Empirical Studies on the Location of Economic Activity and its Consequences

Henry G. Overman

London School of Economics and Political Science

PhD. Thesis 1999

Acknowledgements

Many thanks to Danny T. Quah for supervision.

Several people collaborated with me on the papers that comprise this thesis. Alex Heath co-authored Chapter 2. She introduced me to the AYS data set and, more importantly, to the joys of applied econometrics. Discussions with her also made me think about the issues that I deal with in Chapter 3. Diego Puga co-authored Chapter 4. He sparked my interested in economic geography questions and thus helped shape the direction of my current and future research. He also helped with all the LaTeXbits and pieces that significantly reduced the burden of putting together this thesis. Yannis Ioannides co-authored Chapters 5 and 6. He also, voluntarily, read through the entire thesis and suggested many changes which significantly improved the text.

Duncan Black, Gilles Duranton, Maia Guell–Rotllan, Louise Keely, James Richardson and Geoff Shuetrim have all read and commented on various parts of the thesis. I am grateful to all of them for their time and interest.

Thanks to the United Kingdom Economic and Social Research Council for funding. I have received additional logistic and financial support from The European Institute and The Centre For Economic Performance. In the last year I have also been employed, first by the Economics Department, and then by the Department of Geography and Environment, at the London School of Economics. I am grateful to staff in both organisations and both departments for their continued support.

Finally, I am fortunate in having many close friends and family. Many thanks to all of them for their support and company over the years.

Abstract

The thesis looks at the determinants of the location of economic activity and the impact that different location patterns can have on economic and social outcomes.

Chapter 1 provides a brief introduction and summary of the remaining chapters. Chapter 2 looks at neighbourhood effects on school drop out rates using data from the Australian Youth Survey. We identify two different types of neighbourhood effects. First, teenagers are more likely to drop out if the average drop out rate in the neighbourhood is high. Second, teenagers are more likely to drop out if they live in neighbourhoods with a high percentage of adults with vocational qualifications. Chapter 3 uses similar data to test for neighbourhood effects at different spatial scales. We find that educational composition of larger neighbourhoods influences drop out rates, possibly reflecting the structure of local labour market demand. We also find that low socio-economic status of the immediate neighbourhood has an adverse impact on drop out rate.

Chapter 4 considers the evolution of European regional unemployment. European regions have experienced a polarisation of their unemployment rates between 1986 and 1996, as regions with intermediate rates have moved towards either extreme. Regions' outcomes have closely followed those of neighbouring regions. This is only weakly explained by regions being part of the same Member State, having a similar skill composition, or broad sectoral specialisation. Even more surprisingly, foreign neighbours matter as much as domestic neighbours. All of this suggests a reorganisation of economic activities with increasing disregard for national borders.

Chapter 5 considers mobility within the US city size distribution. Papers that study city size distributions have concentrated predominantly on the shape of that distribution, while ignoring mobility within the distribution. We develop a series of tools that can be used to study such intra-distribution dynamics and apply them to data from the US. Chapter 6 uses a similar set of tools to examine spatial aspects of the evolution of the US system of cities. We find that some features of that evolution are consistent with theoretical models developed by the new economic geography.

Contents

List of Figures

List of Tables

1

Introduction

1. The location of economic activity and its consequences

The chapters in this thesis represent empirical studies of a number of seemingly unrelated issues. The first two chapters deal with the importance of neighbourhood effects for education decisions. The third chapter looks at the evolution of regional unemployment rates across Europe. The final two chapters consider the evolution of the US city size distribution. Two key threads connect all of these chapters, however.

First, all of the empirical work is related to the location of economic activity and the implications that this has for economic and social outcomes. Thus, the first two chapters consider whether neighbourhood characteristics can reinforce family and background effects on education decisions. We know that families sort across geographical space, so that neighbourhoods are heterogenous. Understanding whether this sorting has an impact on socio–economic outcomes is important if we want to change these outcomes. Chapter 2 tries to identify the channels through which neighbourhood may affect educational outcomes. Chapter 3 tries to identify at what scale these neighbourhood effects may occur.

The emphasis in Chapter 4 is somewhat different. In the first two chapters, the distribution of families across neighbourhoods is taken as given, and we study the impact that this distribution has on education outcomes. In Chapter 4, we consider the distribution of unemployment rates across European regions. Here, in contrast to the first two chapters, we are interested in understanding what drives the evolution of the distribution of unemployment rates across regions, rather than taking that distribution as given and considering the implications for socio–economic outcomes. Chapter 4 shows that a possible determinant of regional unemployment outcomes may be the way that firms are currently relocating across the European Union. This suggests that the way that firms sort across regions affects unemployment outcomes for those regions. A similar process of sorting by families across neighbourhoods leads to the underlying distribution that we take as given in Chapters 2 and 3.

Chapters 5 and 6 are concerned with the distribution of city sizes in the US. Chapter

5 concentrates on examining how that distribution evolves over time, and how we can characterise that evolution. Chapter 6 considers the importance of spatial features for understanding the evolution of the distribution characterised in Chapter 5. The evolution of the city size distribution is driven by the relocation of firms and workers across cities, and from rural to urban environments. Studying the city size distribution helps us understand the economic mechanisms governing these location and relocation decisions.

The second feature that unifies the chapters is that they need to address a number of similar econometric issues. First, all of the chapters deal with situations where there are interactions between the units under consideration (teenagers, regions, cities). Further, these interactions are partially governed by distance. In the first two chapters this interaction is direct – the drop out behaviour of a teenager in one neighbourhood may influence the drop out behaviour of other teenagers in the same neighbourhood. In Chapter 4, the interactions are a result of the fact that the underlying location decisions of both firms and workers partly determine the evolution of unemployment rates. Given that these decisions are interlinked, the result is interaction between the unemployment rates of neighbouring regions. Something similar occurs when we consider the evolution of the city size distribution. Theoretical models that explicitly consider the spatial structure of the urban system suggest that this spatial structure governs the evolution of individual cities within that system. There are thus complex feedbacks between cities within the urban system.

There is a large spatial econometrics literature which tries to deal with interactions governed by distance. The work in this thesis is clearly related to that literature. However, in contrast to much of that literature, the interactions between units are one of the key aspects of each of the processes that we want to capture. That is, the spatial correlation is not just a nuisance effect that we want to condition out using the tools developed by the spatial econometric literature.

A second common feature, is that we are often interested in the distribution of activities across space. Sometimes we will find it informative to consider those distributions directly, rather than reducing the issue to one of understanding the behaviour of a representative unit within that distribution. Theoretical work on the issues that we consider, suggests that spatial interactions may lead to multiple equilibria for the outcomes of interest. That is, units with initially similar characteristics may see very different outcomes. Outcomes that may not be a–priori predictable on the basis of observable characteristics. In these situations, no unit can be characterised as the representative agent, and we are forced to study the distribution directly.

Having outlined a number of themes that link the chapters, the remainder of this introduction provides a brief summary of the individual chapters.

2. The influence of neighbourhood effects on education decisions in a nationally funded education system

This chapter considers empirical evidence on the importance of geographical neighbourhoods for social and economic outcomes. Casual observation suggests that children who grow up in 'bad neighbourhoods' tend to have worse outcomes on a range of social indicators. They accumulate less human capital, drop out of school earlier and have a higher risk of involvement in criminal activity. Young women are more likely to get pregnant in their teenage years, and tend to form single parent households after the birth of their child. However, the fact that neighbourhood characteristics appear to be related to individual behaviour may result from the tendency of families with similar characteristics to live close to each other[1]. Sorting across neighbourhoods leads to a *correlation* between neighbourhood characteristics and drop out rates. The neighbourhood effects that we want to capture are ones where neighbourhood economic and demographic characteristics *cause* changes in drop out behaviour. This chapter considers teenage drop out rates to examine whether concentrations of poorer families in bad neighbourhoods may exacerbate individual and family effects.

Empirical papers studying the effects of neighbourhood characteristics on socio–economic variables have predominantly used US data. We argue that the local nature of the US schooling system means that neighbourhood effects on education decisions may act through fiscal or social channels. We use data for a nationally funded public schooling system to identify neighbourhood effects in an environment where the level of school funding is independent of neighbourhood composition.

We identify two different types of neighbourhood effects on school drop out. First, teenagers are more likely to drop out if the average drop out rate in the neighbourhood is high. Second, teenagers are more likely to drop out if they live in neighbourhoods with a high percentage of adults with vocational qualifications. The fact that neighbourhood effects appear to operate through these two channels has interesting implications for policy. The existence of endogenous effects suggests that one–off interventions may push neighbourhoods towards a better self–sustainable equilibrium. The policy implications of the second finding are less clear and depend on the mechanism through which these effects operate. We are unable to distinguish whether the results reflect local labour market conditions or the importance of local information networks. However, our results in Chapter 3 suggest that the former is the most likely mechanism.

3. Neighbourhood effects in small neighbourhoods

This chapter considers the existence and the scale of neighbourhood effects on the drop out decisions of Australian teenagers. We deal with two related questions. First, does the concentration of poorer families in poor neighbourhoods amplify individual and

[1] That is, to sort across neighbourhoods according to socio–economic criterion.

family effects on drop out tendencies? Second, at what spatial scale might such effects occur? That is, do neighbourhood effects depend on the socio–economic composition of the immediate or the larger locality?

We use data on a sample of Australian teenagers to test for neighbourhood effects on school drop out rates. The data allows us to test for neighbourhood effects at two different spatial scales. First, we have information on postcodes, which are larger neighbourhoods, often corresponding to school catchment areas. Second, we have collection district data which define much smaller local neighbourhoods.

We find that educational composition of the larger neighbourhood can influence the drop out rate. We also find that low socio–economic status of the immediate neighbourhood has an adverse impact on drop out rate. The combined evidence from small and large neighbourhoods suggest that the large neighbourhood result on education composition is most likely to reflect the impact of local labour market demand. This poses an interesting problem for policy makers – how best to deal with school drop out when this may reflect rational choices in the context of local labour market conditions. The small neighbourhood results also have interesting policy implications. They suggest that government policy may need to consider the socio–economic composition of quite small geographical areas if it considers interfering in the market to create greater income mixing within neighbourhoods.

