STRATHCLYDE

DISCUSSION PAPERS IN ECONOMICS

THE EFFECTS OF AGGLOMERATION ON WAGES: EVIDENCE FROM THE MICRO-LEVEL

BY

BERNARD FINGLETON AND SIMONETTA LONGHI

NO. 11-24

DEPARTMENT OF ECONOMICS
UNIVERSITY OF STRATHCLYDE
GLASGOW

The effects of agglomeration on wages: evidence from the micro-level

Bernard Fingleton[1]

Simonetta Longhi[2]

This version, May 2011

Abstract : This paper estimates individual wage equations in order to test two rival non-nested theories of economic agglomeration, namely New Economic Geography (NEG), as represented by the NEG wage equation and urban economic (UE) theory , in which wages relate to employment density. The paper makes an original contribution by evidently being the first empirical paper to examine the issue of agglomeration processes associated with contemporary theory working with micro-level data, highlighting the role of gender and other individual-level characteristics. For male respondents, there is no significant evidence that wage levels are an outcome of the mechanisms suggested by NEG or UE theory, but this is not the case for female respondents. We speculate on the reasons for the gender difference.

Keywords : urban economics, new economic geography, household panel data.

JEL classifications : C21, C31, C33, D1, O18, R20, R12

1. Introduction

Recent papers have suggested that models deriving from urban economics (UE) may provide a better explanation of spatial variation in wage levels over short distances than the New Economic Geography (NEG) wage equation (Combes, Duranton and Overman, 2005, Brakman, Garretsen, and Van Marrewijk, 2009, Fingleton, 2011). Somewhat in contrast Fujita, Krugman and Venables(1999) emphasise the generality of the processes embodied in NEG, regardless of spatial scale, in other words NEG is a 'one-size-fits-all' model. This comes across from the Preface of the their seminal book, which emphasises 'how a common approach …can be applied to a wide variety of issues in regional, urban and international economics'. However, although this 'one-size-fits-all' approach has been subject to criticism on empirical grounds, current evidence supporting the superiority of the rival UE model as a basis for modelling localised wage variation is compromised somewhat by being based on areal units which are unable to allow full identification of individual-level heterogeneity and its influence on wage levels. In order to build on, and advance beyond, the current state-of-the-art, and to revisit the debate surrounding the respective virtues of NEG and UE, in this paper we

[1] Department of Economics, University of Strathclyde. Email bf100@cam.ac.uk. Corresponding author.
[2] Institute for Social and Economic Research (ISER), University of Essex. Email slonghi@essex.ac.uk.

examine data at the lowest possible level of spatial aggregation, namely the individual level, allowing us to take account of various individual level variables and also, by means of fixed effects in a panel data model, to also control for unobserved individual level heterogeneity. Our paper takes its cue from the observation by Garretsen and Martin (2011), in the recent special issue of the *Journal of Economic Geography*, that 'geographical economists have started to expand their field by building on and incorporating new insights from adjacent fields like international economics, labour economics and urban economics, in order to take account, for instance, of firm and worker heterogeneity, knowledge spillovers, different types of transport cost assumptions, the use of micro-data and more generally a much more detailed analysis of agglomeration economies'. Likewise, Ottaviano(2011), Venables(2011) and Combes et. al. (2011) emphasise the potential role to be played by 'micro-heterogeneity' across people (and firms) in our understanding of agglomeration economies.

An additional consideration is the problem of accounting for the endogeneity of key variables, an adequate solution to which has hitherto proved elusive or costly. Our analysis is based on data from the British Household Panel Survey, and in order to link households to centres of employment, we take advantage of the commuting flow data available in the UK 2001 census, but this introduces an additional endogenous element to our analysis. The question of the choice, validity and appropriateness of the instruments needed to produce consistent estimates remains an important and difficult-to-solve conundrum. We present in this paper what we believe are some novel solutions to the selection of instruments, using historical data from 1861 British census and data on the location of early railways, all of which we believe gave the impetus for additional urban development and the focal points for contemporary agglomeration processes.

2. Theory

The theory of the rival models has been recently sketched by Fingleton(2011). In this paper we use this summary of the rival theories as the background to our empirical analysis. Our first theory, namely UE, for our purposes is best represented in the work of Abdel-Rahman and Fujita (1990), and Fujita and Thisse(2002, page 102), although different set-ups leading to the same reduced form are given in Combes, Mayer, and Thisse (2008) and Brakman, Garretsen, and Van Marrewijk (2009). Assume that a final sector (C) exists in which the market structure is one of perfect competition. None the less increasing returns occur as a consequence of firms within the intermediate (monopolistic or M) sector providing inputs to the final sector. This is because firms in the M sector, which have the sole input labour, are characterised by a fixed labour requirement s and a marginal labour requirement a, thus giving increasing internal returns to scale. We assume, without loss of generality, that the final sector comprises a single firm, and that this has the following production function

$$Q = ((E^C)^{\beta} I^{1-\beta})^{\alpha} L^{1-\alpha} \tag{1}$$

indicating that final sector production depends on the number of *C* labour units E^C, on the level of composite services I from the M firms, and on the amount of land L. Assume that production is per unit area, so that $L=1$, then if $\alpha < 1$ the model includes the effects of congestion (Ciccone and Hall, 1996) on the level of production. Given that I is solely a function of the size of the labour force in the M sector, E^M, then it follows that[3]

$$Q = ((E^C)^{\beta} I^{1-\beta})^{\alpha} = \phi E^{\gamma} \tag{2}$$

[3] See for example Fujita and Thisse(2002), Fingleton and López-Bazo (2003).

in which $E = E^M + E^C$, ϕ is a complex function of constants α, β, μ and s, and $\gamma = \alpha[1+(1-\beta)(\mu-1)]$ with the elasticity of substitution in the CES production function for the intermediate sector equal to $\sigma = \frac{\mu}{\mu - 1}$. If the intermediate sector is relevant ($\beta < 1$), has some monopoly power ($\mu > 1$) and congestion is sufficiently weak ($\alpha << 1$), then it may turn out that γ > 1 and hence there are increasing returns to scale. Taking the wage rate as the derivative

$$w = \frac{\partial Q}{\partial E^C} = \frac{Q\alpha\beta}{E^C} \tag{3}$$

then with $E^C / E = \beta$,

$$\frac{wE}{Q} = \alpha \tag{4}$$

and

$$\ln(w) = \ln(\phi) + \gamma \ln(E) + \ln(\alpha) - \ln(E)\ . \tag{5}$$

Adding disturbances ξ to capture unobserved random effects, and with κ absorbing the constants, we can test the null hypothesis that $\gamma - 1 = 0$ via the regression model

$$\ln(w) = k_1 + (\gamma - 1)\ln(E) + \xi \tag{6}$$

The major advantage of this reduced form, compared with the rival NEG theory, is the relatively small number of assumptions. As mentioned in Fingleton(2011), the requirement for the basic UE model is simply total employees per square km (E) and the wage rate w. However we are here working in a panel data context, and hence with time varying w and E.

