimation by OLS is unbiased and efficient. All regressions at the regional level in Table 5 include the spatial lags of both productivity and the unemployment rate. Columns (1) and (2) consider only contemporaneous values for the variables in differences ( $l = 0$ ) and exclude a lagged dependent variable ( $\gamma = 0$ ). No restrictions are imposed on the parameters. The Q tests reject the absence of autocorrelation in the residuals. Columns (3) and (4) therefore add a lag of the differenced independent

**Table 5:** Dynamic wage curve estimation assuming homogeneous slopes, including a spatial lag of productivity and the unemployment rate. Regressions using the regional share in aggregate hours worked as weights. LR-elast. (weigh.): the long run wage curve elasticity. The row LR-elast. (unw.) separately reports the elasticity for unweighted regressions.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$\Delta \ln(w_{rt})$	$\Delta \ln(w_{ct})$	$\Delta \ln(w_{rt})$	$\Delta \ln(w_{ct})$	$\Delta \ln(w_{rt})$	$\Delta \ln(w_{ct})$	$\Delta \ln(wsh_{rt})$	$\Delta \ln(wsh_{ct})$
$\Delta \ln(prod_{rt})$	0.639*** (16.98)		0.632*** (16.68)		0.628*** (16.55)			
$\Delta \ln(prod_{ct})$	0.394*** (9.59)	1.036*** (16.34)	0.407*** (9.87)	1.042*** (17.04)	0.377*** (9.39)	1.004*** (30.90)		
$\Delta \ln(gvap_{ct})$	0.831*** (9.24)	0.831*** (4.35)	0.672*** (6.79)	0.670*** (3.47)	0.783*** (11.26)	0.808*** (5.59)		
$\ln(w_{r,t-1})$	-0.242*** (-9.64)		-0.218*** (-8.66)		-0.231*** (-10.20)			
$\ln(prod_{r,t-1})$	0.173*** (5.95)		0.141*** (4.32)		0.145*** (4.59)			
$\ln(prod_{c,t-1})$	0.0898*** (3.23)	0.221*** (4.14)	0.0984*** (3.20)	0.196*** (3.36)	0.0854*** (2.92)	0.200*** (3.31)		
$\ln(gvap_{c,t-1})$	0.181*** (6.13)	0.146** (2.29)	0.186*** (6.21)	0.152** (2.10)	0.231*** (10.20)	0.200*** (3.31)		
$\ln(u_{r,t-1})$	-0.00695 (-1.40)		-0.01000* (-1.92)		-0.00964* (-1.82)		-0.0223*** (-9.04)	
$\ln(u_{c,t-1})$	-0.0162*** (-2.99)	-0.0220*** (-2.94)	-0.0149*** (-2.67)	-0.0242*** (-3.34)	-0.0140*** (-2.62)	-0.0237*** (-4.65)		-0.0245*** (-5.13)
$\ln(w_{c,t-1})$		-0.195*** (-3.50)		-0.172*** (-2.82)		-0.200*** (-3.31)		
$\Delta \ln(w_{r,t-1})$			-0.165*** (-4.42)		-0.144*** (-3.79)			
$\Delta \ln(prod_{r,t-1})$			0.136*** (3.32)		0.122*** (2.97)			
$\Delta \ln(prod_{c,t-1})$			0.0381 (0.99)	0.172 (1.44)	0.0180 (0.48)	0.124 (1.07)		
$\Delta \ln(gvap_{c,t-1})$			0.329*** (4.63)	0.308** (2.15)	0.362*** (5.34)	0.320** (2.37)		
$\Delta \ln(w_{c,t-1})$				-0.163 (-1.51)		-0.128 (-1.16)		
$\Delta \ln(wsh_{r,t-1})$							-0.164*** (-4.25)	
$\ln(wsh_{r,t-1})$							-0.232*** (-9.99)	
$\Delta \ln(wsh_{c,t-1})$								-0.144 (-1.50)
$\ln(wsh_{c,t-1})$								-0.193*** (-3.24)
Constant	0.589*** (3.37)	0.593* (1.81)	0.340* (1.81)	0.305 (0.84)	-0.241*** (-12.30)	-0.205*** (-4.60)	-0.224*** (-12.20)	-0.205*** (-4.57)
LR-elast. (weigh.)	-0.0960 (-6.767)		-0.114 (-6.641)		-0.102 (-7.547)		-0.0960 (-7.161)	
LR-elast. (unw.)	-0.0740 (-8.004)	-0.113 (-2.676)	-0.0770 (-8.581)	-0.141 (-2.438)	-0.0720 (-10.51)	-0.119 (-2.881)	-0.0630 (-8.921)	-0.127 (-3.064)
N.Obs.	3892	286	3646	268	3646	268	3646	268
Level	region	country	region	country	region	country	region	country
R-sq	0.838	0.891	0.844	0.896			0.303	0.355
Q AR(1) p	0	0.0300	0.873	0.742	0.579	0.786	0.808	0.770
Q AR(2) p	0.0500	0.120	0.170	0.350	0.290	0.270	0.240	0.530
HT I(1) z	-15.44	-4.621	-12.33	-3.756	-13.73	-4.308	-13.65	-4.386

