Neighbourhoods, Households and Income Dynamics: A semi-parametric investigation of Neighbourhood Effects

Anne Bolster[1], Simon Burgess[2], Ron Johnston[3], Kelvyn Jones[3]
Carol Propper[2] and Rebecca Sarker[1]

[1] CMPO, University of Bristol
[2] Department of Economics, University of Bristol, CMPO and CEPR
[3] School of Geographical Sciences, University of Bristol

August 2004

Abstract

Using a unique dataset, we present evidence on income trajectories of people living in micro neighbourhoods. We place bounds on the influence of neighbourhood making as few parametric assumptions as possible. The paper offers a number of advances. We exploit a dataset that is large, representative, longitudinal with very local neighbourhoods. We analyse income growth over one, five- and ten-year windows. We analyse the whole distribution of income growth and track large gainers and losers as well as average outcomes. We consider the appropriate definition of neighbourhood. We find little evidence of a negative relationship between neighbourhood and subsequent income growth.

Address for Correspondence
Simon.Burgess@bristol.ac.uk
Department of Economics
University of Bristol
12 Priory Road
Bristol
BS8 1TN

1. Introduction

The relationship between an individual's locale and her life chances is a much-studied one across the social sciences. This issue is given considerable importance by the high levels of income segregation or clustering measured in the US (see Jargowsky, 1997 and 2003) and Britain (see Dorling and Thomas, 2004). However, as Durlauf's (2004) survey indicates, the empirical work is split – quasi-experimental evidence finds little impact of neighbourhood on adult outcomes, while the bulk of observational studies do. In this paper, the use of a unique dataset allows us to make considerable progress on this question, and to open up new lines of enquiry. We present evidence on the long-term income trajectories of people living in different micro neighbourhoods in Britain. Our dataset allows us to investigate income dynamics, different definitions of neighbourhood, and to allow for a high degree of heterogeneity.

The hypothesis we investigate in this paper is that there are important detrimental neighbourhood effects: that is, that otherwise-identical people living in different areas have different prospects. The core problem in this literature is the identification of such a causal relationship given the selection mechanisms operating to assign poor people to poor neighbourhoods. Our principal aim is to place credible bounds on the influence of neighbourhood while making as few parametric assumptions as possible, and no exclusion restrictions, in the context of a large scale representative observational study.

This paper offers a number of significant advances relative to the literature. First, we exploit a dataset[1] that is representative for Britain, that is longitudinal (following individuals over ten years), and has very local neighbourhood characteristics (down to the nearest few streets, around 500 people). Our neighbourhoods are very local and are more likely to correspond to 'real' neighbourhoods than the ward (in Britain) or census tract (in the US) units commonly used in other analyses. This combination of attributes makes the dataset extremely powerful; we detail all this below[2].

Second, we analyse income trajectories over one-, five- and ten-year windows. Thus we address the extent to which income prospects over the future are related to the nature of the

---

[1] This is the British Household Panel Survey (BHPS).

individual's local neighbourhood? The panel gives us ten years of data for individuals and thus allows us to take dynamics more seriously than previous studies in this context. Our focus is on adults, and household income is our main outcome variable. This is of significant interest in its own right (as the long-standing research on neighbourhoods and poverty attests). But it also serves as a catch-all for other neighbourhood influences. An individual's environment may influence her employment, health, marital status, number of children and so on; if present, these influences are all likely to be reflected in income. We do not condition on these factors so as to allow neighbourhood the maximum influence. Third, rather than simply analysing the mean income growth by neighbourhood type, we analyse the whole distribution and so can track large gainers and losers as well as average outcomes. We use graphical procedures and quantile regression to characterise any changes in this distribution across neighbourhood characteristics.

Fourth, we consider the appropriate definition of neighbourhood. Our data allow us to construct 'bespoke' neighbourhoods around individuals and to consider different spatial scales. We compare the influence of characteristics of a very local definition of neighbourhood with a broader definition. This is usually ignored in other quantitative studies as neighbourhood is defined just by the available data (see Durlauf, 2004, pp. 63 – 5). We also investigate the impact of *neighbouring* neighbourhoods on outcomes. So, for example, what are the outcomes for residents in a poor area nested within a better off area, compared to those in a poor area in a wider poor area? This appears to be a new approach. Dietz (2002) notes that the standard neighbourhood model assumes that "no interaction occurs among neighbourhoods. … Thus neighbourhoods with identical characteristics but dissimilar neighbouring neighbourhoods are considered equivalent." (p. 541).

We find a strong negative contemporaneous correlation between the level of income and the disadvantage of the neighbourhood. Thus at least one of two mechanisms – causality or sorting – is working to generate this pattern. Moving on to the dynamic results, our findings show no evidence of a negative relationship between neighbourhood and subsequent income growth. This is true for one-year, five-year and ten-year changes, for almost all population groups, and at different parts of the income growth distribution. If anything, the results show that the distribution of income growth is shifted up somewhat for individuals starting in poorer

---

[2] The combination of the BHPS with the neighbourhood data has been used before, by Buck (2001), but he does not exploit the longitudinal element of the data that is key to our approach.

neighbourhoods. If we condition on the individual's prior income level, we do isolate a negative association of income change and neighbourhood. We focus on this in the modelling framework, which highlights the role of two factors in interpreting neighbourhood influences – the dynamic adjustment of income and the nature of the housing finance system, particularly in responding to temporary income shocks. In summary, we argue that our results are consistent with at most only a small detrimental neighbourhood effect.

The next section briefly reviews related literature. Section 3 sets out a modelling framework. Section 4 discusses the data in detail, and section 5 presents the results. The final section offers some conclusions.

2. Literature

The excellent survey of the literature on neighbourhood effects available in Durlauf (2004) means that we do not need to provide a lengthy overview here[3]. Instead we summarise Durlauf's findings and highlight some of the issues most relevant to this paper. We also discuss in more detail some recent papers using similar data to ours for Britain.

Durlauf (2004, p. 2) credits the work of Wilson (1987, 1996) with a significant role in the resurgence of interest in neighbourhoods. Manski (2000, p. 122) also alludes to the power of Wilson's argument. Cutler and Glaeser (1997) focus on the effects of neighbourhoods defined by poverty and race and ethnicity. The recognition of the long-run persistence of spatially concentrated areas of poverty has also been important, as has been the refinement of techniques. Theoretical analyses of neighbourhood influences most relevant to this paper are largely based on models of social interactions (see Blume and Durlauf, 2001; Brock and Durlauf, 2003; Manski, 2000). These are based on role model effects or peer group influences, or in Manski's terminology, interactions of expectations or preferences. In our context, this would mean that the observation of individuals with particular income growth paths changes the views of others on what was feasible in their current situation (role models or informational interactions); and that observation of individuals motivated by hard work and financial success, or the reverse, inspires similar preferences among others (peer groups or preference interactions).

---

[3] Dietz (2002) offers a survey more accessible outside economics.

Durlauf (2004) categorises empirical evidence on neighbourhood influences by the research methodology used. The quasi-experimental evidence, provided principally by the Gautreaux programme and the Moving to Opportunity (MTO) demonstration, is very useful in side-stepping some of the identification problems associated with observational studies. Rosenbaum (1995) and others for Gautreaux, and Katz, Kling and Liebman (2001), Ludwig, Duncan and Hirschfield (2001) and Goering, Feins and Richardson (2002) among others for MTO, detail the results. Similarly, Oreopoulos (2003) exploits the random assignment of children to housing projects in Toronto. Durlauf, (2004, pp. 61 – 2) notes that even the quasi-experimental evidence is not unproblematic and not necessarily generalisable. There appears to be agreement that moving to better neighbourhoods generally has some positive effects for children: higher college attendance, better health outcomes and reduced behavioural problems. However, Oreopoulus (2003) finds no long-run effect of a poor neighbourhood on labour market outcomes, and Keels et al (2004) show that the latest MTO results show little impact on school test scores.

However, the results for adults – the chief focus of our study – are more mixed. Katz, Kling and Liebman (2001) and Goering, Feins and Richardson (2002) report little impact on adult economic outcomes. Note that in the former study, the minimum length of time which sample members had lived in their new residence was 9 months, and the average time between assignment and the follow-up survey was only 2.2 years, and it may be that over a longer span, stronger effects would emerge. Indeed the theory suggests that the overall impact of moves to better neighbourhoods on employment is ambiguous: the move may increase access to employment opportunities, and local norms may be more supportive of work and less so of welfare use, but the move may disrupt access to social support networks that would previously have been sources of job information and child care.

Turning to the non-experimental analyses, Durlauf notes that among the mass of studies available, "the bulk of empirical studies … find evidence of their presence" (p. 54). Since most of the studies he discusses refer to adults, albeit typically young ones, this stands in some contradiction to the quasi-experimental evidence noted above. However, because of a variety of statistical and conceptual problems in the studies, Durlauf concludes that in fact they may only provide limited support for the importance of neighbourhood influences.

Many authors have discussed in detail the nature of the econometric challenges facing empirical investigation of neighbourhood effects – for example, the original paper by Manski (1993), Moffitt (2001), Brock and Durlauf (2003), and Durlauf (2004). Identification is the key problem. It is at its most acute in Manski's (1993) reflection problem, when the researcher is trying to estimate the influence on agent A's behaviour of the behaviour of A's group. Less acute but still problematic is the case, relevant to our study, when the neighbourhood/group is chosen by the agent. This selection effect means that simple regressions cannot be straightforwardly interpreted. The usual instrumental variables approaches are available in principle but finding suitable instruments is difficult, and identification by functional form or other parametric assumptions is fragile. In this paper we do not attempt to identify directly a causal relationship. Instead, we exploit the longitudinal nature of our data, combined with its very local scale to put bounds on any neighbourhood influences on life chances. Of course, selection is still an issue, and this is not a general approach – but the results in this case do allow us to argue that neighbourhood influences appear to be unimportant.

In recent papers on neighbourhood influences on income and poverty using the BHPS, Buck (2001) and McCulloch (2001) adopt different approaches to ours. McCulloch (2001) examines the relationship between a ward-level disadvantage index[4] and a number of financial, health, and social support outcomes for individuals using a multi-level approach. The individual outcomes analysed are individual self-assessment of financial difficulties, expectation of financial situation a year ahead, a low income dummy variable, self-rated health status, mental health status, dislike of current neighbourhood, and the level of social support[5]. The sample is based on the first 8 waves of the BHPS (1991-1998) and individuals are included in the dataset up to the point at which they first move home, when they are removed from the estimation sample. This sample selection is an attempt to reduce the extent to which area variations in social and economic outcomes may be attributable to sorting of individuals into areas. However, since we cannot assume that individuals were randomly assigned to wards when first seen in 1991, this clearly does not deal with the problem. Estimation is contemporaneous, with the current outcome regressed on the characteristics of the current area. The sample is treated

---

[4] A ward is an electoral unit of about 12,000 people, and the disadvantage index is a Townsend composite index with 1991 census data.

[5] Social support variable is proportion of negative responses to the following questions: (1) Is there someone who will listen? (2) Is there someone to help in a crisis? (3) Is there someone you can relax with? (4) Is there anyone who really appreciates you? (5) Is there anyone you can count on to offer comfort.

as a large cross-section and the longitudinal element is ignored. An unconditional model[6] shows that area disadvantage increases significantly the probability of negative outcomes for seven out of eight of the outcomes considered (not financial expectations) for both men and women. However, once a number of individual and household characteristics (household type, education, ethnicity, social housing, car access and unemployment) are controlled for, the effect reduces and remains significant for only 4 of the outcomes: employment (not for women), current financial situation, health status, and a dislike of the neighbourhood.

Buck (2001) uses very similar data to McCulloch – the BHPS for years 1991-1999, and the same ward-level Townsend deprivation index using 1991 census data – but examines six different individual outcomes (a non-monetary poverty index, various employment variables, and low household income indicators, whether exit or enter low income status). Buck also finds that adding in additional individual and household controls attenuates the impact of the area, but in this study they remain significant. Buck also investigates the scale at which area effects operates (the area disadvantage index in this case being the local area unemployment rate) using census enumeration districts at various scales (this is the same definition we use, discussed more below). For all outcomes except employment expectations, area effects declined with increasing distance or number of people for which the local area was defined.

3. Modelling framework

In this section we set out an economic model to interpret our results. We set out two assumptions about the nature of heterogeneity in income, and two assumptions about the selection process. The aim is to determine what conditions have to hold for a model with detrimental causal neighbourhood effects to be consistent with our findings. Our definition of neighbourhood effects is that the future life chances of otherwise identical individuals placed in different neighbourhoods will evolve differently.

a) Income Models

Income model 1: static (level) heterogeneity

The first and simplest model we consider for individual *i* at time *t* is:

$$y_{it} = \boldsymbol{b}X_{it} + h(age_{it}) + \boldsymbol{a}Z_{j(i)t} + \boldsymbol{m}_i + \boldsymbol{e}_{it} \qquad (1)$$

---

[6] With only region of residence and interview year controlled for.

where y is income, X individual characteristics, Z the neighbourhood characteristic, m a time-invariant individual effect, and e is a (possibly serially-correlated) error term. We discuss possible correlation of e and Z below. The direct area effect issue we are investigating is whether $\alpha = 0$ or not. Note that we are not modelling neighbourhood characteristics, $Z_{j(i)t}$, as time-varying in themselves. The t subscript indicates that i can change location and hence neighbourhood type through time. Clearly, neighbourhoods do evolve, and in a way that may be correlated with the income dynamics of their inhabitants, but we do not need to model that here to make our point.

Taking a t-period difference yields:

$$\Delta^t y_{it+t} = b\Delta^t X_{it+t} + \Delta^t h\left(age_{it+t}\right) + a\Delta^t Z_{j(i)t+t} + \Delta^t e_{it+t} \qquad (2)$$

Note that the individual fixed effect drops out after differencing. So if there is sorting into neighbourhoods on the income intercept, this is dealt with by looking at income growth.

