# Testing the 'new economic geography' : a comparative analysis based on EU regional data

Bernard Fingleton

Cambridge University

bf100@cam.ac.uk

paper presented at Workshop on

**Spatial Econometrics**

held at

Kiel Institute for World Economics April 8 – 9, 2005

# Testing the 'new economic geography' : a comparative analysis based on EU regional data

Bernard Fingleton
Cambridge University

Abstract

This paper evaluates 'new economic geography' theory by comparing it with a competing non-nested model derived from urban economics. Using bootstrap inference and the J-test, the paper shows that while NEG theory is supported by the data, it needs to be modified to achieve this, and it is not the only, or even the best or simplest, explanation of regional wage variations across the EU.

## 1. Introduction

Interest in economic geography has been stimulated by the introduction of a formal general equilibrium 'new economic geography' (NEG) theory in which increasing returns to scale are an outcome of each agent solving a clearly defined economic problem within the context of a monopolistic competition market structure (Dixit and Stiglitz, 1977). Recents books, notably Fujita, Krugman, and Venables (1999) and Brakman, Garretsen, and van Marrewijk(2001), have help to popularise these developments in geographical economics, and despite some cautious reactions (Neary, 2001), on the whole NEG theorizing has been reasonably widely appreciated among the broader economics and regional science community, helping to establish at a formal level the role of increasing returns, which had long been seen as a key to understanding the spatial concentration of economic activity. Initially, theoretical developments were at the cutting edge of research activity, but more recently we have seen a growing literature aimed at operationalising and testing NEG (see for example Combes and Lafourcade, 2001, 2004, Combes and Overman, 2003, Forslid et. al. 2002, Head and Mayer, 2003, Redding and Venables, 2004, Rice and Venables, 2003). Among this literature is analysis relating to the so-called wage equation, which links nominal wages to market access or potential[1], and which was initially studied by Hanson(1997,1998) and more latterly by Roos(2001), Brakman et. al. (2002), Mion(2003) and Niebuhr(2004) and Niebuhr(2004). The present paper also follows this strand of analysis.

This recent rigorous empirical work has raised some questions about the operationalization, scope and relevance of NEG theory, and in this heightened wave of constructive criticism, I follow Davis and Weinstein (2003) and Head and Ries(2001) by going beyond NEG model fitting, calibration and parameter estimation to examine the success of NEG in the face of a competing explanation. Although Leamer and Levinsohn's (1994) advice is to 'estimate don't test', it is this kind of direct confrontation that is seen as the acid test of whether a theory can be accepted as the superior explanation of empirical reality. In this spirit, the present paper, building

[1] Harris(1954) was the first to use a variant of the market potential concept.

on the work in Fingleton(2003, 2004), confronts NEG with an alternative (simpler) model derived largely from the literature of urban economics (what is referred to as the UE model), to see which of the two provides a better explanation of variations in nominal wage rates across 200 EU regions.

One issue of particular importance here is the fact that the two competing hypotheses are non-nested, meaning that one is not simply a restricted version of the other, comprising a subset of its the explanatory variables. There is a wide literature on the most appropriate way to test non-nested hypotheses, which is not straightforward, and in this paper I make extensive use of bootstrapping the J-test (following Davidson and MacKinnon, 2002a), in order to obtain appropriate reference distributions for the test statistic.

To summarise, in Section 2 of the paper I briefly set out the basis of the relevant theoretical relationship coming from NEG theory, namely the wage equation linking nominal wages to market potential. Section 3 is concerned with an outline of the competing UE hypothesis. In Section 4, additional covariates are introduced as a necessary requirement for unbiased estimation, and in Section 5, estimation methods are considered and the main empirical results of the confrontation between NEG and UE are presented. Section 6 concludes. The Appendix gives supplementary results that are used to support the conclusions reached.

## 2. The NEG model

The wage equation (1) derives from the system of simultaneous equations given by Fujita, Krugman and Venables (1999), linking nominal wages ( $w_i^M$ ) in the monopolistically competitive sector *M* to market access ( $P_i$), where i denotes region. Note that this is a short-run equilibrium relationship based on an assumption that the migration response (say) to real wage differences is slow compared with the instantaneous entry and exit of firms in the *M* sector (usually taken to be industry) so that profits are immediately driven to zero. It is only in the very long-run that we would expect movement to a stable long-run equilibrium resulting from labour migration.

$$w_i^M = P_i^{\frac{1}{\sigma}} \tag{1}$$

$$P_i = \sum_r Y_r (G_r^M)^{\sigma-1} (\overline{T}_{ir})^{1-\sigma} \tag{2}$$

Equation (2) shows that *P* for region i depends on income levels ($Y_r$), *M* prices ($G_r^M$) and transport costs from region i to r ($\overline{T}_{ir}$), where $\sigma > 1$ is the elasticity of substitution of *M* varieties, summing across all regions including i. The competitive sector C (normally characterised as 'agriculture') consists of goods that are freely transported and produced under constant returns, so that *C* wages $w_i^C$ are constant across regions. We assume iceberg transport costs of the form

$$\overline{T}_{ir} = e^{\tau \ln D_{ir}} \tag{3}$$

in which $D_{ir}$ is the straight-line distance between regions i and r. Since some of the regions are quite large, it is infeasible to assume that internal distances are zero. The problem of internal distance estimation was first considered by Stewart(1947), whose solution underlies the convention (Head and Mayer, 2003) that $D_{ii} = \frac{2}{3}\sqrt{\frac{area_i}{\pi}}$ in which *area*$_i$ is area i's area in square miles. I assume that $\tau = 0.1$, so as to avoid large values in the exponentiation. The use of natural logarithm of distance rather than distance *per se* implies a power function, since $(\mathrm{e}^{\tau \ln D_{ir}})^{\beta} = e^{\beta\tau \ln D_{ir}} = D_{ir}^{\tau\beta}$.

The *M* price index $G_i$ is given by

$$G_i^M = [\sum_r \lambda_r (w_r^M e^{\tau \ln D_{ir}})^{1-\sigma}]^{\frac{1}{1-\sigma}} \tag{4}$$

in which the number of varieties produced in region r is represented by $\lambda_r$, which is equal to the share in region r of the total supply of *M* workers.

Income in region r is

$$Y_r = \theta\lambda_r w_r^M + (1-\theta)\phi_r w_r^C \qquad (5)$$

In order to estimate equation (3), I use the share of $C$ workers in each region $(\phi_i)$, and the share of *M* workers ($\lambda_r$), and the expenditure share of *M* goods ($\theta$) is taken as the overall share of total employment in 2000 that is engaged in *M* activities, assuming also that $\theta$ is also the total *M* workers and 1-$\theta$ is the total *C* workers using a suitable metric that equates the overall number of workers to 1.