It is well known that the short-run equilibrium for the NEG model amounts to a handful of simultaneous equations (as shown by Fujita, Krugman and Venables, 1999), one of which is the so-called wage equation. This provides an alternative explanation of spatial variation in wage levels, although it is specifically written in terms of the M sector, which under the standard theory is taken to be 'industry'. The basic wage equation is

$$w_i^M = [\sum_r Y_r (G_r)^{\sigma-1} (T_{ir})^{1-\sigma}]^{\frac{1}{\sigma}} = P_i^{\frac{1}{\sigma}} \tag{7}$$

in which w^M is M sector wages, which depends on market potential P, which is a function of income Y, of the M sector price indices G, of trade costs T_{ir} between locations i and r, and of the elasticity of substitution, σ. Prices G and incomes Y are given by

$$G_i = [\sum_r \lambda_r (w_r^M T_{ir})^{1-\sigma}]^{\frac{1}{1-\sigma}} \tag{8}$$

$$Y_r = \theta\lambda_r w_r^M + (1-\theta)\phi_r w_r^C \tag{9}$$

The share of M workers in location i is denoted by λ_i and ϕ_i is the share of *C* workers. Also, θ is the share of total employment working in the *M* sector, and $1-\theta$ is the C employment share.

Given these it is possible to solve the simultaneous equations to give the short run NEG equilibrium given by the wage equation (7).

We can turn this into a regression equation comparable to (6) by taking logs, and adding and disturbance ε to take account of the fact that we typically would be working with overall, rather than sector specific, wages, hence

$$\ln w = \frac{1}{\sigma}\ln P \quad + \varepsilon \tag{10}$$

Note from the definition of P in equation (7) that the NEG wage equation is significantly more complex than is the equivalent wage equation (6) from UE theory. Moreover it is by definition endogenous, since it depends on w .

Our solution to the system (7, 8, and 9) uses the same approach as described in Fingleton(2006), but as we will be analysing panel data, we calculate P_i for each year of our study, thus λ_i , ϕ_i , θ and overall nominal wages w_i (approximating w_i^M) are time varying, but σ is constant over time and also for consistency with the earlier literature is set equal to 6.25. Likewise trade costs are time-invariant, equal to $T_{ir} = e^{\tau \ln D_{ir}} = D_{ir}^{\tau}$ with D_{ir} equal to the straight line interregional distance between regions i and r, and with $\tau = 0.1$. One immediate concern, which makes explicit the difference between the UE and NEG theories, is the existence, definition and measurement of trade costs. For UE theory it is not an issue, they are irrelevant. For NEG theory, it is one of the problems that have to be solved. We do not delve deeply into this, but simply refer to the discussion in Garretsen and Martin (2010), Bosker and Garretsen(2010), Fingleton and McCann(2007), Fingleton(2005), Redding and Venables(2004) and the related literature cited in these papers.

Moreover for both variables P_i and E_i there are additional issues of endogeneity which need to be resolved to obtain consistent estimates of the panel wage equations, since as presently described, both employment density (UE) and market potential (NEG) relate to regions of employment, not individual households. By taking account of commuting flows, we introduce an additional element of endogeneity as described below.

3. Data

The UK census gives data on commuting travel between 408 (pre-2009) unitary authority and local authority districts (UALADs) covering the surface area of Great Britain. These districts are one of the fundamental spatial building blocks of our analysis[4]. The census data are therefore a 408 by 408 interaction matrix of commuting frequencies. We normalise these frequencies by dividing each cell by its row total, so that the normalised commuting flows sum to 1 across rows. The matrix product of this n by n matrix (OD_2001) and an n by 1 vector gives an n by 1 vector of weighted averages with weights determined by relative commuting frequencies. The main diagonal of the matrix naturally contains the largest weights.

[4] In practice we confine analysis to the 376 UALADs of England and Wales, although as described below we calculate our market potential and commuting weighted variables using all 408 districts.

We also utilize total employment by UALAD over the period 1998 to 2008, together with UALAD shares in M and C sectors, as defined in Fingleton(2006). The employment totals divided by each UALAD area (in sq km) gives the time-varying employment density variable (E) of the UE model. The matrix product of OD_2001 and E gives the first of our explanatory variables, E', which we prefer to E because our data are the outcome of home-based interviews. We are explaining wages by home location, so that what is important is the employment density of the employment centre to which the worker commutes, not employment density in the place of residence. Since we do not have precise knowledge of the specific commuting destination of individuals, we use the information given in the commuting matrix to obtain a per-UALAD weighted average of employment density by year with weights determined by the relative commuting frequency from the UALAD of residence to both itself and all other UALADs.

The UALAD shares in M and C sectors λ_i, ϕ_i, and θ for each year from 1998-2008 are required to obtain our time-varying measure of market potential, and as with E' the matrix product of OD_2001 and *P* gives the explanatory variable P'. Again the important feature of this variable is that it measures access by resident households to market potential, with weights allotted according to relative commuting frequencies from household locations.

In our econometric analysis, we use E' and P' to explain spatial variation in individual residence-based wage levels[5], as described subsequently. As a precursor, we also examine the correlation, at the UALAD level, between E' and P' and mean weekly gross pay[6] of full time[7] workers for UALAD of residence (wage by home UALAD, or w'') for each year from 2002, and between E' and P' and mean weekly gross pay of full time workers by workplace[8] (w). The correlations give an initial rough indication of the relative explanatory power of our variables E' and P'. The correlation matrix[9] (Table 1) shows that $\ln E$ is relatively weakly correlated with wage by home UALAD $\ln w''$, although it is more strongly correlated with ln w, reflecting the essential need to take account of commuting in explaining wage rate variation. The commuting weighted version of employment density (ln E') shows the highest correlation with wages by home UALAD, reflecting the fact that what is important is employment density within commuting distance, not employment density in the home UALAD. While weighting by commuting frequency makes a big difference to the apparent explanatory power of employment density, its effect on market potential is very small, although it does marginally increase the correlation. This marginal impact is because a home location's market potential depends on surrounding locations, so that leafy suburbs of big cities already possess high market potential prior to weighting. The fact that weighting by commuting frequency makes very little difference to the market potential variable is indicated by the very strong linear correlation between ln P and ln P'.

[5] Given in our data from the British Household Panel Survey (BHPS), to be described subsequently.
[6] From NOMIS.
[7] 30 or more hours per week.
[8] These data require a small amount of interpolation to fill missing cells. For wages by place of employment, we interpolate 24 cells out of 4488 for the period 1998 to 2008.
[9] Averaging each correlation matrix for the years 2002 to 2008.

Table 1 : Correlations between home and work-based wages, employment density and market potential by UALAD						
	Ln w''	Ln w	Ln E	Ln E'	Ln P	Ln P'
Ln w''	1.0000	0.7336	0.2362	0.6476	0.6330	0.6469
Ln w		1.0000	0.5242	0.6660	0.6553	0.6539
Ln E			1.0000	0.7190	0.6192	0.6008
Ln E'				1.0000	0.8193	0.8361
Ln P					1.0000	0.9953
Ln P'						1.0000

Our econometric analysis combines the UALAD data on employment density and market potential, namely the variables E' and P', with individual level data obtained from the British Household Panel Survey[10] (BHPS). The data set comprises 10 waves (waves 8 to 17) of interviews over the period 1998 to 2007 (the panel extends outside this time period, but data considerations limited our analysis to these 10 waves), so that overall we have information on 52,042 interviews for England and Wales[11]. The selection of variables from the BHPS is motivated by the typical specification of a Mincerian wage equation, in which wages partly depend on experience and schooling. Many studies have used earnings data from the BHPS, most recently Francesconi et. al. (2011), who give many insights regarding the source and quality of the BHPS data, noting that the BHPS earnings data seems to be equally as reliable as the Family Resources Survey (FRS), which is a special income survey forming the basis of official UK income distribution estimates. They note that our preferred variable, 'usual gross earnings' (w')[12], which measures monthly not hourly earnings, is essentially based on the BHPS variables PAYGL and PAYU , and does include some imputation, but it 'is a measure which is favoured by many analysts'. Table 2 gives the distribution by region of three key variables, where each region's value is the mean, averaging over all respondents (differentiating between males and females) within the region and over time. The GORs (Government Office Regions) used in Table 2 are aggregations of UALADs. These are not the quantities that we use in our econometric models, but nevertheless these data are informative. The correlation between wages, market potential and employment density and the importance of gender and location are apparent from this table.