We allow neighbourhood to have an <u>indirect</u> impact, by allowing it to have an influence on DX and DZ – that is, an impact on employment, marital status etc. and on the evolution of the individual's location. So we assume that DX and DZ change according to a repeated Markov process resulting in:

$$\begin{aligned}\Delta^t X_{it+t} &= f\left(X_{i0}, Z_{i0}\right) + u_i \\ \Delta^t Z_{it+t} &= g\left(X_{i0}, Z_{i0}\right) + w_i\end{aligned} \qquad (3)$$

where the terms u_i and w_i stand for the history of shocks between t and $t + t$. Assuming $f()$ and $g()$ are linear:

$$\begin{aligned}\Delta^t X_{it+t} &= f_1 X_{i0} + f_2 Z_{i0} + u_i \\ \Delta^t Z_{it+t} &= p_1 X_{i0} + p_2 Z_{i0} + w_i\end{aligned} \qquad (4)$$

We assume that the errors are uncorrelated with Z_{i0} and each other. Substituting into (2) yields:

$$\Delta^t y_{it+t} = bX_{i0} + aZ_{j(i)0} + \left[u_i + w_i + \Delta^t e_{it+t}\right] \qquad (5)$$

Note that $Dh(age)$ is invariant across i if $h()$ is linear, and just drops into X_{i0} if not. The error terms are uncorrelated with X_{i0} and Z_{i0}. The initial conditions problem is that Z_{i0} might be correlated with m but this was differenced out; any correlation with De remains. The coefficients b and a are defined as follows: using (2) and (4) we find that $b = bf_1 + ap_1$ and $a = bf_2 + ap_2$. The hypothesis of no neighbourhoods effects can then be thought of as both $\alpha = 0$ (no direct impact) <u>and</u> $f_2 = 0$ (no indirect impact); i.e. that $a = 0$ in (5).

Income model 2: dynamic (growth) heterogeneity

It seems possible that individuals differ in unobserved ways that affect income growth as well as simply the level. To allow for a broader incorporation of heterogeneity, we modify the income equation (1):

$$y_{it} = \boldsymbol{b}X_{it} + \boldsymbol{m}_{0i}h(age_{it}) + \boldsymbol{a}Z_{j(i)t} + \boldsymbol{m}_{1i} + \boldsymbol{e}_{it} \qquad (6)$$

This allows for heterogeneity in the age-income growth profiles. After differencing and substituting in again from the Markov chain, this will add an individual fixed effect into the equivalent to (4):

$$\Delta^t y_{it+t} = bX_{i0} + aZ_{j(i)0} + \left[\boldsymbol{m}_{0i} + \boldsymbol{u}_i + \boldsymbol{w}_i + \Delta^t \boldsymbol{e}_{it+t}\right] \qquad (7)$$

Now this is problematic in the standard selection way as Z_{i0} and $\boldsymbol{m}_{0i}$ may be correlated. That is to say, an observed correlation between $\boldsymbol{D}y$ and Z might arise through direct or indirect neighbourhood effects (the parameter a), or through selection (through Z depending on $\boldsymbol{m}_0$). Thus, if there is neighbourhood sorting on the age-income slope, then isolating a causal relationship is still problematic even in differences.

Income model 2 refinement: dynamic (growth) heterogeneity, and response heterogeneity.

Finally, we can allow for heterogeneous direct responses (heterogeneous treatment effects) of individuals to any neighbourhood effects. We re-write the income model as:

$$y_{it} = \boldsymbol{b}X_{it} + \boldsymbol{m}_{0i}h(age_{it}) + \left(\overline{\boldsymbol{a}} + \boldsymbol{a}_i\right)Z_{j(i)t} + \boldsymbol{m}_{1i} + \boldsymbol{e}_{it} \qquad (8)$$

So individuals may respond differently to the neighbourhood characteristic. Working through the algebra, and imposing a common effect of X variables:

$$\Delta^t y_{it+t} = \overline{b}X_{i0} + a_i Z_{j(i)0} + \left[\boldsymbol{m}_{0i} + \left(b_i - \overline{b}\right)X_{i0} + \boldsymbol{u}_i + \boldsymbol{w}_i + \Delta^t \boldsymbol{e}_{it+t}\right] \qquad (9)$$

where b_i and a_i follow straightforwardly as before. There are now further problems in that Z_0 may be correlated with the X_0 term in the composite error term, a form of the initial conditions problem.

b) Selection models

Both selection models work through the housing market. We assume that housing in each location has a price, and that this price depends on Z. Households sort into districts on the basis

of their income (see Epple and Sieg, 1999, for an example of such a model). Differences in tastes could be added with no extra insight. We argue below that the important issue is whether location depends on permanent income only or also responds to temporary shocks.

Selection model 1: permanent income and housing

The assumption in this case is that either housing finance depends on permanent income only, or moving costs are too high to allow moving for any temporary income shock. In fact, housing finance is likely to depend in part on expected <u>future</u> income, and the only predictable component of that is permanent income. Either way, the assumption is that temporary income falls in income do not trigger moves to a worse location.

$$Z_{j(i)t} = \boldsymbol{g}\left(\overline{X}_i, \boldsymbol{m}_i\right) + noise \qquad (10)$$

Note the dependence of Z on $\boldsymbol{m}$ which is at the root of the selection problem in estimating the impact of neighbourhood on the income process. To explain actual movement across areas, a more complex model would allow for slow learning of a true underlying $\boldsymbol{g}(.)$, and for liquidity constraints, but this suffices to make our point.

Selection model 2: temporary income shocks and housing

In this model we allow temporary income shocks – $\boldsymbol{e}_{it}$ – to matter for location. The assumption here is that either the housing finance system cannot support temporary falls in income, or that temporary falls are mistaken for permanent falls, or that moving costs are very low. But permanent and temporary income may have different impacts on location:

$$Z_{j(i)t} = \boldsymbol{g}\left(\overline{X}_i, \boldsymbol{m}_i\right) + \boldsymbol{d}\left(\boldsymbol{e}_{it-1}, \boldsymbol{e}_{it-2}, ..\right) + noise \qquad (11)$$

The key point here is that the dependence of $Z_{j(i)t}$ on $\boldsymbol{e}_{it-1}$ ties any mean reversion or 'rebound' in income to neighbourhood. In this context, mean reversion implies that $E\left(\Delta y_t \mid y_{t-1} < \overline{y}\right) > 0$; that is, individuals with a low income draw in *t-1* are more likely to get a positive income change in period *t*. The relevance for the results reported below is that if individuals with a low income draw in *t-1* move to a poor location, then this positive income change is associated with the poor location.

In the analysis below, we examine income level and growth in relation to initial neighbourhood (Z_0), after accounting for some initial characteristics X_0. This is to be interpreted as the

counterpart to (1) and (5) in model 1 or (8) and (9) in model 2. We also estimate models including a lagged dependent variable, as a method for dealing with income dynamics. First, however, we describe the dataset.

4. Data

The key components of our data are individuals and areas.

(a) Individuals

We use the first ten waves of the British Household Panel Survey (BHPS) covering 1991-2000. The first wave of the BHPS was designed as a nationally representative sample of the population of Great Britain living in private households in 1991, and had a sample size of over 5,500 households covering over 10,000 people. On-going representativeness of the (non-immigrant) population has been maintained by using a following rule typical of household panel surveys: at the second and subsequent waves, all original wave 1 sample members (OSMs) are followed (even if they move house, or if the household splits up), and there are annual interviews with all adult members of all households containing either an OSM, or an individual born to an OSM whether or not they were members of the original sample. New panel members who subsequently stop living with an OSM are, however, not followed and interviewed again. Thus, for example, if a non-OSM married an OSM at wave 2, and the partnership subsequently dissolved, the OSM is followed, but the non-OSM is not. BHPS cross-sectional weights are applied. We select a pooled sample of 109,026 individual observations who form 59,620 household observations. Of these 109,026 individuals, 92689 have non-missing household income data. Endogenous sample attrition is an issue. On item non-response, individuals with missing income data are not very dissimilar to those with complete income data for all years in terms of employment status, housing tenure and sample mean income. They are slightly more likely to come from poorer neighbourhoods (around 15% of households have missing income in the richest areas compared to around 16.3% in the poorest). Individuals typically have a time-series of data for which income data may be missing for a small number of years. We also set to missing a small number of household income observations with very low income – below £3000 per year.

The outcome variable in this paper is net[7] annual household income, deflated to January 2001 prices and equivalised using the McClements scale, before housing costs. This variable has been constructed by Bardasi, Jenkins, Rigg (2003) using data from the BHPS at each wave. This net income figure includes net income from employment, investments, pensions, benefits and other transfers. We analyse the levels of household income, one-year percentage changes in income (dated t to $t+1$), and five-year non-rolling[8] changes in income (t to $t+5$). We analyse income trajectories over the whole sample window. This is measured by the slope coefficient of the regression of household income against time, computed separately for each individual, for all individuals with more than seven observations.

We also look at earnings; these are defined as: net annual labour income[9] deflated to January 2001 prices. It is an individual level variable. As for household income, we analyse the levels of earnings, one-year percentage changes in earnings (dated t to $t+1$), five-year non-rolling changes in earnings (t to $t+5$), and earnings trajectories over the whole sample window.

(b) Areas

Many studies are forced to use rather large scale areas to capture neighbourhood effects. We create a set of 'bespoke neighbourhoods' for each individual at each point in time, a procedure developed independently by Tunstall *et al* (2000) and Buck (2001). The smallest of these is based on the characteristics of the people in the nearest few streets; larger ones are constructed to reach certain size thresholds. Each individual's home postcode is matched to an enumeration district (ED). These are the smallest data units for which UK census data are made available – at the 1991 census they contained around 500 people. Adjacent districts are then identified according to the distance between their population-weighted centroids, and enumeration districts are aggregated into the bespoke neighbourhood until the required population or distance threshold is crossed. We identify different spatial scales defined by population (the nearest n people to the respondent's home address, where n = 500, 1000, 2000, 5000, 10000).

---

[7] Of national but not local taxes.
[8] I.e. If the individual is present in all ten waves, he/she will have two observations for this variable: at wave 1 and wave 5. All other individuals have a non-missing value for this variable for the first wave observed they are observed, subject to income being also observed in the fifth wave following entry to the sample.
[9] Both the household income and individual earnings variable are for the period of 1st September of year t-1 to 31st August of year t.

Having defined the bespoke neighbourhoods, we characterise them using Census 1991 data for these ED sets. Eighteen variables are available describing the socio-economic and demographic character of the ED. We construct a composite index of area characteristics at each of the spatial scales using principal components analysis (see Johnston et al 2004). The first principal components is essentially a measure of disadvantage: it is increasing in unemployment, proportion living in local authority rented accommodation, having no car, long-term sick, single parent families, and decreasing in the proportion owning their homes outright, and employed in professional or managerial occupations (see Appendix 2 for details). We label our two area variables with the scale variable (people) and the radius (500 and 10,000): *p500f1* and *p10kf1*. These variables are scaled to have zero mean and unit standard deviation.

We have the location data for each individual at each date, so we can track them as they change neighbourhood from year to year. The area descriptors are only available at one date (the 1991 Census) so our characterisation of each area is fixed in time. To the extent that areas can change, we would only be measuring with error the true state of a neighbourhood at a later date than 1991. However, the characteristics of most areas remain relatively constant over substantial periods with only a few – such as those subject to 'gentrification' – experiencing significant change in a short period. The advantage of using the fixed indicator is that we do not have to model a process of neighbourhood change that would likely be correlated with the focus of interest, the dynamics of individual income. We discuss in a later section the consequences of possible measurement error induced by having a non-time-varying indicator.

We also use BHPS information to see how closely these census-based variables correspond to features of neighbourhoods that are often cited in qualitative studies of neighbourhoods. For some (not all) waves the BHPS collects data on respondents' views of their neighbourhood and their dwelling. These factors are listed in Table 1, along with their correlation coefficient with *p500f1*. Neighbourhood variables such as perceptions of the degree of vandalism, graffiti, street violence and car crime correlate quite highly with our neighbourhood measure. This helps to paint a picture of what a low *p500f1* neighbourhood looks like, and give us some confidence that the rather abstract neighbourhood characteristic measure derived from factor analysis does correspond quite well with more subjective measures of neighbourhood quality. Interestingly, whilst the accommodation characteristics are also significantly associated with *p500f1*, the absolute value of the correlation is very low – all under 0.1 for the physical features

of the dwelling *per se*. This implied heterogeneity of dwelling quality within very small neighbourhoods previews the heterogeneity of income growth we report below.

Appendix 1 shows a summary of the variables we use, with descriptive statistics.

5. Results

We present our results using distributional graphs and quantile regressions. We impose little structure on the pattern of outcomes over areas. The idea is to allow for heterogeneous responses to neighbourhood conditions[10].

The graphs group neighbourhoods into centiles of *p500f1* and plot the quartiles of the outcome distribution against the ranked centiles, with the most deprived at the right of the graph. Each centile contains approximately 1000 observations for income levels, though there are fewer than this because of missing income data, and clearly for the five-year (resp. trajectory) analyses there are fewer observations available as each individual can appear only twice (resp. once). The nature of the neighbourhood variable plotted in these graphs is such that each centile is necessarily homogeneous in terms of the disadvantage score, *p500f1*. However, they are not necessarily close in space, as in general, each centile will be made up of EDs from different areas. For example, the top centile will contain EDs from the richest areas around Britain, though spatial autocorrelation will mean a closer spatial relationship than random assignment. The quantile regressions use the raw *p500f1* values and all observations. We use deflated, equivalised household income in all analyses (other than in the sub-section on earnings).