Rather than the conventional definition[2], in this paper the *M* sector is identified with the Market Services, with all other activities assumed to be competitive (*C*), since it seems reasonable to characterise the sector as a set of small firms operating under monopolistic competition, producing differentiated varieties under internal increasing returns to scale. This choice is also based on the precedence set in the earlier UE literature (Rivera-Batiz, 1988, Abdel-Rahman and Fujita, 1990). The assumption is that with free entry and exit to the sector and profits continually being driven to zero, there are numerous start-ups so that fixed start up costs are a prominent part of many firms' costs structure, and their small equilibrium size means that internal economies do not become negligible. For example, assume that typical firm t has a single input, labour $(L)$, so that its total cost function is $L = s + am(t)$ in which the fixed labour requirement is *s* and the marginal labour requirement $a$, and the equilibrium output is $m(t)$. Although as $m(t)$ increases, returns to scale (defined as average cost divided by marginal cost) fall asymptotically to 1, typically $m(t)$ is small. Hence it seems reasonable to choose a 'sector' typified by small firms using labour as a predominant input, firms freely entering and leaving the market, and competitive pressure giving a zero profit equilibrium. Additionally, it turns out (see Table 1 notes) that identifying

[2] Manufacturing is assumed to have increasing returns to scale in many theoretical and applied papers, for example Forslid et. al. (2002) use evidence from the presence of scale economies in different industrial sectors provided by Pratten (1988).

the $M$ sector with industry does not alter our conclusion regarding the relative explanatory power of our two competing hypotheses.

Defining the $M$ sector enables us to obtain quantities for $\phi_i$, $\lambda_r$ and $\theta$, but in order to calculate equation (1) it is also necessary to have data on wage rates $w_i^M$ in the $M$ sector and wages $w^C$ for the C sector. Unfortunately these data are not available, and I have therefore used the overall wage level ($w_i^o$) as a proxy for $w_i^M$. The basis of the empirical analysis is therefore annual compensation by NUTS 2 region, data which are produced by Cambridge Econometrics(CE) using the EUROSTAT REGIO database and EUROSTAT national accounts. In the theory, C wages do not vary with region, and I approximate them by assuming that $w^C = MEAN(w_r^o)$. Allowance is made subsequently for the measurement errors these assumptions introduce into our analysis.

For the UK part of the EU compensation data, I use a more credible source of compensation data, namely the New Earnings Survey(NES)[3] giving regional weekly wage rates in pounds sterling. This follows from the fact that the compensation in euros provided in the CE database is an exact linear function of total GVA per worker, so I prefer to replace this by the direct survey data of the NES. Compatibility with the other EU regions was achieved by multiplying each region's NES wage rate by the ratio of overall UK annual euro compensation per employee to the UK overall weekly wage rate[4]. The resulting euros wage data for UK regions gives a total UK wage bill which is exactly equal to the total in the CE database. In fact using the entire CE compensation dataset has a negligible effect on the results obtained. In exactly the same way, the German compensation data for NUTS 2 regions is obtained by scaling the NUTS 1 region wage by the ratio of the output per worker in the NUTS 1 and NUTS 2 regions.

[3] This is an annual employer-based survey carried out by the Office of National Statistics. The data are gross weekly pay for male and female full time workers irrespective of occupation, and are available on the NOMIS website (the Office for National Statistics' on-line labour market statistics database).
[4] Equal to pounds per week times employment for each region to give the overall UK wage bill, then divided by total UK employment.

3. THE UE model

The UE model is the same as that set out in Fingleton (2003), following Rivera-Batiz, (1988), Abdel-Rahman and Fujita, (1990) and Ciccone and Hall(1996), so in order to save space I simply sketch its main features here. The model again divides the economy into an *M* and a *C* sector, with the same characteristics as outlined above, and under the model the *M* sector provides inputs to *C*'s production which have the effect that internal scale economies in the *M* sector translate into external economies to the *C* sector that are increasing in the density of economic activity. This then leads to a reduced form with wages as a function of the density of employment in the area, and thus in this way we have a competing (UE) hypothesis for regional wage variation.

To see this in a little more detail, assume that the production technology for the C sector is a Cobb-Douglas production function

$$Q = (E(C)^{\beta} I^{1-\beta})^{\alpha} L^{1-\alpha} = [f(E)]^{\alpha} L^{1-\alpha} \tag{6}$$

in which $L$ is land, $E(C)$ is the level of $C$ labour units, $E = E(C) + E(M)$, and $I$ is the level of composite services ($I$) derived from the $M$ sector, determined by a CES sub-production function under monopolistic competition. Production is per unit of land, hence $L = 1$, and from this it is possible to show (Fujita and Thisse, 2002, p 102) that the level of $C$ production is defined by the total number of labour units $E$ (in both $C$ and $M$ sectors), in this case per unit area, thus

$$Q = (E(C)^{\beta} I^{1-\beta})^{\alpha} = \phi E^{\gamma} \tag{7}$$

in which $\phi$ is a function of other constants and $\gamma$ is the elasticity where

$$\gamma = \alpha[1 + (1-\beta)(\mu - 1)] \tag{8}$$

The UE model therefore capture increasing returns to the density of activity given by $E$, reflecting the increased variety of $M$ services, so long as $\gamma > 1$. Equation (6) shows that whether or not we see increasing returns depends on services being sufficiently important to final production, which is indexed by the magnitude of $\beta$ < 1, and on the amount of internal scale economies to producer services ($\mu$ > 1). It also depends on congestion effects (1-$\alpha$ < 1) being sufficiently small so as not to overcome the other two factors (Ciccone and Hall, 1996).

The direct comparability of the UE and NEG models depends on both acting as competing explanations for nominal wage rates. For the UE model, the wage rate is the outcome of assuming an equilibrium allocation of production factors so that the coefficient $\alpha$ is equal to the share of $Q$ that goes to $E$ (rather than the other factor $L$), in other words using standard equilibrium theory and equating the wage rate to the marginal product of labour, we obtain

$$w = \frac{\alpha Q}{E} \tag{9}$$

Substituting into equation (7), we obtain

$$\ln(w) = \ln(\alpha\phi) + (\gamma - 1)\ln(E) \tag{10}$$

It is apparent that the UE hypothesis makes no reference to market potential, which depends on transport costs, transport cost mediated price index variations and income variations across regions. The position of a region in relation to other regions is of no consequence, and it is the internal conditions within each region, which are important. Both theories depend on Dixit-Stiglitz monopolistic competition theory, but the $M$ variety elasticity of substitution $\sigma = \mu/(\mu - 1)$ only enters the UE reduced form (10) via the 'returns to scale' parameter $\gamma$, in contrast in NEG $\sigma$ appear*s* in various ways. It is both the coefficient on $P$ in the reduced form (1), and it also determines $P$ , crucially controlling the magnitude of distance cost effects via $\overline{T}_{ir}^{1-\sigma}$ .

4. The extended model specifications

In modelling the wage data, under NEG theory market access (*P*) is the principal explanatory variable, but there are also other ancillary effects that also need to be taken account of in order to allow unbiased estimation. Similarly, under UE theory the wage rates depends primarily on the density of employment (*E*), but will in practice depend also on other factors. I assume that for both hypotheses one of the principal causes of wage rate differences between regions is regional variation in labour efficiency, which is assumed to depend on schooling (*S*) and on technical skills (*T*) acquired at the place of work. In the analysis below I therefore include the variables *S* and *T* to capture efficiency variations across the EU regions.

The schooling variable *S* is the share of the population aged 25-59 with a high level of educational attainment in 1999, as provided for EU NUTS 2 regions by Eurostat's Labour Force Survey. The technical skill variable *T* is represented by the International Patent Classification patents per capita (averaged over 1985-1995) by EU NUTS 2 region that is available from REGIO, which broadly reflects regional variations in R&D activity and therefore workers with computing and information technology skills. Full technical details of data availability, definitions and methodologies are given in the *Regions: Statistical Yearbook* published by Office for Official Publications of the European Communities. First I discuss in more detail what these variables imply by comparing them with supplementary data available for the UK.