4. Unemployment clusters across European regions and countries

In the decade up to the mid 1980s, the average European unemployment rate was rising. However, differences in unemployment rates across European regions were very stable, with regional labour forces adjusting just enough to offset ongoing changes in regional employment. In this chapter we start by showing that the evolution of the regional distribution of unemployment rates over the last decade has been quite different. The average European unemployment rate was the same, 10.7%, in 1996 as in 1986, and the decade separating them could be thought of as covering a full cycle in unemployment rates. Yet during this decade there has been a polarisation of unemployment rates across the regions of the EU. To go beyond the limited conclusions that can be drawn from comparing summary statistics over time, we look at the evolution of the shape of the whole distribution of European unemployment rates. We also track the outcomes of individual regions. Regions that in 1986 had a low unemployment rate relative to the EU average still tended have a relatively low unemployment rate in 1996. Similarly, regions that in 1986 had a relatively high unemployment rate still tended have a relatively high unemployment rate in 1996. However, regions with intermediate initial unemployment rates had mixed fortunes. Some saw a marked fall in their relative unemployment rate, while others saw it rise, and still others saw it roughly unchanged.

We show that this process has been driven by changes in regional employment rather than by changes in demographic structure or labour market participation. There has been some labour force adjustment to regional employment changes. Regions with relatively low unemployment rates have typically experienced above average labour force growth, while regions with relatively high unemployment rates have generally experienced a below average increase, or a fall, in their labour force. However, this adjustment has been insufficient to prevent the polarisation of European unemployment rates.

We use two complementary techniques, one parametric, one nonparametric, to examine these alternative explanations. The nonparametric technique involves grouping regions by some common characteristic (like State Membership, or similar skill composition) and then examining the similarity of unemployment outcomes within groups. This technique has the distinct advantage that it allows for different regional characteristics to matter to different degrees for different parts of the distribution. Its main disadvantage is that it only allows one to consider a single factor at a time. To ensure that our results are robust in this respect, we finish with a more standard parametric analysis. This also allows us to consider the importance of cross border effects.

Both the parametric and nonparametric techniques show that regions' unemployment outcomes have closely followed those of neighbouring regions. This is only weakly explained by regions being part of the same Member State, having a similar skill composition, or broad sectoral specialisation. Remarkably, we find that neighbouring regions across national borders are as important as domestic neighbours in determining unemployment outcomes. The clusters of high and low unemployment that have emerged over the last decade show little respect for national borders.

The fact that unemployment outcomes are so much more homogenous across neighbours, foreign and domestic, than across regions in the same Member State also tells us something about the spatial dimensions of the emerging clusters of high and low unemployment in Europe. The average Member State has 13.6 regions, while the average neighbourhood has 5.6 regions. Hence these are clusters of typically less than one half of the size of the average Member State of the European Union, but often extend across national borders and include regions from more than one Member State.

That also has important implications for policy. European regional policy has traditionally targeted mainly regional differences in income per capita, but is increasingly shifting its focus towards tackling regional differences in unemployment rates. There is a clear empirical reality underlying this change in emphasis – in contrast to the divergence of unemployment rates across European regions, differences in regional incomes per capita are narrowing. But there is one important additional difference. While inequalities in incomes per capita exhibited a core–periphery gradient, unemployment clusters are more localised and emerging in both the core and the periphery of the EU. There is strong political opposition to tackling these growing unemployment rate differences through increased labour mobility. Recent location theories suggest that the self–reinforcing nature of agglomerations will make these hard to break once they become established. However, given that the unemployment clusters we find are of not very large size and scattered across Europe, it may be politically viable as well as more efficient to implement policies that accept some clustering and larger mobility within a neighbourhood.

5. The cross–sectional evolution of the US city size distribution

Studies of the distribution of city sizes have tended to concentrate on the shape of that distribution. We argue that this emphasis neglects other important features of the distribution – most noticeably the nature of intra–distribution dynamics. This neglect has lead to the development of theoretical models which are capable of generating the external shape of the city size distribution, but that may rely on unrealistic assumptions on intra–distribution dynamics.

This chapter considers tools that allow us to characterise the nature of intra–distribution dynamics for the city size distribution. In comparison to existing studies, our work has two main advantages. First, our empirical tools do not require us to discretize the city size distribution before studying the intra–distribution dynamics. We are thus able to study characteristics of the evolution of the city size distribution which may be disguised by techniques which require us to discretize that distribution. Second, we are able to characterise intra–distribution dynamics using statistics that are directly comparable across different urban systems, even when those urban systems differ in size.

We use these tools to give benchmark figures for the degree of mobility within the US city size distribution. We also use them to consider the degree of mobility within different regional sub–systems. We find that different regions show different degrees of intra–distribution mobility. More surprisingly, we find that the largest 'top tier' cities appear to show more mobility, than a collection of large 'second tier' cities. Usually, these two tiers would be absorbed within one discrete state when we use techniques that discretize the distribution.

The results on regional and tier sub–systems also throw up questions for the literature that tries to model the economic mechanisms that may govern the evolution of urban systems. Are their economic forces that can explain the apparent differences in the nature of intra–distribution mobility between different regional sub–systems? More interestingly, what explains the apparent stability of the second tier of cities relative to the other two top tiers?

6. The spatial evolution of the US urban system

This chapter attempts to examine empirically some of the spatial aspects of the evolution of the US system of cities. The evidence that we consider falls in to two broad categories.

First, we consider spatial features of the system about which theory is relatively silent. This includes, for example, evidence on the co–evolution of the distribution of market potentials and relative city sizes. Second, we consider spatial features of the system with which theory deals more directly. This includes, for example, evidence on the relationship between relative growth rates and relative market potentials. This second set of results could be characterised as 'tests' of the new economic geography. Care is needed here, however, as the results that we get are also consistent with other models of the evolution of the urban system.

The key empirical implication common to the newer theoretical frameworks is a prediction that the dynamic evolution of wages and population reflects spatial considerations. Theory suggests that there are complex interactions between spatially dispersed economic agents, with those interactions partly governed by distances between the location of those agents. We use tools developed by Quah (1996, 1997*a*,*b*) to characterise some key aspects of that evolution. We consider the relationship between the distributions of city sizes, market potential and wages. We find that there is no simple relationship between the distributions of any of these variables. Further, our empirical technique allows us to see that these complex relationships evolve over time.

We also use these techniques to examine the relationship between city growth rates and market potential. We then estimate a parametric specification which allows us to compare a number of predictions from different theoretical models of the evolution of the urban system. Initial parametric results suggest that there is a negative relationship between city size and market potential if we do not take in to account own lagged city size. Once we allow for own lagged city size, there is a positive relationship between market potential and city growth. Own lagged city size has a negative effect. By far the most robust parametric result relates to the ratio of lagged own city size to market potential. When cities are small relative to their market potential they grow faster.

This result is consistent with a theoretical models advanced as part of the new economic geography. In particular, the results are consistent with models of the urban system that combine new economic geography findings on inter–metropolitan distance with older notions of intra–metropolitan congestion. However, if the results are driven by the own lagged city size variable, then these results may also be consistent with other theoretical models that only emphasise congestion effects within cities.

2

The Influence of Neighbourhood Effects on Education Decisions in a Nationally Funded Education System

with Alex Heath

1. Introduction

This chapter considers empirical evidence on the importance of geographical neighbourhoods for social and economic outcomes. Casual observation suggests that children who grow up in 'bad neighbourhoods' tend to have worse outcomes on a range of social indicators. They accumulate less human capital, drop out of school earlier and have a higher risk of involvement in criminal activity. Young women are more likely to get pregnant in their teenage years, and tend to form single parent households after the birth of their child. However, the fact that neighbourhood characteristics appear to be related to individual behaviour may result from the tendency of families with similar characteristics to live close to each other[1]. This chapter considers teenage drop out rates to examine whether concentrations of poorer families in bad neighbourhoods may exacerbate individual and family effects.

Traditionally, following human capital theory, an individual's education decisions have been treated as a function of personal characteristics, the family environment and macroeconomic conditions[2]. Recently, however, there has been a rapid increase in the number of empirical studies analysing how the immediate geographical

[1] That is, to sort across neighbourhoods according to socio-economic criterion.

[2] The classic reference is Becker (1993) For a recent study treatment using Australian data see Miller and Volker (1987).

environment affects behaviour, above and beyond the effects of family background and macroeconomic conditions. This recent expansion in the empirical literature investigating the existence of neighbourhood effects has, in part, been driven by the availability of data with contextual information. That is, data allowing individuals to be located in relatively small geographic areas which can be thought of as neighbourhoods. Most of these data sources are concerned with the circumstances of those living in inner city areas and the suburbs of major US cities. Comparable evidence from outside the US is not readily available[3]. This chapter provides such evidence, for a nationally funded school system using data from the early 1990's.

We use Australian data to examine the earliest education decision available to Australian teenagers – whether or not to complete high school. We combine the Australian Youth Survey (AYS) with neighbourhood data derived from the 1991 Australian census to create a data set of individuals, with information on their personal characteristics, family background and their immediate environment. Although the results are interesting in their own right, they also have implications for our understanding of neighbourhood effects in a much wider context. One of the main problems facing studies of neighbourhood effects using US data is that the locally funded nature of the US education system makes it difficult to distinguish between true neighbourhood effects and differential tangible inputs into the schooling system. In contrast, the distribution of funding across Australian secondary schools is relatively equitable and virtually independent of neighbourhood composition.

Why should policy makers be interested in socio–economic neighbourhood effects? If neighbourhood effects exist, the ability of families to sort across neighbourhoods may lead to costs for families in low income neighbourhoods that outweigh the benefit to families in high income neighbourhoods. Policy makers may want to intervene to ensure that externalities arising from the presence of neighbourhood effects are internalised. The subsequent increase in efficiency could increase welfare for all. In addition, equity considerations suggest that, 'A system that allows the accidents of geography and birth to determine the quality of education received by an individual is inimical to the idea of equal opportunity in the market place'[4]. The policy response will depend on the mechanisms through which neighbourhoods influence education outcomes, as well as the strength of these neighbourhood effects.