[10] Available to registered users from the University of Essex Data Archive.
[11] We have excluded the panel for Scotland because of data limitations. Also our analysis is restricted working age people (16-59; 16-64) and to employees, thus excluding the self-employed.
[12] Note that we drop one prime to signify that these are individual rather than district wage rates.

Table 2 : Mean values of key quantities by gender						
Region	w' (m)	w' (f)	P' (m)	P'f)	E' (m)	E' (f)
inner London	2584.5294	1930.1276	15915.841	16264.885	17187.406	16558.211
outer London	2532.0992	1797.1468	14736.621	14400.115	9707.852	9874.2536
rest of South East	2224.7121	1612.7589	11991.816	12152.663	2781.1845	2750.0431
South West	1893.231	1342.1788	9783.0648	10097.007	711.00207	738.87098
East Anglia	1875.5236	1265.5857	9846.3993	10591.273	974.81758	941.84007
East Midlands	1793.4305	1378.2085	11001.559	10983.387	794.11442	801.12415
West Midland conurbation	1735.2434	1256.7967	11000.11	10412.604	1449.2159	1454.626
rest of West Midlands	1863.5184	1398.7208	10625.072	11069.454	578.7194	622.5279
Greater Manchester	1881.3974	1639.0742	10906.554	11404.08	1324.2585	1423.7009
Merseyside	1936.1838	1292.1316	10309.538	11198.281	899.07457	928.70055
rest of North West	1850.9187	1510.6885	10068.882	9972.3477	608.65312	622.34631
South Yorkshire	1717.2631	1312.597	11182.266	10872.861	540.93747	552.11226
West Yorkshire	1805.1408	1407.1729	10551.84	10356.856	562.92345	581.40123
rest of Yorkshire and Humberside	1833.9364	1470.2301	9442.2548	9750.0063	372.73763	382.73737
Tyne & Wear	1615.5951	1455.059	8724.9941	8940.0079	1043.8569	1078.8811
rest of North	1842.5888	1372.405	8654.3219	9181.3532	331.20526	367.75231
Wales	1725.2956	1319.8593	10072.057	10303.542	341.58409	339.7878
Total	1935.326	1462.5881	10887.011	11152.661	1903.2682	2096.575

In our wage equation, we capture the effect of experience via the BHPS variables age of respondent and age-squared (designated age and age2), anticipating a positive coefficient on age and a negative coefficient on age-squared, thus giving a quadratic relationship between experience (age) and wage level. Additionally, we include 8 Standard Occupational Classification dummies (SOC1 to SOC8), and dummy variables indicating whether the respondent has children (Kids) and whether the respondent is 'married' (Married)[13]. We also include 9 year dummies to capture year-specific national factors that might have an impact on wages (e.g. inflation etc.). Importantly, we also include region dummies in many specifications so as to correctly identify the effects of market potential and employment density separately from other unspecified sources of spatial economic heterogeneity. Other unobserved sources of individual heterogeneity are captured by the fixed effects in our model specification. This means that one important variable, gender (Male, female), is not identified. We give special attention to this variable by estimating our fixed effects panel

[13] Strictly married or cohabiting (as opposed to widows, singles, divorced etc.).

separately for men and for women. We therefore have data on 21 individual time-varying explanatory variables and one dependent variable ln w'. In order to be able to supplement this suite of explanatory variables by our two rival time-varying theory-derived measures of the individual's 'economic environment', namely E' and P', which are available for UALADs, it was necessary to link UALADs to respondents' places of residence and also to the wave of the survey. The confidential information on respondent locations was accessible to us on licence and we were therefore, for each individual at each point in time, able to provide individual-specific measures of market potential and employment density.

An important consideration in our analysis is the endogeneity associated with E' and P'. This has several sources. For example, it is likely that commuting frequencies will be a consequence of wage levels, with high wage centres attracting more commuters. This in turn will influence our variables which are of course a weighted function of commuting frequencies. Moreover it is likely that measurement error will be factor, particularly in the values obtained for market potential, because this depends on an unknown parameter σ, and on assumptions about the definition of C and M sectors (see Section 2). Moreover, the definition of market potential shows that it depends on wage levels, so that there is potentially a two-way interaction between our dependent and independent variables. Although market potential is calculated using the wage data by UALAD of employment discussed above, nevertheless it is seems likely that endogeneity involving individual level wages and market potential will occur. Likewise, in the case of commuting-weighted employment density E', one might anticipate that this will be an effect of w' as well as being a cause, since workers may be attracted to locations with high wages. In addition commuting frequencies will again depend on wages so the effect of wages is embodied within E' for this reason also. Because of these considerations, to achieve consistent estimates, we need to rely on appropriate instrumental variables.

4. Preliminary Estimates

Our preliminary analysis sets the scene for our more rigorous subsequent econometrics in which we endeavour to take full account of endogeneity. In this section we provide information that contributes to the overall understanding and conceptualization of the relationship between E', P' and w'. Our analysis compares the relative efficacy of (log) market potential and (log) employment density in explaining variations in individual (log) wages(ln w'), controlling for individual-level covariates (time constant variables, plus a quadratic age function, dummies for children, 'marriage', occupational classification and panel wave). Of the time constant variables, we give special attention to gender and we therefore split our analyses between male and female respondents in England and Wales in full time occupation. In all of our analyses we find that we are constrained to separate models for males and females because time-invariant variables such as gender are not identified in our fixed effects specifications[14].

[14] We tested random effects models, which identify time-invariant variables, but according to the Hausman specification test parameter estimates are not consistent. Likewise the Hausman-Taylor (1981) estimator for error-components models, which identifies time invariant covariates and permits correlation with unobserved individual effects, supports our main conclusions set out below, but also fails the Hausman specification test.

In order to obtain an initial insight, we ignore the endogeneity issues and fit fixed effects panel models (in effect OLS models fitted to deviations from individual means) which are summarised in Tables 3 and 4. They are simply presented as a precursor to valid inference on the basis of consistent estimates. Because our rivals are non-nested, this means that we are unable to constrain a parameter to zero to reduce from one to the other, allowing a simple test of the null that the constrained parameter is truly zero. To allow non-nested rivals to be tested, Hendry(1995) suggests a data generating process (DGP) in which both rival theories are combined as an artificial nesting model(ANM), of which both rivals are special cases. Given rivals A and B, we are interested in whether A encompasses, or explains the results of, B, and vice versa. So if dropping A from the ANM model produces a significant loss of fit, and dropping B does not, then in effect A is explaining the ANM results and therefore the B results embodied within the ANM.