Pan-European educational attainment measures are undoubtedly subject to variations due to varying national standards. Moreover there may be doubt that the labour force survey data measures educational attainment with sufficient accuracy. In fact we get a good indication of the quality of the data used by comparing it with UK census data on the proportion of the population (aged 18 and over) with no educational qualifications[5]. The (Pearson product moment) correlation between the 1991 NUTS 2

[5] Unfortunately at the time of writing we only have access to the 1991 census data at the NUTS 2 level, which is a 10% sample provided by the NOMIS database. In order to justify our correlation of S with 1991 UK census data, we observe that at a different level of spatial resolution the 1991 and 2001 censuses give essentially the 'same' distribution. Comparing the 1991 and 2001shares with no qualifications for the 408 unitary authority and local authority districts in Great Britain, we find that while the average population share with no qualifications has fallen dramatically, there exists a strong linear correlation ($r = 0.872$) between the 1991 and 2001 census data sets.

census data on the shares with no qualifications and the NUTS 2 level pan-European labour force survey data indicator (S) is equal to -0.948.

Similarly, our interpretation of IPC patents per capita as a proxy for $T$ is supported by a fairly strong correlation (r = 0.654) at the UK NUTS 2 level with the location quotient based on data from the year 2000 annual business enquiry employee analysis for the two digit sectors 72 (computing and related activities) and 73 (research and development). The assumption is that the workers in these sectors have a high level of computing and related skills which enhances their efficiency.

I also assume that there are various national-level factors relating to differences between countries in labour efficiency, which I capture by country-specific dummy variables. However these dummy variable undoubtedly also represent the net effect of various other country-specific effects, such as any remaining differences employment law and minimum wages, working hours regulations and exchange rates, and so on, so that the national dummies are in effect catch-all variables helping to account for a large portion of the variance in wage rates and hopefully ruling out misspecification bias due to omitted variables. The final specifications are therefore

$$\text{H1:NEG} \quad \ln w^o = b_0 + b_1 \ln P + b_2 S + b_3 T + dummies + \xi \qquad (11)$$

$$\text{H2:UE} \quad \ln w^o = c_0 + c_1 \ln E + c_2 S + c_3 T + dummies + \Psi \qquad (12)$$

There are other wage equation specifications that these can be related to. For instance, analysing data for smaller areal units than the NUTS 2 regions used here, one may wish to explicitly include the effects of commuting in the model, as in Fingleton(2003,2004), allowing the level of worker efficiency in an area to depend on workers resident outside the local area. More generally, a wage equation falls out from the version of NEG theory developed by Helpman(1998) and Hanson(1998). While this has essentially the same micro-foundations as Fujita et. al.(1999), a non-tradable consumption good (housing services) replaces the perfectly traded competitive sector (or 'agriculture' in Fujita et. al., 1999). Brakman, Garretsen and Schramm(2004) develop this approach, creating reduced forms quite similar to equation (11), including district specific control variables (dummies) comparable to the variables *S*, *T* and the dummy variables used here. With regard to equation (12), similar specifications are the outcome of adding variables equivalent in effect to *S*, *T* and the dummy variables to the basic UE specification linking wage rates with employment density. Combes et. al. (2004) for example exploit a large database to control for worker skill differences, emphasising the effect of endogenous interactions (skilled workers attracted to high wages and well as high wages dependent on skilled workers) and the role of amenity difference between areas.

Clearly there are other variables that could be introduced to sharpen the models or replace the variables actually used. For instance Neibuhr(2004) uses the share of total population with qualifications or work experience in science and technology occupations, and also introduces variables such as local amenities (climate etc), sectoral composition (GVA shares in markets services, etc) and border effects. However in a cross-sectional model, if these were significant causes of spatial variation, their omission would show up as significant residual autocorrelation since the dependent variable is itself significant spatially autocorrelated (Moran's I = 0.7508, with standardised values = 13.80, using equation 10). It turns out (see the empirical analysis below) that the simple specification adopted leaves no significant residual autocorrelation, so there does not appear to be any significant omitted variable. It appears as though any remaining unexplained variation can be treated as stochastic error, as represented by spherical disturbances $\xi, \Psi$ with variances $\Omega^2$ and $\Phi^2$ respectively.

It might also be argued that technological externalities have been omitted from the discussion thus far, with the exception of the inclusion of congestion effects in the UE model. There is a growing body of evidence that other un-priced factors will also affect productivity and wage rates, notably as a result of spillover effects relating to knowledge and its enhanced rate of generation and transmission (see for example Audretsch and Feldman, 1996, and Breschi and Lissoni, 2001). The essential idea here is that firms investing in knowledge production will be unable to completely capture the benefits of their investment, which will spill over as an external economies to other firms employing skill-enhanced job-migrants. I therefore assume that the presence of a high proportion of workers who are associated with research and development, knowledge generation and production and transmission, as represented by the variables *S* and *T*, will be associated with additional externalities which boost labour efficiency levels and wage rates, capturing in an indirect way the more elusive technological externalities associated with knowledge flows. These spillovers are likely to be primarily confined within local labour market areas within the EU, since job-migration is much easier than household migration, for various cultural and economic reasons. The NUTS 2 regions are essentially formal (administrative) regions rather than functional regions, and although it is possible that such spillovers will cross regional boundaries so that efficiency levels and hence wage rates in neighbouring regions may tend to be correlated, our tests for residual spatial autocorrelation do not detect them, probably because of they are absorbed by the country dummy effects.

## 5. Estimation methods and results

The initial results in this section show that the NEG model, augmented by labour efficiency variables and catch-all dummies, accounts for a very large proportion of wage variation across the EU NUTS 2 regions. Table 1 summarises the results of fitting equation (11) by 2sls, using the assumed value $\sigma = 6.25$ to construct $P_i$. The value assumed is at the mid-point of the range of empirical estimates given by Head and Mayer(2003). Similarly Head and Ries(2001) and Feenstra(1994) suggest a range from about 4 to 9.

I use 2sls because of the endogeneity and measurement error embodied within the market potential measure *P*, which itself depends on the $w^o$. The issue of endogeneity has been given careful consideration by Mion(2004), who reviews the adoption by Hanson(1998) of higher levels of spatial aggregation (US States rather than counties) of the right hand side endogenous variables to break the link between the county-level right hand side variables and the disturbances. However Mion(2004) argues that this entails an information loss and does not guarantee exogeneity. To overcome endogeneity, Mion(2004) exploits the time dimension, assuming that dynamics occur because of sluggish adjustment to equilibrium rather than fully contemporaneous simultaneity, and picks up spatial dependence via and endogenous and exogenous spatial lags such as *Wx*, where *W* is the n by n matrix with cell (i,j) equal to the reciprocal of distance between regions i and j, post-multiplied by (say) the n by 1 vector *x*. In the spatial econometrics literature, one typically sees similar (cross-sectional) spatial lag models written as $\ln(w^o) = \rho W \ln(w^o) + Xb + \upsilon = (I - \rho W)^{-1}(Xb + \upsilon)$, in which *X* is the n by k matrix comprising the unit vector and k-1 regressors and *b* is a k by 1 vector of parameters, $\rho$ is a scalar parameter and *v* is a vector of well-behaved disturbances. In this case the inversion requires a non-singular matrix, but $I - \rho W$ becomes singular for instance at the points $\rho = 1/e_{max}$ and $\rho = 1/e_{min}$, where $e_{max}$ and $e_{min}$ are the largest positive and negative eigenvalues of *W*. The situation here is fairly similar to the existence of unit roots in time series (Fingleton, 1999), so that $\rho$ estimates outside the stable envelope $1/e_{min} < \rho < 1/e_{max}$ can result in a potentially 'explosive' or non-convergent model with unknown properties. While maximum likelihood estimation (using the correct likelihood function) restrains $\rho$ to within the stable range, this is not guaranteed by other estimation methods. It is therefore questionable whether the estimation carried out by Mion(2004) is appropriate.