The structure of this chapter is as follows. The next section discusses the existing theoretical and empirical literature regarding the existence of neighbourhood effects, and the mechanisms through which they may operate. Section 3 provides information about the data used in the subsequent analysis. In particular, we detail the Australian Youth Survey and how it links to the 1991 Australian Census data. Other

[3]Notable exceptions include Robertson and Symons (1996) and Meghir (1997) who use data for the UK, although Meghir (1997) does not directly consider neighbourhood effects, and Robertson and Symons (1996) concentrate on peer group effects.

[4](Fernandez and Rogerson, 1988, p136).

practical considerations such as neighbourhood definitions and variable selection are also highlighted. Section 4 explores the importance of neighbourhood effects for the school leaving (drop out) decisions of Australian teenagers. The results suggest that there are significant exogenous and endogenous social effects on Australian drop out rates. Policy conclusions are drawn in Section 5.

2. The existing literature

A number of papers suggest mechanisms through which neighbourhood effects might arise. Namely, in situations where some aspects of the individuals information set may depend on location, or where an individual's payoff or optimal strategy may be influenced by the action of others in their neighbourhood. These ideas provide theoretical support for the existence of neighbourhood effects. Establishing empirical support for neighbourhood effects has proved difficult. Data availability, measurement and identification problems have all dogged attempts to test for neighbourhood effects. We return to these issues below.

Struefert (1991) presents a theory of role models to explain why teenage education decisions may be affected by neighbourhood composition. He assumes that children infer the returns to effort at school by examining the outcomes of adults in their neighbourhood, and base their education decisions on this information. Thus, the distribution of education across neighbourhoods can influence the education decisions of future generations.

Montgomery (1991) provides an alternative 'social networks' explanation. His model assumes that the unemployed have different productivity levels, but that, without further information, they are observationally equivalent to potential employers. By introducing a social structure in which workers with similar productivity levels are more likely to associate with each other, it becomes possible for employers to increase the probability of hiring a high productivity worker by employing people recommended by current high productivity workers. By increasing information flows, social networks relieve adverse selection problems and increase efficiency. To the extent that social networks are localised it is possible that some neighbourhoods will provide their job seeking residents with better job information networks than others. For example, high unemployment areas are likely to have less active job information networks, which will decrease the probability of receiving job offers and may decrease the incentives to leave school early[5]. We will see that some of the evidence that we find in this paper may be consistent with such theories of social networks.

These two models help explain how the composition of the neighbourhood may affect an individual's decisions. We label these spillovers exogenous neighbourhood

[5]This may also help explain the increasing concentration in Australia of unemployment in low status neighbourhoods between the 1976 and 1991 Censuses (Gregory and Hunter (1995)). For further evidence on this see Heath (1998).

effects. We are also interested in considering endogenous neighbourhood effects, where the propensity of staying on at school is an increasing function of other teenagers' propensities to stay on at school. In the terminology of Cooper and John (1988), we are interested in the existence of strategic complementarities. The presence of strategic complementarities also raises the possibility of multiple equilibria across otherwise identical neighbourhoods. Banerjee and Besley (1990) use this idea to model the importance of peer effects on education achievement. These endogenous effects are also implicit in the ethnographic evidence described in Akerlof (1997), which suggests that an individual's payoffs to completing school can be severely diminished if peer group members do not complete school.

A rapidly growing literature has also found empirical support for the existence of neighbourhood effects. Jencks and Mayer (1990) provide a detailed survey of the early literature. One of the best sources of data for looking at the influence of the neighbourhoods on education outcomes is the 1968 sample of the University of Michigan Panel Study of Income Dynamics (PSID) combined with the 1970 Census Fifth Count for Zip Codes. This provides a sample of young male heads of household who were 23–32 years old in 1978 and who were living with at least one of their parents in one of 188 Standard Metropolitan Statistical Areas in 1968. The neighbourhood data consist of a number of socio–economic indicators recorded by five digit zip code.

Two representative papers are Datcher (1982) and Corcoran, Gordan, Laren, and Solon (1992). Both find strong intergenerational links between father's 1968 income and son's subsequent economic status. However, neither report a strong impact of neighbourhood variables on son's income over and above family background effects. Corcoran *et al.* (1992) conclude that a likely reason for these problems is the presence of measurement error and omitted variable bias.

Crane (1997) finds evidence of neighbourhood effects which are especially important in low income neighbourhoods. He suggests that the extremely bad outcomes observed in inner city areas of major US cities, can be explained by epidemic or contagion effects, triggered after some critical level of social problems is reached. After this point, outcomes in these neighbourhoods deteriorate rapidly as susceptibility to these problems increases. He tests this hypothesis by estimating a piecewise linear logit model using the Census Bureau's 1970 Public Use Microdata Samples, and finds that the probability of dropping out of school is much higher than background characteristics suggest for teenagers in the lowest 5% of the neighbourhood distribution.[6]

Case and Katz (1991) explicitly allow for the possibility of strategic interaction between agents in their analysis of the influence of neighbourhoods on the outcome of youths in low income neighbourhoods in inner city Boston. They look at the influence of peer behaviour and the characteristics of older members of the neighbourhood on several

[6]Borjas (1992) and Borjas (1995) use similar data to test for the importance of 'ethnic capital' on the human capital decisions of individuals from different ethnic groups. This work on the impact of ethnic group outcomes on individual decisions has close parallels with the neighbourhoods literature.

outcome variables including teen pregnancy, drug abuse, church attendance, involvement in crime and drop out rates. They find that there are significant neighbourhood effects, even after a large array of family background characteristics are taken into account. Interestingly, they find that child behaviour is strongly influenced by similar behaviour of the neighbourhood adult population. High rates of neighbourhood crime bias children towards criminality, high neighbourhood church attendance biases children in other, more saintly, directions.

3. The AYS and Australian neighbourhoods.

The Australian Youth Survey is compiled by the Australian Department of Employment, Education and Training[7]. The data covers the period from 1989 to 1994. The first wave, sampled in 1989, consists of 5350 sixteen to nineteen year olds. In each subsequent year roughly 1500 sixteen year olds are interviewed for the first time, and all other panel members are re–interviewed where possible. Our sample includes teenagers who were in the final year of high school, or were in the same cohort but left school at an earlier stage. In this sample, the probability of leaving school early is 30 percent, which is consistent with aggregate retention rates over this period.

Extensive individual and family background information is collected, including details of educational outcomes and labour market experience for both the respondent and the members of their household. Unfortunately, parent's income is not well measured. Child reported income figures are available, but the response rate is relatively low and the quality of the data is questionable[8]. There is, however, detailed information about the occupational status of parents and their education levels, both of which are likely to be good proxies for income, especially permanent income. These variables are also likely to provide information about the parents' likely attitudes to education. Information on other important variables is also available, including the number of siblings and the type of school attended.

Most importantly for our purposes, the AYS provides detailed geographic information. As well as providing information about which state the respondent lives in, and the section of state[9] the respondent lived in before they were 14 years old, the AYS allows us to identify individuals' geographic neighbourhoods in most years. In 1989 and 1990 the information is recorded by 1986–defined collection districts (CD), which are small neighbourhoods containing, on average, 465 individuals. The postcode where the interview took place is available for re–interviewees in 1991 and all people interviewed from 1992 to 1994. Postcodes are larger than CDs, but there is a mapping from 1986–defined CDs to 1991 defined postcodes. The average postcode has 5558 residents over the age of 15 years. The largest postcode has a population of 62885; the smallest

[7]Now known as the Department of Employment, Workplace Relations and Small Business.

[8]For further discussion see Dearden and Heath (1996).

[9]Section of state is categorised as either capital city, other city, country town or rural area.

has less than a hundred residents. The distribution is highly skewed with 90 percent of postcodes with fewer than 15131 residents.

The following analysis is restricted to major urban areas for two reasons[10]. The first is that the Australian Bureau of Statistics (ABS) introduces sampling error into small postcodes to ensure confidentiality. By excluding non-major urban areas, most affected postcodes will be excluded. The second reason is that the concept of the neighbourhood underlying the economic models above is related to physical proximity. Consequently, low density population areas, such as rural areas, do not conform easily to the concepts underlying our analysis.

The childhood postcode is defined as the postcode where the individual was interviewed when they were 16, as this is the earliest recorded neighbourhood information. The postcode information for 16 year-olds is missing in 1991, and these individuals are allocated their 17 year old postcode from the subsequent interview. This is also done for the 17 year olds in 1989 to increase the available sample. Childhood postcodes are only defined if the children are living with one or both of their parents. This is standard practice in the literature, but may cause biases if the decision to move out from the family home is a function of the endogenous variable. For example, if children who leave school at 16 are also more likely to move out then we under sample this group of children. It should be noted here, that respondents who reported that they had spent most of their life until 14 overseas are excluded from the analysis. This does not significantly affect any of the results.

We have information on a range of neighbourhood characteristics at the postcode level from the 1991 Australian Census. This includes information about male and female education attainment, household and personal income, and labour force status. We also have a neighbourhood socio-economic status (SES) variable based on 1991 Census data which was constructed at CD level by Hunter (1996).

Further details of the data, and the variables that we use, are given in Appendix A.

4. Empirical model and results

The purpose of our analysis is to estimate the effects of neighbourhood on education decisions. In particular, we look at the first free education decision available to Australian teenagers: whether to complete high school, or to leave at the legal minimum age of 15 years. In Section 4.1 we develop an empirical model within the framework presented by Manski (1993, 1995) to formalise the different mechanisms through which neighbourhood effects could operate. In Section 4.2 we present the results of estimation that ignores neighbourhood effects. In Sections 4.3 and 4.4 we add neighbourhood variables and discuss the importance of neighbourhood effects for our understanding

[10]Major urban areas are defined as cities with greater than 100 000 in population.

of teenage education decisions. In Section 4.5 we discuss alternative interpretations of these results.

4.1 A common framework

Our empirical work is based on the model originally proposed by Manski (1993, 1995), summarised as follows:

$$y_i^* = \alpha + z_i'\beta + E(z|x_i)'\gamma + \delta E(y^*|x_i) + \varepsilon_i; \tag{2.1}$$

where y_i^* is the underlying propensity to leave school before the final year of high school for individual i; z_i are the personal background and family characteristics of individual i; x_i is the postcode neighbourhood of individual i; $E(z|x_i)$ are the average characteristics of the individuals in that neighbourhood; $E(y^*|x_i)$ is the probability of being an early school leaver in that neighbourhood; and ε_i is the error term which contains all the unobserved factors which affect individual i's propensity to leave school before the final year.