Table 3 : Fixed effects panel estimates of key quantities : male full time employees, England & Wales

				With region dummies		
Dependent variable ln w'	ln E'	ln P'	ln E' & ln P'	ln E'	ln P'	ln E' & ln P'
ln E' Est.	.0257798	------	.0146792	-.0020259	------	-.0143287
ln E' s.e.	.0068916	------	.0119856	.0109422	------	.0130975
t-ratio	3.74	------	1.22	-0.19	------	-1.09
F-prob.	0.0002	------	0.2207	0.8531	------	0.2740
ln P'	------	.1797224	.0953336	------	.1217869	.1879034
ln P' s.e.	------	.0484243	.0842175	------	.0918544	.1099527
t-ratio	------	3.71	1.13	------	1.33	1.71
F-prob.	------	0.0002	0.2577	------	0.1849	0.0875
F-prob $\ln E' \& \ln P'$	------	------	0.0005	------	------	0.2283
R-squared within	0.3115	0.3115	0.3116	0.3156	0.3157	0.3157
R-squared between	0.0341	0.0331	0.0346	0.0423	0.0441	0.0433
R-squared overall	0.0695	0.0687	0.0704	0.0800	0.0820	0.0812

The individual covariate coefficients have not been given here and in subsequent tables in order to save space.

Table 4 : Fixed effects panel estimates of key quantities : female full time employees, England & Wales						
				With region dummies		
Dependent variable ln w'	ln E'	ln P'	ln E' & ln P'	ln E'	ln P'	ln E' & ln P'
ln E' Est.	.0538981	------	.0598124	.037413	------	.0364668
ln E' s.e.	.0073175	------	.0129679	.0122935	------	.0147864
t-ratio	7.37	------	4.61	3.04	------	2.47
F-prob.	<0.0001	------	<0.0001	0.0023	------	0.0137
ln P'	------	.294398	-.0499587	------	.1854029	.0143748
ln P' s.e.	------	.0510748	.090434	------	.1037975	.1248133
t-ratio	------	5.76	-0.55	------	1.79	0.12
F-prob.	------	<0.0001	0.5807	------	0.0741	0.9083
F-prob $\ln E' \& \ln P'$	------	------	<0.0001	------	------	0.0097
R-squared within	0.3939	0.3928	0.3939	0.3997	0.3994	0.3997
R-squared between	0.1137	0.1044	0.1136	0.1194	0.1150	0.1194
R-squared overall	0.1557	0.1486	0.1555	0.1617	0.1586	0.1617

Both Tables 3 and 4 provide some initial evidence that each of our rivals seems to have a highly significant effect, as indicated by the t-ratios. In Table 3, when considered separately, both ln E' and ln P' appear to be highly significant. However the right hand panels of Table 3 show that when region dummies[15] are also included in the specification, thus giving a much tougher test, the individual significance of ln E' and ln P' disappears. This suggests that these variables are capturing spatially varying or 'environmental' effects aspects other than employment density or market potential, and controlling for these via the region dummies it is evident that ln E' and ln P' have no 'effect' on individual male wages. One would anticipate that If NEG theory was the dominant explanation of individual wage variation, it would retain significance given the presence of the employment density variable, with the latter failing to carry any additional explanatory information, but when they are combined together as an artificial nesting model (ANM) neither emerges as significant, either with or without the presence of region dummies. It is evident from Table 3 that for males neither rival encompasses the other; neither rival stands out as significant given the presence of the other. For male respondents, dropping ln E' from the ANM inclusive of region dummies does not produce a significant loss of fit[16], given the presence of ln P'. Likewise dropping ln P' gives a similar outcome although there is an indication of weak significance accepting a 10% level of risk (p = 0.087). Neither does eliminating both rivals simultaneously produce a significant loss

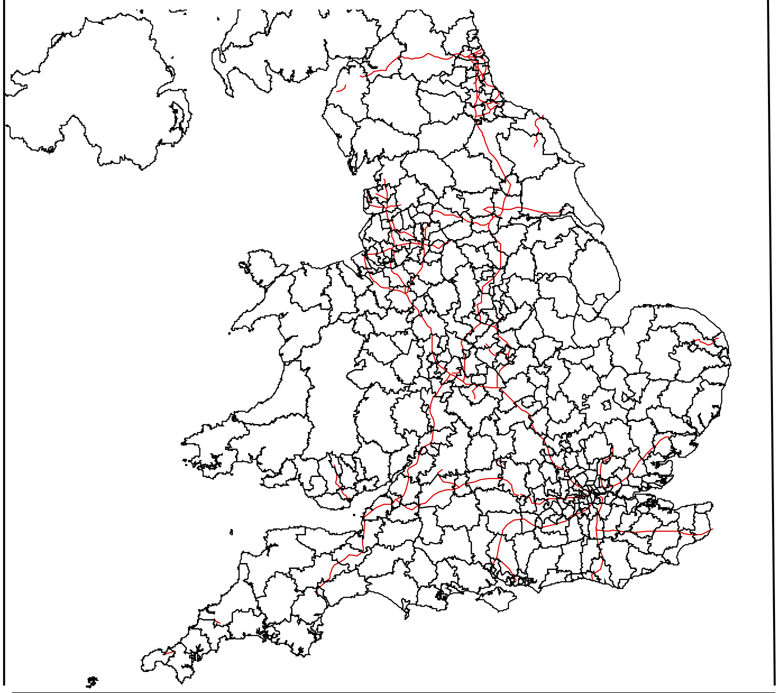
[15] inner London, outer London, rest of South-East, South West, East Anglia, East Midlands, West Midlands conurbation, rest of West Midlands, Greater Manchester, Merseyside, rest of North West, South Yorkshire, West Yorkshire, rest of Yorkshire and Humberside, Tyne & Wear, rest of the North, Wales.

[16] F(1, 17972) = 1.20, Prob > F = 0.2740.

of fit[17], so the lack of individual significance is evidently not the result of collinearity inflating standard errors and reducing significance.

When we consider female respondents, the results are remarkably different (Table 4). Without region dummies, both ln E' and ln P' are nominally highly significant, and the artificial nesting model strongly supports the hypothesis that ln E' is the dominant variable and ln P' does not seem to carry any significant additional explanatory information. Our more rigorous model which includes the region dummies reaffirms these outcomes, although ln P' considered alone is only marginally significant (p = 0.074). Even with region dummies, ln E' retains its apparent significance in the presence of ln P', as shown by column 7 of Table 4. Given region dummies, dropping ln E' produces a significant loss of fit[18] whereas dropping ln P' does not[19].

Figure 1 : The 1845 rail network in England and Wales



5. Instruments

Our selection of instruments is conditioned by a number of factors. First, we need a sufficient number to allow overidentification and therefore to test the exogeneity of the instruments. Given that we will simultaneously introduce both rival variables, E' and P', in some of our model specifications, we need at least three excluded instruments. However we do not want

[17] F(2, 17972) = 1.48, Prob > F = 0.2283.
[18] F(1, 118408) = 6.08, Prob > F = 0.0137.
[19] F(1, 11840) = 0.01, Prob > F = 0.9083.

too many instruments, because of the possibility that a large number may overfit the endogenous variables and lead to incorrect inferential decisions (Roodman, 2009). In other words a multiplicity of instruments may fit the endogenous variable so well that the fitted values used in the second stage may still contain the endogenous component of variation that we are attempting to expunge, thus leading to bias. A second issue is similar to that encountered by the need to merge UALAD and BHPS data earlier, the problem of different spatial units. Thirdly, we wish to avoid weak instruments, which itself leads to bias and size distortion (Stock, Wright and Yogo, 2002, Stock and Yogo, 2005), so our instruments should be sufficiently correlated with the endogenous regressors E' and P' while remaining orthogonal to the disturbances.

