

Area disparities in Britain: understanding the contribution of people vs. place through variance decompositions

Article (Published Version)

Gibbons, Stephen, Overman, Henry G and Pelkonen, Panu (2014) Area disparities in Britain: understanding the contribution of people vs. place through variance decompositions. *Oxford Bulletin of Economics and Statistics*, 76 (5). pp. 745-763. ISSN 0305-9049

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/59970/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Area Disparities in Britain: Understanding the Contribution of People vs. Place Through Variance Decompositions*

STEPHEN GIBBONS†, HENRY G. OVERMAN‡ and PANU PELKONEN§

†*Department of Geography and Environment and Spatial Economics Research Centre, London School of Economics, Houghton Street, London WC2A 2AE, UK
(e-mail: s.gibbons@lse.ac.uk)*

‡*Department of Geography and Environment and Spatial Economics Research Centre, London School of Economics, Houghton Street, London WC2A 2AE, UK
(e-mail: h.g.overman@lse.ac.uk)*

§*Department of Economics, University of Sussex, Brighton BN1 9RH, UK
(e-mail: P.O.Pelkonen@sussex.ac.uk)*

Abstract

This article considers methods for decomposing wage variation into individual and group specific components. We discuss the merits of these methods, which are applicable to variance decomposition problems generally. The relative magnitudes of the measures depend on the underlying variances and covariances, and we discuss how to interpret them, and how they might relate to structural parameters of interest. We show that a clear-cut division of variation into area and individual components is impossible. An empirical application to the British labour market demonstrates that labour market area effects contribute very little to the overall variation of wages in Britain.

I. Introduction

...you are under no obligation to analyse variance into its parts if it does not come apart easily, and its unwillingness to do so naturally indicates that one's line of approach is not very fruitful.
R. A. Fisher (1933)

These warnings of R. A. Fisher notwithstanding, there are surely many situations where the decomposition of variance is potentially informative, even if the variance does not

*We thank Phil Murphy, attendees at various seminars, plus two anonymous referees for their contributions. An earlier version of this article circulated as Spatial Economics Research Centre SERC DP0060 with the title 'Wage disparities: people or place' This work was based on data from the Annual Survey of Hours and Earnings, produced by the Office for National Statistics (ONS) and supplied by the Secure Data Service at the UK Data Archive. The data are Crown Copyright and reproduced with the permission of the controller of HMSO and Queen's Printer for Scotland. The use of the data in this work does not imply the endorsement of ONS or the Secure Data Service at the UK Data Archive in relation to the interpretation or analysis of the data. This work uses research data sets which may not exactly reproduce National Statistics aggregates.

JEL Classification numbers: J31, R11.

come apart easily. This article considers methods for decomposing the variance of an individual outcome into group components, where this individual outcome is determined by both individual characteristics and the characteristics of the group to which the individual belongs and where individuals can sort across groups. In particular, we are interested in the importance of individual characteristics vs. group membership in explaining variation in individual outcomes. Such decompositions are of interest in a wide variety of contexts. For example, labour economists have long been concerned with the role sorting on individual characteristics plays in explaining differences in wages between groups of workers (particularly wage differences across industries). See, for example, Krueger and Summers (1988), Gibbons and Katz (1992) and Abowd, Kramaz and Margoli's (1999). Another example occurs in urban economics when considering the extent to which sorting matters for individual and area disparities in wages. See, for example, Duranton and Monastiriotis (2002), Taylor (2006), Dickey (2007), Mion and Naticchioni (2009), Dalmazzo and de Blasio (2007) and Combes, Duranton and Gobillon (2008). Education economics provides another example, where the interest is in finding out the contribution of schools to the variance of pupil test scores (e.g. Kramarz, Machin and Ouazad, 2009). Despite this widespread and growing interest in decompositions by geography, industries, firms, schools and other groups, the alternative methods and their interpretations have not been very clearly articulated in previous literature.

The first part of this article discusses several methods for gauging the contribution of group effects to the dispersion of outcomes across individuals, where these group effects are treated as fixed effects.¹ To make the discussion concrete, we focus on wages as the individual outcome, and consider the role of individual characteristics and a group effect associated with the labour market area of the individual. The methods we discuss are based on decomposing the variance of individual wages into between-group and within-group components in order to estimate the share of the individual variance attributable to group effects. These methods are easily extended to consider the share of *group* level variation attributable to group, as opposed to individual, effects. We consider the relationships between these different methods and how their results might be interpreted. The approaches that we consider are not novel, but the issues surrounding their application have not received careful attention in the literature. The second part of the article applies these decompositions to data on employee wages in Britain. The results here illustrate that area effects make only a small contribution to overall wage inequality, and that observed disparities between mean wages in different areas are primarily due to sorting.

II. Variance decompositions

Assume we have data on wages w_i and individual characteristics x_i and that individuals are located in one of J areas.² Consider the (Mincerian) linear regression model for individual log wages:

¹ By this we mean that we will allow the effects of area and individuals to be correlated. Researchers in some disciplines like to use multi-level/hierarchical models (Goldstein, 2010) to provide variance decompositions, but these estimators assume that the various levels (e.g. areas, individuals) are uncorrelated random effects, which does not seem at all desirable in many contexts.

² We assume that individuals live and work in the same area.

$$\ln(w) = x'\beta + d'\gamma + \varepsilon, \quad (1)$$

where ε are random errors, d is a $J \times 1$ column vector with j th element equal to one if the worker is in area j and γ is a $J \times 1$ vector of parameters. Variable $d'\gamma$ is the 'area effect' for workers in area j . The vector of individual characteristics x could consist of observable variables (e.g. age), or be a vector of individual dummy variables, so that $x'\beta$ is an 'individual effect'.³ Unless needed, we suppress individual i (and time t) subscripts and all constant terms (so variables are in deviation-from-mean form). It simplifies notation to define $\delta \equiv d'\gamma$ and $\chi \equiv x'\beta$ and rewrite equation (1) as:

$$\ln w = \chi + \delta + \varepsilon. \quad (2)$$

Our aim is to assess the contribution of area effects $\delta \equiv d'\gamma$ to the variance of $\ln w$. The area effect associated with an area j gives the expected gain in log wages that a randomly chosen individual could expect from being assigned to area j (relative to some arbitrary baseline area).

Decomposing the total variance

Estimation of equation (1) by OLS (or some equivalent method) gives coefficient estimates $\hat{\beta}$, $\hat{\gamma}$ and residuals $\hat{\varepsilon}$. We can decompose the variance of $\ln w$ as follows:

$$\text{var}(\ln w) = \text{var}(\hat{\chi} + \hat{\delta} + \hat{\varepsilon}) = \text{var}(\hat{\chi}) + \text{var}(\hat{\delta}) + 2\text{cov}(\hat{\chi}, \hat{\delta}) + \text{var}(\hat{\varepsilon}), \quad (3)$$

where $\text{var}(\cdot)$ is the sample variance, $\text{cov}(\cdot)$ the sample covariance, $\hat{\delta} \equiv d'\hat{\gamma}$ and $\hat{\chi} \equiv x'\hat{\beta}$ and $\hat{\varepsilon} \equiv y - d'\hat{\gamma} - x'\hat{\beta}$. Dividing through by $\text{var}(\ln w)$ gives:

$$\frac{\text{var}(\hat{\chi})}{\text{var}(\ln w)} + \frac{\text{var}(\hat{\delta})}{\text{var}(\ln w)} + \frac{2\text{cov}(\hat{\chi}, \hat{\delta})}{\text{var}(\ln w)} + \frac{\text{var}(\hat{\varepsilon})}{\text{var}(\ln w)} = 1. \quad (4)$$