Thus $E(z|x_i)$ captures exogenous neighbourhood effects, and $E(y^*|x_i)$ captures endogenous neighbourhood effects.

4.2 Individual effects

We start our analysis by estimating the model assuming that neighbourhood effects are not important (i.e. assuming that $\gamma = \delta = 0$ in Equation 2.1). This specification has been considered in earlier literature and has been quite successful in explaining teenage education decisions (Miller and Volker (1987)). Because we do not observe the propensity to leave school early but the final decision, which is a binary variable, we estimate this model using probit. The results are presented in the first two columns of Table 2.1.

Because the probit model is non–linear, the estimated coefficients will provide information about the direction of the effect an independent variable has on the probability of leaving school early, but the magnitude of the effect depends on where the probability is evaluated. To facilitate comparison we present results in marginal effects form. The marginal effect can be interpreted as the impact a one unit change in the variable will have on the probability of leaving school early, given that the probability is initially evaluated at the sample mean. For dummy variables, marked with an asterisk, the reported marginal effect will be the change in the probability of being an early school leaver if the individual has that characteristic rather than the reference characteristic given by the omitted group. The results in Table 2.1 are expressed in marginal effects form.

Note that most of the variables have the expected effect. Males are 8 percentage points more likely to leave school early than females, and older cohort members are

	Individual Effects Coeff.	Individual Effects t–stat	Reduced Form Coeff.	Reduced Form t–stat	Structural Form Coeff.	Structural Form t–stat	Sample Average
Neighbourhood							
Average personal income			0.000	0.39	0.000	−0.43	5601.48
Proportion grad qual			0.001	0.31	0.015	2.28	0.13
Proportion trade qual			0.014	3.02	0.016	3.41	0.14
Unemployment rate			0.005	1.52	0.001	0.23	0.11
Endogenous effect					0.021	2.36	0.30
Personal Background							
Male*	0.080	5.59	0.081	5.65	0.080	5.60	0.50
Age	0.093	6.48	0.092	6.43	0.092	6.40	17.11
Number of Siblings	0.009	1.77	0.009	1.85	0.008	1.73	2.02
English not 1^{st} language							
English good*	−0.080	−3.12	−0.072	−2.76	−0.072	−2.75	0.10
English poor*	−0.137	−2.09	−0.129	−1.94	−0.126	−1.89	0.01
Born overseas*	−0.077	−3.15	−0.076	−3.12	−0.075	−3.08	0.12
School							
Catholic*	−0.125	−7.32	−0.120	−6.96	−0.119	−6.94	0.23
Other non–government*	−0.170	−7.10	−0.161	−6.48	−0.160	−6.46	0.09
Parent Characteristics							
Fathers occ. status @14	−0.002	−3.51	−0.002	−3.48	−0.002	−3.47	29.72
Mother's occ. status @14	0.000	−0.26	0.000	−0.29	0.000	−0.27	19.56
Father not emp @14*	−0.017	−0.50	−0.016	−0.45	−0.012	−0.36	0.05
Mother not emp @14*	0.004	0.20	0.004	0.18	0.003	0.14	0.36
Father not present @14*	0.059	2.26	0.060	2.30	0.061	2.32	0.16
Mother not present @14*	0.443	10.84	0.445	10.82	0.444	10.79	0.05
Father has							
degree*	−0.055	−2.17	−0.052	−2.04	−0.052	2.01	0.17
trade qualifications*	0.020	0.95	0.016	0.79	0.017	0.84	0.17
other post-secondary*	−0.039	−1.55	0.045	−1.78	−0.046	−1.83	0.10
Mother has							
degree*	−0.091	−3.56	−0.083	−3.22	−0.087	−3.39	0.13
trade qualifications*	0.019	0.52	0.017	0.49	0.018	0.49	0.04
other post-secondary*	−0.025	−1.19	−0.025	−1.17	−0.026	−1.22	0.15
Section of state							
Other capital city*	0.014	0.70	−0.005	−0.22	−0.001	−0.07	0.17
Rural area*	0.045	1.03	0.025	0.57	0.028	0.63	0.03
Country town*	0.026	0.99	0.008	0.29	0.010	0.39	0.08
State							
Victoria*	−0.025	−1.33	0.001	0.05	−0.004	−0.17	0.24
South Australia*	−0.048	−1.90	−0.037	−1.32	−0.053	−1.86	0.10
Western Australia*	−0.008	−0.32	−0.008	−0.29	−0.010	−0.33	0.11
Queensland*	−0.072	−3.35	−0.045	−1.79	−0.048	−1.89	0.17
Tasmania*	−0.005	0.09	0.043	0.79	0.015	0.27	0.02
ACT*	−0.119	−3.43	−0.087	−2.19	−0.095	−2.40	0.04
No. of observations	4401		4401		4401		
Log Likelihood	-2256.80		-2249.20		-2246.41		
Pseudo R^2	0.159		0.162		0.163		
Test overall significance	854.28	$\sim \chi^2(34)$	869.44	$\sim \chi^2(38)$	874.99	$\sim \chi^2(39)$	

Note that time dummies have been included in estimation but are not reported.

Table 2.1. Neighbourhood effects regression results

more likely to leave school early than younger ones, perhaps reflecting the effects of repeating earlier school years. Teenagers with more brothers and sisters are more likely to leave school early possibly reflecting financial constraints.

Teenagers who do not have English as their first language are significantly more likely to complete high school than teenagers who are born overseas. These effects are independently significant, so that teenagers who were born overseas and did not learn English as their first language are 15.7 percentage points more likely to complete high school, if they judge themselves to speak English well and are 21.4 percentage points more likely to stay on if the were born overseas and have poor English skills. This may reflect different attitudes to education, but may also reflect the relatively poor prospects these teenagers may face in the high unemployment youth labour market. Teenagers who attend a government school are 17 percentage points more likely to leave school early than their counterparts attending a Catholic school and are 12.5 percentage points more likely to leave than teenagers at private schools. Table?

Parents' characteristics are important for explaining the decision to leave school early. Teenagers with fathers who have higher status jobs are less likely to leave high school early. Teenagers from single parent families are significantly less likely to complete high school. Parents with degree qualifications are much more likely to have children who complete high school: degree qualified fathers and mothers decrease the chances of leaving by 5.5 and 9.1 percentage points respectively.

There are initial indications that location has an influence on teenage decision making, as the section of state and state variables are jointly significant. If, however, the neighbourhood has some influence on a teenager's school leaving decision, this will not be captured fully by the variables we have included, and will instead enter the error term. In Figure 2.1, we plot the actual and predicted probabilities of leaving school early against the proportion of the neighbourhood with graduate qualifications. We are interested in whether the difference between these two probabilities varies systematically across neighbourhoods.

To construct this heuristic measure of spatial correlation, we assign individuals to neighbourhoods and then calculate the average neighbourhood drop out rate by smoothing across neighbourhoods ranked according to proportion of neighbourhood with graduate qualifications. The non–parametric smoothing procedure we use is loess (Cleveland (1993)). Loess applies a weighting scheme to individuals on either side of the target individual so that more similar individuals receive higher weights[11]. The final step in the procedure is to use these weights to estimate a weighted least squares regression, centred on each individual in turn, of the outcome variable, eg the observed

[11]The weights are derived from a tricube function:

$$T(u) = \begin{cases} (1 - |u|^3)^3 & \text{for } |u| < 1 \\ 0 & \text{otherwise} \end{cases}$$

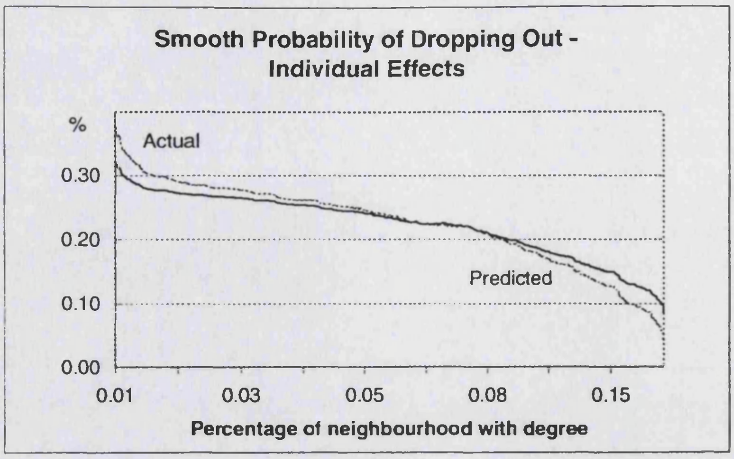


Figure 2.1. Individual effects - predicted versus actual drop out behaviour

decision of whether or not to leave school early, on the variable defining the ranking of the individuals. The smoothed outcome for the centre individual is then calculated as the predicted value of the outcome from the regression.

The underlying rationale, is that neighbourhoods with similar education compositions should show similar drop out behaviour. This will be true, whether or not exogenous neighbourhood effects matter, providing that neighbourhoods with similar proportions of degree holders have similar compositions with respect to other characteristics. This procedure allows us to maintain the full sample as we can estimate actual neighbourhood drop out rates for large and small neighbourhoods when only the zero–one drop out decision is observed. We loose some information on endogenous social effects however, because neighbourhoods with unusually high drop out rates only form part of a weighted average when calculating the predicted and actual drop out rates. Thus, to the extent that endogenous neighbourhood effects are important, the figure will underestimate the systematic nature of the errors. We return to this issue later.

Figure 2.1 shows that a systematic difference between the smoothed actual and predicted probabilities of leaving school early does exist: the individual effects model is under–predicting the probability of being an early school leaver in less educated neighbourhoods and over–predicting this probability for high education neighbourhoods. Although family background explains a large amount of the absolute difference in the probability of being an early school leaver, it is not the whole story[12].

[12] We have tried to account for the unusual shape of the tails by allowing for a non–linear effect of parental background. Children living in the lowest (highest) ranked neighbourhoods are, presumably, more likely to have both parents with very low (high) education levels. To see whether it is this effect driving the tails, we interact the parent secondary dummies and the parent degree dummies. However, standard chi–tests show that both variables are insignificant. Further, including both additional dummies, does not change the shape of the predicted drop out probabilities (perhaps unsurprising, given that both variables are insignificant).