Robust t-statistics in parentheses. \*: p<0.1; \*\*: p<0.05; \*\*\*: p<0.01. Cross-sectional and year dummies are included. HT-test: reject H0 of no-cointegration for z < -1.65.



variables ( $l = 1$ ), and of differenced wages, the dependent variable ( $j = 1$ ). Again no parameter restrictions are imposed. The Q tests no longer reject the absence of residual autocorrelation. In this specification the effect of an increase in productivity growth in all regions (considering spatial and time lags) in the region level analysis equals  $(0.632+0.407+0.141+0.0984)/(1-(-0.165))=1.042$ ; with a 95 percent confidence interval of  $[0.987,1.096]$ . Wages are close to dynamically homogeneous in productivity. For prices the elasticity in response to a change in inflation is  $(0.672+0.329)/(1+0.165)=0.86$   $([.71,1.01])$ . The fact that the effect of prices is smaller than 1 and estimates are less precise compared to productivity may reflect the fact that only a national price deflator is used for lack of regional data. In the long-run equilibrium in levels, productivity and prices have estimated elasticities of 0.85  $[0.69,1.01]$  and 1.10  $[0.996,1.20]$ .

The specifications shown in columns (5) and (6) impose dynamic and long run homogeneity in prices and productivity. The results are qualitatively not very different from the unconstrained ones in columns (3) and (4). Given their elaborate spatial and temporal lag structures, and given that they pass the specification tests, specifications (3)-(4) and (5)-(6) are the preferred dynamic wage curve estimates.

As a robustness check, the specifications in columns (7) and (8) impose strict homogeneity in prices and productivity within each period, by considering changes in the wage share rather than its individual components. The fact that the addition of a lag of the differenced wage share is needed to control for residual autocorrelation may be indicative of a specification error.

It is found throughout that, also when considering more elaborate dynamics and spatial lags, irrespective of whether or not homogeneity properties are imposed, the long-run wage curve is estimated to be steeper when using national data compared to regional data. The wage curve elasticity is estimated to be between 17 and 32 percent higher at the country level when comparing to regional regression while weighting regions by their share in national hours worked, and 50 to 100 percent without weighting.

**Wage curve elasticity: interpretation** The NIWRU or NAWRU corresponds to the unemployment rate at which wage growth or acceleration is 0. Being a specific level of the unemployment rate, at least its unit of measurement is easily interpretable. In comparison, the dimensionless slope of a wage curve is a more abstract concept. To aid interpretation, some back-of-the-envelope reasoning can be used to translate the wage curve elasticity into a statement on a specific level of the unemployment rate. The absolute value of the wage curve elasticity corresponds to the level of the unemployment rate with a specific wage pressure: it is the level at which a 1 percentage point change in the unemployment rate gives rise to a 1 percent change in the wage share. E.g. a wage elasticity of -0.1 implies that at an unemployment rate of  $\text{abs}(-0.1)$  or 10 percent, a percentage point decrease to an unemployment rate of 9 percent leads to an increase in the wage share by approximately 1 percent. The wage share is relatively stable and bounded, such that a change in the wage share by 1 percent is actually relatively large. It turns out that the variation in the wage share and unemployment rates is such that taking the absolute value of the elasticity also corresponds to the level of

the unemployment rate at which, as a rule of thumb, a one standard deviation in the unemployment rate approximately leads to a standard deviation in the wage share (or 78 percent of a standard deviation to be precise).<sup>9</sup>

If we call one standard deviation in the wage share significant, the unemployment rate corresponding to absolute value of the wage curve elasticity may be called the ‘significant wage pressure rate of unemployment’, or SWPRU. In this view the differences in the wage curve elasticities which are observed depending on the level of aggregation are quite large: In the preferred specification of columns (3) and (4) of Table 5 the estimated SWPRU is 11.4 percent on the regional level versus 14.1 percent using aggregated country level data. If one prefers two standard deviations of the wage share as a definition of what constitutes significant wage pressure, these values are halved, to unemployment rates of 5.7 percent and 7 percent respectively. These unemployment rates are not steady state or ‘natural’ rates to which the economy would return. They are levels of the unemployment rate at which a further tightening of the labour market would lead to a specific amount of upward pressure on wages.

**Two more indications of distributional aggregation bias, revisited** Now briefly reconsider the two additional predictions from the discussion in sections 3.3 and 3.4.