In order to try to ensure exogeneity, we chose our instruments from data gathered more than 130 years previously, partly by using data from the 1861 British Census[20]. The chosen instruments were selected to indicate those locations that were rapidly expanding as cities at the height of the industrial revolution, and as significant causes of present-day city locations we anticipate that they will correlate strongly with market potential and employment density, which are at their maximum in the large densely populated cities of today. On the other hand, we do not expect our instruments to be causally related to current wage levels. We test these assumptions via the standard diagnostic tests reported subsequently.

The census variables adopted as instruments are population change from 1851 to 1861(*pop_ch*), the share[21] of male employment in manufacturing in 1861 (*m_manuf_sh*), and the number of people born in Ireland per thousand population(*irld_pt*) in each location in 1861. The level of population change identifies rapidly growing locations of the mid-19th century, thus indicating where urbanization and localization externalities were strong, and where it was likely to continue into the future. The importance of rapid growth in this era is readily identified in the built environment of many of Britain's major cities today, which possess a significant Victorian legacy. The boom industry of the mid 19th century British city was manufacturing, and we are able to also pick up the growth points in the urban system by means of the *m_manuf_sh* instrument. We complement this by the *irld_pt* instrument, which identifies the rapidly expanding centres of employment which were particularly the destinations of many people displaced by the Irish famine of 1845 and 1852.

One additional instrument is the number of railway lines existing in 1845 in each locality[22]. Figure 1 show the distribution of railways against the back ground of UALADs, and from this it is apparent that England and Wales already had a communications network linking the main urban centres, together with some more remote railway lines such as in Cornwall, that were related to mining activity. Nevertheless the distribution of railways largely reflects the distribution of the main urban centres and indicates potential growth points for future urban development.

[20] The data are taken from a large dataset which has been assembled from a number of sources by David Gatley. This is part of part of the Great Britain Historical GIS, developed by Ian Gregory.

[21] As a share of the total of men working the following sectors identified in the 1861 Census : agriculture & farming, mining and brick-making, building, manufacturing, transport & storage, dealing, commercial service, general labour, public service and domestic service.

[22] We are grateful to Robert Schwartz for providing these data, which was created by a team led by Jordi Marti-Hennebourg. More detail of these data is provided at http://www.mtholyoke.edu/courses/rschwart/railways

Thus far we have not examined the issue of matching the spatial units, the 'districts', used for the 1861 census variables and also the matching of the rail network vector, to the respondents' household locations. To do this we use the UALAD boundary system as an intermediate stage, first mapping our district-based variables and our railway occurrence vector to UALADs and then subsequently mapping the resulting UALAD-based variables to household locations, as was done for the variables E' and P'. The initial step is to translate the Census and railway data to the UALAD system of spatial units. First let us consider the 1861 Census data which is organised by 638 districts covering England and Wales and the Channel islands. Fortunately district grid references are available, and in 156 cases we are able to match precisely districts and UALADs, which possess the same names and fairly precisely overlapping locations.

Given these 156 matching locations, we can estimate regression models in which the x and y UALAD coordinates are the dependent variables and (a cubic function of) the x and y coordinates of the 1861 Census districts are the explanatory variables. Both regressions, estimated by OLS, report R-squareds in excess of 0.99. Given the estimated regression coefficients, we then predict the x and y UALAD coordinates of all 636 Census districts on the basis of the cubic functions of their district coordinates. Given that both sets of spatial units, districts and UALADS, are now on the same coordinate system, we next calculate the distances between each district and (English and Welsh) UALAD, which is a 636 by 376 matrix.

This distance matrix is then the basis of the mapping of the 1861 Census variables into the UALADs, by first calculating a weighting matrix by taking the reciprocal of distance[23] to the power 4 (chosen to ensure that remoter locations carry effectively zero weight), subsequently normalised so that rows sum to 1. The mapping is the outcome of the matrix product of this weighting matrix and the 636 by 1 vectors of Census variables, the result being 376 by 1 vectors giving 1861 Census variable values in each UALAD. In the case of the 1845 rail network, we use the intersect function of the GIS software package Arcview to give a count of the number of segments of railway than occurs within each UALAD. This picks out the major centres of industrial and mining activity in 1845, for example, Stockton-on-Tees, which was at the forefront of railway technology in the early industrial revolution, has 6 lines. There are 11 lines in the inner London boroughs, and 17 in the outer London boroughs, thus indicating that even by 1845 railway technology had diffused from early centres of innovation such as the Stockton-on-Tees to Darlington line[24] to become a passenger and freight service connecting major cities of England and Wales.

6. Results using instrumental variables

The results presented above are preliminary in the sense that we did not take account of the endogeneity of $\ln E'$ and $\ln P'$. In order to obtain consistent estimates of their effects, we instrument $\ln E'$ and $\ln P'$ using the instrumental variables described above, namely population change (*pop_ch*), Irish-born residents per thousand (*irld_pt*), the share of male workers in manufacturing (*m_manuf_sh*) and the number of railway lines in 1845 (*railines*). Table 5 gives the

[23] Zero distances were handled by mapping the variables in the coincident locations directly into the UALADs.
[24] Constructed from 1826 onwards, these 26 miles of track became the template for railway systems throughout the world. The initial railways carried coal from various mines in the North East of England, subsequently extending and evolving from mainly horse drawn to mainly steam locomotion, as well as introducing timetables, signalling systems and passenger services.

outcomes for male respondents, reaffirming our suggestion based on Table 3 that neither variable has a significant effect on male wages. While there is an indication of individual significance from the specifications without region dummies, the inclusion of dummies demonstrates that we cannot infer that individual male wage variations depend either on market potential or employment density.

The results for females remain distinctly different, as apparent from Tables[25] 6 and 7. Column 2 of Table 6 shows the outcome of estimating the $\ln E'$ model using the instruments deemed to be exogenous. It is apparent that we fail to reject the Sargan null with a test size of 0.05, although the result is close to significance. Assuming the orthogonality of instruments *railines*, *pop_ch* and *irld_pt*, we infer that $\ln E'$ is a significant cause of female wage variation. Column 3 of Table 6 indicates that $\ln P'$ is also strongly associated with female wage variation, but in this case we cannot assume consistent estimation because of the rejection of the Sargan null of orthogonality. The fourth column of Table 6 shows the outcome of attempting to use orthogonal instruments. In this case we have constructed an additional instrument equal to the product of *railines* and *pop_ch*, namely *rail_pch*. Although our instruments now do pass the Sargan test, we believe there is a weak instrument problem. Our test statistic is the Cragg-Donald Wald F statistic which is referred to the critical values[26] given by Stock and Yogo (2005). The test statistic equals 15.804 which lies between the critical values for 15% and 10% maximal size (nominal size plus size distortion), which are equal to 11.59 and 19.93 respectively. The indication of some size distortion suggests that we should not rely on the column 4 estimates in Table 6. Columns 5 and 6 provide more rigorous tests because of the addition of the regional dummies to the set of regressors. The regressor sub-set is seen to be orthogonal to the errors, and we see that both $\ln E'$ and $\ln P'$ retain their significance.

Table 7 shows the outcome of confronting the two non-nested rivals for the female wage data. Column 2 uses all four instruments and points to the dominance of $\ln E'$, but the instruments collectively fail the Sargan test. Column 3 employs an orthogonal subset of three instruments and points to the insignificance of $\ln P'$ given the presence of $\ln E'$. The strongest evidence supporting the causality of $\ln E'$ and its encompassing, possibly, of $\ln P'$ is provided by column 4, where it's significance is maintained despite the presence of the region dummies. On the other hand, we see that the p-value for the $\ln P'$ t-ratio of 1.67 is 0.094, which hints that this variable may be carrying some additional explanatory information. It is apparent that UE theory is outperforming NEG theory for female employees, although the sign and near significance of $\ln P'$ means that we cannot dismiss NEG entirely.