In this paper I assume full adjustment to short-run equilibrium and approach the issue of endogeneity from a non-dynamic perspective, relying on the presence of exogenous covariates *S*, *T* and the country dummies. The assumption that *S* and *T* are exogenous contrasts with Redding and Scott's (2003) model of endogenous interaction of schooling (human capital formation) and wages, which proposes that workers upgrade from unskilled to skilled by considering the benefits of this transition, in terms of additional wage rates, in relation to the cost of acquiring additional skills. However in the present context it seems reasonable to assume exogeneity. Not only does the *S* and *T* data pre-date the wages data, and therefore cannot be determined by current wage rates, but they are also undoubtedly affected by factors other than wage differentials, such as Government and EU policy initiatives, institutions and social and cultural differences[6].

[6] Redding and Scott (2003) acknowledge the importance of institutions and policies for economic development by also formalising the indirect effects of less tangible influences operating through agricultural intensity, industry production costs and education costs. For instance the equilibrium

Additionally I utilize the so-called three group method which was initially introduced in the context of measurement error problems (see for instance Koutsoyiannis, 1977, Kennedy, 2003) but which usefully provides an extra instrument that is quite highly correlated with *P* and yet presumably remain uncorrelated with the errors. This is based on the rank order of *P*, with the instrument taking values –1,0,1 according to whether *P* is in the bottom third, middle, or upper third of its ranking.

The first stage of 2sls involves regressing *P*, as defined by equation (2), on *S*, *T*, *WS*, *WT*, $I^{NEG}$, $WI^{NEG}$, and the 14 country dummies. The exogenous spatial lags *WS*, *WT* and $WI^{NEG}$ are the result of multiplying variable *S*, *T* and $I^{NEG}$ by the standardised contiguity matrix[7] *W*, which is derived from the 200 by 200 contiguity matrix $W^*$, hence

$$\begin{aligned} & W_{ij}^* = 1, \qquad i \longleftrightarrow j \\ & W_{ij}^* = 0, otherwise \\ & W_{ij} = \frac{W_{ij}^*}{\sum_j W_{ij}^*} \end{aligned} \tag{13}$$

in which $\longleftrightarrow$ indicates that regions i and j are contiguous. In other words cell i of vector *WS* , for example, is the weighted average of *S* in regions that are contiguous to region i, with weights equal to the reciprocal of the number of regions contiguous to i.

The second stage uses the fitted values of this first stage regression as the instrumental variable for *P*, giving the results in Table 1.

Table 1 near here

supply of skilled workers is reduced by higher agricultural productivity, increased by lower production costs due perhaps to technology transfer, and increased by lower education costs.

[7] This matrix is also used throughout for the spatial autocorrelation tests.

Table 1 shows that the estimated value $\hat{\sigma}$ = 2.329 differs from the assumed mid-point value $\sigma$ = 6.25 used to construct $P_i$ . The approximate 99% confidence interval for $\hat{\sigma}$ is 1.552 to 4.665, although this interval is rather narrow because of the very close fit of the model. Table 2 shows the outcome of re-estimating the NEG model with different assumed values for $\sigma$, indicating that it is only when $\sigma$ reaches approximately 12 does the assumed value lie within the 99% confidence envelope. Although an elasticity of substitution of this magnitude is outside the range normally associated with NEG models, this range is mainly based on $M$ defined as industry, and therefore very diverse varieties with, presumably comparatively low substitutability compared with market services. I find that assuming has $\sigma$ = 12 produces a well-fitting model with goodness-of-fit statistics (Table 3) similar to those obtained by assuming $\sigma$ = 6.25.

Table 2 near here

Table 3 near here

The competing model H2:UE given by equation(12) is fitted in almost precisely the same way as the NEG model, with an instrumental variable for employment density ($E$) to allow for any endogeneity caused by employment levels depending on wage rates. I use the same set of instruments as for equation (11) but with $I^{UE}$ , $WI^{UE}$ in place of $I^{NEG}$ , $WI^{NEG}$, having been calculated in the same way. Table 4 summarises the results, showing that H2:UE performs equally as well as H1:NEG in explaining the variation in wage rates, and although there is an indication of positive residual autocorrelation, it is not significant at conventional levels. The coefficient estimate $\hat{c}_1$ = 0.064766 means that $\hat{\gamma}$= 1.064766, implying increasing returns to scale with a 1% increase in employment density associated with a 1.06% increase in the wage rate.

Table 4 near here

The two competing models of Tables 3 and 4 both account for almost the same proportion wage rates variance, and both show that the regressors $S$, $T$ and either ln $P$ or ln $E$ are significant. However the hypotheses are non-nested, in other words the explanatory variables of one are not a subset of the explanatory variables of the other, so it is not possible to simply test the models by restricting parameters. In general, with non-nested hypotheses, inferential methods used to test nested hypotheses becomes inappropriate ( Cox, 1961,1962, Pesaran, 1974, and Pesaran and Deaton, 1978). In order to overcome this problem, I use the comparatively simple[8] Davidson and MacKinnon(1981,1982) J-test applied to 2sls estimation. This involves estimating the H2:UE model to obtain fitted values $\ln \hat{w}^o_{UE}$, which are then added as an auxiliary variable to the maintained H1:NEG model, giving equation (14). If the coefficient on the added variable is not significantly different from expectation under the maintained hypothesis, then we do not reject H1. However, the non-symmetry of the test means that rejecting H1 does not imply that H2 is true, and vice versa. It could turn out that both H1 and H2 are falsified. We also need to test the opposite case, first estimating H1:NEG to obtain the fitted values $\ln \hat{w}^o_{NEG}$ which then becomes an auxiliary variable under the maintained H2:UE model, as in equation (15).

$$\ln w^o = b_0 + b_1 \ln P + b_2 S + b_3 T + b_4 \ln \hat{w}^o_{UE} + dummies + \xi \tag{14}$$

$$\ln w^o = c_0 + c_1 \ln E + c_2 S + c_3 T + c_4 \ln \hat{w}^o_{NEG} + dummies + \Psi \tag{15}$$

One problem with this approach is that the reference distributions for the t-ratios on the auxiliary fitted variables $\ln \hat{w}^o_{NEG}$ and $\ln \hat{w}^o_{UE}$ are unknown, and not simply N(0,1), which tends to over-reject the null. Fan and Li(1995), Godfrey(1998), MacKinnon(2002c) and Davidson and MacKinnon(2002a,b) suggest the bootstrap J-test to obtain a better measure of the true size of the J-test, and this has been suggested by Godfrey (1983) and Pesaran and Weeks(1999) for non-nested linear regressions estimated[9] by 2sls. Taking H1:NEG as the maintained hypothesis, for example, I first use 2sls to fit equation (14) to obtain

$$\hat{J}_1 = \frac{\hat{b}_4}{s.e.(\hat{b}_4)} \tag{16}$$

[8] There is an extensive literature dedicated to non-nested hypothesis tests, including the Mizon and Richards(1986) encompassing test, although none are as straightforward as the J-test.