Firstly, the wage curve at the country level is predicted to be more convex in the aggregated data compared to the regional series (as illustrated in Figure 1). This was observed to be the case in the more basic regressions reported in Table 3. It turns out that this also is (modestly) the case using the more elaborate specifications considered here. Adding a squared regional unemployment rate in the same specification (column 3) results in a coefficient of between -0.0496 and -0.057 depending on weighting and whether or not the country level unemployment rate is included. At the country level the coefficient on the squared unemployment rate is -0.0460. Repeating this exercise with the unemployment in levels rather than logs results in coefficients on the squared unemployment rate between 0.17 and 0.35, compared to 0.46 at the country level. While these differences are in line with the predictions, they seem small. Aggregation does not turn a concavity in logs at the regional level into a convex relationship at the country level.

Secondly, including the sample equivalent of the omitted term  $\ln E[(u_r/u_c)^b]$  with  $b = -0.1$  in equation (6), for the specification reported in column (5) of Table 5, lowers the estimated long-run wage curve elasticity at the country level from -0.141 to -0.129

---

<sup>9</sup> $\ln(wsh) = \beta \ln(u_r)$  implies  $\frac{\Delta wsh}{wsh} \approx \beta \frac{\Delta u_r}{u_r}$ . Filling in a one percent increase in the wage share and a one percentage point decrease in the unemployment rate shows that  $0.01 = \beta \frac{-0.01}{u_r}$  holds for  $u_r = -\beta = \text{abs}(\beta)$  since  $\beta < 0$ . The interquartile range of the wage share is [0.43-0.56] with a median value of 0.525. The median of the country level standard deviations in the wage share over time is 0.0125, or 0.0238 relative to the median. For the unemployment rate, the standard deviation over time in absolute terms is 0.0186. Bravely rounding both 0.0238 and 0.0186 to 0.02 gives  $0.02 = \beta \frac{-0.02}{u_r}$ , showing that for European regional data, a wage curve elasticity of for example -0.1 implies that, crudely, a standard deviation in the unemployment rate is predicted to lead to a standard deviation in the wage share, when the unemployment rate is 0.1 or 10 percent (7.8 percent without rounding). A steeper wage curve estimate would imply that this amount of wage pressure for a given decrease in unemployment is reached at a higher level of unemployment.

(compared to -0.114 for the employment weighted regression using the same data at the regional level). This further suggests that distributional aggregation bias is responsible for a significant part of the observed difference between the regional and country level results, but some of the gap remains.

## 4.2 NAWRU estimation

The lagged level of wages is highly significant in the error-correction based wage curve regressions, both at the regional and country level (see for example Table 5). This suggests that there is no single level of the unemployment rate at which wages are stable. Wages (or the wage share) rather stabilise at a level that depends on the level of unemployment, as described by a long-run wage curve. At least for European data and the time period under consideration, imposing  $\alpha_1 = 0$  in equation (1) amounts to estimating a mis-specified model. Estimating NAWRUs without validating this assumption is widely practised.

Estimating the NAWRU using aggregated data suffers from other potential sources of bias, even under the assumption of mean-scaled regional unemployment rates. As shown in equation (6), the variances of the shocks to the regional and country level wage equation appear in the constant term of the aggregate log-linear wage curve. If these quantities are fixed in time, estimation of slope parameters (such as the long run wage curve elasticity) is not affected. These variance terms would be captured by the constant term in the regressions based on aggregate data, however, and as such directly enter the country level NAWRU calculation (see equation 3). The regional NAWRU estimation based on equation (4) is not affected. If the variation in the shocks to the regional wage equations is larger in downturns, this would cause a country level NAWRU estimate to be overly pro-cyclical. The effect of the variance of the error in the macro-equation acts in the opposite direction.

It is still interesting to see whether there is a difference between estimation on the regional versus country level for the NAWRU, even if it will be hard to pinpoint which type of aggregation bias is at work. Table 6 shows the NAWRU estimation results for four groups of estimates. The first two columns start from the unconstrained dynamic wage curves reported in columns (3) and (4) of Table 5. To estimate the NAWRU all the variables in levels are removed except for the regional unemployment rate and the region and country level fixed effects. Also the year dummies are removed. The dynamics are rich with a lagged dependent variable and a lag of changes in productivity and prices. Notice that also in this specification the estimated short run elasticity in productivity and prices is close to 1 (summing up the coefficients on the differenced terms for this variables and dividing by one minus the coefficient on the lagged differenced wages). The median of the region and country level estimated NAWRUs are reported near the bottom of the table. The difference between the region-level median NAWRU of 8.168 and the country level estimate of 8.632 is about 0.5 percentage points, or 6 percent in relative terms. Columns (3) and (4) impose dynamic homogeneity in prices and productivity. The difference between the region and country level estimate remains about the same.