[25] The results in these tables mainly use a sub-set of the four instruments in order to pass the Sargan test.
[26] The critical values depend on the number of instruments and on the number of included endogenous variables.

Table 5 : Panel IV estimates of key quantities : male full time employees, England & Wales				With region dummies		
Dependent variable ln w'	ln E'	ln P'	ln E' & ln P'	ln E'	ln P'	ln E' & ln P'
ln E' Est.	.0372599	------	-.0160177	.035437	------	.0415879
ln E' s.e.	.0169723	------	.0428763	.0269506	------	.0344429
t-ratio	2.20	------	-0.37	1.31	------	1.21
F-prob	0.0281	------	0.7087	0.1885	------	0.2273
ln P'	------	.4791933	.6418167	------	.2881572	-.1778286
ln P' s.e.	------	.1878841	.4741789	------	.484929	.619914
t-ratio	------	2.55	1.35	------	0.59	-0.29
F-prob	------	0.0108	0.1759	------	0.5524	0.7742
F-prob $\ln E' \& \ln P'$	------	------	0.0362	------	------	0.4044
Excluded instruments	*pop_ch irld_pt railines m_manuf_sh*	*pop_ch irld_pt railines m_manuf_sh*	*pop_ch irld_pt railines m_manuf_sh*	*pop_ch irld_pt railines m_manuf_sh*	*pop_ch irld_pt railines m_manuf_sh*	*pop_ch irld_pt railines m_manuf_sh*
Sargan	3.297	1.596	1.454	0.118	1.495	0.036
p-value	0.3480	0.6603	0.4833	0.9896	0.6834	0.9822
R-squared within	0.3114	0.3101	0.3091	0.3151	0.3155	0.3150
R-squared between	0.0388	0.0485	0.0499	0.0465	0.0462	0.0448
R-squared overall	0.0749	0.0861	0.0873	0.0840	0.0841	0.0821
				See Appendix Table A2	See Appendix Table A3	
Note : estimates obtained using STATA commands xtivreg and xtivreg2 , as described in Schaffer (2010) and Schaffer and Stillman (2010),.						

Table 6: Panel IV estimates of key quantities : female full time employees, England & Wales				With region dummies	
Dependent variable ln w'	ln E'	ln P'	ln P' •Note1	ln E'	ln P'
ln E' Est.	.0978588	------	------	.0898075	------
ln E' s.e.	.0206672	------	------	.0331926	------
t-ratio	4.73	------	------	2.71	------
F-prob	<0.0001	------	------	0.0068	------
ln P'	------	.9459282	3.234297	------	2.26336
ln P' s.e.	------	.242684	1.125114	------	.8836997
t-ratio	-----	3.90	2.87	-----	2.56
F-prob		0.0001	0.0040		0.0104
Excluded instruments	*railines pop_ch irld_pt*	*railines pop_ch*	*rail_pch railines*	*railines pop_ch irld_pt*	*railines pop_ch*
Sargan	5.561	6.526	1.054	4.699	0.024
p-value	0.0620	0.0106	0.3045	0.0954	0.8758
R-squared within	0.3920	0.3846	0.2253	0.3988	0.3791
R-squared between	0.1233	0.1117	0.0721	.1248	0.0970
R-squared overall	0.1586	0.1440	0.0769	0.1632	0.1237
• Note1 : Weak identification test (Cragg-Donald Wald F statistic): 15.804				See Appendix Table A1	

Table 7: Panel IV estimates of key quantities : female full time employees, England & Wales			
			With region dummies
Dependent variable ln w'	ln E' & ln P'	ln E' &ln P'	ln E' &ln P'
ln E' Est.	.1226048	.1346289	.0726038
ln E' s.e.	.0305228	.0307455	.0350848
t-ratio	4.02	4.38	2.07
F-prob	0.0001	<0.0001	0.0385
ln P'	-.166714	-.5189208	1.532537
ln P' s.e.	.3012775	.3212722	.9161183
t-ratio	-0.55	-1.62	1.67
F-prob	0.5800	0.1063	0.0944
F-prob $\ln E' \,\&\ln P'$	<0.0001	<0.0001	0.0069
instruments	*pop_ch irld_pt railines m_manuf_sh*	*railines irld_pt pop_ch*	*railines pop_ch irld_pt*
Sargan	12.830	2.954	1.803
p-value	0.0016	0.0857	0.1793
R-squared within	0.3915	0.3922	0.3860
R-squared between	0.1231	0.1136	0.1108
R-squared overall	0.1573	0.1514	0.1390

For the female data, we evaluate the effect of employment density on wages controlling for the other variables, using the preferred model summarised by Table 6, column 5 and detailed in Appendix Table A1. Employment density is a separate cause of wage variation distinct from the significant regional effects captured by the region dummies. We find that doubling employment density, which is equivalent to migrating from East Midlands to inner London, raises female wages by ln($2^{0.0898075}$) = 6.22%. By comparison, having children reduces female wages by about 8% and 'marriage' raises wages by 1.7%. Also evident is the significant quadratic relationship between wages and age, as is typical of many wage equations (the test statistic equals 618.82 which is highly significant when referred to χ_2^2) , and a significant occupational category effect (177.94 $>> \chi^2_{8,0.05}$) . Also of course there are unobserved time-invariant effects such as start-of-period educational attainment which are picked up by the fixed effects of our panel model, and the regional dummies give 55.41 $>> \chi^2_{16,0.05}$.

In contrast for males, Table A2 indicates that marriage raises wages by 2.86% and again there is a quadratic relationship between age and wages(1176.8 $>> \chi^2_{2,0.05}$) . The occupational dummies are also significant (test statistic 119.63 > $\chi^2_{8,0.05}$). Although there is evidently no causal effect of ln E' per se, the collective significance of the region dummies (52.33 $> \chi^2_{16,0.05}$) indicates that the male respondent's 'environment' does have an effect. Very similar results are obtained for the model with ln P', as shown in Table A3.

7. Conclusions

This paper has taken up the challenge to examine agglomeration process using micro level data as has been emphasized in the recent special issue of the *Journal of Economic Geography*. In particular we have responded to the focus given by the special issue editors, Garretsen and Martin(2011), on the 'need to work with micro-data'. Ours is we believe the first paper to actually do this in practice, and we have mixed findings. The evidence supporting the impact of market potential on wages, as envisaged by NEG theory, is very weak when we look at individual wage rate variations over small distances. In contrast the externalities associated with our rival urban economic theory appear to be more relevant as a cause of wage variation, but only for female workers. With regard to males, neither of the theoretical processes we have focussed on has an effect on wage levels. That is not to say that other mechanisms in the economic environment, within which each male respondent is embedded, do not have an effect, as shown by the highly significant set of regional dummies. As we have shown, the other factors affecting wage levels of males and females are also different. For women, a key issue is the impact of having children, which is clearly associated with reduced earnings. For both men and women, marriage seems to count, though more so for men, and occupational status and age also important factors for both groups. While there is a multiplicity of individual-level causes, some unobserved though controlled for, it remains the case that where you live also has a significant effect on wage levels. For women, given that the degree of proximity to dense employment centres is a significant factor affecting wages, we can speculate that despite working full time, on average more women than men carry out home-making duties, and therefore for women the spatial arrangement of job and home becomes a crucial issue. For men this is evidently less important. Indeed it is revealing that marriage is associated with higher wages, possibly because it motivates and permits men to earn more given that home-duties typically tend to be more the woman's role. Thus for men, it appears that they can travel further and use this spatially flexibility to maximise incomes in a way that is less possible for more spatially constrained females.