Using $R^2(\ln w; x, d)$ to denote the R -squared from a regression of $\ln w$ on x and d we get:

$$\frac{\text{var}(\hat{\chi})}{\text{var}(\ln w)} + \frac{\text{var}(\hat{\delta})}{\text{var}(\ln w)} + \frac{2\text{cov}(\hat{\chi}, \hat{\delta})}{\text{var}(\ln w)} = R^2(\ln w; x, d), \quad (5)$$

Note that $R^2(\ln w; x, d)$ provides one potential estimate of the contribution of area effects to the variance in $\ln w$. However, interpretation of this overall R -squared statistic in such a way requires the assumption that *all* variance in individual characteristics (including within-area variation) is *caused* by area effects, implying for example that area effects determine all differences in individual skills between areas. In most contexts, this assumption is surely too extreme, because individuals sort across areas through migration, and area differences in skills are at least in part driven by this sorting. The principle advantage of having data on worker characteristics and workplace location, and panel data that follows individuals over time as they move across areas, is, of course, that we can use equation (5) to obtain more informative estimates of the contributions of area vs. individual effects in $R^2(\ln w; x, d)$. The next sections set out the alternative ways of doing this in greater detail.

³ A more general extension to panel data is simple and discussed in Appendix A.

The raw variance share

A more plausible upper estimate of area share is obtained if we assume area causes any differences in *average* individual characteristics in each area. Then, the contribution of area is:

$$\frac{\text{var}(\tilde{\delta})}{\text{var}(\ln w)} = R^2(\ln w; d), \quad (6)$$

where $\tilde{\delta} \equiv d'\tilde{\gamma}$, $\tilde{\gamma}$ is the coefficient vector and $R^2(\ln w; d)$ the R -squared from the regression of $\ln w$ on area dummies d , with no control for x [contrast $\hat{\delta} \equiv d'\hat{\gamma}$ from equation (1)]. We refer to this as the raw variance share (RVS). RVS is simply the between-group variance in log wages divided by the total variance in log wages. If average individual characteristics vary by area and are correlated with, but not *caused* by, area effects then $\tilde{\gamma}$ is an upward biased estimate of γ and RVS provides an over-estimate of the area share of variance.

The correlated variance share

A second possibility, based on the decomposition of total variance in equation (4) is:

$$\frac{\text{var}(\hat{\delta})}{\text{var}(\ln w)}, \quad (7)$$

where $\hat{\delta} \equiv d'\hat{\gamma}$ comes from a regression of $\ln w$ on area dummies d and individual characteristics x [i.e. OLS estimation of equation (1)]. That is, the area share is the ratio of the variance of area effects to the variance of log wage. This seems like an intuitive measure of the contribution of area effects. Note, however, that interpretation is complicated because even though equation (1) controls for individual x , area effects $\hat{\delta}$ will still be correlated with x and $\hat{\chi}$ if there is associative sorting across areas. This means that $\text{var}(\hat{\delta})/\text{var}(\ln w)$ excludes the direct contribution of sorting [which enters through the covariance term in equation (5)], but captures any indirect effect that sorting may have on the *variance* of the area effects themselves. That is, it captures the contribution of area effects *including* those induced by area composition but not the effects, if any, that area has in determining individual characteristics. We show this in more detail below in the section on Interpretations. For this reason, we refer to this measure as the correlated variance share (CVS). The ratio $\text{var}(\hat{\chi})/\text{var}(\ln w)$ provides an analogous CVS estimate of the variance share attributable to individual characteristics, conditional on the area effects.

Uncorrelated variance share

We can also estimate the contribution of the components of area effects that are uncorrelated with individual characteristics. There are a number of equivalent methods for doing this that give a statistic that is often referred to as the semi-partial R -squared [and by Borcard and Legendre (2002) as 'Fraction a' partitioning]. It is also related to the partial R -squared, as shown in Appendix B. We refer to this as the uncorrelated variance share (UVS). Analogous methods can be used to estimate the corresponding UVS attributable to individual characteristics, but we describe only the process for area effects below.

The UVS measures the amount of variation in log wage that is explained by the part of the area effect that is uncorrelated with individual characteristics that are included in x . Appendix C shows that these three methods of calculating the UVS are equivalent. Appendix D also provides a decomposition of total variance similar to equation (4) but based on the UVS rather than the CVS.

These equivalent methods for estimating UVS are:

R-squared method

This method involves three steps: (a) regress $\ln w$ on x and d to get $R^2(\ln w; x, d)$ to obtain an R -squared that estimates the share of overall variance explained by both individual characteristics and area effects; (b) regress $\ln w$ on *just* x alone to get $R^2(\ln w; x)$, which is the share of overall variance explained by just the individual characteristics; (c) calculate $UVS = R^2(\ln w; x, d) - R^2(\ln w; x)$ to obtain the share attributable to the components of that are correlated with area effects but uncorrelated with individual characteristics

ANOVA (partial sum of squares) method

An alternative is to compute the Regression Sum of Squares and Total Sum of Squares, which are routinely computed by ANOVA software, and calculate:

$$UVS = R^2(\ln w; x, d) - R^2(\ln w; x) = \frac{RSS(\ln w; x, d) - RSS(\ln w; x)}{TSS(\ln w)}, \quad (8)$$

where RSS is regression sum of squares and TSS the total sum of squares, so the numerator is the partial sum of squares of d .

Partitioned regression method

(a) Regress $\ln w$ on x and d and obtain the regression predictions of area effects $\hat{\delta} \equiv d' \hat{\gamma}$ and individual effects $\hat{\chi} \equiv x' \hat{\beta}$; (b) regress $\hat{\delta}$ on $\hat{\chi}$ and obtain a regression coefficient estimate $\hat{\theta}$ and the uncorrelated residual area components $\hat{\omega} = \hat{\delta} - \hat{\theta} \hat{\chi}$; (c) regress $\ln w$ on the residual $\hat{\omega}$ from (b) and calculate $UVS = R^2(\ln w; \hat{\omega})$.

Balanced variance share

To consider the explanatory power of individual characteristics and group membership, Abowd *et al.* (1999) suggest reporting the standard deviation (across workers) of the *effect* of a variable and the correlation of the variable with wages as yet another variance analysis. The effect of a variable is calculated by multiplying the coefficient on that variable by the value of that variable for each observation. The variable is said to have large explanatory power when the effect has a large standard deviation and is highly correlated with wages. Combes *et al.* (2008) perform such an analysis to consider the importance of individual vs. area in explaining area wage disparities.

These two components (the standard deviation of the effect and the correlation with wages) can be used to construct another variance share measure. Using $\hat{\delta} \equiv d' \hat{\gamma}$, $\hat{\chi} \equiv x' \hat{\beta}$,

$s(\cdot)$ to denote sample standard deviation and $r(\cdot)$ the Pearson correlation coefficient note that:

$$\begin{aligned} \frac{s(\hat{\chi})}{s(\ln w)}r(\ln w, \hat{\chi}) + \frac{s(\hat{\delta})}{s(\ln w)}r(\ln w, \hat{\delta}) &= \frac{\text{cov}(\ln w, \hat{\chi})}{\text{var}(\ln w)} + \frac{\text{cov}(\ln w, \hat{\delta})}{\text{var}(\ln w)} \\ &= \frac{\text{var}(\hat{\chi})}{\text{var}(\ln w)} + \frac{\text{var}(\hat{\delta})}{\text{var}(\ln w)} + \frac{2\text{cov}(\hat{\chi}, \hat{\delta})}{\text{var}(\ln w)} = R^2(\ln w; x, d). \end{aligned} \quad (9)$$

Hence:

$$\frac{s(\hat{\delta})}{s(\ln w)}r(\ln w, \hat{\delta}) = \frac{\text{cov}(\ln w, \hat{\delta})}{\text{var}(\ln w)} = \frac{\text{var}(\hat{\delta}) + \text{cov}(\hat{\chi}, \hat{\delta})}{\text{var}(\ln w)} \quad (10)$$

provides this other measure of the area share of the variance of $\ln w$. An analogous equation to equation (10) would yield a measure of the individual share in the variance of $\ln w$. The decomposition simply apportions the sorting component, $2\text{cov}(\hat{\chi}, \hat{\delta})$ in equation (4), equally to area effects and to individual characteristics, so we call this the balanced variance share (BVS). Note that there is no obvious justification for splitting the covariance in this way, apart from to guarantee that the area and individual variance shares add up to the total R -squared $R^2(\ln w; x, d)$.