**Table 6:** NAWRU estimation. The regressions exclude all variables in levels (such as year dummies) except for the lagged unemployment rate and cross-sectional specific intercepts. The reported NAWRUs are the medians over the region and country specific NAWRU estimates.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$\Delta \ln(w_{rt})$	$\Delta \ln(w_{ct})$	$\Delta \ln(w_{rt})$	$\Delta \ln(w_{ct})$	$\Delta \ln(wsh_{rt})$	$\Delta \ln(wsh_{ct})$	$\Delta \ln(wsh_{rt})$	$\Delta \ln(wsh_{ct})$
$\Delta \ln(w_{rt-1})$	-0.269*** (-6.62)		-0.254*** (-6.20)					
$\Delta \ln(prod_{rt})$	0.626*** (15.79)		0.627*** (16.00)					
$\Delta \ln(prod_{ct})$	0.342*** (8.35)	0.969*** (19.30)	0.340*** (8.37)	0.966*** (33.37)				
$\Delta \ln(gvap_{ct})$	0.552*** (6.04)	0.551*** (3.27)	0.741*** (9.92)	0.758*** (5.00)				
$\Delta \ln(prod_{rt-1})$	0.211*** (5.34)		0.200*** (4.82)					
$\Delta \ln(prod_{ct-1})$	0.0918** (2.35)	0.301*** (3.07)	0.0864** (2.19)	0.266*** (2.85)				
$\Delta \ln(gvap_{ct-1})$	0.483*** (6.25)	0.457*** (3.26)	0.513*** (6.54)	0.474*** (3.21)				
$\ln(u_{rt-1})$	-0.0274*** (-11.24)		-0.0227*** (-9.36)		-0.0237*** (-12.15)		-0.0248*** (-9.89)	
$\Delta \ln(w_{ct-1})$		-0.264*** (-2.85)		-0.232** (-2.53)				
$\ln(u_{ct-1})$		-0.0298*** (-6.01)		-0.0234*** (-4.78)		-0.0254*** (-3.81)		-0.0261*** (-3.34)
$\Delta \ln(wsh_{rt-1})$					-0.230*** (-3.41)		-0.245** (-2.23)	
$\Delta \ln(wsh_{rt-2})$							-0.0983 (-0.78)	
$\Delta \ln(wsh_{ct-1})$						-0.228 (-0.68)		-0.239 (-1.44)
$\Delta \ln(wsh_{ct-2})$								-0.0916 (-0.72)
Constant					-2.585*** (-72.83)	-2.533*** (-22.60)	-2.573*** (-74.14)	-2.518*** (-22.75)
NAWRU (weigh.)	8.168		6.676		6.883		7.129	
NAWRU (unw.)	8.013	8.632	6.704	7.143	7.070	7.203	7.207	7.257
N.Obs.	3646	268	3646	268	3646	268	3400	250
Level	region	country	region	country	region	country	region	country
R-sq	0.862	0.910			0.866	0.843	0.871	0.846
Q AR(1) p	0.0100	0.668	0.0700	0.655	0.780	0.853	0.433	0.932
Q AR(2) p	0	0.710	0	0.750	0	0.250	0.610	0.960
HT I(1) (z<-1.65)	-13.56	-3.728	-14.58	-4.211	-15.25	-4.373	-14.25	-4.120

Whereas there was little proof of residual autocorrelation in specifications (3) to (6) of Table 5, removing the level variables introduces significant residual autocorrelation in the regional series. This residual autocorrelation does not readily disappear when considering additional lags.

Columns (5) and (6) in Table 6 consider the wage share as the dependent variable, imposing dynamic homogeneity within every time period. The results can be compared to the specifications in columns (7) and (8) of Table 5. The difference between the median of the NAWRUs estimated at the regional (weighted) and country level is less than 5 and 2 percent in these specifications. Adding two lags of the dependent variable is required to remove the autocorrelation in the residuals (columns 7 and 8). For this specification the difference between the national and regional estimate becomes very small.

It is not straightforward to pick a preferred specification in Table 6, given the omission

of the highly significant variables in levels, given the appearance of residual autocorrelation and the multiple lags of the dependent variable that are needed to capture it, and knowing that the constant may be capturing the variation in shocks to the region and country level wage equations. Overall, the difference between the regional and country level NAWRU is smaller than the bias observed for the wage curve elasticity; ranging from 6 to 7 percent in the regressions allowing for rich dynamics, to less than 5 percent when imposing strict homogeneity in prices and productivity by considering wage shares.