4.3 Reduced form

If we take expectations of Equation 2.1, conditional on the individual's neighbourhood we obtain $E(y^*|x_i)$ as a linear function of $E(z|x_i)$. Substituting, out for $E(y^*|x_i)$, we obtain Equation 2.2.

$$y_i^* = \frac{\alpha}{(1-\delta)} + z_i'\beta + \frac{E(z|x_i)'(\gamma - \delta\beta)}{(1-\delta)} + \varepsilon_i. \tag{2.2}$$

Given personal characteristics and family background variables, we can test for the presence of neighbourhood effects by including average neighbourhood characteristics in the standard probit framework. We will not be able to separately identify the coefficients γ or δ, and therefore, we cannot distinguish between endogenous and exogenous neighbourhood effects in this reduced form specification. The results from estimating Equation 2.2 are presented in columns 3 and 4 of Table 2.1.

In general, the size and significance of the marginal effects of personal characteristics and family background variables do not change noticeably. The variables which are most affected by the presence of the neighbourhood variables are, unsurprisingly, the section–of–state and state variables. The neighbourhood variables indicate that an individual is more likely to leave school early if the proportion of people in the neighbourhood with vocational qualifications is higher, and to a lesser extent, if the neighbourhood unemployment rate is higher.

Figure 2.2 compares the smoothed actual probability of leaving school early with the smoothed probability of leaving school early predicted by the reduced form estimation. Including neighbourhood effects has reduced the systematic error between the predicted and actual probabilities of leaving school early, although the reduced form model is still under–predicting the probability of leaving school early for the low education neighbourhoods. Thus, there appears to be some support for the Crane (1997) hypothesis of epidemic effects in low status neighbourhoods.

There are several possible explanations for the importance of the proportion of the neighbourhood with vocational qualifications. One possible explanation, is that it captures the extent and usefulness of the job information network available to a teenager contemplating leaving school early. To the extent that vocationally trained adults are aware of jobs that offer opportunities to early school leavers, teenagers with access to this network will have higher expected benefits of leaving school early than teenagers in neighbourhoods without a high proportion of vocationally trained adults.

Another related possibility is that the proportion of vocationally trained adults represents the level of local labour demand, and therefore the probability of an early school leaver securing a job. For this effect to be operating, it would be necessary to argue that the proportion of vocationally trained adults is a better proxy for the local demand for unskilled labour than the local unemployment rate. This is not an unreasonable hypothesis, however, the distinction between these two channels cannot

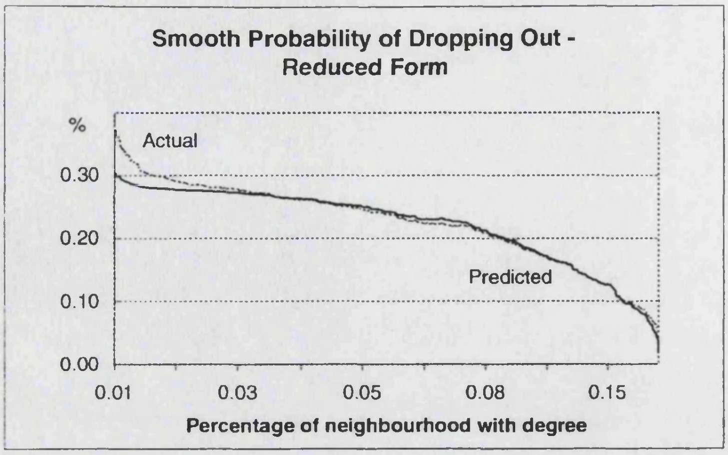


Figure 2.2. Reduced form - predicted versus actual drop out behaviour

be resolved in the current context, especially given potential multicollinearity problems (see below).

The final possibility is that there is a role model effect, similar to the model presented by Struefert (1991). In this model the probability of leaving school early increases as the number of high earning, highly educated role models in the neighbourhood decreases. This model is based on the underlying assumption that the returns to completing high school and undertaking further education are higher than they are for leaving school before the completion of Year 12. This is true in Australia (see Gregory and Vella (1996)). However, Dockery and Norris (1996) present evidence that suggests the returns to completing an apprenticeship are also high in Australia. If a teenager's information set includes a large number of adults receiving relatively high returns on their vocational training, and there are relatively few adults to demonstrate the returns to graduate education, this will naturally bias them towards leaving school early to find an apprenticeship.

4.4 Structural form

Although we have established that neighbourhood effects appear to be present, we cannot determine whether these are exogenous or endogenous effects. If we had sufficient observations in each neighbourhood, we could estimate Equation 2.1 using sample estimates of $E(y^*|x_i)$. Due to the small number of observations per neighbourhood, however, we must use information from individuals in 'similar' neighbourhoods to calculate a sample estimate, $\hat{E}(y^*|x_i)$. Again, we can smooth across neighbourhoods to construct this estimate. Because $\hat{E}(y^*|x_i)$ is calculated from the sample, it is potentially correlated with the error for each individual in that neighbourhood[13]. This is a side

[13]This is independent of the fact that we have smoothed over neighbourhoods, although smoothing lowers the correlation between the endogenous variable and the error term.

effect of the very feedback structure that we are trying to capture.

We try to solve this problem by finding suitable instruments for $\hat{E}(y^*|x_i)$. As always, a good instrument should be correlated with the average probability of drop out in the neighbourhood, but should not affect the individual's decision to drop out. We choose the average number of siblings in the neighbourhood as an instrument, because it should be correlated with the average probability of peers drop out, but should not have an effect on the individual's decision beyond this. Again, the small number of individuals in each neighbourhood suggests that we use a smoothed sample average of the number of siblings in the neighbourhood.

The results of estimating Equation 2.1, instrumenting $\hat{E}(y^*|x_i)$ with the smoothed average number of siblings, are presented in columns 5 and 6 of Table 2.1. The estimated marginal effects of the variables capturing personal characteristics and family background do not change noticeably. The proportion of the neighbourhood with trade qualifications has remained positive and significant and the marginal effect is comparable to that estimated in the reduced form specification. The coefficient on the variable included to capture the endogenous effects, $\hat{E}(y^*|x_i)$, is positive and significant.

However, the estimated marginal effects and significance of the other neighbourhood composition variables change markedly when moving from the reduced to the structural form. The local unemployment rate changes from being marginally significant to being insignificant, and the marginal effect of the proportion of the neighbourhood with graduate qualifications, which had a perverse sign in the reduced form regression, has become larger and significant. To some extent this was to be expected if the reduced form parameters were capturing both neighbourhood effects[14].

Again, we can provide a heuristic check on whether endogenous effects matter, by looking at the difference between actual smoothed probabilities of leaving school and the smoothed probabilities predicted from the structural form model. Figure 2.3 provides further support for the Crane (1997) hypothesis of epidemic effects in low status neighbourhoods because the inherently non–linear nature of the endogenous neighbourhood variable has improved the ability of the model to predict the probability of teenagers leaving school early in low status neighbourhoods. Although this specification has reduced the difference between actual and predicted smoothed probabilities at the low end of the education distribution, these differences have increased slightly at the upper end of the distribution relative to the reduced form specification.

In summary, it appears that significant neighbourhood effects influence a teenager's

[14]When estimating the structural form excluding the proportion of the neighbourhood with graduate qualifications and the local unemployment rate, the endogenous effect variable is estimated to have a marginal effect of 0.4 of a percentage point with a t–statistic of 1.23, and the proportion of the neighbourhood with vocational qualifications has a marginal effect of 0.8 percentage points and remains significant. This confirms the intuition that the significance of the proportion of graduate qualifications in the neighbourhood is spurious, but also makes it difficult to assess the importance of endogenous neighbourhood effects.

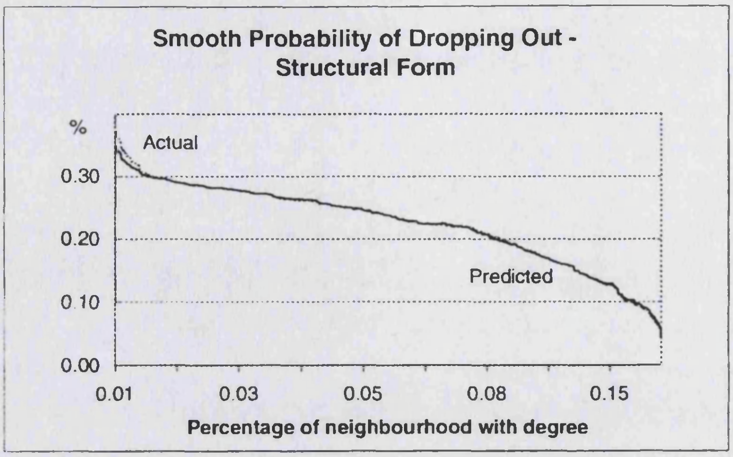


Figure 2.3. Structural form - predicted versus actual drop out behaviour

decision of whether or not to complete high school. Neighbourhood composition affects this decision through the proportion of the neighbourhood with vocational qualifications. There is also some evidence for the presence of endogenous neighbourhood effects. Although multicollinearity problems make it difficult to separate the two effects, the structural form model appears to be better at explaining actual school leaving behaviour over the whole distribution of neighbourhoods than the reduced form model. However, we note that these effects are dominated by personal characteristics and family structure. Before concluding, we briefly consider possible objections to our interpretation of the results as demonstrating the existence of neighbourhood effects.

4.5 Have we really found neighbourhood effects?

One common objection to empirical analysis of neighbourhood effects is that the neighbourhood composition variables may just be picking up omitted individual level variables such as parents' attitudes. There are two responses to this objection. The first is that omitted background variables are more likely to be correlated with the large number of included background variables than with neighbourhood variables.

The second response is that the mechanisms by which such effects are supposed to occur is difficult to specify. Our interpretation can only be affected by an omitted variable, which is positively correlated with the proportion of the neighbourhood with vocational training, as a negatively correlated variable would induce negative bias which serves to strengthen our case. It is difficult to imagine an omitted individual level variable which increases the probability of a teenager leaving school early, and is more highly correlated with the proportion of vocationally trained adults in the neighbourhood than with any individual level variables.