[9] Davidson and MacKinnon (2002a) show why bootstrapping the J-test almost always works well compared with the ordinary J-test, even when assumptions of normal errors and exogenous regressors do not hold.

and then refer this statistic to its reference distribution obtained by resampling the residuals[10] under the maintained hypothesis.

Table 5 near here

The estimates in Table 5 are the result of fitting equation (14), with $\sigma$ = 12 and with $\ln \hat{w}^o_{UE}$ the outcome of fitting equation (12) (as summarised by Table 4). For the reference distribution I randomly re-sample with replacement from the vector of residuals produced by the maintained hypothesis H1:NEG. To achieve this, commencing with the equation (11) estimates, I calculate the 2sls residual vector $\hat{\xi} = \ln w^o - \ln \hat{w}^o$ and resample this $B^*$ times to give $\hat{\xi}_B$ , where $B$ = 1…$B^*$ denotes the bootstrap sample number. From this I calculate, for $B$ = 1… $B^*$, $\ln w^o_B = A\hat{b} + \hat{\xi}_B$ , in which $A$ is an n by k matrix with columns 1, ln$P$ , $S$, $T$ and the 14 country dummies and $\hat{b}$ is the k by 1 vector of 2sls estimates given by Table 3, plus the coefficients for the country dummies. First the resulting vectors $\ln w^o_B$ ($B$ = 1…$B^*$ ) are used as the dependent variable to estimate the UE model equation (12) by 2sls which provides fitted values $\ln \hat{w}^o_B$. Second I obtain the set of $B^*$ t-ratios ($\hat{J}_1$ s) by introducing $\ln \hat{w}^o_B$ (in place of $\ln \hat{w}^o_{UE}$ ) as the ancillary variable in equation (14), which is estimated by 2sls. Note that for each of the $B^*$ samples, a completely new set of 2sls estimates are obtained, with the log market potential ( $\ln P_{iB}$ ) and its instrument recalculated using $\ln w^o_B$ ($B$ = 1…$B^*$ ). The $B^*$ $\hat{J}_1$ s are an appropriate reference distribution for testing the significance of the t-ratio given as 5.834 in Table 5.

Figure 1 shows the $\hat{J}$ reference distribution for $B^*$ = 999, which clearly illustrates how the N(0,1) distribution would lead to over-rejection of the maintained hypothesis. The reference distribution has mean equal to 2.107 and variance equal to 0.9123,

---

[10] Davidson and MacKinnon(2002a) recommend scaling the residuals by multiplying by $\sqrt{n/(n-k)}$ , but with n = 200 and k = 18, this amounts to 1.048, which has a negligible effect.

hence 5.834 would be an extreme occurrence under the maintained hypothesis. We therefore have quite strong evidence that H1:NEG should be rejected.

Figure 1 near here

$$\hat{J}_2 = \frac{\hat{c}_4}{s.e.(\hat{c}_4)} \tag{17}$$

While we have rejected H1 using H2, this does not imply that H2 is true, and it is Entirely possible that H2 could be rejected by H1, in which case neither NEG nor UE would be acceptable. In order to test this proposition, I therefore treat H2:UE as the maintained hypothesis and look at the significance of $\ln \hat{w}^o_{NEG}$ in equation (15), where $\ln \hat{w}^o_{NEG}$ is the vector of fitted values given by equation (11). The resulting estimates of equation (15) are in Table 6.

The $\hat{J}_2$ reference distribution is obtained using the same method as for $\hat{J}_1$ but using the Table 4 estimates rather than Table 3. In this case $\ln w^o_B = A\hat{b} + \hat{\Psi}_B$, in which $A$ is an n by k matrix with columns 1, ln$E$ , $S$, $T$ and the 14 country dummies and $\hat{b}$ is the k by 1 vector of estimates given by Table 4. Hence the vectors $\ln w^o_B$ ($B$ = 1…$B^*$ ) lead to the $\hat{J}_2$ reference distribution, again with each $\hat{J}_2$ provided by 2sls estimation in which $P$ is re-calculated for each sample.

Table 6 near here

The $\hat{J}_2$ reference distribution given in Figure 2 has a mean equal to 3.204 and variance equal to 0.9956, so the observed t-ratio of 4.298 is quite close to the expected value and with an upper tail probability of 0.14 could have been generated by randomly re-sampling the residuals from the maintained model.

Figure 2 near here

The evidence I have presented suggests that of the two competing hypotheses, it is UE which stands up better when confronted with the competing hypothesis, and UE falsifies the (augmented) NEG model even though the latter fits the data extremely well. There is some other evidence (See Appendix) that suggests that the NEG model is also tenable, but this refers to NEG models with less plausible elasticities of substitution. For example if we adopt the central value from the typically published range ($\sigma = 6.25$) then there is a significant difference between the estimated value (Table 1) and the assumed value, which places a question mark against the assumed value. Ignoring the difference between estimated and assumed $\sigma$ and repeating the above analysis (see Appendix), does also confirm the main interpretation of the data presented here, that the UE model rejects the NEG model, although in this case there is some indication that the (questionable) NEG model also marginally rejects UE. In the appendix I also give results based on assuming $\sigma = 20$, which is atypical of the published estimates although it does produce not dissimilar estimated and assumed $\sigma$ values. In this case, while the NEG model does not reject the UE model, thus supporting the main thesis of this paper, neither is the NEG model rejected by the UE model (see Appendix). This result however is consistent with our interpretation that UE dominates NEG as a hypothesis, since with $\sigma = 20$, market access (ln $P$) , is strongly linearly correlated with employment density (ln $E$) (see Table 2), so in this case the principal explanatory variables of the two competing hypotheses are very similar.

## 6.Conclusion

New economic geography theory has led to a considerable amount of recent work developing operational NEG models that give weight to empirical as well as theoretical concordance. In this paper I show that an econometric model motivated by NEG theory accounts for a very large proportion of the variation in wage rates across 200 NUTS 2 region of the EU. However, the main contribution of this paper is that it tests the validity and scope of NEG against a competing theory. Given two competing models of more or less equal explanatory power, it seems correct to favour the simpler of the two, in this case the UE hypothesis. To summarise, the principal finding of this paper is that when we directly confront the two hypotheses, it is UE rather than NEG that dominates. On the whole NEG is rejected by UE, whereas UE tends not to be refuted by NEG.

While they do support the underlying theory, the UE model estimates (Table 4) also indicate significant effects due to the variables other than employment density, a point emphasised in Coombes et. al.(2004) who conclude that high local wages are primarily the outcome of high-skill workers gathering in dense local labour markets, and that employment density has a secondary, but significant role. The elasticity for employment density (6.5%) obtained in this paper is consistent with what has been shown elsewhere, for instance Rosenthal and Strange (2004) indicate that the literature has commonly reported elasticities (for productivity) in the range of about 3-8%. Coombes et. al.(2004) prefer the lower bound of their range of about 2-6% depending on data and specification, and the preferred estimates in Fingleton(2004) are 1.4-1.8%. With regard to the NEG model, the fact that its empirical performance evidently improves as $\sigma$ increases points to relatively limited interaction as a determinant of actual market potential, which depends increasingly on the internal market potential of each area as $\sigma$ grows larger. This corresponds to other findings in the literature (e.g. Roos, 2001, Hanson, 1998, Brackman et. al., 2002), which have tended to show mainly localised interaction between areas.