After first looking at income levels, we analyse short-run (one-year) income changes and medium-run income changes (5- and 10-year changes). We also look at the role of lagged income, and the earnings component of income. As robustness checks we discuss the appropriate definitions of neighbourhoods and measurement error. Finally, we summarise and interpret the results using the modelling framework above.

---

[10] Moffitt (2001) gives examples of how the same neighbourhood can affect different people in different ways.

(a) Income levels

Figure 1a shows a clear gradient of household income levels and neighbourhood disadvantage. The unit of observation here is a household-year, and one observation is taken per household[11] per year. The distribution of household income is both shifted down and compressed in poorer neighbourhoods. The median varies by about a factor of two between the poorest and least poor neighbourhoods; the upper quartile by more than two, and the lower quartile by relatively little. The second panel of the Figure shows that the same pattern holds at the radius of the nearest 10,000 people (in fact, the pattern is repeated at all scales). Similarly, the variance of household income within each centile of *p500f1* (10,875) is much larger than the variance of the mean income across centiles (2,938). This is remarkable given that the building blocks of these centiles are very small scale areas; and shows that income heterogeneity is large even at a very local level.

These observations are confirmed by the quantile regressions shown in Table 2a: the neighbourhood disadvantage index has a negative coefficient at all the quartiles, and shows a larger negative gradient through the 75th percentile than through the 25th percentile. The Table also shows a very similar set of coefficients using the broader neighbourhood variable, *p10kf1*. This finding is repeated for income changes below[12]. The size of the coefficients can be judged relative to the distribution of income levels; the neighbourhood factor is scaled to have unit standard deviation, and a unit change in this is associated with a change of around £2000 in income, compared to a mean income level of £16,880 and a standard deviation of £11,701. This is not trivial, equivalent to about one sixth of the standard deviation in income.

This unconditional picture may be generated by life-cycle or other effects. We therefore condition on a set of fixed characteristics (age, qualifications, gender, and year), compute the residuals from the OLS regression and repeat the graph and quantile regression procedures. We explicitly do not condition on a range of other factors that affect income, such as employment status and marital status. This is because we want to allow neighbourhood to potentially influence these outcomes, and so do not wish to condition them out. The graph in Figure 1c is little changed from Figure 1a. The quantile regressions in Table 2a shows significantly lower

---

[11] For the conditional results reported below, we use the head of household's characteristics.
[12] It partly reflects spatial autocorrelation – we have discussed this 'fractal' nature of the neighbourhood factors elsewhere (Johnston et al, 2004).

coefficients on the neighbourhood variable, about half the size at the median, but these are still statistically very significant.

Table 2b presents evidence split by household structure and by housing tenure. The results show the same overall pattern is apparent for all groups, but we see that the impact of neighbourhood is lowest for the old, and for single parents. The other family structure groups are all similar and show a stronger relationship with neighbourhood. Splitting by housing tenure groups, the relationship between neighbourhood type and income level is weaker for those in rented accommodation.

(b) Short run changes

The quartiles of the one-year percentage change in household income from t to $t+1$ are plotted against the area disadvantage index at t in Figure 2. The unit of observation here (and for the other change analyses) is an individual-year, as individuals can change household. The figure shows a very clear result: the distribution of short-run changes in income is approximately the same across the whole distribution of neighbourhood types. In the richest few percent of neighbourhoods in Britain, the one year income change runs from an upper quartile of about 20% to a lower quartile of –10%; in the poorest neighbourhoods, it has the same distribution. Throughout the distribution of neighbourhood types, the upper quartile centres around approximately 15-20%, the median around 0-5%, and the lower quartile around –10%[13]. The same pattern is true in the second panel of the graph using centiles of the broader *p10kf1* neighbourhood variable. Again the within-variance (across centiles of *p500f1*) of 60.1 dominates that of the between-variance, equal to 2.3, again indicating the similarity of the average of the income change distributions.

These visual impressions are confirmed by the quantile regressions in Table 3a. The coefficients on *p500f1* show that the 25th percentile is flat, and the 50th and 75th percentiles are in fact gently rising. This picture is striking – there is no sign of any negative relationship between neighbourhood disadvantage and household income growth. By contrast, the results show a set of distributions across areas with essentially the same median and lower quartile, but an increasing upper quartile as neighbourhood disadvantage worsens. This is also true for

---

[13] Lines at the 5th and 95th percentiles of the income growth distribution are flat also, though the latter is very noisy.

the variable characterising the household's wider environment, *p10kf1* – the coefficients are a little smaller and less well determined, but follow the same pattern. To gauge the quantitative significance of these, a one standard deviation change in the neighbourhood disadvantage score implies the one-year percentage change in income is 0.2 higher. This can be compared to a mean change of 8.5 and a standard deviation of 60.

It is clear that looking at changes in terms of amounts (£) rather than percentage changes would produce a different picture. Given the symmetric change apparent in Figure 2, we would see both larger increases and larger decreases in amount changes in richer areas. This would therefore yield a negative (resp. positive) slope at the upper (lower) quartile, and little difference at the median. We focus on percentage changes partly because one can infer from that what absolute changes would be like (but not *vice versa* so easily), and partly because much economic modelling of income uses log income and change log income is approximately equivalent to a percentage change in income.

We also consider the residuals after conditioning on the same set of variables as above. These might be more important for income changes. For example, it may be that young people disproportionately live in poorer areas, and also tend to experience faster income growth. This positive effect may offset negative area influences, thus masking the true impact of locale. In fact the bottom panel of Figure 2 and the results in Table 3a show that conditioning has little effect on the outcome[14].

Running the regressions for different sub-groups, reported in Table 3b, we see that there is no relationship between neighbourhood and one year income growth for the over 60s and for single parents[15]. For couples, the relationship is generally positive and spreading the distribution out as neighbourhood worsens. By housing tenure groups, there is similarly a strong positive effect for home owners, and no effect for renters.

---

[14] It would clearly be of relevance to ask how long individuals have lived in their starting neighbourhoods. This raises two problems, however. First, the data on this in the BHPS is rather noisy, and not always consistent between years. Second, since elapsed time in the present area is clearly an endogenous variable, modelling this takes us away from our aim of imposing minimal structure.
[15] These characteristics are defined for date *t*.

(c) Medium run changes

We now consider five-year and ten-year changes in income. While one-year growth rates may seem too short run to capture any neighbourhood influences, these time spans cover significant periods of people's lives. Sample sizes are now smaller as each individual can now appear at most twice in the five-year analysis, and once in the ten-year (see above for details). Figure 3 repeats our standard graph for the quartiles of five-year percentage income changes (*t* to *t+5*) against neighbourhood disadvantage *t*. The smaller sample size per centile results in the choppy nature of the graph. However, the same result is apparent: the distribution of income change is about the same at all levels of neighbourhood disadvantage. The lower quartile is approximately –20 to –10%, the median is approximately 0% to 15%, and the upper quartile is around 30 to 50%. Again this is true at different spatial scales.

The quantile regressions in Table 4a confirm the visual impression that the distribution of five-year changes is shifting up and increasing in variance as the area becomes poorer. The slope of the area index is positive and statistically significant at the 5% level for all three quartiles, and the upper quartile has a larger slope than the other quartiles, indicative of a fanning out of the distribution of income changes in poorer areas. As before, the results using the broader definition give similar but slightly smaller coefficients. The quantitative significance of the estimates is also in line with the short-run changes. The effect at the median of 1.8 and at the upper quartile of 4.1 from a one standard deviation in neighbourhood disadvantage can be contrasted with a distribution of 5 year changes with mean 20.7 and standard deviation of 72.3.

Controlling for the same fixed characteristics gives the same pattern as the unconditional five-year change data. These are shown in the bottom panel of Figure 3 and in Table 4a and confirm the unconditional data. They show that the distribution of medium-run income increases slightly with area disadvantage at all quartiles, controlling for age, gender and individual human capital.

The smaller sample size limits what inference can be made from analyses of the distribution of five-year changes by area type for each of the different household composition types and tenure types shown in Table 4b, but we again find differences between the older group and single parents on the one hand (no relationship), and couples on the other (positive relationship).

We now utilise the full longitudinal capability of the BHPS and look at income change over the entire sample window. Since we did not want to restrict the sample to those who were always present, we do not take a 10-year difference for some, a 9-year difference for others etc.. We estimate trend income growth for each individual separately from a regression of income against time individual-by-individual. We plot the distribution of this coefficient across neighbourhoods, so there is one observation per person (regardless of the significance of the coefficient). While different in one regard, the pattern is similar to the five-year changes pattern. Figure 4 and the quantile regressions in Table 5 show that the lower quartile and the median are increasing as area worsens (as for five-year changes), but now the upper quartile decreases. One result of this is that the variance of the ten year changes is much lower for those starting in poorer areas. As before, the results using the broader area definition, and conditioning on our set of fixed characteristics, are largely unchanged. The size of the coefficients also are small – a one standard deviation change in the neighbourhood factor is associated with a change in the trajectory of 21 at the median[16], compared to a mean trajectory of 281 and a standard deviation of 1261.8.

(d) A specific illustration

We can illustrate these national results by focussing on three cuts through the distribution of neighbourhood types. We take observations in the following ranges: $5^{th} - 15^{th}$ percentiles of *p500f1*, $50^{th} - 60^{th}$ percentiles and the $80^{th} - 90^{th}$ percentiles. For each range we estimate kernel densities of income level, and the one-year and five-year income change. These are presented in Figure 5. They illustrate very clearly that while income *level* is very strongly related to neighbourhood type, income *growth* is not at all related.

(e) The role of lagged income

The results so far have shown that conditioning on time-invariant individual characteristics (human capital) makes little difference to the estimated influence of neighbourhood. In this section we consider the impact of conditioning on lagged household income. This immediately takes us away from our aim of imposing few parametric assumptions since estimating the coefficient on a lagged dependent variable is problematic in this context. Disentangling

---

[16] The units here are the same as for the level of income, namely deflated, equivalised pounds.

dynamic adjustment from individual effects is well known to be difficult (Nickell, 1981). Measurement error also becomes more important.

We include the lagged level of income in the initial conditioning equation for the income growth rates. We take the residuals from these as before and repeat the above procedures on these series. The results are in Table 6 with the earlier results repeated for convenience. Comparing the second and third rows of each of the three sets of results shows the impact of including the lagged income level. Two points are clear. First, in each case the coefficient on the neighbourhood variable at the median changes from positive to negative, and significantly so. Second, the coefficients remain relatively small: a one standard deviation change in neighbourhood disadvantage is associated with a 1.2 point change in one-year growth (mean = 8, standard deviation = 60), a 2.1 point change in five-year growth (21, 73) and a 12.8 change in trajectory (281, 1262). We return to interpret this below.

To make a closer connection to the literature on low income exits, we re-run the quantile regressions on a low income sample[17]. The results are in the final row of each set in Table 6. Perhaps unsurprisingly there are differences at the upper quantile, and some effect over the five year growth horizon.

(f) Earnings

One key component of household income is earnings. We report in Table 7 the results of repeating our earlier analyses on individual earnings. This is not meant as a neighbourhoods-based analysis of earnings as that would clearly require taking account of local labour markets. But it complements the tables for household income as earnings may be seen as a prime channel through which neighbourhood effects operate. We use individual earnings as the dependent variable and do not correct for within-household earnings correlation. We retain zero earnings observations as zero, not missing, since changes between positive and zero earnings reflect real transitions[18]. We see a strong correlation between neighbourhood disadvantage and the level of earnings at the median and also the upper quartile. This is as expected. Turning to the earnings changes, we see negative coefficients at the median everywhere, significant three out of six times, but always small. The final rows of the table also

---

[17] Specifically, income in the base year is below the standard poverty line – 60% of household median income.
[18] Note that 37.3% of the 104973 observations have zero earnings and 21.2% of the 7280 trajectories are zero.

show the compression of earnings change in the poorer neighbourhoods seen in household income above. Thus earnings changes are giving a hint of a negative association with neighbourhood disadvantage.

(g) Neighbourhood definitions

We have used both local and broad definitions of neighbourhoods above, and seen that both are related to outcomes, but the former more strongly. It is also useful to see whether there is any impact of a household's broader neighbourhood, conditional on the immediate locale.

Table 8 presents results for estimating together the local (*p500f1*) and wider (*p10kf1*) area characteristics. However, since they are highly correlated we re-parameterise this as *p500f1* and (*p10kf1 – p500f1*) to reduce multicollinearity. The results show two things. First, the coefficients on *p500f1* are barely changed from the earlier tables. Second, the additional impact of the wider area is essentially zero for income growth, while positive for the level of income.

We present some results on the interactions of area effects at different spatial scales. Taking individuals in neighbourhoods of 500 people at a particular disadvantage level, we look at the effect of different disadvantage levels in the surrounding wider neighbourhood of 10000 people (*p10kf1*). The range of encompassing neighbourhoods around very poor inner neighbourhoods is rather restricted, but not around neighbourhoods in the third quartile – see Figure 6, which shows a reasonable spread around each.

Starting with income levels, Figure 7 graphs the quartiles of household income across centiles of *p10kf1*, for sub-sets defined by *p500f1* being in a particular narrow range; table 9 presents the accompanying quantile regression. The level of disadvantage of a household's wider area is generally negatively associated with household income, conditional on being in a particular type of immediate local area. The median coefficients were significant for households in bands from the 80th centile. These changes are not trivial, and possibly reflect the spatial correlation of small areas nested within larger ones. Looking at these nested area effects for one-year and five-year income changes, we find similar results to the previous sub-sections. An individual's wider area has no adverse effect on the distribution of her income changes at either the one year or five year horizon. This can be seen from the insignificance of the slope of *p10kf1* in the

income change quantile regressions shown in Table 9, and the graphs of Figures 8 and 9 for one-year and five-year income changes respectively.