Contribution of area effects to area disparities

So far, we have focused on deriving the contribution of area effects to overall wage disparities. We can, however, use the RVS, CVS, UVS or BVS to back out the contribution of area effects to *area*, rather than overall individual disparities. The ratio CVS/RVS or $\text{var}(\tilde{\delta})/\text{var}(\hat{\delta})$ from equations (6) and (7) gives the contribution of area excluding the direct contribution of sorting. The ratio UVS/RVS or $[R^2(\ln w; x, d) - R^2(\ln w; x)]/R^2(\ln w; d)$ gives the contribution of area excluding both the direct and indirect contribution of sorting. BVS/RVS provides a final measure which adds in an additional contribution attributable to the covariance. The interpretations follow directly from the properties of CVS, UVS and BVS.

III. Comparisons and interpretation

We have outlined several methods for calculating the individual and group shares of variance. In this section, we consider the relative magnitudes of the measures and discuss how their interpretation depends on the underlying determinants of individual differences in outcomes.

Magnitudes

In Appendix C [equation (C4)], we show that $R^2(\ln w; x) = (1 + 2\hat{\theta} + \hat{\theta}^2)\text{var}(\hat{\chi})/\text{var}(\ln w)$ where $\hat{\theta}$ is the coefficient from the regression of $\hat{\delta} \equiv d'\hat{\gamma}$ on $\hat{\chi} \equiv x'\hat{\beta}$. An analogous derivation gives:

$$\begin{aligned} \text{RVS} &= R^2(\ln w, d) = (1 + 2\hat{\rho} + \hat{\rho}^2)\text{var}(\hat{\delta})/\text{var}(\ln w) \\ &= ((1 + \hat{\rho}^2)\text{var}(\hat{\delta}) + 2\text{cov}(\hat{\chi}, \hat{\delta}))/\text{var}(\ln w), \end{aligned} \tag{11}$$

where $\hat{\rho}$ is the coefficient from a regression of $\hat{\chi} \equiv x' \hat{\beta}$ on $\hat{\delta} \equiv d' \hat{\gamma}$. Comparing with equation (7), it is immediate that $\text{RVS} = (1 + 2\hat{\rho} + \hat{\rho}^2)\text{CVS}$. Note that unless $-2 < \hat{\rho} < 0$ or $\text{var}(\hat{\delta}) = 0$, $\text{RVS} > \text{CVS}$. That is, when individual characteristics are positively correlated with area effects (i.e. $\hat{\rho} > 0$), RVS is greater than CVS. In other words, the R -squared from a regression of $\ln w$ on area dummies will attribute to area effects some of the contribution of individual characteristics. The same is true when sorting is negative, and strong relative to the variance of area effects (i.e. $\hat{\rho} = \text{cov}(\hat{\chi}, \hat{\delta})/\text{var}(\hat{\delta}) < -2$). In the range $-2 < \hat{\rho} < 0$, negative sorting tends to cancel out the contribution of area effects when estimated using RVS. When $\hat{\rho} = -1$, negative sorting exactly offsets area effects, so regression of $\ln w$ on area dummies would yield an RVS of zero, whereas CVS would be positive.

Appendix C [equation (C5)] shows that the uncorrelated variance share is:

$$\text{UVS} = \frac{\text{var}(\hat{\delta}) - \hat{\theta}^2 \text{var}(\hat{\chi})}{\text{var}(\ln w)}, \tag{12}$$

where $\hat{\theta}$ is the coefficient from the sample regression of $\hat{\delta} \equiv d' \hat{\gamma}$ on $\hat{\chi} \equiv x' \hat{\beta}$. Comparing the numerators in the UVS in equation (12) and CVS in equation (7):

$$\text{var}(\hat{\delta}) \geq \text{var}(\hat{\delta}) - \hat{\theta}^2 \text{var}(\hat{\chi}) \Rightarrow \text{CVS} \geq \text{UVS}, \tag{13}$$

with equality if $\hat{\delta}$ is uncorrelated with $\hat{\chi}$. Therefore the CVS gives an area share that is at least as large as that obtained by the UVS.

As we said before, the BVS adds $\text{cov}(\hat{\chi}, \hat{\delta})$ to the CVS. Hence, with positive covariance between $\hat{\chi}$ and $\hat{\delta}$, the BVS for area share is bigger than the CVS. From equation (11) it is also clear that the RVS for area share is bigger than BVS, when there is positive sorting (covariance). With negative sorting the BVS for area share is smaller than CVS, but it may be larger or smaller than UVS or RVS. Finally, all methods give the same result if $\hat{\chi}$ and $\hat{\delta}$ are uncorrelated. Naturally, these comparisons follow through to the decomposition of the variance of $\ln w$ into the components attributed to area plus observable characteristics, vs. unobservable factors. For example, with positive sorting, adding up the BVS for area [$\text{BVS}(\hat{\delta})$] and BVS for individual characteristics $\text{BVS}(\hat{\chi})$ minimises the share of the variance attributed to unobservables (i.e. it is the standard R -squared) and attributes as much as possible to $\hat{\chi}$ and $\hat{\delta}$. At the other extreme, adding $\text{UVS}(\hat{\delta})$ to $\text{UVS}(\hat{\chi})$ provides a more conservative estimate of the share attributable to area and individual characteristics based on the orthogonal components of $\hat{\chi}$ and $\hat{\delta}$ alone, and so assigns a greater share of $\ln w$ to unknown components.

Given that all of the methods give a different answer when there is associative sorting of individuals across groups it is natural to ask how we should interpret the different calculations and whether we should prefer any specific method. This depends on the underlying factors determining individual and area outcomes as we now show.

Interpretations

In order to gain some understanding of what contributes to each of these variance shares, it is helpful to fully specify a system of linear equations which determine individual wages, area effects and individual characteristics. Continuing to use $\delta \equiv d'\gamma$ and $\chi \equiv x'\beta$:

Individual wages:

$$\ln w = \chi + \delta + \varepsilon \quad (14)$$

Area effects:

$$\delta = \bar{\chi}\lambda + \mu \quad (15)$$

Individual characteristics:

$$\chi = \delta\varphi + \nu \quad (16)$$

The equation for individual wages is identical to that used earlier in Section II. As shown in equation (15) we allow area effects to be determined by average area composition (where the mean of χ is taken across individuals, within areas) and by exogenous variation μ (e.g. climate). Equation (16) allows individual characteristics to be determined by area and by exogenous individual specific variation ν (e.g. innate ability). Specifying equation (16) in terms of χ rather than x simplifies later notation.