The omitted variable problem is further complicated by the possibility of endogenous sorting. This will arise if there are omitted variables, such as school quality, which directly affect the probability of leaving school early, but also have an indirect effect on neighbourhood composition through the location decisions made by families on the basis of school quality. Thus, the omitted variable will be correlated with the neighbourhood composition variables, which we have treated as exogenous. Again, there are two possible responses.

The first is that we would expect an omitted variable which causes families to sort, to be more correlated with the endogenous neighbourhood effect, which we have instrumented for. Second, to the extent that the unobserved variable affecting the location decision of families is more highly correlated with the proportion of the neighbourhood with graduate qualifications than with the proportion of the neighbourhood with vocational training, we would expect the positive bias to be greater for this variable. This effect is not apparent in the results presented in Table 2.1.

5. Conclusions

This chapter examines the factors that affect a teenager's decision to leave school early. In particular, we consider whether higher rates of early school leaving in some neighbourhoods is the result of 'clustering' of families with characteristics which discourage school completion, or whether the neighbourhood has an independent effect. We find that, although personal characteristics and family background variables explain much of the distribution of early school leaving behaviour across neighbourhoods, these variables are not enough.

We find evidence of significant exogenous neighbourhood effects. Specifically, we find that a larger proportion of vocationally trained adults in the neighbourhood increases the probability of a teenager leaving school early, even when the qualifications of each parent have been controlled for. We suggest that the most plausible explanation for the presence of this effect is that this variable is a proxy for the extent and usefulness of local job information networks or local labour market characteristics, which may affect the balance of costs and benefits to these teenagers of staying on at school. We also find some evidence for the presence of endogenous neighbourhood effects, which arise when the schooling decisions of other teenagers in the neighbourhood affect an individual's decision.

Distinguishing between endogenous and exogenous neighbourhood effects is important, because the policy implications of these two types of effects are quite different. Theoretical results suggest that endogenous feedback mechanisms, such as endogenous neighbourhood effects can lead to multiple equilibria even for initially identical neighbourhoods. One–off expenditures that reduce the rate of drop out may have long run benefits if the endogenous feedback mechanism pushes the neighbourhood to a new equilibrium. In contrast, policies that attempt to affect school decisions by changing neighbourhood composition may have to be ongoing if endogenous sorting in future periods pushes the neighbourhood configuration back to its old equilibrium.

It is also important to bear in mind that while the analysis in this chapter argues that neighbourhood effects influence teenage education decisions, we have only suggested possible mechanisms through which exogenous neighbourhood effects operate. Further research is necessary to identify the exact channel through which these socio–economic effects operate.

3

Neighbourhood Effects in Small Neighbourhoods

1. Introduction

This chapter considers the existence and the scale of neighbourhood effects on the drop out decisions of Australian teenagers. We deal with two related questions. First, does the concentration of poorer families in poor neighbourhoods amplify individual and family effects on drop out tendencies? Second, at what spatial scale might such effects occur? That is, do neighbourhood effects depend on the socio–economic composition of the immediate or the larger locality?

Put simply, neighbourhood effects occur when geographical location matters over and above personal characteristics. That is, when children from otherwise identical families, with identical abilities etc, show different drop out propensities as a function of the type of neighbourhood that they live in. There are various theories that suggest why location may affect socio–economic outcomes. Peer group effects may mean that children in worse neighbourhoods come under greater pressure from peers to drop out of school and engage in other activities. Alternatively, information mechanisms may be important, whereby children in worse neighbourhoods may be unable to correctly assess the returns to education by observing the adults around them. We return to other possible explanations below.

At an aggregate level, it is obvious that worse neighbourhoods have higher drop out rates. However, the fact that individual behaviour appears to be related to neighbourhood characteristics may result from the tendency of families with similar characteristics to live close to each other. Thus, identifying these effects is difficult, because people sort across geographical space according to characteristics that matter for socio–economic outcomes. Sorting across neighbourhoods leads to a *correlation* between neighbourhood characteristics and drop out rates. The neighbourhood effects that we want to capture are ones where neighbourhood economic and demographic characteristics *cause* changes in drop out behaviour. The problem is exacerbated by the fact that we do not know a-priori at what scale these neighbourhood effects may occur. To separate out the effects of sorting, we need information on both individual and neighbourhood characteristics. To analyse the spatial extent of neighbourhood effects we need information on the characteristics of both large neighbourhoods and the smaller neighbourhoods that make up those large neighbourhoods.

In this chapter, we use Australian data to examine the earliest education decision available to Australian teenagers – whether or not to complete high school. We combine the Australian Youth Survey with neighbourhood data derived from the Australian Census to create a data set of individuals with information on personal characteristics, family background and immediate geographical environment. For the entire sample, we can place teenagers in geographical neighbourhoods that roughly correspond to school catchment areas. For a smaller sub-sample, we can identify where the family live within these larger neighbourhoods. We use census data on the socio-economic conditions in both the larger and the smaller neighbourhoods to test for the presence of neighbourhood effects. The data set has a number of key advantages. First, the sample is relatively recent (1989 to 1994). Most other empirical studies use data from far earlier time periods. Second, the neighbourhood data is from the Australian Census conducted in 1991. This means that neighbourhood variables do not have to be constructed from the sample, but are actual population values that reflect the 'true' socio-economic characteristics of the neighbourhood. In addition, these characteristics are measured relatively near the start of the sample period, thus reducing potential endogeneity problems[1].

A rapidly growing empirical literature has considered the existence of neighbourhood effects. Jencks and Mayer (1990) provide a detailed survey of the early literature. Most of this work uses US data. There are relatively few papers which deal with the existence of neighbourhood effects in a public school system. Likewise, there is very little work that looks directly at the issue of the scale at which neighbourhood effects matter.

Many US studies on education outcomes use the 1968 sample of the University of Michigan Panel Study of Income Dynamics (PSID) combined with the 1970 Census Fifth Count for Zip Codes[2]. This provides a sample of young male heads of household who were 23–32 years old in 1978 and who were living with at least one of their parents in one of 188 Standard Metropolitan Statistical Areas in 1968. The neighbourhood data consists of a number of socio-economic indicators recorded by three or five digit zip code. Crane (1991) uses data from the Census Bureau's 1970 Public Use Microdata Samples to test for epidemic effects triggered after some critical level of social problems is reached. For this study, the neighbourhood variables were calculated from data on around 1500 nearest neighbour families. Case and Katz (1991) study the influence of neighbourhoods on the outcome of youths in low income neighbourhoods in inner city

[1]These can arise if families sort across neighbourhoods in response to these characteristics.

[2]See for example Datcher (1982) and Corcoran *et al.* (1992).

Boston. These four frequently cited studies all use different neighbourhood definitions and provide little information on the size characteristics of the neighbourhood that they consider. Our data allows for two neighbourhood definitions which, although based on statistical data collection areas, actually correspond to neighbourhood concepts that may be considered important in determining drop out propensities. This allows us to explicitly consider the scale at which neighbourhood effects may occur.

The chapter is structured as follows. Section 2 describes the data. In particular, we discuss how we link the Australian Youth Survey to the 1991 Australian Census data. We also consider neighbourhood definitions and variable availability. Section 3 sets out our empirical model, and explores the importance of neighbourhood effects for the school leaving decisions of Australian teenagers. Section 4 concludes.

2. Data and definitions

The Australian Youth Survey is compiled by the Australian Department of Employment, Education and Training[3]. The data covers the period from 1989 to 1994. The first wave, sampled in 1989, consists of 5350 sixteen to nineteen year olds. In each subsequent year, roughly 1500 sixteen year olds are interviewed for the first time, and all other panel members are re–interviewed where possible.

The AYS provides detailed geographic information for all respondents. As well as providing information about which state the respondent lives in, and the section of state[4] the respondent lived in before they were fourteen years old, the AYS allows individuals to be located by their geographic neighbourhood in most years. In 1989 the information is recorded by 1986–defined collection districts (CD), which are small neighbourhoods containing, on average, 465 individuals. The postcode where the interview took place is available for re–interviewees in 1991 and all people interviewed from 1992 to 1994. Postcodes are significantly larger than CDs, but there is a mapping from 1986 defined CDs to 1991 defined postcodes. The average postcode has 5558 residents over the age of 15 years. The largest postcode has a population of 62885; the smallest has less than a hundred residents. The distribution is highly skewed with 90 percent of postcodes with fewer than 15131 residents.[5]

Although postcode and CD areas may not correspond exactly to some consistent notion of neighbourhood we still use them to define the neighbourhood of respondents. Postcode areas are viewed as defining some larger neighbourhood, CD areas as defining some smaller neighbourhood. Postcode areas often correspond closely to school catchment areas. CD areas are somewhat more arbitrary, but their small size means that they reflect the immediate geographical neighbourhood well. In addition, using these as our neighbourhood definitions means that we can get very detailed information

[3]Now known as the Department of Employment, Workplace Relations and Small Business.

[4]Section of state is categorised as either capital city, other city, country town or rural area.

[5]Loosely, CDs might correspond to blocks and postcodes to wards.

on a whole range of socio-economic indicators at two different neighbourhood levels. Further, this data does not need to be constructed from the sample, but instead can come from population values obtained through the census. Finally, socio-economic indicators at both the large and small neighbourhood level are likely to be highly correlated with the same indicators for 'correctly' specified neighbourhoods.

We have information on a range of neighbourhood characteristics at both the CD and the postcode level from the 1991 census. This includes information about male and female educational attainment, household and personal income and labour force status. We also have a neighbourhood socio-economic status (SES) variable which was constructed from 1991 census data at the CD level by Hunter (1996). As the name suggests, this variable is constructed to provide an indicator of the socio-economic conditions in a neighbourhood as a function of a number of characteristics including income, labour force status and educational composition. It is thus a neighbourhood equivalent of the individual SES variable that we also have available. Using this socio-economic status index has one key advantage – it captures the combined impact of a variety of neighbourhood characteristics that tend to be highly collinear. When we try to include these variables separately, the collinearity leads to high standard errors on the individual coefficients. Using the socio-economic index helps reduce this multicollinearity problem.[6]

Our sample includes teenagers who were in the final year of high school, or were in the same cohort but left school at an earlier stage. In this sample, the probability of leaving school early is 30 percent, which is consistent with aggregate retention rates over this period. Extensive individual and family background information is collected, including details of educational outcomes and labour market experience for both the respondent and the other members of their household. Unfortunately, parental income is not well measured. Child reported income figures are available, but the response rate is relatively low and the quality of the data is questionable. There is, however, detailed information about the occupational status and education levels of both parents, variables which are likely to be good proxies for income, especially permanent income. These variables are also likely to provide information about parental attitudes to education. Information on other important variables are also available, including the number of siblings and the type of school attended.