(h) (Mis-)Measuring neighbourhood influence?

Defining a measure of potential neighbourhood influence is not straightforward either conceptually or practically. Our measure is unlikely to perfectly characterise the essence of living in a poor neighbourhood, and so we consider the importance of measurement error, first in a static context and then in the dynamics. In fact, our use of principal components derived from 18 different Census variables means that we will capture the broad thrust of the data. The first principal component, used above, explains 31% of the variation. Since the regressions we report are bivariate ones, a simple correction can be applied for any potential degree of measurement error. Given an estimated slope coefficient on the neighbourhood characteristic of *a*, the true parameter $\boldsymbol{a}$ can be recovered as $\boldsymbol{a} = a.r$, where *r* is the 'reliability' factor given by $r = \left(1 + \boldsymbol{s}_u^2 \big/ \boldsymbol{s}_Z^2\right)^{-1}$, where $\boldsymbol{s}_u^2$ is the variance of the measurement error and $\boldsymbol{s}_Z^2$ is the variance of the true neighbourhood characteristic[19]. Picking a value for *r* allows one to calculate the degree of attenuation of our estimates. The central point is that our estimated effects are so low in absolute quantitative terms, that even doubling them does not produce an economically large effect.

Turning to the dynamic effects, over the decade covered by our sample, neighbourhoods will have been changing in an unmeasured way as we can only characterise them once, at the Census date of 1991. There are two factors that reduce the impact of this problem. First, the use of principal components minimises the problem since it averages out individual measures to produce an overall characterisation. If we were able to repeat this annually, it would vary less than any one individual measure. Second, we know from other research that areas in Britain do not vary much in relative socio-economic terms over quite long horizons (see Dorling et al, 2000, for an extreme example of this). Nonetheless, we would expect some attenuation of the estimated effect over the period as our neighbourhood measure becomes more out-of-date. To check the scale of this, we re-ran the 5-year income change quantile regressions separately for the two 5-year tranches, 1991 – 1995 and 1996 – 2000. The state of the neighbourhood is

---

[19] Strictly this is valid for OLS rather than quantile regression. But since we are only guessing at values of *r* this is unlikely to add much further inaccuracy.

measured correctly for the first tranche and not for the second. The results for the three quartiles for the first tranche are {2.17 (25th percentile), 2.06 (50th percentile), 3.68 (75th percentile)} and for the second {2.52, 0.80, 4.20}; these can be compared to the whole period estimates of {2.21, 1.83 and 4.10} from Table 4a. Thus while the attenuation is apparent, particularly at the median, estimating for the period when the neighbourhood attributes are best measured only produces a slightly higher number, and still a positive one.

A second sense in which the dynamics are mis-measured is that households will have been exposed to neighbourhood influences for varying periods of time. This clearly might matter: an individual located with particular peer groups and role models for a year may be less likely to be affected than someone located there for longer. But modelling the joint income and neighbourhood mobility processes would require a set of structural assumptions that takes us away from the approach taken in this paper. However, one simple way to gauge the likely impact is the following: using the second five-year tranche, instead of taking the neighbourhood measure for the individual's location at 1995 as the independent variable, we use the average value over all locations s/he lived in 1991 – 1995. This will induce bias, since the mobility process is likely to be correlated with the income dynamics. In fact, the results show very little difference: {1.65 (25th percentile), 0.45 (50th percentile), 4.15 (75th percentile)} compared to the numbers in the preceding paragraph for the second tranche.

(i) Summary and Interpretation

The data suggest the following:
- A strong correlation of neighbourhood disadvantage and household income,
- Very little correlation at the median of neighbourhood disadvantage with growth in household income,
- At the upper quartile of the income growth distribution, one- and five-year growth appears to be higher in poorer neighbourhoods, and also (five-year growth) at the lower quartile,
- Controlling for human capital makes little difference to these conclusions,
- The effects found in income growth – positive changes – are quantitatively small,
- In each case, the influence of the wider neighbourhood (nearest 10,000 people) is absolutely smaller than that of the closer neighbourhood (nearest 500 people),

- Splitting the sample by demographic structure or residential tenure, we find no effect of neighbourhood on income growth for those most likely to be on fixed incomes (the over 60s and single parents), and small positive effects for couples and for home owners.

- Investigating the income adjustment process and controlling for the lagged level of income does produce a negative coefficient for income growth.

The implications of the modelling framework for understanding these results are set out in Table 10. If income model 1 is generating the data, then the results capture the true neighbourhood effect since looking at income changes differences out the assumed additive heterogeneity, and it is zero. The combination of income model 2 and selection model 1 suggest that the neighbourhood effect is positive. If income model 2 and selection model 2 are generating the data, then there is scope for the existence of a detrimental neighbourhood effect. Such an effect plus any growth selection effect must be offset by the rebound effect generated by temporary income shocks impacting on location.

The results including the lagged income level suggest that there may be some basis for this, and we now consider these different results. Having set out the income and selection models above in general terms, we strip the model down to link it more transparently to these results. Suppressing subscripts, we can write a simplified model for income as:

$$y = \boldsymbol{b}X - \boldsymbol{a}.Z + (1 - \boldsymbol{r})y_{-1} + u \qquad\qquad (12)$$

where all coefficients are positive, u is the error term, and $\boldsymbol{a}$ is the coefficient of interest capturing the causal plus growth selection effects. Note that we can write (12) in error-correction form, $\Delta y = -\boldsymbol{r}(y_{-1} - \overline{y})$, where $\overline{y} = (\boldsymbol{b}X - \boldsymbol{a}Z)/\boldsymbol{r}$ is the long run mean of y. The role of $\boldsymbol{r}$ as the rebound parameter is then clear: a shock taking y_{-1} below its mean is followed in expectation by a rise in income. Add to this a simple selection model, $Z = -(1/\boldsymbol{g})y_{-1}$, or inverting gives $y_{-1} = -\boldsymbol{g}.Z$.

We can interpret the three parameters as: $\boldsymbol{r}$ – the 'rebound' or error-correction parameter, $\boldsymbol{g}$ – the sorting parameter, and $\boldsymbol{a}$ – the causal plus growth selection effect. Bringing them together and substituting out for the lagged income level,

$$\Delta y = \boldsymbol{b}X + (\boldsymbol{r}\boldsymbol{g} - \boldsymbol{a})Z + u \qquad\qquad (13)$$

Thus we can see that in a model without lagged income the coefficient on Z is a combination of rebound and selection on levels, and any causal plus growth selection effect. This reconciles

the positive effect found in the bulk of the results above (at the median) with the negative coefficient once the lagged income level is controlled for.

We should be cautious in taking the estimates of (12) at face value as it is known to be difficult to correctly estimate lagged adjustment parameters in cross-section (and short panels): issues such as the nature of the unobserved heterogeneity, heterogeneity in the error-correction coefficient and income measurement error are important. In particular, individual fixed effects are likely to reduce the role of lagged income. In line with the approach of this paper, we leave these results as measuring the maximum possible influence of neighbourhood.

6. Conclusions

Our aim in this paper has been to provide credible bounds on the size of any neighbourhood influences on economic outcomes, by eschewing questionable exclusion restrictions and making as few parametric assumptions as possible. Our data also allow us to address two questions typically ignored in the 'neighbourhood effects' literature. The first is the appropriate definition of neighbourhood, including the appropriate spatial scale and the role of neighbouring neighbourhoods. We study both a very local definition – around the nearest 500 people – and a much broader one – 10,000 people. We show that both matter but that the former has a stronger association with individual outcomes. We also show that conditional on the very local area, the broader one has little effect. The second issue is income dynamics, using the ten year panel on individual outcomes. Contrary to a causal neighbourhood effects story, we find little association between neighbourhood disadvantage and subsequent income growth; in fact a small but significant positive association is estimated. This is true for one-, five- and ten-year income changes. Investigation of the income adjustment process suggests that this might be due to a 'rebound' effect – that is mean reversion in the income process, plus (re)location in the basis of temporary income shocks. If these are important, then they could mask detrimental neighbourhood influence.

We estimate that the highest value this could take is about one percentage point on one-year income change and two percentage points on five-year change. Any selection process on income growth heterogeneity, or any revision downwards of the size of the rebound effect with more structural estimation of the income process will reduce this. We have allowed for a lot of

heterogeneity of response, and conditioned on simple life cycle and human capital measures. This is our estimated upper bound on the role of neighbourhood.

These results suggest that the segregation of poor individuals into poor neighbourhoods may not have any substantial long-term detrimental effects. The distribution of income growth rates for people living in such areas is indistinguishable from that for people living in rich areas. Of course, this does not necessarily remove an area-basis for policy. The high levels of clustering may mean that the most efficient way of targeting individual policies is on an area basis. Nevertheless, the results support the idea that the main sources of low income are to be found in earnings, employment and demographics, not in neighbourhood characteristics.

References

Bardasi, E., Jenkins, S.P. and Rigg, J. A. (2003) *Documentation for Derived Current and Annual Net Household Income Variables, BHPS Waves 1-10.* http://www.dataarchive.ac.uk/doc/3909%5Cmrdoc%5Cpdf%5C3909userguide.pdf

Blume, L. and Durlauf, S. (2001) *The Interactions-Based Approach to Socioeconomic Behaviour.* In *Social Dynamics*, S. Durlauf and H. P. Young, eds, Cambridge: MIT press.

Brock, W. and Durlauf, S. (2001) Interactions-Based Models, in Handbook of Econometrics, Volume 5, J. Heckman and E. Leamer (eds), Elsevier Science B.V.

Brock, W. and Durlauf, S. (2003) *A Multinomial Choice Model with Social Interactions.* In *The Economy as a Evolving Complex System III*, L. Blume and S. Durlauf, eds., Oxford University Press.

Buck, N. (2001) *Identifying Neighbourhood Effects on Social Exclusion.* Urban Studies, 38, 2251-2275.

Cutler, D. and Glaeser, E. (1997) *Are Ghettos good or bad?* Quarterly Journal of Economics vol. 112 pp. 827 – 872.

Dietz, R. D. (2002) *The Estimation of Neighborhood Effects in the Social Sciences: An Interdisciplinary Approach.* Social Science Research, 31, 539-575.

Dorling, D., Mitchell, R., Shaw, M., Orford, S. and Davey Smith, G. (2000) *The Ghost of Christmas Past: health effects of poverty in London in 1896 and 1991* British Medical Journal, vol. 321, pp. 1547 – 1551.

Dorling, D. and Thomas, B. (2004) People and Places: A 2001 Census atlas of the UK Policy Press, Bristol

Durlauf, S. (2004) *Neighborhood Effects.* Handbook of Regional and Urban Economics, Volume 4, Economics, J.V. Henderson and J.-F. Thisse, eds., forthcoming

Epple, D. and Sieg, H. (1999) *Estimating Equilibrium Models of Local Jurisdictions* Journal of Political Economy vol. 107, pp. 645 – 681.

Goering, J., Feins, J. and T. Richardson (2002) *A Cross-Site Analysis of Initial Moving to Opportunity Demonstration Results.* Journal of Housing Research, 13, 1, 1-30.

Jargowsky, P. A. (1997) *Poverty and Place: Ghettos, Barrios, and the American City.* New York: Russell Sage Foundation.

Jargowsky, P. A. (2003) *Stunning Progress, Hidden Problems: The Dramatic Decline of Concentrated Poverty in the 1990s* Living Cities Census Series, Center on Urban and Metropolitan Studies. Washington, DC: The Brookings Institution

Johnston, R., Jones, K., Burgess S., Propper, C., Sarker, R. and Bolster, A. (2004) *Fractal Factors? Scale, Factor Analysis & Neighbourhood Effects.* Geographical Analysis, 36.

Katz, L., Kling, J. and Liebman, J. (2001) *Moving To Opportunity in Boston: Early Results of a Randomized Mobility Experiment.* Quarterly Journal of Economics, 116, 2, 607-654.

Keels, M., Duncan, G., Deluca, S., Mendenhall R., and Rosenbaum (2004) *15 years later: Can residential mobility programs provide a long-term escape from neighbourhood segregation, crime and poverty?* Mimeo.

Ludwig, J., Duncan, G., J. and Hirschfield, P. (2001) *Urban Poverty and Juvenile Crime: Evidence from a Randomized Housing-Mobility Experiment*. Quarterly Journal of Economics. 116, 2, 655-680.

McCulloch, A. (2001) *Ward-level Deprivation and Individual Social and Economic Outcomes in the British Household Panel Study*. Environment and Planning A, 33, 667-84.

Manski, C., (1993) *Identification of Endogenous Social Effects: The Reflection Problem*. Review of Economic Studies, 60, 3, 531-542.

Manski, C., (2000) *Economics Analysis of Social Interactions*. Journal of Economic Perspectives, 14, 3, 115-136.

Moffitt, R., (2001) *Policy Interventions, Low-level equilibria, and social interactions*. In *Social Dynamics*, S. Durlauf and H. P. Young, eds., Cambridge: MIT Press

Nickell, S., J. (1981) *Biases in Dynamic Models with Fixed Effects*. Econometrica, 48, 879-897.

Oreopoulos, P. (2003) *The long-run consequences of living in a poor neighborhood'* Quarterly Journal of Economics. Vol. CXVIII pp. 1533 – 1575.

Rosebaum, J., (1995) *Changing the Geography of Opportunity by Expanding Residential Choice: Lessons from the Gatreaux Program*. Housing Policy Debate, 6, 1, 231-269.

Taylor, M. F., (ed). with Brice, J., Buck, N. and Prentice-Lane, E. (2003) *British Household Panel Survey User Manual Volume A: Introduction, Technical Report and Appendices*. Colchester: University of Essex.

Tunstall, H., Johnston, R. J., Rossiter, D. J., Pattie, C. J., MacAllister, I. and Dorling, D. F. L. (2000) *Geographical scale, the 'feel-good factor' and voting in the 1997 general election in England and Wales*. Transactions, Institute of British Geographers vol. NS25, pp. 51-64.