Averaging equation (16) at the area level gives area-mean individual characteristics $\bar{\chi} = \delta\varphi + \bar{\nu}$, and it is the correlation between these and area effects δ that biases estimations of the relative contributions of individual and area effects. Part of this correlation arises because of the interdependency between χ and δ explicit in equation (15) and (16) when $\lambda \neq 0$ and/or $\varphi \neq 0$. However, there is an additional correlation, present even when $\lambda = 0$ and $\varphi = 0$ if $\text{cov}(\mu, \bar{\nu}) \neq 0$. That is, if exogenous area effects (μ) and area-mean exogenous individual characteristics ($\bar{\nu}$) are correlated through sorting processes that are otherwise unrelated to wages (e.g. if higher skill people prefer warmer climates).

Note that all this assumes that the underlying sources of variance across areas are the exogenous additive components ($\mu, \bar{\nu}$) and rules out multiple equilibria that arise in nonlinear theoretical models. Even in this very restricted world, it turns out to be difficult to provide firm statements about the share of area effects or whether RVS, BVS, CVS or UVS provides the most appropriate empirical measure.

In the Appendix E, we show that equations (14), (15) and (16) can be used to get expressions for the variance of area effects, mean individual characteristics and their covariance in terms of the underlying structural errors and parameters:

$$\text{var}(\delta) = \{(\lambda^2 \text{var}(\bar{\nu}) + \text{var}(\mu) + 2\lambda \text{cov}(\bar{\nu}, \mu)) / (1 - \lambda\varphi)^2\}, \quad (17)$$

$$\text{var}(\bar{\chi}) = \{\text{var}(\bar{\nu}) + \varphi^2 \text{var}(\mu) + 2\varphi \text{cov}(\bar{\nu}, \mu)\} / (1 - \lambda\varphi)^2, \quad (18)$$

$$\text{cov}(\bar{\chi}, \delta) = \{\varphi \text{var}(\mu) + \lambda \text{var}(\bar{\nu}) + (1 + \lambda\varphi) \text{cov}(\mu, \bar{\nu})\} / (1 - \lambda\varphi)^2. \quad (19)$$

If area affects individual characteristics ($\varphi \neq 0$) equations (18) and (19) show that the exogenous component of area effects (μ) enters into $\text{var}(\bar{\chi})$ and $\text{cov}(\bar{\chi}, \delta)$. Similarly, if area composition determines area effects ($\lambda \neq 0$), equations (17) and (19) show that the exogenous component of individual characteristics ($\bar{\nu}$) enters into $\text{var}(\delta)$ and $\text{cov}(\bar{\chi}, \delta)$. (Positive) sorting, $\text{cov}(\mu, \bar{\nu}) > 0$ increases $\text{var}(\delta)$ if area effects are increasing in $\bar{\chi}$ ($\lambda > 0$) and increases $\text{var}(\bar{\chi})$ if

individual characteristics are increasing in δ ($\varphi > 0$). Positive sorting also increases $\text{cov}(\bar{\chi}, \delta)$, with the increase greater if both $\varphi > 0$ and $\lambda > 0$. These results regarding $\bar{\chi}$ carry through to the distribution of individual characteristics χ because $\text{cov}(\chi, \delta) = \text{cov}(\bar{\chi}, \delta)$ and $\text{var}(\chi) = \text{var}(\bar{\chi}) + \text{var}(\zeta)$, where ζ represents all the within-area individual components of wages. In short, $\text{var}(\bar{\chi})$, $\text{var}(\delta)$ and $\text{cov}(\bar{\chi}, \delta)$, which are observable, cannot be attributed to one or other of the potentially unobservable exogenous components (μ, ν) that are the underlying sources of variance across areas in the simple system represented in equations (14), (15) and (16).

Turning to the four variance shares, it is useful to repeat the formulas:⁴

$$\begin{aligned} \text{RVS} &= \frac{(1 + \hat{\rho}^2)\text{var}(\hat{\delta})}{\text{var}(\ln w)} + \frac{2\text{cov}(\hat{\chi}, \hat{\delta})}{\text{var}(\ln w)}, & \text{BVS} &= \frac{\text{var}(\hat{\delta})}{\text{var}(\ln w)} + \frac{\text{cov}(\hat{\chi}, \hat{\delta})}{\text{var}(\ln w)}, \\ \text{CVS} &= \frac{\text{var}(\hat{\delta})}{\text{var}(\ln w)}, & \text{UVS} &= \frac{\text{var}(\hat{\delta}) - \hat{\theta}^2 \text{var}(\hat{\chi})}{\text{var}(\ln w)}. \end{aligned}$$

It can be seen that equations (17)–(19) can, in principle, be used to express the four variance shares in terms of the underlying structural components. It should be clear, however, that even in this very basic linear world there is no simple relationship between these variance shares and the underlying structural components. Which variance share is most appropriate depends on a complex interplay between: (a) the underlying structural component in which we are interested; (b) the magnitudes and signs of the coefficients determining area effects and individual characteristics and (c) sorting.

One simple case occurs when composition does not cause area effects ($\lambda = 0$), in which case CVS gives the variance share of the area shocks μ . Correspondingly, when area does not cause changes in individual characteristics ($\rho = 0$), $\text{var}(\hat{\chi})/\text{var}(\ln w)$ gives the share of individual characteristics in log wages. Note too that when $\rho = \lambda = 0$, $\text{cov}(\chi, \delta)$ provides an unbiased estimate of the degree of sorting but otherwise combines the covariance caused by sorting and by the causal influence of area on characteristics (and vice-versa). In all other contexts, pinning down the contribution of specific sources of area-level variations is not possible, but as discussed above RVS, BVS, CVS and UVS can at least be used to provide a bounding exercise to guide us as to the likely contributions.

IV. Application: wage disparities in Britain

In this section, we use the four variance shares to look at the contribution of individual and area effects to wage variation in Britain. We improve on existing UK work on regional inequalities (Duranton and Monastiriotis, 2002; Taylor, 2006; Dickey, 2007) by working with functional labour market areas (rather than administrative boundaries) at a smaller spatial scale and by using panel data to control for unobserved individual characteristics. Outside of the UK, a number of existing studies have already used individual data to study

⁴ In these expressions, remember that $\hat{\rho}$ is the coefficient from the regression of $\hat{\chi}$ on $\hat{\delta}$ [i.e. $\text{cov}(\hat{\chi}, \hat{\delta})/\text{var}(\hat{\delta})$] not the ‘structural’ coefficient ρ in equation (16)] and $\hat{\theta}$ is the coefficient from the regression of $\hat{\delta}$ on $\hat{\chi}$ [i.e. $\text{cov}(\hat{\chi}, \hat{\delta})/\text{var}(\hat{\chi})$] not the ‘structural’ coefficient λ in equation (15)]. We ignore this complication in the text, but note that the problems of linking variance shares to structural components persist even in the unlikely case that sample estimates equal their population values.

spatial sorting and agglomeration economies. See, for example, Mion and Naticchioni (2009), Dalmazzo and Blasio (2007) and Combes *et al.* (2008). This article is most closely related to the last of these but pays more attention to the issue of the share of disparities attributable to area effects using the insights developed above.