## 5 Dynamic heterogeneity

As argued by [Pesaran and Smith \(1995\)](#) and [Pesaran \*et al.\* \(1999\)](#) in the context of linear models, ignoring heterogeneity in the slope parameters between micro-units can induce residual autocorrelation when pooling or aggregating. This leads to bias in the presence of lagged dependent variables even for  $T \rightarrow \infty$ . Two alternatives to pooling are suggested: mean-group estimation and pooled mean group estimation. These methods start from the individual micro-series and cannot be used with aggregated data. [van Garderen \*et al.\* \(2000\)](#) show that the derivation of aggregation bias under parameter heterogeneity and nonlinearity is quite complicated. Unbiased estimation using aggregate data requires introducing additional higher order terms, even under restrictive assumption of mean-scaling which is clearly violated in the data. I will therefore not attempt to repair the aggregate regressions or try to uncover how much of the difference between the region and country level analysis is due to dynamic aggregation bias.

An important reason to consider these methods, however, is to make sure that the pooling between regions and countries which was used throughout is not the underlying cause of the observed difference in the country and region level analysis. The expected bias when pooling under dynamic heterogeneity is for the coefficients to tend to 0. A possible explanation for the lower slopes in the regional data could therefore be that the dynamic aggregation bias is larger for the pooled regional regressions. I show in this section that this is not the case: also using pooled mean group and mean group estimation the wage curve elasticities and the NAWRU are higher when using country level data compared to regional data.

The mean group estimator considers each underlying time series separately and averages over (functions of) parameter estimates. Given the limited length of the time series, it is likely that these estimates will be inefficient and quite noisy. A particular concern is the fact that both the wage curve elasticity in the error-correction specification and the NAWRU require taking ratio's, which leads to erratic estimates (see also [Holden and Nymoer, 2002](#)). Only parsimonious specifications are therefore considered to preserve sufficient degrees of freedom, and use the median as a more robust centrality estimate. I also consider the more efficient pooled mean group estimator of [Pesaran \*et al.\* \(1999\)](#), where only short run parameters are allowed to differ between the micro-units while some of the slope parameters for the long-run effects are assumed to be shared.

Given the lower efficiency of these methods and the fact that the main conclusions are not altered, the results of this section should be considered a robustness check, excluding pooling under dynamic parameter heterogeneity as an explanation of the observed difference between region and country level estimation of wage pressure in the economy.

### Dynamic heterogeneity and the wage curve

Table 7 shows the result of estimating a simplified dynamic wage curve using the pooled mean group estimator and the mean group estimator.<sup>10</sup> To preserve degrees of freedom I exclude lags of the independent variable  $l = 0$ , assume contemporaneous dynamic homogeneity in differences and levels such that the differenced price and productivity levels can be brought to the left hand side to consider changes in the labour share of income. Using the pooled mean group estimator and constraining the coefficient on the unemployment rate to be shared among regions or countries gives a long run wage curve elasticity of -0.086 on the country level, compared to elasticities ranging from -0.061 to -0.069 at the regional level, depending on whether regional weights are used and a spatial lag of unemployment is included. It is important to note that this specification allows for an idiosyncratic long-run wage curve elasticity through variation in the coefficient on the lagged wage share between the micro-units (see equation (2)). Since the insignificance of the unemployment rates in the first column is likely due to co-linearity, column (2) repeats the analysis at the regional level excluding the spatial lag of the unemployment rate, leading to similar estimates.

Columns (4) to (6) use pooled mean group estimation but rather restrict the coefficient on the lagged wage share in levels to be identical across the micro-units, and allow the coefficient on the unemployment rate to differ between regions and countries. This results in long run wage curve elasticities between -0.055 and -0.076 on the region level versus -0.127 at the country level.

There is substantial residual autocorrelation in these specifications, however, which does not disappear when including additional lags. Therefore consider the equation in differences in columns (7) and (8).<sup>11</sup> Estimation using the mean group estimator allows all parameters to vary between regions and countries. The estimated wage curve elasticities are smaller using this specification, but remain substantially higher for the country level analysis.

The difference between the estimated wage curve elasticity when using regional and aggregated data remains substantial with the pooled mean group and mean group estimators, with differences of around 25 to 70 percent when comparing country level estimates to the regional estimates using the regional share in country level hours worked as weights.

---

<sup>10</sup>The `xtdcce2` command from [Ditzen \(2018\)](#) in Stata is used to perform the analysis.

<sup>11</sup>The lagged dependent variable is omitted in the differenced equations. Including it results in very similar regional elasticity estimates, but increases the country level elasticity to -0.15.