8. References

Abdel-Rahman, H., Fujita, M. (1990) Product variety, Marshallian externalities, and city sizes. *Journal of Regional Science,* 30: 165-183.

Bosker, M., Garretsen, H. (2010) Trade costs in empirical New Economic Geography. *Papers in Regional Science,* 89: 485-511.

Brakman, S., Garretsen, H., Van Marrewijk, C. (2009) Economic Geography within and between European Nations: The Role of Market Potential and Density across Space and Time. *Journal of Regional Science*, 49: 777-800.

Ciccone, A., Hall, R. E. (1996) Productivity and the density of economic activity. *American Economic Review*, 86: 54-70.

Combes, P.P., G. Duranton and H.G. Overman (2005) Agglomeration and the adjustment of the spatial economy. *Papers in Regional Science*, 84: 311-349.

Combes, P-P., Mayer, T., Thisse, J.F. (2008) *Economic Geography: the Integration of Regions and Nations*, Princeton: Princeton University Press.

Combes, P.-P., Duranton, G., Gobillon, L. (2011) The identification of agglomeration economies. *Journal of Economic Geography*, 11: 253–266.

Fingleton, B. (2005) Towards applied geographical economics: modelling relative wage rates, incomes and prices for the regions of Great Britain. *Applied Economics,* 37: 2417-2428.

Fingleton, B. (2006) The new economic geography versus urban economics: an evaluation using local wage rates in Great Britain. *Oxford Economic Papers,* 58: 501-530.

Fingleton, B. (2011) The empirical performance of the NEG with reference to small areas. *Journal of Economic Geography,* 11: 267-279.

Fingleton B., Lopez-Bazo, E. (2003) Explaining the distribution of manufacturing productivity in the EU regions. In B. Fingleton (ed.) *European Regional Growth.* Berlin: Springer-Verlag, 375 – 410.

Fingleton, B., McCann, P. (2007) Sinking the iceberg? On the treatment of transport costs in new economic geography. In B. Fingleton (ed.) *New Directions in Economic Geography.* Cheltenham: Edward Elgar, 168-203.

Francesconi, M., Sutherland, H., Zantomio, F. (2011) A comparison of earnings measures from longitudinal and cross-sectional surveys: evidence from the UK. *Journal of the Royal Statistical Society A*, 174: 297-326.

Fujita, M., Krugman, P. R., Venables, A. J. (1999) *The Spatial Economy : Cities, Regions, and International Trade.* Cambridge, MA: MIT Press.

Fujita, M., Thisse, J-F. (2002) *Economics of Agglomeration.* Cambridge: Cambridge University Press.

Garretsen, H., Martin, R. (2010) Rethinking (new) economic geography models: taking geography and history more seriously. *Spatial Economic Analysis,* 5(2): 127-160.

Garretsen, H., Martin, R. (2011) The *Journal of Economic Geography* a decade on : where do we go from here? *Journal of Economic Geography,* 11: 207-213.

Gatley, D. (1996) *A User Guide to the 1861 Census and Vital Registration Data Base*. Staffordshire University: The Knowledge, Organizations and Society Research Group.

Harris, C. (1954) The market as a factor in the localization of industry in the United States. *Annals of the Association of American Geographers*, 64: 315-348.

Hausman, J.A., Taylor, W.E. (1981) Panel data and unobservable individual effects. *Econometrica*, 49: 1377-1398.

Hendry, D.F. (1995) *Dynamic Econometrics.* Oxford: Oxford University Press Oxford.

Ottaviano, G. I. P. (2011) 'New' new Economic Geography: firm heterogeneity and agglomeration economies. *Journal of Economic Geography*, 11: 231–240.

Redding, S. J., Venables, A. J. (2004) Economic geography and international inequality. *Journal of International Economics,* 62: 53-82.

Roodman, D. (2009) A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics*, 71: 135-158.

Schaffer, M. (2010) Xtivreg2: Stata module to perform extended IV/2SLS, GMM and AC/HAC, LIML and k-class regression for panel data models. *Statistical Software Components, Boston College.* Department of Economics, http://econpapers.repec.org/RePEc:boc:bocode:s456501.

Schaffer, M.E., Stillman, S. (2010) Xtoverid: Stata module to calculate tests of overidentifying restrictions after xtreg, xtivreg, xtivreg2 and xthtaylor. http://ideas.repec.org/c/boc/bocode/s456779.html.

Stock, J.H., Yogo, M. (2005) Testing for Weak Instruments in Linear IV Regression, in Identification and Inference for Econometric Models. In D.W.K Andrews and J.H. Stock (eds) *Essays in Honor of Thomas Rothenberg*. Cambridge: Cambridge University Press, 80-108.

Stock, J. H., Wright, J. H., Yogo, M. (2002) A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, 20(4): 518–29.

Venables, A. J. (2011) Productivity in cities: self-selection and sorting. *Journal of Economic Geography*, 11: 241–251.

9. Appendix

Table A1: Specification and estimates for the dominant model of female wages

Ln w'	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ln E'	.0898075	.0331926	2.71	0.007	.0247512	.1548637
Kids	-.0844895	.0087378	-9.67	0.000	-.1016153	-.0673637
age	.0682164	.0081528	8.37	0.000	.0522373	.0841956
age2	-.0009508	.0000382	-24.86	0.000	-.0010258	-.0008758
Married	.0171632	.0082331	2.08	0.037	.0010267	.0332996
SOC2	-.0250495	.0121371	-2.06	0.039	-.0488378	-.0012613
SOC3	-.0404932	.0100692	-4.02	0.000	-.0602284	-.020758
SOC4	-.0776477	.0085633	-9.07	0.000	-.0944314	-.060864
SOC5	.0094178	.0217805	0.43	0.665	-.0332713	.0521069
SOC6	-.0875885	.01239	-7.07	0.000	-.1118724	-.0633046
SOC7	-.123773	.0128608	-9.62	0.000	-.1489797	-.0985662
SOC8	.0175156	.0191726	0.91	0.361	-.020062	.0550933
SOC9	-.076969	.0150899	-5.10	0.000	-.1065446	-.0473933
Y2	.0620884	.0106515	5.83	0.000	.0412118	.0829651
Y3	.1385968	.0169027	8.20	0.000	.1054682	.1717254
Y4	.2119586	.0243009	8.72	0.000	.1643298	.2595874
Y5	.267738	.0317092	8.44	0.000	.2055891	.3298869
Y6	.3333285	.0392882	8.48	0.000	.2563251	.4103319
Y7	.3959798	.0467012	8.48	0.000	.3044471	.4875125
Y8	.4661746	.0542009	8.60	0.000	.3599427	.5724065
Y9	.5243501	.061708	8.50	0.000	.4034046	.6452955
Y10	.583144	.0693397	8.41	0.000	.4472408	.7190473
reg1	-.1313366	.1398414	-0.94	0.348	-.4054206	.1427475
reg2	-.1814615	.1158273	-1.57	0.117	-.4084788	.0455558
reg3	-.117929	.084105	-1.40	0.161	-.2827718	.0469139
reg4	-.0605741	.0581484	-1.04	0.298	-.1745428	.0533946
reg5	-.3011046	.071516	-4.21	0.000	-.4412734	-.1609359

reg6	-.1625923	.0578139	-2.81	0.005	-.2759054	-.0492793
reg7	-.1207753	.0797112	-1.52	0.130	-.2770065	.0354558
reg8	-.0556856	.0563296	-0.99	0.323	-.1660897	.0547185
reg9	-.0760971	.0773657	-0.98	0.325	-.2277311	.075537
reg10	.0407252	.0840304	0.48	0.628	-.1239715	.2054218
reg11	-.0317677	.0572576	-0.55	0.579	-.1439906	.0804552
reg12	-.1575003	.0712416	-2.21	0.027	-.2971313	-.0178694
reg13	-.1022551	.0681211	-1.50	0.133	-.23577	.0312599
reg14	.0472133	.066794	0.71	0.480	-.0837005	.1781272
reg15	.1652104	.0908338	1.82	0.069	-.0128205	.3432414
reg16	.0311142	.0663279	0.47	0.639	-.098886	.1611145