We use data for 1998–2008 from the Annual Survey of Hours and Earnings (ASHE) and its predecessor the New Earnings Survey (NES). NES/ASHE are constructed by the Office of National Statistics (ONS) based on a 1% sample of employees on the Inland Revenue PAYE register for February and April (Office for National Statistics, 2011). The sample is of employees whose National Insurance numbers end with two specific digits and workers are observed for multiple years (up to 11 years given the time span of our sample). The sample is replenished as workers leave the PAYE system (e.g. to retirement) and new workers enter it (e.g. from school). NES/ASHE include information on occupation (Standard Occupation Classification, SOC), industry (Standard Industrial Classification, SIC), whether the job is private or public sector, the workers age and gender and detailed information on earnings. We use basic hourly earnings as our measure of wages. NES/ASHE do not provide data on education but information on occupation works as a good proxy for our purposes. NES/ASHE provide national sample weights but as we are focused on sub-national travel to work areas (TTWA) data, we do not use them in the results we report below.

ASHE provides information on individuals including their home and work postcodes, while the NES only reports work postcodes.⁵ We use the National Statistics Postcode Directory (NSPD) to assign workers to TTWA using their work postcode, allowing us to use the whole NES/ASHE sample over our study period. Given the way TTWA are constructed (so that 80% of the resident population also work within the area) the work TTWA will also be the home TTWA for the majority of workers. To ensure reasonable sample sizes, our analysis divides Britain into 157 ‘labour market areas’ of which 79 are single ‘urban’ TTWA and 78 are ‘rural areas’ created by combining TTWAs.⁶

We derive estimates of RVS, CVS, UVS and BVS, from regression analysis of individual wages, based on equation (1) above, but adapted to include individual fixed effects and time dummies to take advantage of the panel dimension of our data:

$$\ln w_{ijt} = \alpha_i + \lambda_t + x'_{it}\beta + d'_{jt}\gamma + \varepsilon_{ijt}, \quad (20)$$

where j indexes areas, α_i are individual fixed effects, λ_t are time dummies, x_{it} is a vector of time varying individual characteristics, and everything else is as in equation (14). β captures the impact of individual characteristics controlling for area. Likewise, γ captures the impact of area controlling for both time invariant unobserved individual characteristics (α_i) and time varying observed individual characteristics (x_{it}). These parameter estimates allow us to obtain estimates of $\hat{\delta} = d'\gamma$ and $\hat{\chi} = x'\hat{\beta}$ from which we can calculate RVS, CVS, UVS and BVS.

⁵The NES data goes back to 1970. However, work postcodes are only available since 1997 and are required in order to provide a consistent series of labour market areas, because the TTWA definitions provided in the NES/ASHE data change over time. At the time of the analysis the 1997 work postcode data was unreliable, which is the reason our panel starts in 1998, and we focus only on recent patterns of area disparities.

⁶We reached this classification in three steps: (a) we identified the primary urban TTWAs as TTWA centred around, or intersecting urban-footprints with populations of 100,000 plus; (b) we identified TTWA with an annual average NES/ASHE sample size greater than 200 as stand-alone non-primary-urban TTWA (e.g. Inverness); and (c) we grouped remaining TTWA (with sample sizes below 200) into contiguous units (e.g. North Scotland).

Including individual fixed effects means that identification of the area fixed effects comes from movers across areas. As usual, we cannot rule out the possibility that shocks to ε_{it} are correlated with the decision to move so the area effect estimates need to be interpreted with some caution. In the absence of random allocation (or a policy change that as good as randomly assigns people) tracking individuals and observing the change in wages experienced when they move between areas is the best we can do to identify causal area effects.⁷ Estimating the area effects from movers in our data means we need to assume fixed-across-time area effects, given the frequency of moves and the length of the panel, and the number of TTWAs.⁸ We shall show, however, that dropping individual fixed effects and controlling only for observables provides area effects estimates that are very stable over time. Given this, and as individual unobserved characteristics are important for explaining wages, our preferred specification incorporates individual effects but assumes area effects are fixed across time. Some basic descriptive statistics are provided in Appendix F.

We start by estimating equation (20) with an increasing set of control variables to show how allowing for sorting across areas affects the magnitude of estimated area effects. To summarize the distribution of effects, we first report the percentage change in wages when moving between different parts of the area effects distribution: the minimum to maximum and to mean, mean to maximum, the 10th to the 90th percentile and the 25th to the 75th percentile. Table 1 reports results based on several different specifications. The first row reports results from equation (20) when only including time dummies (which gives the upper bound estimates of area effects). The second row reports results when the observable variables are a set of age dummies, a gender dummy and a set of one digit occupation dummies. The third row uses a set of age dummies, a gender dummy, two digit occupation dummies, industrial dummies (three digit SIC) and dummies for public sector workers, part time workers and whether the worker is part of a collective agreement. Results using individual fixed effects are reported in rows four and five. In row four, we simply include year dummies and individual effects. Row five uses individual effects, year dummies, age dummies and one digit occupation dummies.

If we ignore the role of sorting (i.e. do not control for α_i and x_{it}) then the differences in area average wages look quite large. Moving from the worst to the best area, average wages increase by just over 60%; from the minimum to the mean by a little under 20%; and from the mean to the maximum by just over 35%. Of course the minimum and maximum represent extremes of the distribution. The move from the 10th to the 90th percentile sees average wages increase by 22%, from the 25th to 75th percentile by just over 10%. Introducing a limited set of observable characteristics (row 2) to control for sorting substantially reduces estimated area differences. A larger set of individual characteristics (row 3) reduces the estimated differences further as does allowing for individual fixed effects, with or without additional observable characteristics (rows 4 and 5). Interpreting observed area differences as area effects considerably overstates the impact on wages that occurs as individuals move

⁷ Note, that we track individuals in the panel for up to 10 years, so the wage changes we observe for individuals moving between areas are not just the one-off changes when the move occurs, but include longer-run changes (e.g. due to differences in wage growth rates between areas).

⁸ Theoretically, we could still allow for such year on year changes in TTWA specific effects, but estimating them requires movers in and out of all areas in every year which turns out to be too demanding given our sample sizes.

TABLE 1

Distribution of area effects ($\hat{\gamma}$)

<i>Controls</i>	<i>Min–max</i> (%)	<i>Min–Mean</i> (%)	<i>Mean–max</i> (%)	<i>p90–p10</i> (%)	<i>p75–p25</i> (%)
Year dummies	61.7	18.2	36.8	22.0	10.6
+ age, gender, soc-1	36.6	9.0	25.4	12.7	6.4
+ age, gender, soc-2, sic-3, public, union, part-time	29.0	6.5	21.0	9.7	4.2
Individual fixed effects and year dummies	20.3	7.6	11.8	8.9	4.4
+ age, soc1	17.3	5.9	10.8	7.3	3.8

Notes: Authors own calculations based on NES/ASHE. Specifications are described in detail in the text (soc-1, soc-2 signify one and two digit occupation; sic-3 signifies three digit industry code). Results for 157 areas. Based on 1,511,163 observations (305,752 individuals)

from bad to good areas. Specifically, across the comparisons we report, ignoring the role of sorting overstates area effects by a factor of three.

It is worth noting that the comparison with the figures reported in Combes *et al.* (2008) suggests that the sorting is even more important in driving area disparities in Britain than it is in France. Their nearest equivalent to Table 1 above suggests that the max/min difference without any controls for individual effects is 74% (their table 3, p. 730) falling to 38% when sorting of workers across areas is taken into account. Our figures show max/min differences of 61.7% falling to 17.3% once we account for sorting. Similarly, their 90/10 differences start at 21% and fall to 14%–16% after controlling for sorting, while in Britain the corresponding figures from Table 1 are 22% falling to just 3.8%.⁹ Evidently, sorting matters a lot in Britain, with a considerably larger share of the general disparity in wages between areas explained by sorting than in France. As discussed above, the contribution of area effects to overall (and area) wage disparities depends not only on the size of specific effects but also on the overall distribution of area effects and on the distribution of individuals across those areas. We suggested four alternative variance decompositions for measuring this contribution. Table 2 reports results for the contribution of area differences to overall and area wage disparities using these four measures. We calculate them from regression specifications using the same individual characteristics as in Table 1. For comparison, the second panel of the table reports the contribution of the individual characteristics using the different measures while the third panel reports the contribution of area effects to area disparities. In each panel, we also report the standard deviation of the estimated area effects ($\hat{\delta}$) or individual effects ($\hat{\gamma}$) and their correlation with ln wages, to mimic the presentation given in Combes *et al.* (2008) Table 2 for France. As discussed above in equation (10), the BVS can be derived by dividing the standard deviation of the estimated effect by the standard deviation of ln wages and multiplying the result by the correlation between the estimated effect and ln wages.