Wilson, W. J., (1987) *The Truly disadvantaged*. Chicago: University of Chicago Press.

Wilson, W. J., (1996) *When Work Disappears: The World of the New Urban Poor*. New York: Knopf.

Table 1: Correlation of neighbourhood characteristic (*p500f1*) with BHPS neighbourhood and accommnodation questions

Variable description	Correlation with *p500f1*	
	N*T	N
Worry about being affected by crime	-0.04	-0.12
Extent of concern about crime	-0.17	-0.21
Feel safe walking alone at night	0.18	0.13
Extent of graffiti on walls	-0.41	-0.41
Extent of teenagers hanging about	-0.31	-0.29
Extent of drunks/tramp s on street	-0.29	-0.26
Extent of vandalism	-0.37	-0.37
Extent of racial insults/attacks	-0.24	-0.22
Extent of homes broken into	-0.21	-0.24
Extent of cars stolen/broken into	-0.26	-0.27
Extent of people attacked on street	-0.30	-0.28
Likes present neighbourhood	0.20	0.19
Vandalism or crime	-0.25	-0.27
Noise from neighbours	-0.16	-0.16
Street noise	-0.12	-0.12
Accommodation:		
Pollution/environmental problems	-0.07	-0.06
Shortage of space	-0.08	-0.09
Not enough light	-0.05	-0.06
Lack of adequate heating	-0.09	-0.07
Condensation	-0.09	-0.10
Leaky roof	-0.02	-0.03
Damp walls, floors etc	-0.07	-0.09
Rot in windows, floors	-0.05	-0.06

Note: All the correlations shown have p-values below 0.00; the first column uses all the data, whereas the second column picks out data from the first year each individual was observed. The questions on crime were asked in year 1997 only, the accommodation questions were asked in years 1996 to 2000, and the "Likes present neighbourhood" question was asked in years 1991-2000. Some of the variables are coded on a four point scale, for example the perception of crime levels; others such as satisfaction with accommodation are binary responses.

Table 2a: Quantile regression of income level on neighbourhood characteristic

Dependent variable: household income

Unit of observation: household-year

Scale of neighbourhood characteristic	Sample size	Coefficient on neighbourhood characteristic by quantile			p-value of F test:		
		q25	q50	q75	q25= q50	q50= q75	q25 = q50 = q75
p500f1	50223	-1407.9***	-2292.2***	-3422.0***	0***	0***	0***
p10kf1	50223	-1171.2***	-1955.8***	-2937.4***	0***	0***	0***
Conditional on individual characteristics							
p500f1	49910	-461.27***	-946.08***	-1757.11***	0***	0***	0***

Notes:
1. * significant at 10%; ** significant at 5%; *** significant at 1%
2. Unweighted regression.
3. Conditional regression is for the household head.

Table 2b: Quantile regression of income level on neighbourhood characteristic (*p500f1*) by household characteristics

Dependent variable: household income

Unit of observation: household-year

Household characteristic	Sample size	Coefficient on neighbourhood characteristic by quantile			p-value of F test:		
		q25	q50	q75	q25= q50	q50= q75	q25 = q50 = q75
All	50223	-1407.9***	-2292.2***	-3422.0***	0***	0***	0***
Household composition							
All > 60 yrs of age	13967	-488.4***	-1055.7***	-2116.7***	0***	0***	0***
Single adult, no kids	6127	-2023.0***	-2759.6***	-3473.3***	0***	0***	0***
Single Parent	3203	-485.5***	-1053.3***	-1746.6***	0***	0***	0***
Couple, no kids	8590	-2249.8***	-2913.4***	-3816.7***	0***	0***	0***
Couple, with kids	12383	-1747.3***	-2175.5***	-3058.4***	0***	0***	0***
Multiple Adult household	5798	-1434.5***	-1753.2***	-2660.8***	0***	0***	0***
Tenure							
Owned outright	11971	-864.6***	-1773.7***	-3029.0***	0***	0***	0***
Owned with mortgage	21118	-1360.8***	-1947.0***	-2770.3***	0***	0***	0***
Local Authority rented	9659	-321.0***	-323.5***	-242.2***	0.96	0.23	0.48
Housing Associated rented	2394	-377.0***	-378.3***	-537.2**	0.99	0.46	0.76
Other rented	4849	-864.0***	-1221.0***	-1124.9***	0***	0.67	0.01***
Rural							
Rural	16445	-1329.9***	-2137.0***	-3752.5***	0***	0***	0***
Urban	33778	-1509.4***	-2415.9***	-3381.5***	0***	0***	0***

Notes:
1. * significant at 10%; ** significant at 5%; *** significant at 1%
2. Unweighted regression.
3. Conditional regression is for the household head.

Table 3a: Quantile regression of one-year income change on neighbourhood characteristic

Dependent variable: One-year change in household income
Unit of observation: individual-year

Scale of neighbourhood characteristic	Sample size	Coefficient on neighbourhood characteristic by quantile			p-value of F test		
		q25	q50	q75	q25 = q50	q50 = q75	q25 = q50 = q75
P500f1	69580	-0.05	0.20***	1.23***	0.01***	0***	0***
P10kf1	69580	0.13	0.14*	0.87***	0.88	0***	0***
Conditional on individual characteristics							
P500f1	69282	-0.02	0.32***	1.22***	0***	0***	0***

Notes:
1. * significant at 10%; ** significant at 5%; *** significant at 1%
2. Unweighted regression.
3. Conditional one-year changes were obtained as the residuals from a regression on age12, age12 squared, maxquals, maxquals*age12, gender, and year dummies. Unweighted regression.
4. Income changes were capped at 500% because of a small number of very large outliers.

Table 3b: Quantile regression of one-year income change on neighbourhood characteristic (*p500f1*) by household characteristics

Dependent variable: One-year change in household income
Unit of observation: individual-year

Household characteristic	Sample size	Coefficient on neighbourhood characteristic by quantile			p-value of F test		
		q25	q50	q75	q25 = q50	q50 = q75	q25 = q50 = q75
All	69580	-0.05	0.20***	1.23***	0.01***	0***	0***
Household composition (FAMSTAT)							
All over 60 yrs of age	14695	0.51*	0.18	0.41	0.17	0.40	0.32
Single adult, no kids	4455	-1.16*	-0.37*	-0.04	0.16	0.58	0.26
Single Parent	2912	-0.72	-0.3	2.03*	0.59	0.01***	0.05**
Couple, no kids	13417	0.1	0.12	1.15***	0.94	0***	0***
Couple, with kids	21850	-0.56**	0.36**	1.64***	0***	0***	0***
Multiple Adult household	11978	0.64**	0.37**	0.21	0.33	0.57	0.55
Tenure							
Owned outright	15962	0.87***	-0.03	-1.39***	0***	0***	0***
Owned with mortgage	34066	0.78***	0.40***	0.59***	0***	0.23	0***
Local Authority rented	11183	-0.73**	-0.11	0.97*	0.02**	0.02**	0***
Housing Associated rented	2688	-0.71	0.47	1.91	0.13	0.16	0.10*
Other rented	5533	-0.72	0.55	4.12***	0***	0***	0***
Rural							
Rural	23416	-0.11	0.12	0.76***	0.37	0.02**	0.04**
Urban	46164	-0.19	0.19**	1.32***	0***	0***	0***

Notes:
1. * significant at 10%; ** significant at 5%; *** significant at 1%
2. Unweighted regression.
3. Income changes were capped at 500% because of a small number of very large outliers.

Table 4a: Quantile regression of five-year income change on neighbourhood characteristic

Dependent variable: Five-year change in household income
Unit of observation: individual-year

Scale of neighbourhood characteristic	Sample size	Coefficient on neighbourhood characteristic by quantile			p-value of F test		
		q25	q50	q75	q25 = q50	q50 = q75	q25 = q50 = q75
P500f1	10936	2.21***	1.83***	4.10***	0.34	0***	0***
P10kf1	10936	1.64***	1.40***	2.39***	0.62	0.21	0.31
Conditional on individual characteristics							
P500f1	10900	2.36***	2.11***	4.39***	0.49	0***	0***

Notes:

1. * significant at 10%; ** significant at 5%; *** significant at 1%

2. Unweighted regression.

3. Conditional five-year changes were obtained as the residuals from a regression on age12, age12 squared, maxquals, maxquals*age12, gender, and year dummies. Unweighted regression.

4. Income changes were capped at 500% because of a small number of very large outliers.

Table 4b: Quantile regression of five-year income change on neighbourhood characteristic (p500f1) by household characteristics

Dependent variable: Five-year change in household income
Unit of observation: individual-year

Household characteristic	Sample size	Coefficient on neighbourhood characteristic by quantile			p-value of F test		
		q25	q50	q75	q25 = q50	q50 = q75	q25 = q50 = q75
All	10936	2.21***	1.83***	4.10***	0.34	0***	0***
Household composition (FAMSTAT)							
All over 60 yrs of age	2000	0.44	-0.02	-0.21	0.62	0.91	0.86
Single adult, no kids	673	2.76	-0.09	1.62	0.21	0.60	0.36
Single Parent	421	-0.06	1.61	-1.72	0.63	0.57	0.73
Couple, no kids	2266	2.30***	0.45	2.09	0.06*	0.20	0.08*
Couple, with kids	3633	3.02***	3.45***	4.47***	0.60	0.34	0.58
Multiple Adult household	1913	2.72*	2.08**	3.82**	0.65	0.20	0.39
Tenure							
Owned outright	2424	1.05	-1.41*	-1.72	0***	0.83	0.01
Owned with mortgage	5655	2.69***	2.07***	4.94***	0.36	0***	0***
Local Authority rented	1670	-0.77	0.91	3.55*	0.07*	0.14	0.09*
Housing Associated rented	306	5.41**	2.29	9.13*	0.15	0.13	0.08*
Other rented	877	4.20*	7.14*	16.5	0.43	0.27	0.46
Rural							
Rural	3607	3.83***	1.32	2.43*	0***	0.36	0***
Urban	7329	1.58***	1.96***	4.60***	0.42	0***	0***

Notes:

1. * significant at 10%; ** significant at 5%; *** significant at 1%

2. Unweighted regression.

3. Income changes were capped at 500% because of a small number of very large outliers.

Table 5: Quantile regression of income trajectory (slope coefficient) on neighbourhood characteristic

Dependent variable: Whole window trajectory in household income
Unit of observation: individual

Scale of neighbourhood characteristic	Sample size	Coefficient on neighbourhood characteristic by quantile			p-value of F test		
		q25	q50	q75	q25 = q50	q50 = q75	q25 = q50 = q75
p500f1	6571	77.6***	21.36**	-61.57***	0***	0***	0***
p10kf1	6571	67.54***	6.09	-47.63***	0***	0***	0***
Conditional on individual characteristics							
p500f1	6548	92.36***	24.05**	-29.86**	0***	0***	0***
Starting age							
15-25 years	1293	50.55*	-16.89	-68.96	0.01***	0.17	0.01***
26-59 years	3980	95.04***	25.42**	-59.06***	0***	0***	0***
60+ years	1298	44.92***	8.13	-4.36	0.01***	0.62	0.03**

Notes:

1. * significant at 10%; ** significant at 5%; *** significant at 1%

2. Unweighted regression.

3. Starting age refers to age of individual on 1st December of year first observed.

4. Conditional trajectory data were obtained as the residuals from a regression on age12, age12 squared, maxquals, maxquals*age12, gender, and year dummies. Unweighted regression.

Table 6: Role of lagged income and neighbourhood characteristics

	Sample	Conditional on:		Coefficient on *p500f1* by quantile		
				q25	q50	q75
		Individual characteristics	Lagged Income			
One-year income change						
p500f1	All (69580)			-0.05	0.20***	1.23***
p500f1	All (69282)	Yes		-0.02	0.32***	1.22***
p500f1	All (69282)	Yes	Yes	-1.60***	-1.16***	-1.14***
p500f1	Low income (9660)	Yes	Yes	-1.12**	-1.77***	-5.22***
Five-year income change						
p500f1	All (10936)			2.21***	1.83***	4.10***
p500f1	All (10900)	Yes		2.36***	2.11***	4.39***
p500f1	All(10900)	Yes	Yes	-1.33***	-2.05***	-1.31*
p500f1	Low income (1403)	Yes	Yes	-2.88	-3.01	-8.56***
Income Trajectory						
p500f1	All (6571)			77.6***	21.36**	-61.57***
p500f1	All (6548)	Yes		92.36***	24.05**	-29.86**
p500f1	All (6197)	Yes	Yes	47.91***	-12.83***	-83.00***
p500f1	Low income	Yes	Yes			

Notes:

1. * significant at 10%; ** significant at 5%; *** significant at 1%
2. Unweighted regression.
3. Conditional one-year changes were obtained as the residuals from a regression on age12, age12 squared, maxquals, maxquals*age12, gender, and year dummies. Unweighted regression.
4. Income changes were capped at 500% because of a small number of very large outliers.
5. Low income sample contain individuals with NETINC less than 60% of UK household median income.

Table 7: Quantile regression of earnings on neighbourhood characteristic

	Sample size	Coefficient on neighbourhood characteristic by quantile			p-value for F test:		
		q25	q50	q75	q25 = q50	q50 = q75	q25 = q50 = q75
Level							
Unconditional	104973	0.00	-1656.92***	-2764.30***	0***	0***	0***
Conditional	104396	-20.04	-534.90***	-1254.02***	0***	0***	0***
One-year change							
Unconditional	51644	-1.31***	-0.24***	0.68	0***	0.94	0.01***
Conditional	51408	-3.03***	-1.36***	0.34*	0***	0***	0***
Five -year change							
Unconditional	8898	-2.91	-0.30	5.51***	0.16	0***	0***
Conditional	8864	-3.03**	-2.50*	2.78***	0.64	0***	0***
Income trajectory							
Unconditional	7280	69.19***	-	-81.03***	0***	0***	0***
Conditional	7246	58.54***	-14.34***	-46.63***	0***	0***	0***

Notes:

1. * significant at 10%; ** significant at 5%; *** significant at 1%

2. Unweighted regression.

3. A dashed sign in place of the coefficient implies the coefficient could not be estimated.

Table 8: Quantile regression: contribution of neighbourhood characteristic at nearest 500 people (p500f1) and nearest 10,000 people scale (p10kf1).