These conclusions should of course be qualified. One important caveat is that they are based on the analysis of comparatively small regions, whereas NEG originates from international trade theory, so that one might anticipate that it would be less relevant to understanding small-scale regional wage variations. It appears that what is important for wage variations among NUTS 2 regions of the EU are the links between competitive industry and market services providers, the increasing variety of which in the larger denser cities imparts increasing returns to scale with employment density. Wages also seem to depend on the efficiency of the labour force, plus technological spillover externalities, captured by measures of schooling and technical skills, plus national-levels effects that are picked up by country dummy variables. In this set up, at this scale of analysis, while NEG also works well, it is unnecessary.

Appendix

Table 1A near here

The $\hat{J}_1$ reference distribution generated assuming that $\sigma$ = 6.25 has mean equal to 1.162 and variance equal to 0.8998, indicating that the empirical t-ratio of 6.053 in Table 1A would be an extreme occurrence under the NEG maintained hypothesis. We therefore have strong evidence by which to reject H1:NEG.

Table 2A near here

With UE as the maintained hypothesis, and again assuming $\sigma$ = 6.25 for the competing NEG hypothesis, the $\hat{J}_2$ reference distribution has a mean equal to 1.908 and variance equal to 1.014. Since 3.912 is only exceeded in 2.803% of samples, using conventional significance levels this is sufficient to also falsify H2:UE, but this inference is relatively weak compared with the very strong rejection of H1:NEG.

Table 3A near here

Table 4A near here

With UE as the maintained hypothesis and the competing NEG model estimated with $\sigma$ =20 (Table 3A), I obtain the $\hat{J}_2$ reference distribution (mean = 4.595, variance = 1.023). Since 43.14% of $\hat{J}_2$s exceed the t-ratio of 4.754 (Table 4A), UE is not rejected by the NEG hypothesis.

With NEG as the maintained hypothesis, the Table 5A t-ratio is exceeded by 15.92% of $\hat{J}_1$ s ( mean = 4.131, variance = 0.9471). This itself provides no strong evidence to reject the NEG model, but as I argue above, the results are also consistent with the UE hypothesis since with $\sigma$ =20 the data generated by NEG and UE are quite similar.

Table 5A near here

References


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

Anselin L and H H Kelejian (1997) Testing for spatial error autocorrelation in the presence of endogenous regressors *International Regional Science Review* 20 153-182

Audretsch D B and M P Feldman (1996) R&D spillovers and the geography of innovation and production *American Economic Review* 86 630-40

Brakman S, Garretsen H and C van Marrewijk (2001) *An Introduction to Geographical Economics* Cambridge University Press Cambridge

Brakman S Garretsen H and M Schramm (2002) *The empirical relevance of the new economic geography: testing for a spatial wage structure in Germany* CESifo Working Paper 395 Munich

Brakman S Garretsen H and M Schramm (2004) The spatial distribution of wages : estimating the Helpman-Hanson model for Germany *Journal of Regional Science* 44 437-466

Breschi S and F Lissoni (2001) Localised knowledge spillovers vs innovative milieux : knowledge "tacitness" reconsidered *Papers in Regional Science* 80 255-73

Ciccone A and R E Hall (1996) Productivity and the density of economic activity *American Economic Review* 86 54-70
Ciccone A(2002) Agglomeration effects in Europe *European Economic Review* 46 213-227

Combes, P-P, Duranton G and L Gobillon (2004) *Spatial Wage Disparities: Sorting Matters! CEPR DP* 4240

Combes P-P and M Lafourcade (2001) *Transportation costs decline and regional inequalities : evidence from France CEPR DP* 2894

Combes P-P and M Lafourcade (2004) *Trade costs and regional disparities in a model of economic geography: structural estimations and predictions for France* Unpublished paper available from http://www.enpc.fr/ceras/combes/

Combes P-P and H Overman (2003) The spatial distribution of Economic activity in the EU *CEPR DP* 3999

Cox D R (1961) Tests of separate families of hypotheses *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* 1 105-123

Cox D R (1962) Further results on tests of separate families of hypotheses *Journal of the Royal Statistical Society B* 24 406-424

Davidson R and J MacKinnon (1981) Several tests for model specification in the presence of alternative hypotheses *Econometrica* 49 781-793

Davidson R and J MacKinnon (1982) Some non-nested hypothesis tests and the relations among them *Review of Economic Studies* 49 551-565

Davidson R and J MacKinnon (2002a) Bootstrap J tests of nonnested linear regression models *Journal of Econometrics* 109 167-193

Davidson R and J MacKinnon (2002b) Fast double bootstrap tests of nonnested linear regression models *Econometric Reviews* 21 419 – 429

Davis D R and D E Weinstein (2003) Market access, economic geography and comparative advantage: an empirical test *Journal of International Economics* Volume 59 Issue 1 January 2003 Pages 1-23

Dixit A and J E Stiglitz (1977) Monopolistic competition and optimum product diversity *American Economic Review* 67 297-308

Fan Y and Q Li (1995) Bootstrapping J-type tests for non-nested regression models *Economics Letters* 48 107-112

Feenstra R C (1994) New Product Varieties and the Measurement of International Prices *The American Economic Review* 84 157-177

Fingleton B (1999) Spurious spatial regression: some Monte-Carlo results with a spatial unit root and spatial cointegration *Journal of Regional Science*, 39 1-19

Fingleton B (2003) Increasing returns: evidence from local wage rates in Great Britain *Oxford Economic Papers* 55 716-739

Fingleton B (2004) *The new economic geography versus urban economics : an evaluation using local wage rates in Great Britain* Mimeo Department of Land Economy Cambridge University

Forslid R Haaland J and K-H Midelfart Knarvik (2002) A U-shaped Europe? A simulation study of industrial location *Journal of International Economics* 57 273-297

Fujita M, Krugman P R and A Venables (1999) *The Spatial Economy : Cities, Regions, and International Trade* MIT press Cambridge Massachusetts

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

Godfrey L G (1998) Tests for non-nested regression models: some results on small sample behaviour and the bootstrap *Journal of Econometrics* 84 59-74

Godfrey L G (1983) Testing Non-Nested Models After Estimation by Instrumental Variables or Least Squares *Econometrica* 51 355-366

Hanson G H (1997) Increasing returns, trade, and the regional structure of wages *Economic Journal* 107 113-133

Hanson G H (1998) Market potential, increasing returns and geographic concentration NBER Working Papers 6429, National Bureau of Economic Research

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

Head K and T Mayer (2003) The empirics of agglomeration and trade *CEPR DP* 3985

Head K and J Ries(2001) Increasing returns versus national product differentiation as an explanation for the pattern of US-Canada trade *American Economic Review* 91 858-876

Helpman E (1998) The size of regions, pp 33-54 in Pines D, Sadka E and Zilcha I (eds) *Topics in Public Economics* Cambridge University Press Cambridge

Huriot J-M and J-F Thisse (Eds) (2000) *Economics of Cities* Cambridge University Press Cambridge

Keeble D Owens P and Thompson C (1981*) The influence of peripheral and central locations on the relative development of regions* Final Report on behalf of the DG for Regional Policy of the EU and UK Department of Industry

Kennedy P (2003) *A Guide to Econometrics* 5th Edition Blackwell Oxford

Koutsoyiannis A (1977) Theory of Econometrics 2nd Edition Macmillan London

Leamer E E and J Levinsohn (1994) International Trade Theory: The Evidence NBER Working Papers 4940, National Bureau of Economic Research, also published in *Handbook of International Economics*, volume 3 1995, Eds. G. Grossman and K. Rogoff, North Holland Amsterdam