**Table 7: Dynamic wage curve: pooled mean group and mean group estimation.** The table reports the unweighted regression results. The weighted long-run wage curve elasticity (LR-elast (weigh.)) is calculated by duplicating observations in proportion to their share in the national hours worked and therefore a standard error is omitted.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$\Delta \ln(wsh_{rt})$	$\Delta \ln(wsh_{rt})$	$\Delta \ln(wsh_{ct})$	$\Delta \ln(wsh_{rt})$	$\Delta \ln(wsh_{rt})$	$\Delta \ln(wsh_{ct})$	$\Delta^2 \ln(wsh_{rt})$	$\Delta^2 \ln(wsh_{ct})$
$\Delta \ln(wsh_{r,t-1})$	-0.00841 (-0.53)	-0.000855 (-0.05)		-0.0346** (-2.22)	-0.0211 (-1.37)			
$\ln(wsh_{r,t-1})$	-0.339*** (-23.01)	-0.339*** (-22.75)		-0.338*** (-10.10)	-0.309*** (-11.90)			
$\ln(u_{c,t-1})$	-0.0365 (-0.22)		-0.0864 (-1.27)	-0.0315* (-1.83)		-0.127*** (-6.31)		
$\ln(u_{r,t-1})$	-0.0300 (-0.23)	-0.0607* (-1.94)		-0.0308* (-1.78)	-0.0546*** (-8.13)			
$\Delta \ln(wsh_{c,t-1})$			-0.000313 (-0.01)			-0.0303 (-0.47)		
$\ln(wsh_{c,t-1})$			-0.280*** (-6.41)			-0.217** (-2.21)		
$\Delta \ln(wsh_{r,t-1})$							-1.006*** (-56.16)	
$\Delta \ln(u_{r,t-1})$							-0.00479 (-0.89)	
$\Delta \ln(u_{c,t-1})$							-0.0346*** (-5.25)	-0.0498*** (-4.37)
$\Delta \ln(wsh_{c,t-1})$								-0.930*** (-16.38)
LR-elast. (weigh.)	-0.0690	-0.0630		-0.0760	-0.0640		-0.0390	
LR-elast. (unw.)	-0.0660 (-0.863)	-0.0610 (-1.941)	-0.0860 (-1.265)	-0.0620 (-7.631)	-0.0550 (-8.133)	-0.127 (-6.315)	-0.0390 (-8.988)	-0.0500 (-4.373)
pooled:	llogurc llogur	llogur	llogurc	llogshare	llogshare	llogsharec		
N.Obs.	3646	3646	268	3646	3646	268	3646	268
Level	region	region	country	region	region	country	region	country
Q AR(1) p	0.529	0.337	0.748	0.0680	0.378	0.781	0.301	0.837
Q AR(2) p	0	0	0.270	0	0	0.140	0.230	0.940
HT I(1) z	-16.18	-16.31	-7.034	-17.19	-17.83	-7.358	-16.52	-7.094

Robust t-statistics in parentheses. \*: p<0.1; \*\*: p<0.05; \*\*\*: p<0.01. Cross-sectional and year dummies are included  
HT-test: reject H0 of no-cointegration for z < -1.65.

## Dynamic heterogeneity and the NAWRU

As emphasised in section 4.2, estimation of the constant term in the macro-level wage equation suffers from different biases, even under the restrictive assumption of mean-scaled regional unemployment rates which was sufficient to allow unbiased estimation of the wage curve elasticity. This implies that various biases are likely to affect NAWRU estimation using aggregate data, since the constant term directly enters the NAWRU estimation. Also here, dynamic heterogeneity and nonlinearity would require the addition of higher order terms for unbiased estimation under the assumption of mean-scaled regional unemployment rates. I will therefore not attempt to fix the aggregate regression or uncover the relative size of these biases, but stick to excluding that the observed lower NAWRU estimate using regional data is due to the pooling of dynamic heterogeneous series.

Table 8 considers mean group estimation of the NAWRU on the regional and country level. Columns (1) and (2) consider changes in the wage share, allowing for a lag of the



*Table 8: Mean group NAWRU estimation*

	(1)	(2)	(3)	(4)
	$\Delta \ln(wsh_{rt})$	$\Delta \ln(wsh_{ct})$	$\Delta \ln(wsh_{rt})$	$\Delta \ln(wsh_{ct})$
$\Delta \ln(wsh_{r,t-1})$	-0.156*** (-9.71)		-0.338*** (-10.43)	
$\Delta \ln(wsh_{r,t-2})$			-0.344*** (-10.27)	
$\Delta \ln(wsh_{r,t-3})$			-0.182*** (-5.64)	
$\Delta \ln(wsh_{r,t-4})$			-0.207*** (-7.98)	
$\ln(u_{r,t-1})$	-0.0255*** (-13.06)		-0.0483*** (-9.25)	
$\Delta \ln(wsh_{c,t-1})$		-0.136** (-2.16)		-0.258** (-2.02)
$\Delta \ln(wsh_{c,t-2})$				-0.443*** (-3.91)
$\Delta \ln(wsh_{c,t-3})$				-0.187* (-1.77)
$\Delta \ln(wsh_{c,t-4})$				-0.202** (-1.98)
$\ln(u_{c,t-1})$		-0.0359*** (-5.50)		-0.0481*** (-2.69)
Constant	-2.671*** (-35.17)	-2.525*** (-22.86)	-2.137*** (-6.54)	-1.334 (-1.16)
NAWRU	7.234	7.439	6.964	7.936
NAWRU (weigh.)	7.234		6.964	
N.Obs.	3646	268	2908	214
Level	region	country	region	country
R-sq	0.765	0.756	0.477	0.463
Q AR(1) p	0.174	0.960	0.00300	0.205
Q AR(2) p	0	0	0.0100	0.530
HT I(1) (z<-1.65)	-16.35	-6.465	-14.97	-6.204