Scale of neighbourhood characteristic	Sample size	Coefficient on neighbourhood characteristic by quantile		
		q25	q50	q75
Income level				
p500f1	50223	-1475.9***	-2392.0***	-3631.8***
p10kf1-p500f1		-188.7***	-421.9***	-795.4***
One-year income change				
p500f1	69580	0.03	0.20**	1.25***
p10kf1-p500f1		0.33*	0.02	0.09
Five -year income change				
p500f1	10936	2.35**	1.88**	4.03***
p10kf1-p500f1		0.16	0.36	-0.63
Income trajectory				
p500f1	6571	84.48***	17.71*	-65.42***
p10kf1-p500f1		27.60*	-15.73	-13.12

Notes:
4. * significant at 10%; ** significant at 5%; *** significant at 1%
5. Unweighted regression.

Table 9: Quantile regression on wider area characteristics (*p10kf1*) conditional on area type of local neighbourhood

Income level

	Centile of *p500f1* equals:				
	70-74	**75-79**	**80-84**	**90-94**	**95-100**
P10kf1	-174.19	485.04*	-575.92***	-856.30***	-550.07***
Constant	13169.00***	12460.82***	12638.93***	12443.12***	11538.39***
Observations	2503	2520	2519	2571	3027

One-year income change

	70-74	**75-79**	**80-84**	**90-94**	**95-100**
P10kf1	-1.08***	2.27***	0.47	-0.54	0.43
Constant	2.22***	1.20***	2.28***	2.35***	1.79***
Observations	3268	3198	3312	3262	3765

Five -year income change

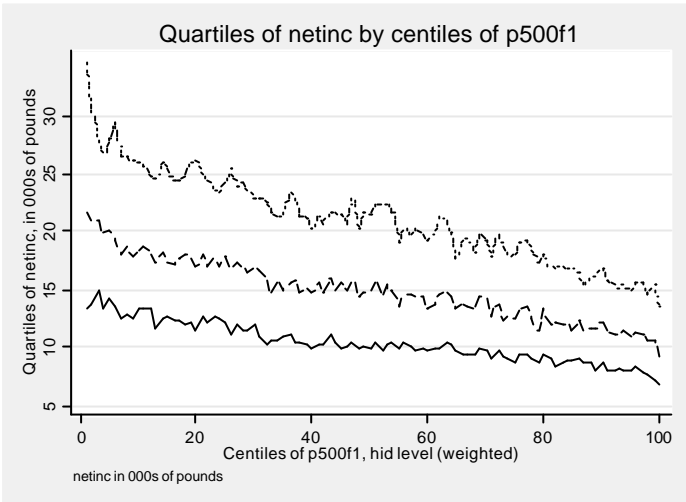|	**70-79**	**80-89**	**90-100**
P10kf1	3.29	-0.06	2.35*
Constant	7.46***	7.12***	7.12***
Observations	877	890	998

Notes:

1. Absolute value of t statistics in brackets.

2. * significant at 10%; ** significant at 5%; *** significant at 1%.

3. *p10kf1* increases as area becomes more deprived.

Table 10: Interpreting the Results

		Income model	
		(1)	(2)
Selection model	(1)	Looking at income growth removes the fixed effect so estimation reveals the true causal neighbourhood effect to be zero. The correlation in levels arises purely from selection.	Estimated impact is the standard combination of causal neighbourhood effect and selection effect. If there is a selection effect (as levels data suggest) then the neighbourhood effect is <u>positive</u>.
	(2)	Looking at income growth removes the fixed effect so estimation reveals the true causal neighbourhood effect to be zero. The correlation in levels arises purely from selection.	This is the only box where our data are consistent with a detrimental neighbourhood effect. A small selection effect in levels plus the neighbourhood effect are offset by a large 'rebound' effect. There may of course be less of a correlation between 'growth' heterogeneity and location as between 'levels' heterogeneity and location.

Figure 1: Income levels and neighbourhood

(a) Narrow neighbourhood of 500 people

Quartiles of netinc by centiles of p500f1

netinc in 000s of pounds

(b) Broader neighbourhood of 10,000 people

Quartiles of netinc by centiles of p10kf1

netinc in 000s of pounds
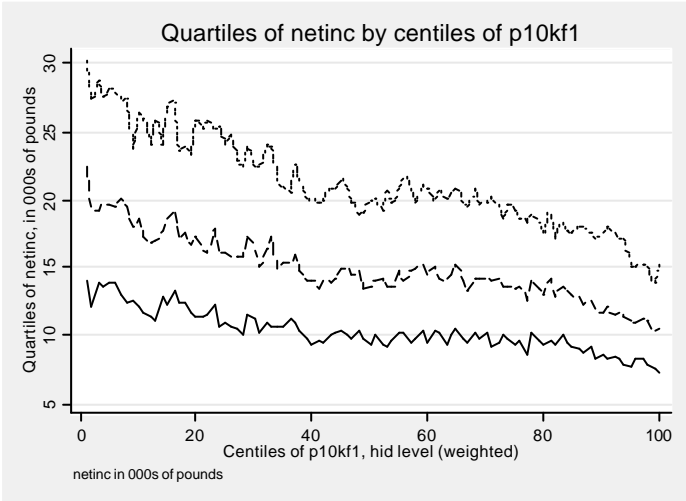
Figure 2: One year income change and neighbourhood
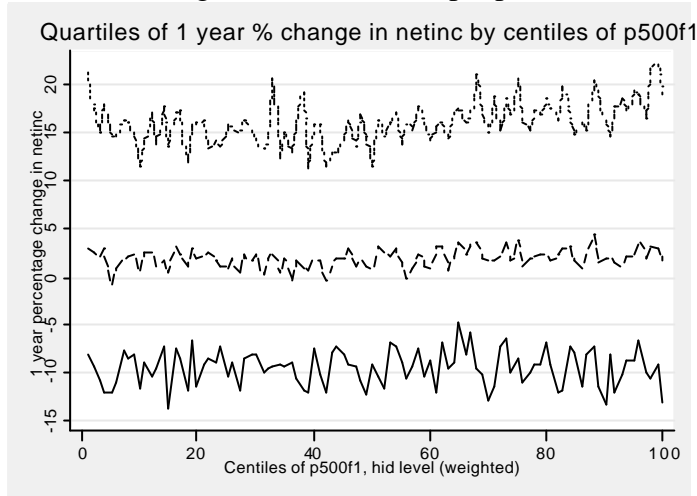
(a) Narrow neighbourhood of 500 people

Quartiles of 1 year % change in netinc by centiles of p500f1

(b) Broad neighbourhood of 10,000 people

Quartiles of 1 year % change in netinc by centiles of p10kf1
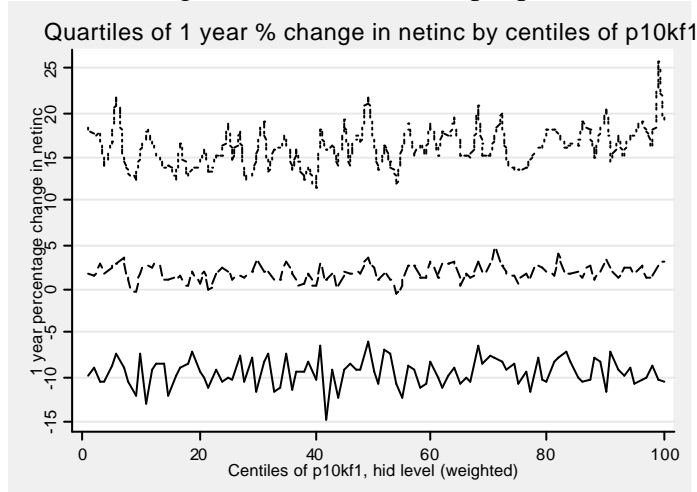
(c) Income change residuals, narrow neighbourhood of 500 people

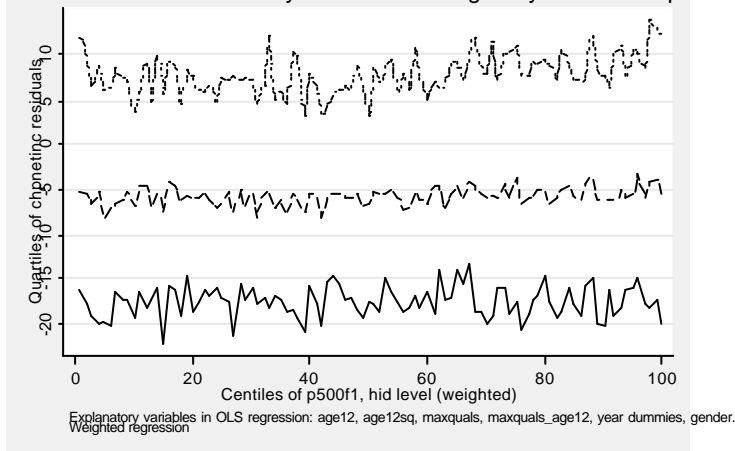Quartiles of residuals of 1 year netinc changes by centiles of p500f1

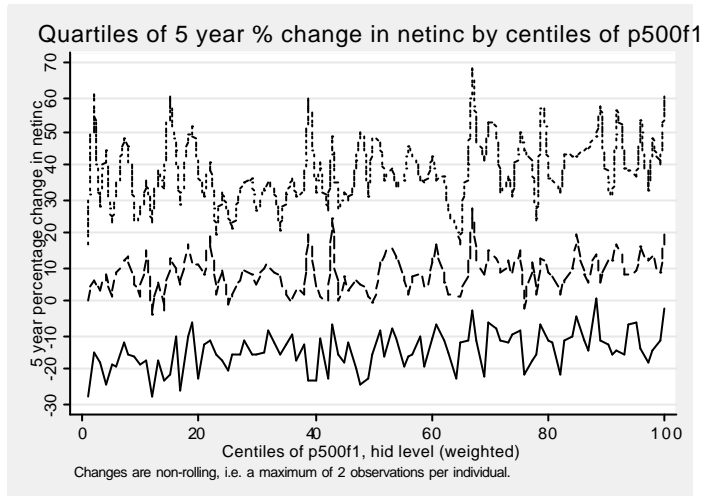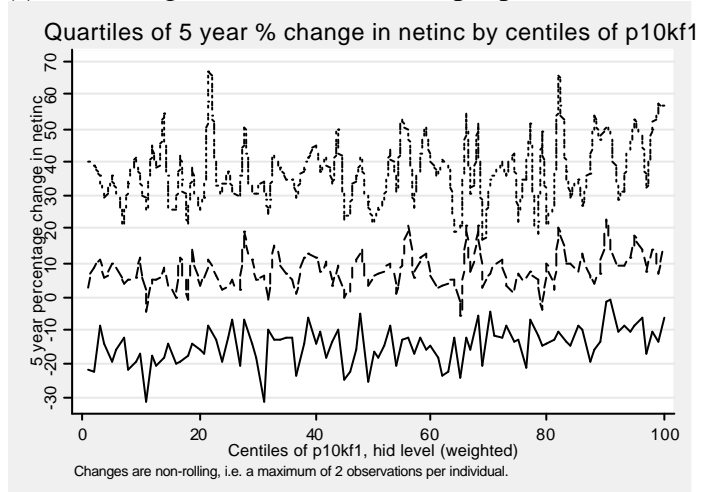Explanatory variables in OLS regression: age12, age12sq, maxquals, maxquals_age12, year dummies, gender.
Weighted regression

Figure 3: Five year income change and neighbourhood

(a) Narrow neighbourhood of 500 people

Quartiles of 5 year % change in netinc by centiles of p500f1

Changes are non-rolling, i.e. a maximum of 2 observations per individual.

(b) Broad neighbourhood of 10,000 people

Quartiles of 5 year % change in netinc by centiles of p10kf1

Changes are non-rolling, i.e. a maximum of 2 observations per individual.