⁹They report 14% in Table 3 but 16% in Table 4 on p. 731, the difference due to a slightly different method of calculation.

TABLE 2
Variance decomposition

	Year dummies only (RVS)	+ age, gender, soc-1	+ age, gender, soc-2, sic-3, public, union, pt	Individual fixed effects and year dummies	+ age, soc-1
	(1)	(2)	(3)	(4)	(5)
<i>Area variance share</i>					
BVS($\hat{\delta}$)	6.0%	3.2%	2.4%	1.1%	0.9%
CVS($\hat{\delta}$)	6.0%	2.8%	2.0%	0.7%	0.6%
UVS($\hat{\delta}$)	6.0%	2.7%	1.3%	0.1%	0.1%
sd($\hat{\delta}$)		0.088	0.075	0.046	0.041
r($\hat{\delta}$, ln w)		0.188	0.169	0.126	0.120
<i>Individual variance share</i>					
BVS($\hat{\chi}$)		57.7%	75.9%	86.8%	88.3%
CVS($\hat{\chi}$)		57.4%	75.5%	86.4%	87.9%
UVS($\hat{\chi}$)		54.0%	73.0%	83.6%	84.9%
sd($\hat{\chi}$)		0.398	0.457	0.489	0.493
r($\hat{\chi}$, ln w)		0.762	0.873	0.933	0.941
<i>Area share of area disparities</i>					
BVS($\hat{\delta}$)/RVS		53.2%	40.4%	18.3%	15.9%
CVS($\hat{\delta}$)/RVS		47.5%	34.1%	12.6%	10.4%
UVS($\hat{\delta}$)/RVS		45.8%	22.3%	1.3%	1.1%

Notes: Authors own calculations based on NES/ASHE. Specifications are described in detail in the text (soc-1, occ-2 signify one and two digit occupation; sic-3 signifies three digit industry code; pt signifies part-time). Based on 1,511,163 observations (305,752 individuals).

The top panel of Table 2, column 1 shows that, even if we ignore the effects of sorting, and simply consider raw area differences in mean log wages, labour market areas only explain 6% of the overall variation in wages across individuals (remember that in the absence of controls for any individual characteristics the variance share of area effects is the same whichever way it is calculated). Moving across the columns in Table 2, panel 1, we progressively add controls for individual characteristics, culminating in a specification with individual fixed effects in columns 4 and 5. Looking down rows within the top panel shows how the variance share changes according to the way we estimate it – BVS, CVS and UVS. As discussed in section III, the correlated variance share provides an upper bound to the combined contribution of exogenous area effects *plus* those induced by area composition (i.e. the area mean individual characteristics). The uncorrelated variance share provides a lower bound to the contribution of exogenous area effects alone. The balanced variance share could be larger or smaller than the correlated variance share, depending on whether there is positive or negative assortative matching of individuals to areas. It is evident from the fact that BVS > CVS > UVS in Table 2, that we have positive sorting across areas. On any of these measures, the contribution of area effects is around 3% or less when controlling for basic observable characteristics (columns 2 and 3) and around 1% or less when controlling for unobservable individual characteristics (columns 4 and 5). In short, the most striking finding from Table 2 is that area effects only play a small role in explaining overall individual wage disparities.

We have shown (in section III) that it is not, in general, possible to determine from these variance shares the exact contribution of sorting, vs. the influence of area composition on area effects, vs. the influence of area effects on area composition. However, comparison of the CVS ($\hat{\delta}$) and UVS($\hat{\delta}$) measures is partly revealing. Looking at the gap between the UVS and CVS in our preferred specifications with individual fixed effects in columns 4 and 5 of panel 1 suggests that area effects arising from area composition account for over half of everything that can be attributed to area effects. To see this, note that $CVS(\hat{\delta}) - UVS(\hat{\delta})$ is more than 50% of either $CVS(\hat{\delta})$ [or even $BVS(\hat{\delta})$], in columns 4 and 5. Even so, this is still a trivial share of the overall individual disparity in wages – roughly 0.5% in our preferred specification [$CVS(\hat{\delta}) - UVS(\hat{\delta})$] in column 5.

In contrast to these results on area effects, the contribution of individual characteristics shown in the second panel of Table 2 is very large. Age, gender and occupation variables alone account for 54–58% of the individual variation in wages. Adding in individual fixed effects drives this share up to between 85% and 88% in the last column, implying that the contribution of individual characteristics is between 100 times (comparing BVS) and 850 times (comparing UVS) bigger than that of area effects. Comparison of $CVS(\hat{\chi})$ and $UVS(\hat{\chi})$ implies that area effects also contribute very little to the variance of individual effects on wages.

Comparing with the results in Combes *et al.* (2008) Table 2, it is again evident that more of the wage dispersion in Britain is attributable to individual characteristics, and less to area effects than is the case in France. Both the standard deviation of area effects (0.14 in France, 0.041 in Britain), and the correlation of these with ln wages (0.34 in France, 0.12 in Britain) are substantially higher in France. Conversely, both the standard deviation of individual effects (0.29 in France, 0.49 in Britain) and the correlation with ln wages (0.80 in France, 0.94 in Britain) are higher in Britain. The BVS for area effects from their work (derived from their table 2) is 13%, whereas the BVS for individual effects is 64%.

The final panel of the table shows that area effects as defined here play a somewhat more important role in explaining the area disparities captured by the simple between-area variance share (the RVS in Table 2). When controlling for basic characteristics, sorting accounts for a little over half of the observed area disparities leaving area effects to account for around 48%. Once controlling for individual fixed effects the upper bound estimate of the contribution of area effects to area disparities is considerably smaller at a little over 10% with the lower bound estimate around 1%. Even so, the dominant contribution to area disparities (one minus the share reported in the final panel of Table 2) must come from individual characteristics, and the way these are distributed across areas.

Table 2 has shown these decompositions based on pooling all years in our data set. The results in Table 3 investigate the stability of the contribution of area effects over time, and shows that this contribution has been stable. For each year, the table reports the contribution of the four area variance shares controlling for the fullest possible set of individual characteristics. For comparison, the first row reports results when pooling across years (taken from Table 2). Repeating other results from Table 2 by year give figures that are similarly stable across years. As discussed above, it is this stability of area effects across time, combined with the importance of individual effects which leads us to view the individual fixed effects constant-across-time area effects specifications reported in the final two columns of Table 2 as our preferred specification.