Table A2: Specification and estimates for the UE model of male wages

Ln w'	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ln E'	.035437	.0269506	1.31	0.189	-.0173853	.0882592
Kids	.0097283	.0069316	1.40	0.160	-.0038574	.023314
age	.0726171	.0072985	9.95	0.000	.0583123	.0869218
age2	-.0010747	.0000314	-34.28	0.000	-.0011361	-.0010132
Married	.0285558	.0083222	3.43	0.001	.0122446	.0448671
SOC2	-.0238691	.010357	-2.30	0.021	-.0441685	-.0035698
SOC3	-.0317964	.0091555	-3.47	0.001	-.0497408	-.013852
SOC4	-.0833291	.0098068	-8.50	0.000	-.1025501	-.0641081
SOC5	-.029976	.0096139	-3.12	0.002	-.0488189	-.0111331
SOC6	-.0454045	.0139937	-3.24	0.001	-.0728317	-.0179774
SOC7	-.0994972	.011968	-8.31	0.000	-.122954	-.0760404
SOC8	-.028246	.009953	-2.84	0.005	-.0477535	-.0087385
SOC9	-.0589747	.0114508	-5.15	0.000	-.0814177	-.0365316
Y2	.0728142	.0095479	7.63	0.000	.0541007	.0915278
Y3	.143809	.015206	9.46	0.000	.1140059	.1736122
Y4	.2150186	.0221601	9.70	0.000	.1715856	.2584517
Y5	.2729843	.0287672	9.49	0.000	.2166017	.3293669

Y6	.3321405	.0356019	9.33	0.000	.262362	.4019189
Y7	.3988613	.0424756	9.39	0.000	.3156106	.4821121
Y8	.4520638	.049251	9.18	0.000	.3555335	.5485941
Y9	.5099758	.0560972	9.09	0.000	.4000273	.6199244
Y10	.5819318	.0630571	9.23	0.000	.4583422	.7055215
reg1	.0781511	.1154893	0.68	0.499	-.1482038	.304506
reg2	.0533558	.0971714	0.55	0.583	-.1370968	.2438083
reg3	.0892361	.0690706	1.29	0.196	-.0461397	.2246119
reg4	.1190076	.0493276	2.41	0.016	.0223274	.2156879
reg5	.1362026	.0592716	2.30	0.022	.0200324	.2523728
reg6	-.0144762	.0481306	-0.30	0.764	-.1088104	.079858
reg7	-.0427612	.0814623	-0.52	0.600	-.2024244	.1169021
reg8	-.1040333	.0539435	-1.93	0.054	-.2097606	.0016939
reg9	.0065891	.0654478	0.10	0.920	-.1216862	.1348644
reg10	-.1702565	.0953716	-1.79	0.074	-.3571814	.0166684
reg11	-.0648925	.0517183	-1.25	0.210	-.1662585	.0364735
reg12	-.0837453	.0627043	-1.34	0.182	-.2066434	.0391528
reg13	-.0048482	.0590585	-0.08	0.935	-.1206007	.1109044
reg14	.0232937	.0530694	0.44	0.661	-.0807203	.1273077
reg15	.0356238	.093512	0.38	0.703	-.1476564	.218904
reg16	-.0002429	.0678508	-0.00	0.997	-.133228	.1327422

Table A3: Specification and estimates for the NEG model of male wages

Ln w'	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ln P'	.2881572	.484929	0.59	0.552	-.6622862	1.238601
Kids	.009324	.006922	1.35	0.178	-.0042428	.0228909
age	.07245	.0073019	9.92	0.000	.0581385	.0867616
age2	-.0010737	.0000319	-33.67	0.000	-.0011362	-.0010112
Married	.0286291	.0083339	3.44	0.001	.012295	.0449632
SOC2	-.0236228	.0103565	-2.28	0.023	-.0439212	-.0033244
SOC3	-.0317725	.0091571	-3.47	0.001	-.0497201	-.013825
SOC4	-.0829448	.0098047	-8.46	0.000	-.1021616	-.0637279
SOC5	-.0300882	.0096272	-3.13	0.002	-.0489572	-.0112193

SOC6	-.0450094	.0139857	-3.22	0.001	-.0724209	-.0175979
SOC7	-.0990152	.011961	-8.28	0.000	-.1224583	-.0755721
SOC8	-.0281802	.0099507	-2.83	0.005	-.0476833	-.0086772
SOC9	-.0588301	.0114482	-5.14	0.000	-.0812682	-.036392
Y2	.0117611	.1053299	0.11	0.911	-.1946817	.218204
Y3	.0187514	.2139374	0.09	0.930	-.4005582	.438061
Y4	.0111833	.3468891	0.03	0.974	-.6687068	.6910734
Y5	-.0127587	.4851223	-0.03	0.979	-.9635809	.9380635
Y6	-.0194931	.5966945	-0.03	0.974	-1.188993	1.150007
Y7	.000301	.6763858	0.00	1.000	-1.325391	1.325993
Y8	-.0173608	.7972342	-0.02	0.983	-1.579911	1.545189
Y9	-.0240389	.9059462	-0.03	0.979	-1.799661	1.751583
Y10	.0019933	.9843818	0.00	0.998	-1.92736	1.931346
reg1	.0533503	.2777323	0.19	0.848	-.4909951	.5976957
reg2	.0354732	.2260709	0.16	0.875	-.4076177	.4785641
reg3	.0887534	.1317984	0.67	0.501	-.1695667	.3470736
reg4	.1393642	.046571	2.99	0.003	.0480867	.2306417
reg5	.1513613	.0622639	2.43	0.015	.0293263	.2733962
reg6	-.0226918	.0758562	-0.30	0.765	-.1713672	.1259836
reg7	-.0359637	.119417	-0.30	0.763	-.2700168	.1980893
reg8	-.1106225	.079547	-1.39	0.164	-.2665317	.0452867
reg9	.0150628	.0905798	0.17	0.868	-.1624704	.1925961
reg10	-.1615054	.1022924	-1.58	0.114	-.3619949	.038984
reg11	-.0685473	.0689454	-0.99	0.320	-.2036777	.0665832
reg12	-.0836199	.0787158	-1.06	0.288	-.2379	.0706601
reg13	.0067116	.0672364	0.10	0.920	-.1250693	.1384926
reg14	.0323668	.0525929	0.62	0.538	-.0707134	.1354469
reg15	.0819322	.0911219	0.90	0.369	-.0966635	.2605278
reg16	.0156261	.0832308	0.19	0.851	-.1475033	.1787554