(c) Income change residuals, narrow neighbourhood of 500 people

Quartiles of residuals of 5 year netinc changes by centiles of p500f1

Explanatory variables in OLS regression: age12, age12sq, maxquals, maxquals_age12, year dummies, gender.
Weighted regression
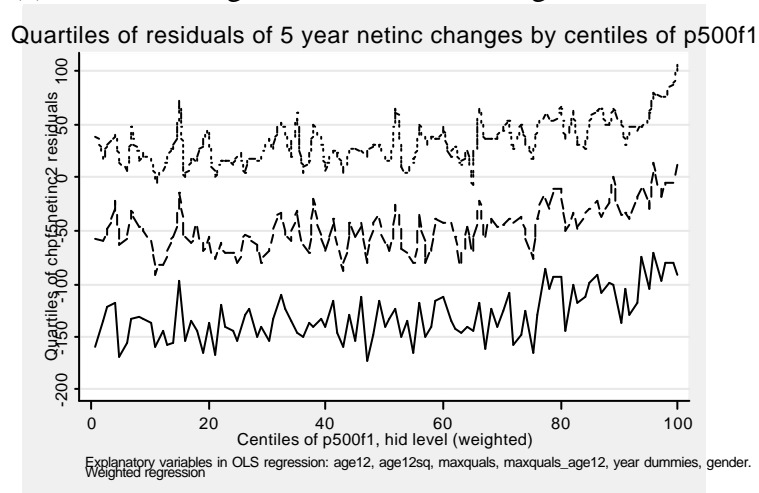
Figure 4: Income Trajectory and neighbourhood

(a) Narrow neighbourhood of 500 people

Quartiles of individual netinc trajectory by starting p500f1 centile

Sample consists of individuals with more than 7 observations only.
7843 observations

(b) Broad neighbourhood of 10,000 people

Quartiles of individual netinc trajectory by starting p10kf1 centile

Sample consists of individuals with more than 7 observations only.
7843 observations
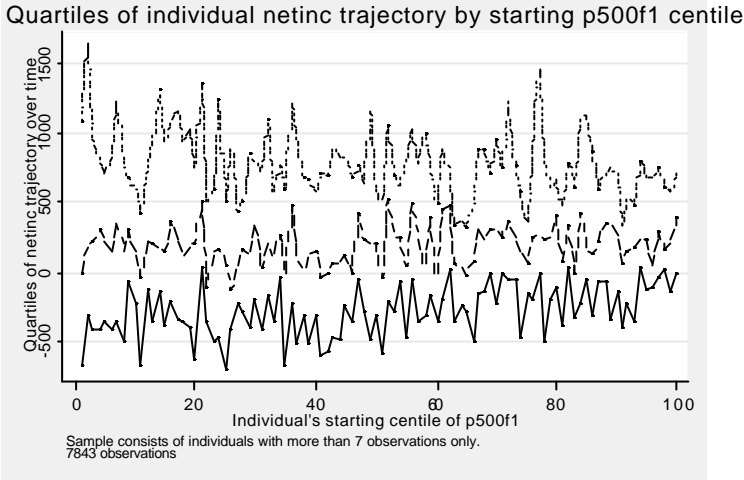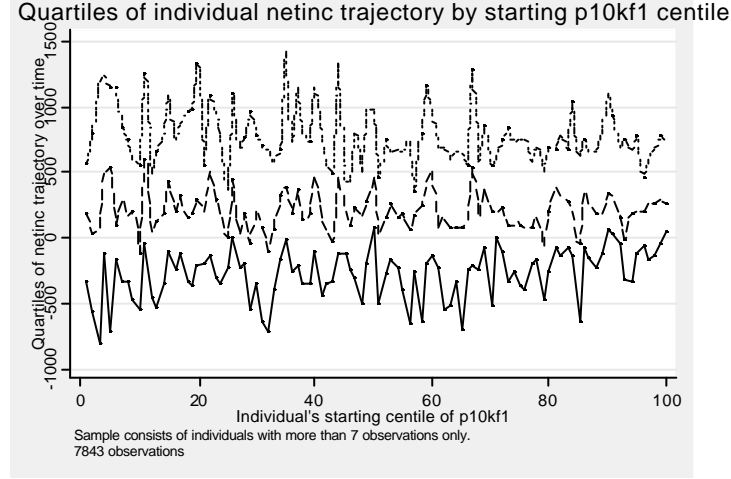
(c) Income trajectory residuals, narrow neighbourhood of 500 people

Quartiles of netinc trajectory residuals by centiles of p500f1

Explanatory variables in OLS regression: age12, age12sq, maxquals, maxquals_age12, year dummies, gender
Weighted regression

44
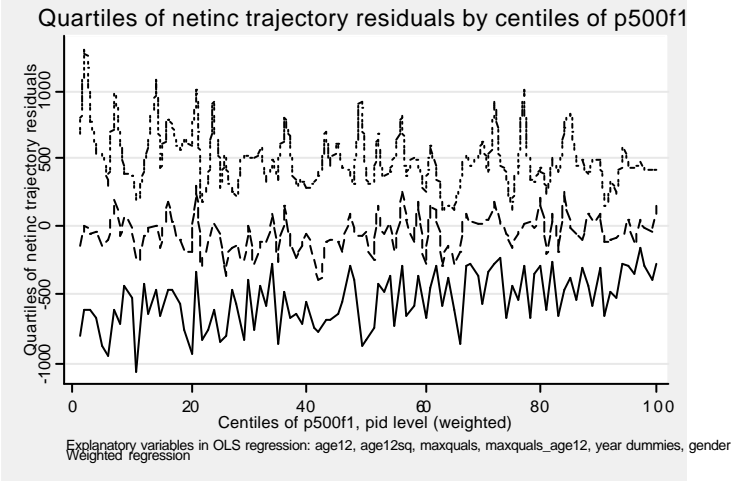
Figure 5: Three cuts through the neighbourhood distribution

(a) Income level

Kernel density esimates of Household income by percentiles of p500f1

(b) One-year income change

Kernel density esimates of 1 year change in Household income by percentiles of p500f1

(c) Five-year income change

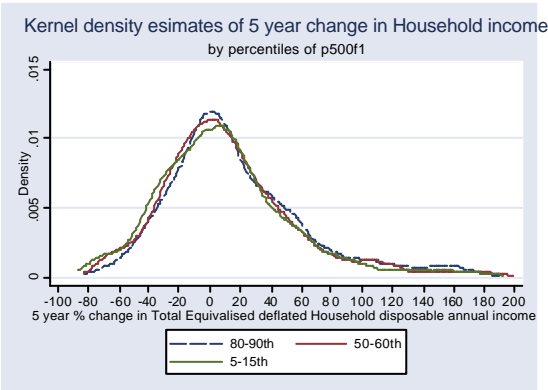Kernel density esimates of 5 year change in Household income by percentiles of p500f1
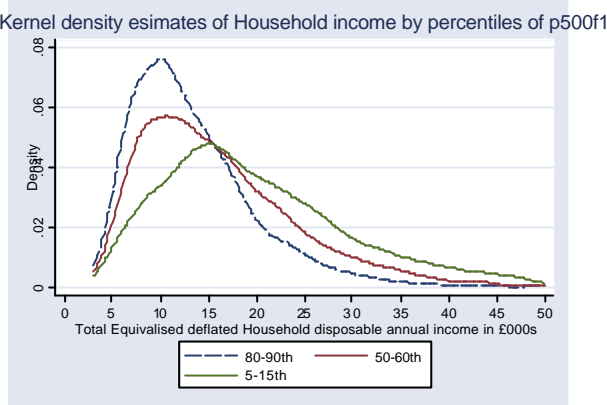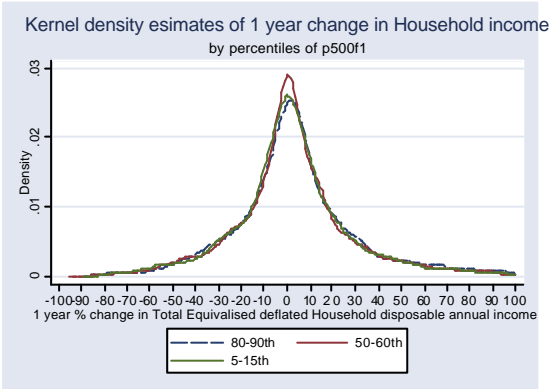
Figure 6: Distribution of centiles of wider area (*p10kf1*) neighbourhood characteristic conditional on nested smaller area (*p500f1*) neighbourhood characteristic.
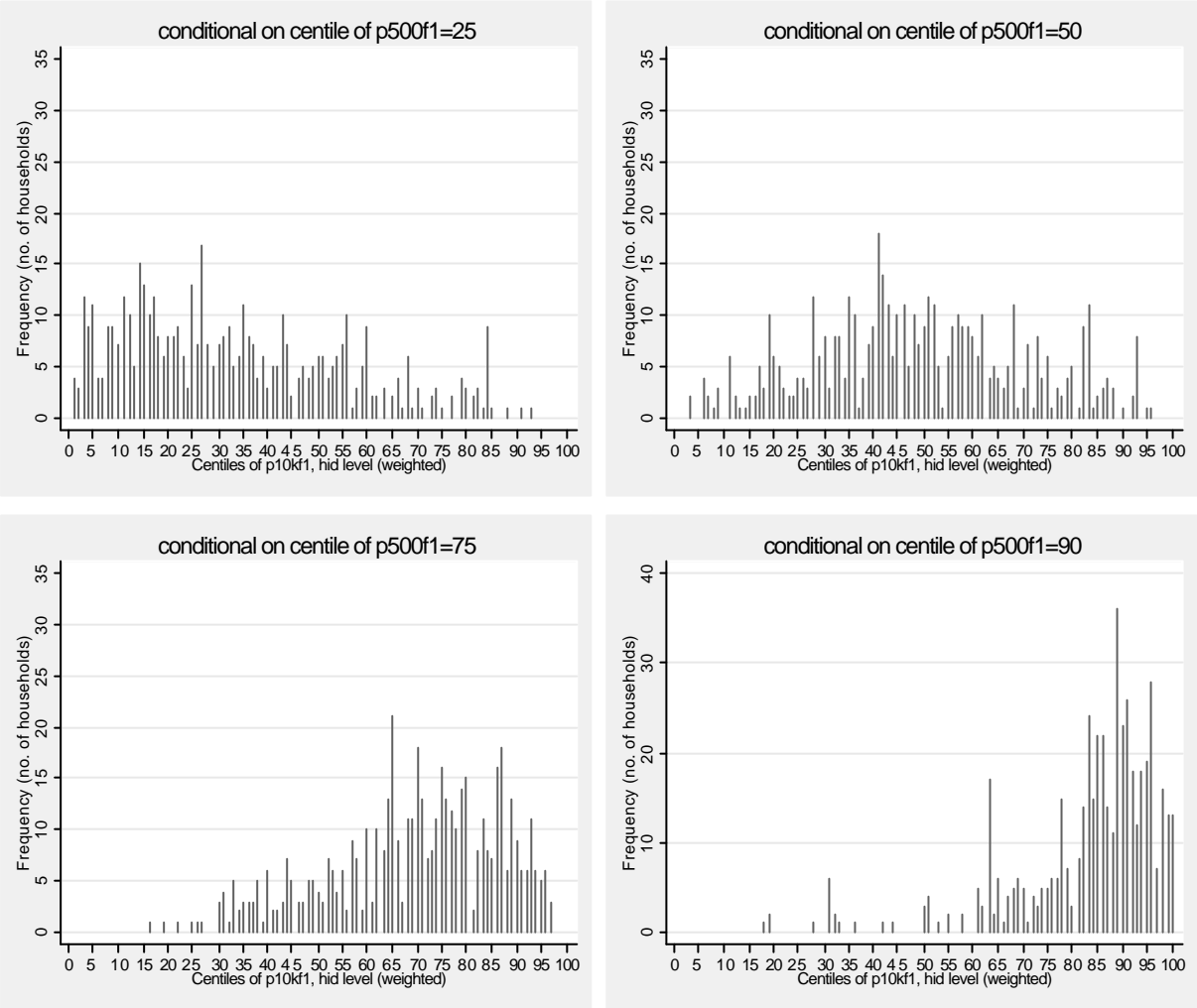
Figure 7: Quartiles of household income by wider area type (centiles of *p10kf1*) conditional on area type of local neighbourhood.
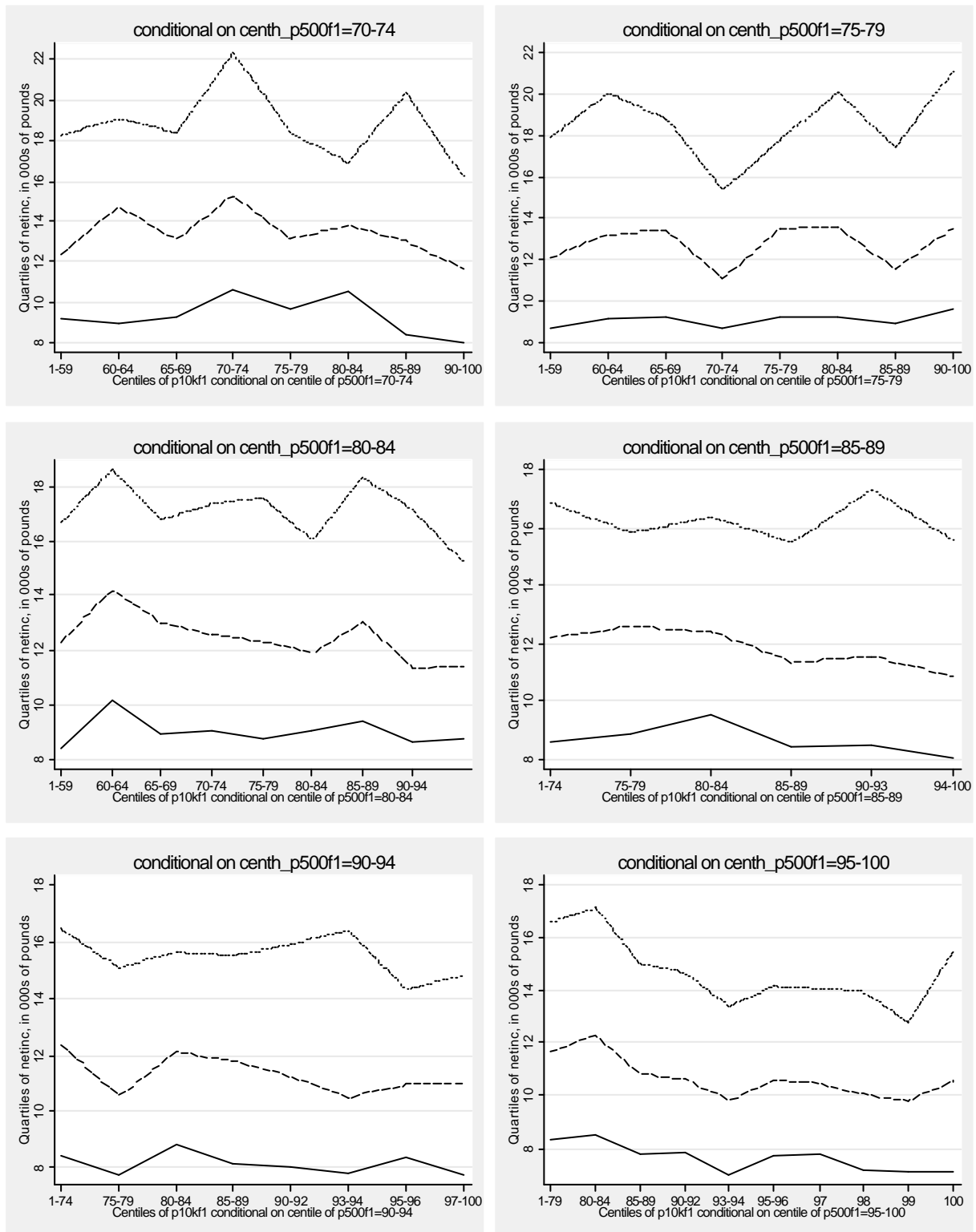
Figure 8: Quartiles of one-year changes in household income by wider area type (centiles of p10kf1) conditional on area type of local neighbourhood.
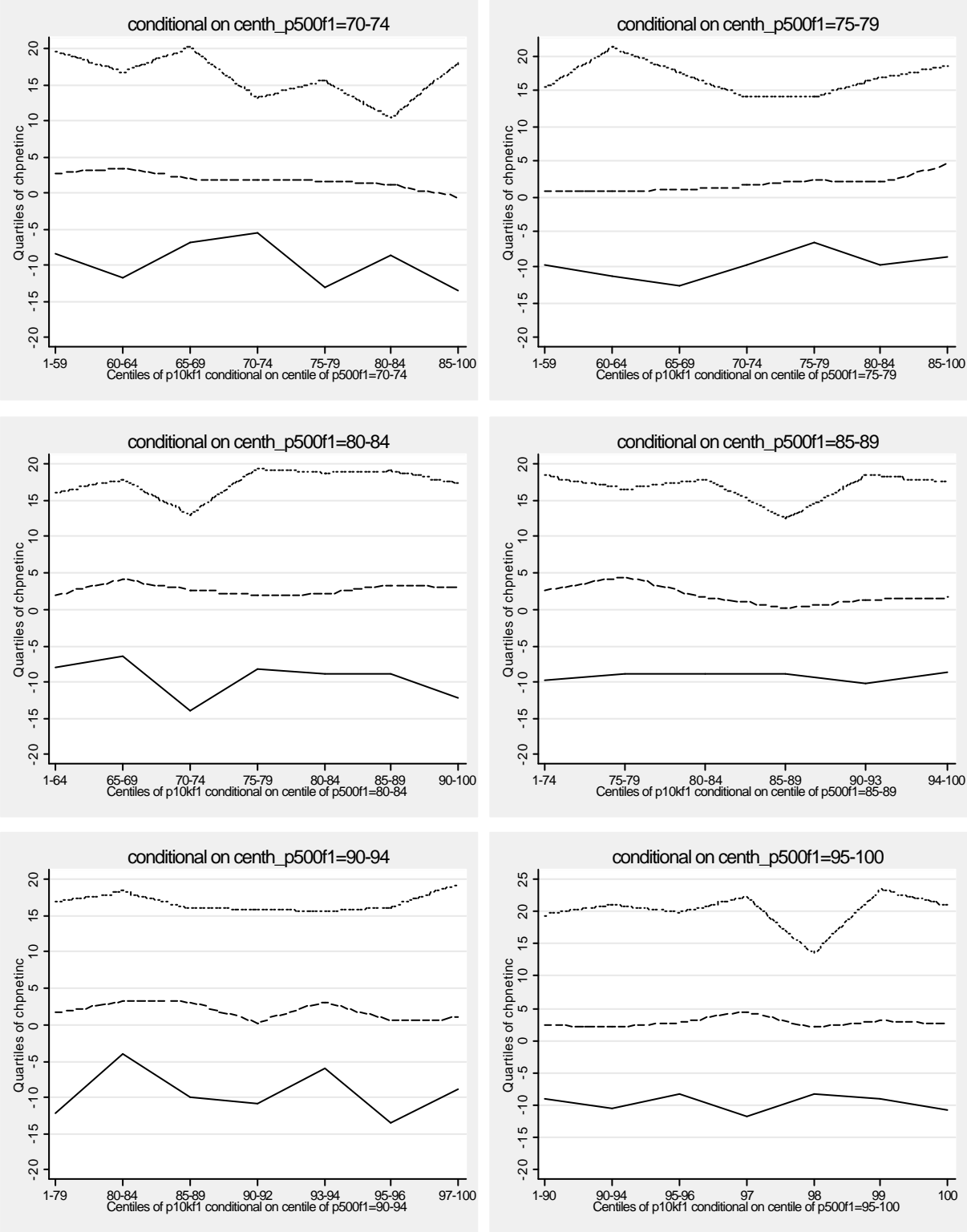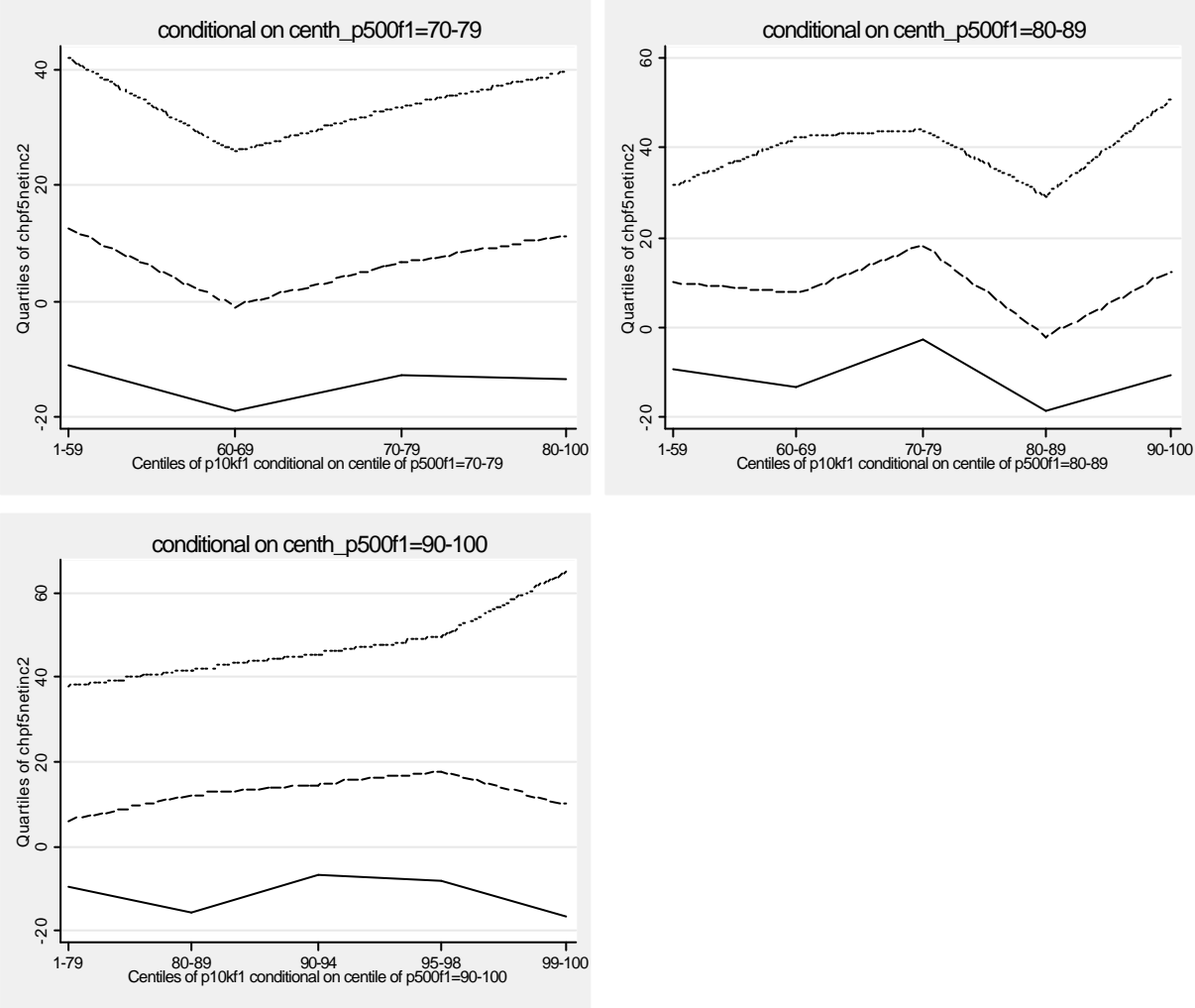
Figure 9: Quartiles of five-year changes in household income by wider area type (centiles of p10kf1) conditional on area type of local neighbourhood.