dependent variable. The coefficients on this lag as well as on the lagged level of the unemployment rate is allowed to differ freely between the micro-units. The estimated difference between the NAWRUs estimated using regional and country level data is quite small at just 0.2 percentage points. The residual autocorrelation does not readily disappear when adding more lags. Columns (3) and (4) show a specification with four lags, which is the maximum number of lags that can be considered while keeping the regional and country level samples identical. Even this number of lags is not sufficient to remove the residual autocorrelation. The difference between the country level and regional NAWRU estimate is about 1 percentage point, or 14 percent in relative terms.

## 6 Discussion and conclusion

This paper shows that two popular measures of wage pressure in an economy, the wage curve and the NAWRU, depend on the level of spatial aggregation of the analysis. Using regional data at the NUTS2 level for the EU for the years 2000 to 2017 wage curve elasticities and (to a lesser extent) NAWRUs that are consistently found to be smaller when estimated on regional data, compared to the same data aggregated at the

country level. This result holds throughout a host of different specifications, including static wage curves and error-correction models, whether controlling for spatial lags of unemployment and productivity or not, and also when using mean-group methods that are robust to pooling under dynamic parameter heterogeneity.

The theoretical and empirical results point to the problematic nature of NAWRU estimation. First, lagged wage (shares) are highly significant in regressions of wage growth including the unemployment rate as an explanatory variable. This points to a relationship between wages and the unemployment rate for which wages are stable, rather than this occurring at a specific level of unemployment. Dropping the level of wages introduces residual autocorrelation which proves hard to control for. Also theoretically the NAWRU estimation on aggregated data is shown to be subject bias through its dependence on the variances of the shocks to the regional and country level wage equations. I suggest taking the unemployment rate corresponding to the absolute value of the wage curve elasticity as an alternative measure of the wage pressure in an economy, since it corresponds to the level of the unemployment rate at which a 1 percentage point decrease in the unemployment rate leads to a 1 percent increase in the labour wage share. For the data this roughly equals the level at which a one standard deviation in the unemployment rate induces a one standard deviation in the wage share.

Bias introduced by pooling or aggregating data with underlying heterogeneity in the dynamic structure at the micro-level has received a lot of attention in the macroeconometric literature starting with [Pesaran and Smith \(1995\)](#). The type of bias which was focused on rather depends directly on the nonlinear relationship between unemployment rates and wages, in combination with basic properties of changes over time in the underlying distribution of regional unemployment rates, as described by for example [Theil \(1954\)](#), [Stoker \(1986\)](#) or [Lewbel \(1992\)](#). Distributional aggregation bias is to be expected given the properties of European regional unemployment rates. Regional unemployment rates matter for local wage setting after controlling for variables at the country level; the relationship between unemployment rates and wages is convex; and changes in regional unemployment rates over time are not mean-scaled. The underlying changes in the regional unemployment rates are such that an upward bias in wage curve estimation using aggregate data is expected: regions with low unemployment rates tend to drive changes in country level unemployment. The nonlinearity of the wage curve implies that the larger wage pressure in those regions will dominate in the aggregate data.

The hypothesis that the more classical distributional aggregation bias is an important driver of the observed difference in wage pressure between regions and countries, is supported by the fact that adding an estimate of the term capturing the effect of non-mean scaled changes in the distribution of regional unemployment changes cuts the gap between the estimates in about half. The relationship between wages and unemployment rates is more convex (in levels) or less concave (in logs), at the aggregate level compared to the regional level, in line with the predictions in presence of distributional aggregation bias.

To base policies only on aggregate data which ignore the underlying regional structure in the relationship between wages and unemployment rates is hazardous. Although

a correctly performed analysis on the national level would capture the observed relationship at the aggregate level, to extrapolate these results to the individual regions in a country is likely to be wrong. Analysis at the aggregate level is overly pessimistic regarding the unemployment rate below which wage pressure builds up, whether estimated through a wage curve or NAWRU. As testified by the results for the unweighted wage curve estimates, wage pressure in an average region is about 50 percent lower than the country level estimates would suggest, and fiscal policies in high-unemployment regions that would succeed in reducing unemployment would lead to less wage pressure than predicted by country level analysis.