MacKinnon J (2002c) Bootstrap inference in econometrics *Canadian Journal of Economics* 35 615-645

Mion G(2004**)** Spatial externalities and empirical analysis: the case of Italy *Journal of Urban Economics* Volume 56 Issue 1 July 2004 Pages 97-118

Mizon G E and J-F Richard (1986) The Encompassing Principle and its Application to Non-Nested Hypotheses *Econometrica 54* 657-678

Neary J P (2001) Of hype and hyperbolas : introducing the new economic geography *Journal of Economic Literature* XXXIX 536-61

Niebuhr A (2004) *Market Access and Regional Disparities* HWWA Discussion Paper 269 Hamburg Institute of International Economics

Ottaviano G I P and J-F Thisse (2001) On economic geography in economic theory: increasing returns and pecuniary externalities *Journal of Economic Geography* 1 153-79

Pesaran H (1974) On the general problem of model selection *Review of Economic Studies* 41 153-171

Pesaran M H and S Deaton(1978) Testing non-nested non-linear regression models *Econometrica* 46 677-694

Pesaran M H and M Weeks (1999) Cambridge Working Papers in Economics 9918 Department of Applied Economics University of Cambridge

Pratten C (1988) A survey of the economies of scale, in Research on the Cost of Non-Europe vol 2, *Studies of the Economics of Integration*, European Commission Luxembourg

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

Redding S and PK Schott (2003) Distance, skill deepening and development: will peripheral countries ever get rich? *Journal of Development Economics* 72 515-541

*Regions: Statistical Yearbook* 2003 Office for Official Publications of the European Communities

Rice P and A J Venables (2003) Equilibrium regional disparities : theory and British evidence *Regional Studies* 37 675-686

Rivera-Batiz F (1988) Increasing returns, monopolistic competition, and agglomeration economies in consumption and production *Regional Science and Urban Economics* 18 125-153

Roos M (2001) Wages and market potential in Germany *Jahrbuch für Regionalwissenschaft* (Review of Regional Research) 21 pp 171-195

Rosenthal S and W C Strange (2004) Evidence on the nature and sources of agglomeration economies in Henderson V and Thisse J-F (eds.) *Handbook of Regional and Urban Economics* volume 4, North-Holland Amsterdam

Stewart JQ (1947) Empirical mathematical rules concerning the distribution and equilibrium of population *Geographical Review* 37 461-85

## Table 1 NEG model estimates

σ = 6.25

| **Parameter**[1] | **2sls estimate** | **st. error** | **t ratio** |
|---|---|---|---|
| constant *($b_0$)* | 8.694769 | 0.310408 | 28.010746 |
| log mkt.pot.($\ln P_i$) *($b_1 = 1/\sigma$)* | 0.429356 | 0.081711 | 5.254596 |
| schooling $S_i$ *($b_2$)* | 0.013898 | 0.001612 | 8.623065 |
| tech.know. $T_i$ *($b_3$)* | 0.376091 | 0.099675 | 3.773192 |
| | | | |
| error variance *($\Omega^2$)* | 0.009588 | | |
| R-squared[3] | 0.9399 | | |
| Correlation[2] | 0.9388 | | |
| Degrees of freedom | 182 | | |
| Residual autocorrelation[4] (z) | -0.07368 | | |

note:

1. The models in these Tables also include 14 national dummy variables, but these estimates are of limited interest and have been omitted.
2. The square of the Pearson product moment correlation between observed and fitted values of the dependent variable.
3. Given by Var( $\hat{Y}$ )/Var($Y$), where $Y$ is the dependent variable.
4. The Anselin and Kelejian (1997) test for residual correlation with endogenous variables and without endogenous lag, using the contiguity matrix.
5. Defining $M$ as manufacturing and construction (or 'industry') and all other sectors as $C$ produces fitted values that are almost identical to those given by this model, with the Pearson product moment correlation coefficient equals 0.9998.

## Table 2 Summary Statistics for various NEG models with different $\sigma$

| $\sigma$ | Upper 99% CL | Lower 99% CL | Residual Sum of Squares | $R^2$ analogue | Corr(ln *P*,ln *E*) |
|---|---|---|---|---|---|
| 3.000 | 1.829 | 0.532 | 1.802 | 0.9368 | 0.5759 |
| $4.900^L$ | 3.239 | 1.064 | 1.771 | 0.9379 | 0.5991 |
| 6.250 | 4.665 | 1.552 | 1.745 | 0.9388 | 0.6181 |
| $7.600^U$ | 6.343 | 2.109 | 1.723 | 0.9395 | 0.6385 |
| 10.000 | 9.320 | 3.139 | 1.691 | 0.9407 | 0.6748 |
| 12.000 | 13.188 | 4.218 | 1.665 | 0.9416 | 0.7031 |
| 15.000 | 17.120 | 5.771 | 1.622 | 0.9431 | 0.7386 |
| 20.000 | 21.093 | 8.144 | 1.572 | 0.9448 | 0.7686 |
| | | | | | |

Note:

U, L = upper and lower bounds of range given in Head and Mayer(2003)

## Table 3 NEG model estimates

$\sigma = 12$

| Parameter | 2sls estimate | st. error | t ratio |
|---|---|---|---|
| constant *($b_0$)* | 9.853781 | 0.094554 | 104.213326 |
| log mkt.pot.($\ln P_i$) *($b_1 = 1/\sigma$)* | 0.156448 | 0.029823 | 5.245808 |
| schooling $S_i$ *($b_2$)* | 0.012318 | 0.001626 | 7.577944 |
| tech.know. $T_i$ *($b_3$)* | 0.368326 | 0.097847 | 3.764303 |
| error variance *($\Omega^2$)* | 0.009147 | | |
| R-squared | 0.9408 | | |
| Correlation | 0.9416 | | |
| Degrees of freedom | 182 | | |
| Residual autocorrelation (z) | -0.03755 | | |

## Table 4 UE model estimates

| Parameter | 2sls estimate | st. error | t ratio |
|---|---|---|---|
| constant *($c_0$)* | 10.134849 | 0.043310 | 234.007942 |
| emp. Density *$\ln E$ ($c_1$)* | 0.064766 | 0.009542 | 6.787265 |
| schooling $S_i$ *($c_2$)* | 0.009134 | 0.001738 | 5.253970 |
| tech.know. $T_i$ *($c_3$)* | 0.247969 | 0.100229 | 2.474030 |
| error variance *($\Phi^2$)* | 0.008869 | | |
| R-squared | 0.9525 | | |
| Correlation | 0.9434 | | |
| Degrees of freedom | 182 | | |
| Residual autocorrelation (z) | 1.780 | | |

## Table 5 NEG as maintained hypothesis

$\sigma = 12$

| Parameter | 2sls estimate | st. error | t ratio |
|---|---|---|---|
| constant *($b_0$)* | 2.904386 | 1.194636 | 2.431190 |
| log market pot. $P_i$ *($b_1 = 1/\sigma$)* | 0.077516 | 0.031681 | 2.446801 |
| schooling $S_i$ *($b_2$)* | 0.003317 | 0.002195 | 1.511089 |
| tech.know. $T_i$ *($b_3$)* | 0.087271 | 0.105613 | 0.826327 |
| $\ln \hat{w}^{o}_{UE}$ *($b_4$)* | 0.696275 | 0.119347 | 5.834050 |
| error variance *($\Omega^2$)* | 0.008439 | | |
| R-squared | 0.9508 | | |
| Correlation | 0.9464 | | |
| Degrees of freedom | 181 | | |