TABLE 3
Area effects: results by year

	RVS (%)	BVS (%)	CVS (%)	UVS (%)
Pooled	6.0	2.4	2.0	1.3
1998	6.2	2.3	1.9	0.8
1999	6.0	2.2	1.9	0.8
2000	6.0	2.2	1.9	0.8
2001	6.3	2.3	2.0	0.9
2002	6.6	2.7	2.3	1.2
2003	6.6	2.7	2.3	1.3
2004	6.6	2.7	2.3	1.3
2005	6.5	2.5	2.1	1.2
2006	6.2	2.4	2.1	1.2
2007	6.8	2.5	2.1	1.1
2008	6.7	2.5	2.1	1.1

Notes: Based on 1,511,163 observations (305,752 individuals). Contribution for pooled lower than average of years because pooled regressions impose time invariant area effects and coefficients on individual characteristics. The pooled estimate of column 1 controls for year effects. All estimates in columns 2 and 3 control for age, gender, 2-digit occupation, 3-digit SIC, public sector, union and part-time.

V. Conclusions

This article has reviewed a number of ways to decompose variance in wages into the contribution from individual and area specific effects. The article collates the methods present in the literature, discusses how different variance decompositions are related to each other and how they should be interpreted, and finally, highlights that a clear-cut division of variation into components is strictly speaking impossible, and that whatever method a researcher chooses, assumptions and caveats will remain.

In particular, we have discussed decompositions that we named RVS, BVS, CVS and UVS. We have shown why disentangling the contribution of area effects to wage distribution is complex, and why an approach based on bounds such as RVS, BVS, CVS and UVS can only provide guidance to the relative importance of area effects.

The RVS is the simple between-area share of the total variance and provides an overall upper bound for the contribution of area effect on total wage variation. It counts the average differences in worker characteristics across areas as part of the area effect. For Britain, this measure has persistently accounted for roughly 6% of the overall wage variation for over a decade. The CVS excludes the direct contribution of sorting, but captures the effect that sorting may have on the variance of the area effect. We interpreted the CVS as an upper bound to the combined contribution of exogenous area effects plus interactions-based spillovers that are linked to the mean characteristics of individuals in each area. CVS in Britain has been around 2%, but less than 1% if individual fixed unobservable effects are controlled for. The alternative BVS decomposition that has appeared in previous work (e.g. Abowd *et al.*, 1999; Combes *et al.* 2008) yields a slightly higher estimate of the contribution of area effects in our setting, because there is positive assortative matching of individuals to areas (i.e. individual wage effects are correlated with area wage effects). However, in general, this measure is quite hard to interpret, as it mixes up the variance attributable to

areas with the covariance between area effects and area-mean individual effects. We argue that CVS is a more natural indicator. Finally, compared to CVS, UVS further excludes the effects that sorting may have on the variation of the area effects, and such provides a lower bound to the contribution of exogenous area effects. With this measure, the area effects explain 1.5% of wage variation in Britain. If individual fixed effects are included, the effect shrinks to less than 0.1%, and this should be considered as the lower bound of the effect of area on variation of wages. These figures are small relative to those found in France using comparable estimates in Combes *et al.* (2008), where (according to the BVS indicator) area disparities account for 13% of the variance in log wages.

Whichever estimate we use, our general finding is that most of the observed regional inequality in average wage in Britain is explained by ‘sorting’ or ‘people’ rather than ‘places’. Our preferred estimates, which include the individual fixed effects, suggest that the contribution of individual characteristics to variation in wages is between 100 to 850 times larger than the contribution of area effects.

Final Manuscript Received: July 2013

References

- Abowd, J., Kramarz, F. and Margolis, D. (1999). ‘High wage workers and high wage firms’, *Econometrica*, Vol. 67, pp. 251–334.
- Abowd, J., Creecy, R. and Kramarz, F. (2002). *Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data*, Cornell University Working Paper.
- Fisher, R. A. (1933) in a letter to L. Hogben 25 February. Printed in *Natural Selection, Heredity, and Eugenics*, p. 218, J.H. Bennett, Clarendon Press, Oxford. 1983 quoted on <http://www.economics.soton.ac.uk/staff/aldrich/fisherguide/quotations.htm> (accessed date: February 2012)
- Borcard, D. and Legendre, P. (2002). ‘All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices’, *Ecological Modelling*, Vol. 153, pp. 51–68.
- Combes, P. P., Duranton, G. and Gobillon, L. (2008). ‘Spatial wage disparities: Sorting matters!’, *Journal of Urban Economics*, Vol. 63, pp. 723–742.
- Dalmazzo, A. and de Blassio, G. (2007). ‘Social returns to education in Italian local labour markets’, *The Annals of Regional Science*, Vol. 41, pp. 51–69.
- Dickey, H. (2007). ‘Regional earnings inequality in Great Britain: evidence from quantile regressions’, *Journal of Regional Science*, Vol. 47, pp. 775–806.
- Duranton, G. and Monastiriotis, V. (2002). ‘Mind the gaps: the evolution of regional earnings inequalities in the U.K., 1982–1997’, *Journal of Regional Science*, Vol. 42, pp. 219–256.
- Gibbons, R. S. and Katz, L. (1992). ‘Does unmeasured ability explain inter-industry wage differences?’, *Review of Economic Studies*, Vol. 59, pp. 515–535.
- Goldstein, H. (2010). *Multilevel Statistical Models*, 4th edn, Wiley-Blackwell, United Kingdom.
- Kramarz, F., Machin, S. and Ouazad, A. (2009). *What Makes a Test Score? The Respective Contributions of Pupils, Schools and Peers in Achievement in English Primary Education*, Centre for Economics of Education CEEDP0102., London School of Economics, London.
- Krueger, A. and Summers, L. H. (1988). ‘Efficiency wages and the inter-industry wage structure’, *Econometrica*, Vol. 56, pp. 259–93.
- Mion, G. and Naticchioni, P. (2009). ‘The spatial sorting and matching of skills and firms’, *Canadian Journal of Economics*, Vol. 42, pp. 28–55.
- Office for National Statistics. (2011) *Annual Survey of Hours and Earnings, 1997–2009: Secure Data Service Access [Computer File]*. UK Data Archive [distributor], Colchester, Essex. March 2011. SN: 6689.
- Ouazad, A. (2008). *A2REG: Stata module to Estimate Models with Two Fixed Effects*, Statistical Software Components. S456942, Boston College Department of Economics, Boston MA, USA.

Taylor, K. (2006). ‘UK wage inequality: an industry and regional perspective’, *Labour: Review of Labour Economics and Industrial Relations*, Vol. 20, pp. 91–124.

Appendix A: Extension to individual and area fixed effects

All the methods can be generalized to allow for multiple effects $\eta_f = h'_f \sigma_f$, where χ and δ are particular cases of η_f , such that equation (2) becomes:¹⁰

$$\ln w = \sum_f \eta_f + \varepsilon = \sum_f h'_f \sigma_f + \varepsilon; \tag{A1}$$

Then the raw variance share for the contribution of any one of these effects η_f is:

$$\frac{\text{var}(\tilde{\eta}_f)}{\text{var}(\ln w)} = R^2(\ln w, h_f),$$

where $\tilde{\eta}_f \equiv h'_f \tilde{\sigma}_f$, $\tilde{\sigma}_f$ is the coefficient vector and $R^2(\ln w, h_f)$ the R -squared from the regression of $\ln w$ on dummies (or other variables) h_f , with no control for $\eta_{\sim f}$, where the notation $\eta_{\sim f}$ means the set of all η effects not including η_f . The variance decomposition of $\ln w$ is:

$$\sum_f \frac{\text{var}(\hat{\eta}_f)}{\text{var}(\ln w)} + \sum_{f \neq g} \sum_{g \neq f} \frac{\text{cov}(\hat{\eta}_f, \hat{\eta}_g)}{\text{var}(\ln w)} + \frac{\text{var}(\hat{\varepsilon})}{\text{var}(\ln w)} = 1, \tag{A2}$$

where $\hat{\eta}_f \equiv h'_f \hat{\sigma}_f$, $\hat{\sigma}_f$ are the coefficients from OLS estimation of equation (A1) and the correlated variance share = $\text{var}(\hat{\eta}_f)/\text{var}(\ln w)$.