The following analysis is restricted to major urban areas, for two reasons. The first is that the ABS introduces sampling error into small postcodes to ensure confidentiality[7]. Second, the type of neighbourhood effects that we may expect in an urban context differ

---

[6]The socio-economic index is constructed using factor scores from principal components analysis. The index is based on several variables: the proportion of the population in Professional, Administrative and Clerical occupations; the proportion of very high income earners; the number of families per house; the proportion of families who own or are purchasing their own home; the percentage of population with various qualifications; and the number of households with more than three cars.

[7]Although the CD data that we have do not suffer from that problem, so we could have reconstructed accurate small postcode data from the weighted average of constituent CDs.

from those that we may expect in a rural context, and our interest lies predominantly with the former.

The large childhood neighbourhood is defined as the postcode where the individual was interviewed when they were 16, as this is the earliest recorded neighbourhood information. The postcode information for 16 year olds is missing in 1991, and these individuals are allocated their 17 year old postcode from the subsequent interview. This is also done for the 17 year olds in 1989 to increase the available sample[8]. Small childhood neighbourhood is defined as the collection district where the individual was living when they were 16. Again, 17 year olds in the 1989 sample were also allocated their childhood CD code. Childhood CD code is only recorded for the first two waves of the sample. The data is missing for all subsequent waves. We are thus left with two samples – an unrestricted sample for whom all postcode data is available; and a restricted sub–sample for whom all CD data is available. Below, we show that the characteristics of the restricted sub–sample are representative of the total sample. In addition, our initial specifications which do not incorporate the CD data allow us to compare the results from the restricted and the unrestricted sample. These results suggest that the restricted sample is representative in terms of behaviour as well. We will return to this issue below.

Childhood postcodes are only defined if the children are living with one or both or their parents. This is standard practice in the literature, but may cause biases if the decision to move out from the family home is a function of the endogenous variable. Thus, if children who drop out are more likely to be living away from home, then we under sample this group of respondents. We also exclude respondents who reported that they had spent most of their life until 14 overseas. In addition we exclude respondents that are married. Neither of these sample restrictions changes the results in any fundamental way.

2.1 How representative is the CD sub–sample?

When we want to consider the importance of small neighbourhood effects we need to restrict the sample, as CD information is only available at the start of this period. Table 3.1 compares the characteristics of this restricted sample with the unrestricted sample. The table gives mean values for a number of key variables. The table suggests that the restricted sample is representative of the total sample. In Sections 3.1 and 3.2 will see that the sub–sample also appears to be representative in terms of drop out behaviour.

[8]Excluding these individuals does not change the results, although it does reduce the accuracy of some point estimates.

	Full sample	Restricted sample
Male	49.7%	52.4%
Number of siblings	2.0	2.0
English good	10.0%	10.1%
English poor	1.0%	1.0%
Catholic school	22.7%	23.1%
Other non-government	9.4%	10.2%
Father's occ. status	29.7	30.7
Mother's occ. status	19.6	19.3
Father not present	15.5%	14.7%
Mother not present	5.0%	4.7%
Father degree	16.6%	17.4%
Father trade qualifications	16.9%	16.5%
Mother degree	12.8%	13.0%
Mother trade qualifications	4.2%	4.2%
Neighbourhood personal income	5601.5	5629.01
Neighbourhood percentage vocational	13.6%	13.3%
Neighbourhood unemployment rate	11.3%	11.5%

Table 3.1. CD sub–sample characteristics

3. Empirical model and results

We want to estimate the effects of two different types of neighbourhood on education decisions. In particular, we will consider the decision on whether or not to complete high school – legally, the first free education decision available to Australian teenagers. In Section 3.1 we present the results of estimation that ignores neighbourhood effects. As outlined above, we do this for both the unrestricted sample and the restricted sample. In Section 3.2 we add neighbourhood variables and discuss the importance of both small and large neighbourhood effects for understanding teenage education decisions. Finally, in Section 3.3 we check the robustness of our small neighbourhood results using a fixed effects logit specification. Effectively, this specification uses large neighbourhood dummies, rather than specific characteristics, to capture the large neighbourhood effects.

3.1 Individual effects

We start by estimating the model assuming that there are no neighbourhood effects. Thus, our basic specification is:

$$y_i^* = \alpha + z_i'\beta + \varepsilon_i; \tag{3.1}$$

where y_i^* is the underlying propensity to leave school before the final year of high school for individual i; z_i are the personal background and family characteristics of individual i; and ε_i is the (normally distributed) error term which contains all the unobserved factors which affect individual i's propensity to leave school before the final year.

Because the observed variable is the zero–one drop out decision, rather than the underlying probability we estimate a probit model. As always, the magnitude of the effects of each variable depend on where the probability is evaluated. We present the results in marginal effects form so the coefficients give the impact of a one unit change in the variable, given that the probability is initially evaluated at the sample mean. For dummy variables, marked with an asterisk, the reported marginal effect will be the change in the probability of drop out if the individual has that characteristic rather than the omitted characteristic. Data appendix A gives definitions of variables and specifies the omitted categories for each group of dummy variables. The results for the individual effects specification for the restricted and the total sample are presented in the first two columns of Table 3.2. Column 1 gives full sample results, column 2 restricted sample results.

We briefly discuss the outcomes for the full sample, before considering the differences between the samples. For the full sample, males are 8 percentage points more likely to drop out than females. Teenagers with more brothers and sisters are more likely to leave school early. Teenagers without English as a first language are significantly more likely to complete high school than teenagers who are born overseas[9]. Teenagers who attend a government school are 17 percentage points more likely to leave school early than their counterparts attending a Catholic school and are 12.5 percentage points more likely to leave than teenagers at other non-government schools. A number of parental characteristics are important for explaining the propensity to drop out. High occupational status for fathers has a positive effect on the probability of staying on, as does having a mother or a father with a degree. Teenagers from single parent families are much more likely to drop out – particularly if it is their mother who is not present in the household.

Table?

Turning now to the restricted sample, we see that the results are broadly comparable. Only two coefficients change sign, and both are insignificant in both the restricted and the full sample estimation. The standard errors of point estimates are increased in the smaller sample and some variables that were significant become insignificant. Most noticeable among these is that the father degree and English as a foreign language variables are no longer significant. However, all other background variables and parental characteristics remain significant with the same sign. The behaviour in the restricted sample would appear to be representative of the total sample.

3.2 Neighbourhood effects

We start by considering the inclusion of large neighbourhood effects. We then consider the inclusion of small neighbourhood effects. Introducing them in this order allows us

[9] Mainly British immigrants.

	Individual		Neighbourhoods			
	(1)	(2)	(3)	(4)	(5)	(6)
Large Neighbourhood						
Average personal income			0.010	−0.002	0.001	
Proportion trade qual.			0.013**	0.015**	0.026**	
Unemployment rate			0.005	0.008	~~0.409~~ 0.049	
Small Neighbourhood						
Average personal income					0.005	0.004
Proportion trade qual.					−0.010**	0.010
SES					−0.008**	−0.006
Personal Background						
Male*	0.080**	0.103**	0.081**	0.102**	0.100**	0.406**
Age	0.093**	0.034	0.093**	0.034	0.035	0.280*
Number of siblings	0.009*	0.025**	0.009*	0.025**	0.025**	0.108**
English not first language						
English good*	−0.080**	−0.046	−0.072**	−0.038	−0.040	−0.377
English poor*	−0.137**	−0.160	−0.129*	−0.153	−0.149	−0.409
Born overseas*	−0.077**	−0.102**	−0.076**	−0.103**	−0.106**	−0.600**
School						
Catholic*	−0.125**	−0.156**	−0.120**	−0.149**	−0.147**	−0.776**
Other non–government*	−0.170**	−0.153**	−0.161**	−0.144**	−0.142**	−0.597**
Parent Characteristics						
Father's occ. status @14	−0.002**	−0.002**	−0.002**	−0.002**	−0.002**	−0.010**
Mother's occ. status @14	−0.0001	0.0002	−0.0002	−0.0002	−0.0002	−0.001
Father not emp @14*	−0.017	−0.030	−0.016	−0.028	−0.031	−0.144
Mother not emp @14*	0.004	−0.022	−0.003	−0.025	−0.230	−0.237
Father not present @14*	0.059**	0.072	0.061**	0.076*	0.072	0.192
Mother not present @14*	0.443**	0.427**	0.445**	0.427**	0.426**	1.999**
Father has:						
degree*	−0.055**	−0.016	−0.052**	−0.011	−0.010	−0.136
trade qualifications*	0.020	0.002	0.016	−0.003	0.002	−0.131
other post-secondary*	−0.039	−0.044	0.045	−0.051	−0.045	−0.220
Mother has:						
degree*	−0.091**	−0.146**	−0.083**	−0.138**	0.122**	−0.687**
trade qualifications*	0.019	−0.005	0.017	−0.004	−0.010	−0.004
other post-secondary*	−0.025	−0.034	−0.025	−0.036	−0.035	−0.298
State						
Victoria	−0.025	−0.074**	0.001	−0.037	−0.024	
South Australia	−0.048*	−0.130**	−0.037	−0.104**	−0.090*	
Western Australia	−0.008	−0.074	−0.008	−0.073	−0.046	
Queensland	−0.072**	−0.143**	−0.045*	−0.100**	−0.088*	
Tasmania	0.005	−0.093	0.043	−0.029	−0.020	
ACT	−0.119**	−0.251**	−0.087**	−0.218**	−0.198**	
Number of obs.	4401	1654	4401	1654	1654	1372
Overall significance	854.2	256.5	869.35	262.5	269.7	159.6
	$\sim\chi(34)$	$\sim\chi(32)$	$\sim\chi(37)$	$\sim\chi(35)$	$\sim\chi(37)$	$\sim\chi(26)$

?