## Table 6 UE as maintained hypothesis

| Parameter | 2sls estimate | st. error | t ratio |
|---|---|---|---|
| constant *($c_0$)* | 2.843273 | 1.696956 | 1.675514 |
| Emp. Density *$\ln E(c_1)$* | 0.044296 | 0.010521 | 4.210097 |
| schooling $S_i$ *($c_2$)* | 0.000487 | 0.002640 | 0.184602 |
| tech.know. $T_i$ *($c_3$)* | -0.027238 | 0.117520 | -0.231772 |
| $\ln \hat{w}^{o}_{NEG}$ *($c_4$)* | 0.712407 | 0.165745 | 4.298210 |
| *error variance ($\Phi^2$)* | 0.008574 | | |
| *R-squared* | 0.9573 | | |
| *Correlation* | 0.9456 | | |
| *Degrees of freedom* | 181 | | |

## Table 1A NEG as maintained hypothesis

$\sigma = 6.25$

| **Parameter** | **2sls estimate** | **st. error** | **t ratio** |
|---|---|---|---|
| constant *($b_0$)* | 2.185597 | 1.113609 | 1.962625 |
| log market pot. $P_i$ *($b_1 = 1/\sigma$)* | 0.256219 | 0.081479 | 3.144585 |
| schooling $S_i$ *($b_2$)* | 0.004063 | 0.002215 | 1.834821 |
| tech.know. $T_i$ *($b_3$)* | 0.079854 | 0.105149 | 0.759441 |
| $\ln \hat{w}^o_{UE}$ *($b_4$)* | 0.694377 | 0.114702 | 6.053745 |
| error variance *($\Omega^2$)* | 0.008359 | | |
| R-squared | 0.9508 | | |
| Correlation | 0.9469 | | |
| Degrees of freedom | 181 | | |

## Table 2A UE as maintained hypothesis

$\sigma = 6.25$

| **Parameter** | **2sls estimate** | **st. error** | **t ratio** |
|---|---|---|---|
| constant *($c_0$)* | 3.539031 | 1.686532 | 2.098407 |
| Emp. Density *$\ln E$($c_1$)* | 0.049477 | 0.010100 | 4.898741 |
| schooling $S_i$ *($c_2$)* | 0.001056 | 0.002672 | 0.395144 |
| tech.know. $T_i$ *($c_3$)* | -0.012935 | 0.118393 | -0.109255 |
| $\ln \hat{w}^o_{NEG}$ *($c_4$)* | 0.643559 | 0.164504 | 3.912105 |
| *error variance ($\Phi^2$)* | 0.008448 | | |
| *R-squared* | 0.9555 | | |
| *Correlation* | 0.9464 | | |
| *Degrees of freedom* | 181 | | |

## Table 3A NEG model estimates

$\sigma = 20$

| Parameter | 2sls estimate | st. error | t ratio |
|---|---|---|---|
| constant *($b_0$)* | 10.224901 | 0.036815 | 277.740422 |
| log mkt.pot.($\ln P_i$) *($b_1 = 1/\sigma$)* | 0.085096 | 0.013837 | 6.150080 |
| schooling $S_i$ *($b_2$)* | 0.010111 | 0.001687 | 5.995522 |
| tech.know. $T_i$ *($b_3$)* | 0.310158 | 0.096943 | 3.199401 |
| | | | |
| error variance *($\Omega^2$)* | 0.008640 | | |
| R-squared | 0.9492 | | |
| Correlation | 0.9448 | | |
| Degrees of freedom | 182 | | |
| Residual autocorrelation (z) | 0.6966 | | |

## Table 4A UE as maintained hypothesis

$\sigma = 20$

| Parameter | 2sls estimate | st. error | t ratio |
|---|---|---|---|
| constant *($c_0$)* | 3.410526 | 1.415081 | 2.410127 |
| Emp. Density *$\ln E$($c_1$)* | 0.036103 | 0.011163 | 3.234032 |
| schooling $S_i$ *($c_2$)* | 0.001937 | 0.002285 | 0.847772 |
| tech.know. $T_i$ *($c_3$)* | 0.030430 | 0.108779 | 0.279741 |
| $\ln \hat{w}^o_{NEG}$ *($c_4$)* | 0.659625 | 0.138750 | 4.754058 |
| | | | |
| *error variance ($\Phi^2$)* | 0.008598 | | |
| *R-squared* | 0.9583 | | |
| *Correlation* | 0.9454 | | |
| *Degrees of freedom* | 181 | | |

## Table 5A NEG as maintained hypothesis

$\sigma = 20$

| **Parameter** | **2sls estimate** | **st. error** | **t ratio** |
|---|---|---|---|
| constant *($b_0$)* | 3.600120 | 1.308687 | 2.750941 |
| log market pot. $P_i$ *($a_1 = 1/\sigma$)* | 0.041895 | 0.016181 | 2.589117 |
| schooling $S_i$ *($b_2$)* | 0.002958 | 0.002192 | 1.349633 |
| tech.know. $T_i$ *($b_1$)* | 0.083665 | 0.106212 | 0.787720 |
| $\ln \hat{w}^{o}_{UE}$ *($b_3$)* | 0.646680 | 0.127698 | 5.064139 |
| error variance *($\Omega^2$)* | 0.008532 | | |
| R-squared | 0.9549 | | |
| Correlation | 0.9458 | | |
| Degrees of freedom | 181 | | |

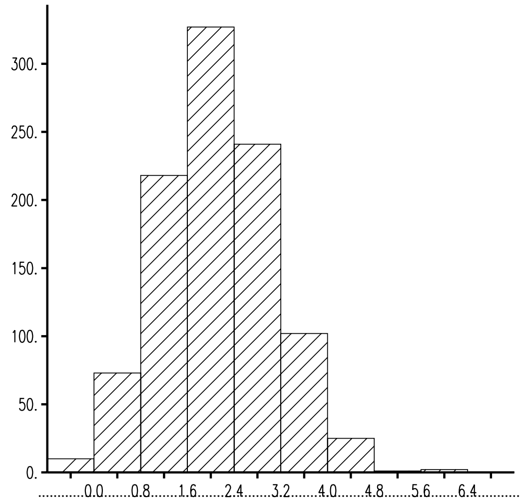


Figure 1

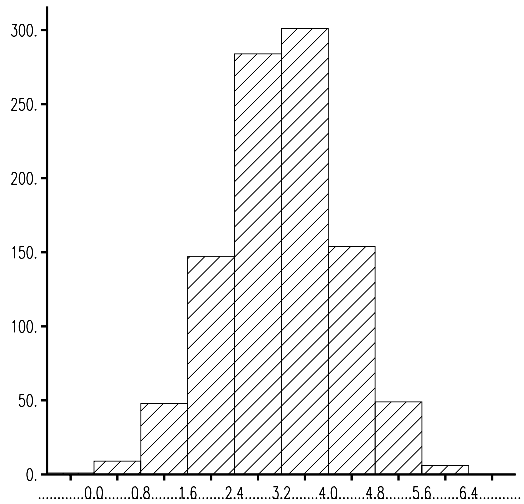


Figure 2