Appendix 1

Variable names and descriptive statistics

Variable Name	Description of variable	Observations	Mean	S.D.	Min.	Max.
Dependent variables						
NETINC	Total household disposable annual income, equivalised using McClements before housing costs index, deflated to Jan 2001 prices using RPI less local taxes. Unit: British pounds.	50223	16879.5	11701.3	3002.3	411060.2
CHPNETINC	Percentage change in NETINC from year t to year t+1	69580	8.5	60.2	-99.0	8635.3
CHPF5NETINC2	Percentage change in NETINC from year t to year t+5, non-rolling variable.	10936	20.7	72.3	-93.1	3046.7
B_NETINC	Household income trajectory: OLS coefficient of NETINC on WAVE for each individual with more than 7 observations.	6571	280.9	1261.8	-8861.3	36347.0
Identifiers						
PID	Cross-wave individual identification number.					
HID	Household identification number.					
YEAR	Indicates year of observation (sample is a panel).				1991	2000
WAVE	Indicates wave or of observation (sample is a panel).				1	10
Individual/household characteristics						
AGE12	Age of individual on 1st December of year of interview.	69580	45.7	18.4	15	98
GENDER	=1 if individual is female	69580	0.53	0.50	0	1
FAMSTAT	Denotes household composition type, types are mutually exclusive.					
	=f60, all household members above 60 years of age on 1st December of interview year.	50068	0.12	0.33	0	1
	=f1a, single adult household.	50068	0.06	0.24	0	1
	=f1ak, single parent household. I.e. single parent and dependent children[20].	50068	0.17	0.38	0	1
	=f2a, household with 2 adults only who form a couple (married or cohabiting).	50068	0.25	0.43	0	1
	=f2ak, household with 2 adults who form a couple and at least one dependent child.	50068	0.12	0.32	0	1
	=fma, 'multiple adult' household: household with either more than one adult with no dependent children i.e. includes households with a couple and non-dependent children.	50068	0.28	0.45	0	1
TENURE	Denotes tenure, types are mutually exclusive.					
	=too, owned outright.	49991	0.24	0.43	0	1
	=tom, owned with mortgage.	49991	0.42	0.49	0	1

---

[20] A dependent child is defined as one aged under 16 or aged 16-18 and in school or non-advanced further education, not married and living with parent (DSS definition), see wDEPCHL in BHPS record wINDALL.

	=tlar, local authority rented.	49991	0.19	0.39	0	1
	=thar, housing associated rented.	49991	0.05	0.21	0	1
	=tr, other rented.	49991	0.10	0.30	0	1
RURAL	=1 (rur) if percentage working in agriculture among nearest 1000 people in local area is greater than 0, = 0 (urb) if percentage working in agriculture among nearest 1000 people in local area is 0.	50223	0.33	0.47	0	1
MAXQUALS	Denotes highest known educational qualification of the individual	69282	3.04	1.45	0	5
	=0 "Still at school"	69282	0.00	0.03		
	=1 "No qualifications"	69282	0.26	0.44		
	=2 "Sub O-level qualifications"	69282	0.09	0.29		
	=3 "O-levels or equivalent"	69282	0.17	0.37		
	=4 "A-levels or equivalent"	69282	0.32	0.47		
	=5 "Higher qualification"	69282	0.17	0.37		

Bespoke Neighbourhood variables

Factor1	Index of neighbourhood characteristics. Increasing in indicators of socio-economic disadvantage.					
P500F1	Factor score 1 of nearest 500 people to household's location.	50223	-0.01	1.00	-2.07	4.95
P2KF1	Factor score 1 of nearest 2000 people to household's location.	50223	-0.01	0.99	-2.19	5.00
P10KF1	Factor score 1 of nearest 10,000 people to household's location.	50223	-0.01	0.99	-2.36	5.21

Notes:

1. Observations numbers for variables NETINC, FAMSTAT, TENURE, RURAL, P500F1, P2KF1, P10KF1 are measured at the household level.

Appendix 2

Table A2. Loadings on the principal component factor of socio-economic disadvantage of bespoke neighbourhood characteristics at the nearest 500 and 10,000 people scale

Census variable	Scale	
	Nearest 500 people	Nearest 10,000 people
% Unemployed	0.823	0.82
% Long term sick aged 16-60/64	0.818	0.853
% Owned outright	-0.619	-0.516
% Local authority	0.861	0.81
% Central heating	-0.365	-0.43
% Exclusive facilities	0.008	0.026
% No Car	0.858	0.838
% Density > 1 person per room	0.495	0.51
% Lone parent	0.768	0.806
% One person pensioner	0.251	0.164
% One person non-pensioner	0.178	0.195
% Black	0.227	0.212
% Indian	-0.015	0.039
% Pakistani & Bangladeshi	0.154	0.245
% Migrant in last year	-0.008	-0.023
% Working in agriculture	-0.165	-0.172
% children (aged 0-15 years)	0.318	0.36
% Professional-managerial	-0.713	-0.802

Notes:

1.Extraction Method: Principal Component Analysis wave by wave. Rotation Method: Oblimin with Kaiser Normalization.