The uncorrelated variance share for any one of these effects η_f is $R^2(\ln w; h_{\sim f}, h_f) - R^2(\ln w; h_{\sim f})$.

Appendix B: Partial R-squared

Another similar measure to UVS is the ‘partial R -squared’. This measures the share of the components of area effects that are uncorrelated with x_i , in the components of $\ln w_i$ that are uncorrelated with x_i . The partial R -squared is:

$$\frac{\text{UVS}_d}{1 - R^2(\ln w; x)} = \frac{R^2(\ln w; x, d) - R^2(\ln w; x)}{1 - R^2(\ln w; x)}. \tag{B1}$$

Appendix C: Proof that three ways of calculating UVS are equivalent

The equivalence of the R -squared and ANOVA methods follows directly from the definition of the partial sum of squares. We now show that the R -squared method and partitioned regression methods are equivalent. Let the sample regression of $\hat{\delta} = d'_i \hat{\gamma}$ on $\chi \equiv x'_i \hat{\beta}$ be:

$$\hat{\delta} = \hat{\theta} \hat{\chi} + \hat{\omega} \tag{C1}$$

¹⁰The stata routine a2reg for two way fixed effects (Ouazad, 2008), based on Abowd, Creecy and Kramarz (2002), is useful for practical implementation in applications including two groups of effects (e.g. individual and area effects). See <http://courses.cit.cornell.edu/jma7/abowd-creecy-kramarz-computation.pdf>.

which implies, by substitution in to the Mincerian wage equation (1):

$$\ln w = \hat{\chi} + \hat{\theta}\hat{\chi} + \hat{\omega} + \hat{\varepsilon} \quad (C2)$$

Then:

$$\begin{aligned} R^2(\ln w; x, d) &= \frac{\text{var}(\hat{\chi})}{\text{var}(\ln w)} + \frac{\text{var}(\hat{\delta})}{\text{var}(\ln w)} + \frac{2\text{cov}(\hat{\chi}, \hat{\delta})}{\text{var}(\ln w)} \\ &= \frac{\text{var}(\hat{\chi})}{\text{var}(\ln w)} + \frac{\text{var}(\hat{\delta})}{\text{var}(\ln w)} + \frac{2\hat{\theta}\text{var}(\hat{\chi})}{\text{var}(\ln w)}, \end{aligned} \quad (C3)$$

where the first line follows from the full variance decomposition equation (4) and the second line follows from equation (C1). Now, from equation (C1):

$$R^2(\ln w; x) = \frac{\text{var}((1 + \hat{\theta})\hat{\chi})}{\text{var}(\ln w)} = \frac{(1 + 2\hat{\theta} + \hat{\theta}^2)\text{var}(\hat{\chi})}{\text{var}(\ln w)}. \quad (C4)$$

So from equation (C2) and (C3) we have:

$$R^2(\ln w; x, d) - R^2(\ln w; x) = \frac{\text{var}(\hat{\delta}) - \hat{\theta}^2 \text{var}(\hat{\chi})}{\text{var}(\ln w)}. \quad (C5)$$

Now in the partitioned regression method:

$$R^2(\ln w; \hat{\omega}) = \frac{\text{var}(\hat{\omega})}{\text{var}(\ln w)}, \quad (C6)$$

where the numerator is:

$$\begin{aligned} \text{var}(\hat{\omega}) &= \text{var}(\hat{\delta}) + \hat{\theta}^2 \text{var}(\hat{\chi}) - 2\text{cov}(\hat{\theta}\hat{\chi}, \hat{\delta}) \\ &= \text{var}(\hat{\delta}) + \hat{\theta}^2 \text{var}(\hat{\chi}) - 2\text{var}(\hat{\theta}\hat{\chi}) \\ &= \text{var}(\hat{\delta}) - \hat{\theta}^2 \text{var}(\hat{\chi}). \end{aligned} \quad (C7)$$

Hence from equation (C4)–(C6) the uncorrelated variance share is:

$$R^2(\ln w; \hat{\omega}) = R^2(\ln w; x, d) - R^2(\ln w; x) = \frac{\text{var}(\hat{\delta}) - \hat{\theta}^2 \text{var}(\hat{\chi})}{\text{var}(\ln w)} \quad (C8)$$

and so the R -squared and partitioned regression methods are equivalent.

Appendix D: Variance decomposition in terms of UVS

As discussed in the main text, it is possible to provide a decomposition of the total variance based on the UVS as well as the more common CVS decomposition in equation (4). Define:

$$\begin{aligned} \text{UVS}_d &= R^2(\ln w; x, d) - R^2(\ln w; x), \\ \text{UVS}_x &= R^2(\ln w; x, d) - R^2(\ln w; d), \\ R^2(\ln w; \tilde{x}, \tilde{d}) &= R^2(\ln w; x, d) - \text{UVS}_d - \text{UVS}_x. \end{aligned} \quad (D1)$$

Then:

$$R^2(\ln w; x, d) = UVS_d + UVS_u + R^2(\ln w; \tilde{x}, \tilde{d}),$$

$$UVS_d + UVS_u + R^2(\ln w; \tilde{x}, \tilde{d}) + \frac{\text{var}(\hat{\varepsilon})}{\text{var}(\ln w)} = 1. \quad (\text{D2})$$

Appendix E: Derivation of reduced form for wages

Substituting the area average of equation (16) for $\bar{\chi}$ in equation (15) and solving for δ gives:

$$\delta = (\delta\varphi)\lambda + \lambda\bar{v} + \mu = (\lambda\bar{v} + \mu)/(1 - \lambda\varphi). \quad (\text{E1})$$

Similarly, taking the area average of equation (16), substituting for δ and solving gives:

$$\bar{\chi} = \delta\varphi + \bar{v} = (\bar{\chi}\lambda)\varphi + \bar{v} + \varphi\mu = (\bar{v} + \varphi\mu)/(1 - \lambda\varphi). \quad (\text{E2})$$

Therefore,

$$\text{var}(\delta) = \{(\lambda^2\text{var}(\bar{v}) + \text{var}(\mu) + 2\lambda\text{cov}(\bar{v}, \mu))/(1 - \lambda\varphi)^2\}, \quad (\text{E3})$$

$$\text{var}(\bar{\chi}) = \{\text{var}(\bar{v}) + \varphi^2\text{var}(\mu) + 2\varphi\text{cov}(\bar{v}, \mu)/(1 - \lambda\varphi)^2\}, \quad (\text{E4})$$

$$\text{cov}(\bar{\chi}, \delta) = \{\varphi\text{var}(\mu) + \lambda\text{var}(\bar{v}) + (1 + \lambda\varphi)\text{cov}(\mu, \bar{v})/(1 - \lambda\varphi)^2\}. \quad (\text{E5})$$

which correspond to equations (17)–(19) in the text.

Appendix F: Basic descriptive statistics

The table below provides basic descriptive statistics for variables of interest. Statistics for two digit SIC and three digit occupation dummies available on request.

<i>Variable</i>	<i>Mean</i>	<i>SD</i>
(log) wage	2.2	0.5
Female	0.5	0.5
Age	39.8	11.9
Part-time	0.2	0.4
Collective agreement	0.6	0.5
Public sector	0.3	0.4
Mover (annual)	0.4	0.5

Notes: Authors own calculations using NES/ASHE. Based on 1,524,322 observations.