## References

- ALBUQUERQUE, P. H. (2003). A practical log-linear aggregation method with examples: heterogeneous income growth in the USA. *Journal of Applied Econometrics*, **18** (6), 665–678.
- ARPAIA, A., KISS, A., PALVOLGYI, B. and TURRINI, A. (2016). Labour mobility and labour market adjustment in the EU. *IZA Journal of Migration*, **5** (1), 21.
- BASSO, G., D'AMURI, F. and PERI, G. (2019). Immigrants, Labor Market Dynamics and Adjustment to Shocks in the Euro Area. *IMF Economic Review*, **67** (3), 528–572.
- BEYER, R. C. M. and SMETS, F. (2015). Labour market adjustments and migration in Europe and the United States: how different? *Economic Policy*, **30** (84), 643–682.
- BLANCHFLOWER, D. G. and OSWALD, A. J. (1994). *The wage curve*. MIT press.
- CAMPBELL, C. M. (2008). An efficiency wage approach to reconciling the wage curve and the Phillips curve. *Labour Economics*, **15** (6), 1388–1415.
- CANOVA, F. (2011). *Methods for applied macroeconomic research*. Princeton university press.
- CARD, D. (1994). *The wage curve*. Cambridge, Mass: MIT Press.
- CHUDIK, A. and PESARAN, M. H. (2015). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. *Journal of Econometrics*, **188** (2), 393–420.
- CLAR, M., DREGER, C. and RAMOS, R. (2007). Wage Flexibility and Labour Market Institutions: A Meta-Analysis. *Kyklos*, **60** (2), 145–163.
- DITZEN, J. (2018). Estimating Dynamic Common-Correlated Effects in Stata. *The Stata Journal: Promoting communications on statistics and Stata*, **18** (3), 585–617.
- HOLDEN, S. and NYMOEN, R. (2002). Measuring Structural Unemployment: NAWRU Estimates in the Nordic Countries. *Scandinavian Journal of Economics*, **104** (1), 87–104.
- JIMENO, J. F. and BENTOLILA, S. (1998). Regional unemployment persistence (Spain, 1976–1994). *Labour Economics*, **5** (1), 25–51.
- KOSFELD, R. and DREGER, C. (2018). Local and spatial cointegration in the wage curve – a spatial panel analysis for german regions. *Review of Regional Research*, **38** (1), 53–75.
- LEWBEL, A. (1992). Aggregation with Log-Linear Models. *The Review of Economic Studies*, **59** (3), 635.
- LIPSEY, R. G. (1960). The Relation between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1862-1957: A Further Analysis. *Economica*, **27** (105).

- NYMOEN, R. and RØDSETH, A. (2003). Explaining unemployment: Some lessons from Nordic wage formation. *Labour Economics*, **10** (1), 1–29.
- PESARAN, M. and SMITH, R. (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics*, **68** (1), 79–113.
- PESARAN, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, **74** (4), 967–1012.
- , SHIN, Y. and SMITH, R. P. (1999). Pooled Mean Group Estimation of Dynamic Heterogeneous Panels. *Journal of the American Statistical Association*, **94** (446), 621–634.
- ROBERTS, J. M. (1997). The wage curve and the Phillips curve. In *Finance and Economics Discussion Paper Series # 57, Federal Reserve Board of Governors*.
- SILVA, J. S. and TENREYRO, S. (2006). The log of gravity. *The Review of Economics and statistics*, **88** (4), 641–658.
- STOKER, T. M. (1986). Simple Tests of Distributional Effects on Macroeconomic Equations. *Journal of Political Economy*, **94** (4), 763–795.
- THEIL, H. (1954). *Linear Aggregation of Economic Relations*. North-Holland.
- VAN GARDEREN, K. J., LEE, K. and PESARAN, M. (2000). Cross-sectional aggregation of non-linear models. *Journal of Econometrics*, **95** (2), 285–331.

## **GETTING IN TOUCH WITH THE EU**

### **In person**

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: [https://europa.eu/european-union/contact\\_en](https://europa.eu/european-union/contact_en)

### **On the phone or by email**

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696, or
- by electronic mail via: [https://europa.eu/european-union/contact\\_en](https://europa.eu/european-union/contact_en)

## **FINDING INFORMATION ABOUT THE EU**

### **Online**

Information about the European Union in all the official languages of the EU is available on the Europa website at: [https://europa.eu/european-union/index\\_en](https://europa.eu/european-union/index_en)

### **EU publications**

You can download or order free and priced EU publications from EU Bookshop at: <https://publications.europa.eu/en/publications>. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see [https://europa.eu/european-union/contact\\_en](https://europa.eu/european-union/contact_en)).

## The European Commission's science and knowledge service

Joint Research Centre

### JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.



**EU Science Hub**

[ec.europa.eu/jrc](https://ec.europa.eu/jrc)



@EU\_ScienceHub



EU Science Hub - Joint Research Centre



EU Science, Research and Innovation



EU Science Hub