Pseudo R^2

Note that time dummies and section of state dummies have been included in the estimation but are not reported. (*) indicates significance at the 10% level; (**) indicates significant at the 5% level. Average personal income is in '000s of dollars. Column (1) reports the same results as column (1) in Table 2.1

Table 3.2. Small & large neighbourhood regression results

to check that the smaller restricted sample is, again, representative in terms of both characteristics and behaviour. The equation that we estimate is now:

$$y_i^* = \alpha + z_i'\beta + E(z|X_i)'\gamma + \varepsilon_i; \tag{3.2}$$

where $E(z|X_i)$ are the average characteristics of the individuals in the large neighbourhood. Remaining notation is as for Equation 3.1. The third and fourth columns of Table 3.2 show the results when we include a number of neighbourhood variables. Column 3 gives the full sample results, column 4 the restricted sample results.

Choosing which neighbourhood variables to include is not easy as, a priori, all of them may have an important influence on drop out effect. However, unsurprisingly, collinearity problems dominate when all of the possible neighbourhood variables are included. Our empirical approach was to start with a large number of neighbourhood variables and test down to a more parsimonious representation. Initially we included the following variables separately for male and females: proportion of neighbourhood with a higher degree; proportion of neighbourhood with a degree; proportion of neighbourhood with various types of diploma: proportion of neighbourhood with skilled vocational qualifications; proportion of neighbourhood with basic vocational qualifications; proportion of neighbourhood with no qualifications; neighbourhood unemployment rate; personal income. We then tested to see if we could combine the male and female variables - we could never reject the hypothesis that the coefficients were the same. Next, we combined the degree qualifications, the vocational qualifications and the other post–secondary qualifications. Again, we could never reject the hypothesis that the coefficients on the sets of variables were the same. We then dropped the no qualifications and the other post secondary qualifications variables which were consistently very insignificant. This left us with percentage of neighbourhood with a graduate qualification; percentage of neighbourhood with a vocational qualification; neighbourhood unemployment rate and average neighbourhood personal income. Both average neighbourhood personal income and percentage with graduate qualification were insignificant, but would appear to be highly collinear. In the end, we present results after dropping the graduate qualification variable. Results are comparable if we drop the neighbourhood income variable instead.

As can be seen from Table 3.2, only one of the neighbourhood variables is significant - proportion of neighbourhood with vocational qualifications is significant at the 1% level. Average neighbourhood personal income is insignificant and neighbourhood unemployment rate is (just) insignificant. The results are somewhat surprising[10]. Our personal prior was that neighbourhood effects would operate through concentrations of either high educated or low educated adults, or through income. The fact that they appear to work through the proportion of adults with vocational qualifications has two interesting interpretations.

Check 5% level

[10]See also Chapter 2.

First, this could reflect the importance of job networks as emphasised by, for example, Montgomery (1991). Young people in these neighbourhoods have access to a larger social network of people that can get them in to jobs where schooling qualifications are not necessarily required. Informational effects may play an additional role – when young people assess the returns to formal education they may use people in their own neighbourhoods to inform that decision. A high proportion of vocationally qualified individuals earning a living from jobs that de–emphasise formal learning may lead to young people forming different opinions about the value of that formal education. The second interpretation is a more classical local labour market interpretation. High concentrations of vocationally qualified adults may indicate neighbourhoods with local labour markets where unskilled job opportunities are more readily available. Given the local nature of labour markets, it may be attractive for children to drop out in neighbourhoods where there are greater job opportunities for unskilled labour. These two channels may obviously interact – school drop outs may find it easier to get connected in to the local labour market when they know a high proportion of adults who work in that market. Notice that the coefficient on neighbourhood unemployment, although (marginally) insignificant, points to somewhat more 'negative' neighbourhood effects. Neighbourhoods with high unemployment rates tend to see higher drop out. This effect is presumably not a result of teenagers dropping out to work in the local labour market (where unemployment is high), but reflects negative feedbacks whereby, for example, a culture of high unemployment leads to high drop out rates and even higher local unemployment. We return to this issue below.

Before introducing small neighbourhood effects variables, we can again compare the results from the restricted sample to those from the unrestricted. From Table 3.2, column 4, we see that the neighbourhood effects, for significant variables, are almost identical. The differences between the individual, family and state effects remain as before. This suggests that, with the exceptions mentioned in Section 3.1, the restricted sample behaviour is representative of the overall sample, particularly when it comes to neighbourhood effects. Table 3.1 reinforces this impression. We see that in terms of neighbourhood characteristics, the restricted sample is highly representative.

We now introduce small neighbourhood effects for the restricted sample. Thus, the equation we now estimate is:

$$y_i^* = \alpha + z_i'\beta + E(z|X_i)'\gamma + E(z|x_i)'\delta + \varepsilon_i; \tag{3.3}$$

where $E(z|x_i)$ are the average characteristics of the individuals in the small neighbourhood. Remaining notation is as for Equation 3.2. Again, we test down from a much broader specification. This time, the process is helped because we have a neighbourhood SES variable, which captures income, occupation and employment characteristics allowing us to use this variable to avoid some of our earlier multicollinearity problems. After testing down, we are left with small neighbourhood variables that are very similar to the large neighbourhood variables. We have CD SES rather than personal income or graduate qualifications, percentage of the CD with vocational qualifications and the unemployment rate of the CD. Even in this parsimonious representation, the unemployment rate remains highly insignificant – so we drop this variable which leaves us with the specification reported in column 5 of Table 3.2.

The results are interesting, and highly informative with respect to the interpretations of the possible neighbourhood effects that we outlined above. First, notice that the significance and the sign of the coefficients on the large neighbourhood variables are unchanged. Second, both small neighbourhood SES and small neighbourhood proportion vocational are significant and have a negative effect on school drop out rate. Consider the negative effect of the proportion vocational education in the small neighbourhood. This suggests that the more classical local labour market interpretation may well be the correct one. A large neighbourhood with a high percentage of vocational educated adults proxies for high local demand for (complementary) unskilled labour. High local labour market demand for unskilled labour alters the incentives to drop out and drop out rates rise accordingly. However, conditional on that, a high concentration of vocational qualified adults in the smaller neighbourhood reduces the drop out rate. Informational networks would appear to play a small part in the effect of vocationally qualified adults on drop out. In fact, a high proportion of vocationally qualified adults in the small neighbourhood would appear to encourage students to stay on at school – possibly so that they can move on to more vocational training. At the same time, a low SES score in the small neighbourhood has a significant impact on school drop out rates. This suggests that there are small clusters of low SES families with much higher drop out rates than we would predict given family background and personal characteristics. Negative neighbourhood feedbacks would appear to occur at the small neighbourhood level acting through the socio–economic composition of that small neighbourhood.

3.3 Fixed effects estimation

In Section 3.2 we tested for the presence of small neighbourhood effects after conditioning on a number of large neighbourhood characteristics. A stronger test for the presence of small neighbourhood effects would involve conditioning out all of the variation in drop out probabilities that may possibly be due to large neighbourhood effects. In the specification in Section 3.2 neighbourhood effects work through neighbourhood average personal income, the proportion of adults with trade qualifications and the neighbourhood unemployment rate. In this section, we want to replace these variables with neighbourhood dummies, so that the dummies capture any difference in average drop out rates between large neighbourhoods, no matter what the cause. To do this, we would need to introduce individual neighbourhood effects to our specification which would condition out the average drop out propensity in the large neighbourhood, leaving small neighbourhood effects to explain the variation within those neighbourhoods.

For discrete dependent models, the choice between fixed and random effects models is somewhat constrained. Introducing fixed effects in to the standard probit specification is problematic. There is no feasible way to remove the heterogeneity from the nonlinear structure (by differencing for example) and with large numbers of cross-sectional units, estimation of the individual neighbourhood dummy coefficients is intractable. Some progress has been made on a probit specification incorporating random effects. However, as with standard linear formulations, we need to assume that the individual random effects are uncorrelated with the other regressors. If neighbourhood effects do occur at the small neighbourhood level, then this assumption is clearly unlikely to hold, precisely because each large neighbourhood is formed from a collection of small neighbourhoods.

Instead, we use a (Chamberlain) conditional (fixed effects) logit specification[11]. This specification represents the simplest way of conditioning out large neighbourhood heterogeneity. The basic idea is to consider the conditional likelihood function, where the likelihood for each set of neighbourhood observations is conditioned on the number of 1s in the neighbourhood. The fixed effects logit specification is:

$$\text{Prob}(y_{ij} = 1) = \frac{e^{\alpha_i + \beta' x_{ij}}}{1 + e^{\alpha_i + \beta' x_{ij}}}; \tag{3.4}$$

where α_i is the fixed effect for large neighbourhood i, and the x_{ij} are the characteristics of individuals that vary within large neighbourhoods. These characteristics include small neighbourhood characteristics as well as background variables.

The conditional likelihood is

$$L^c = \prod_{i=1}^{N'} \text{Prob}\left(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, \ldots, Y_{iN} = y_{iN} \middle| \sum_{n=1}^{N} y_{iN}\right), \tag{3.5}$$

where N' is the number of neighbourhoods and N is the number of teenagers within each neighbourhood.

We must drop two types of neighbourhoods when implementing this procedure. The first are large neighbourhoods with only one observation. The second are large neighbourhoods where behaviour is uniform. That is, large neighbourhoods where everyone drops out, or where everyone stays on. Neighbourhoods with uniform behaviour do not contribute to the conditional likelihood function. We also drop variables that do not vary within groups[12]. See Greene (1997) for more details on implementing the conditional logit model. The restrictions leave us with a sample of 1372. The results for this fixed effects logit specification are reported in Column 6 of Table 3.2.

[11] See Chamberlain (1980).

[12] Specifically the state dummies.

Unfortunately, parameter values are not directly comparable across the two different specifications. However, we can see that the sign and significance of variables remains unchanged, suggesting that our small neighbourhood results are robust to conditioning out all large neighbourhood effects whatever their cause. Most of the individual and family variables have the same sign and significance as the probit results suggesting that they are robust to alternative specifications of the large neighbourhood effects. All of the small neighbourhood variables are insignificant. However, the SES small neighbourhood variable is only just insignificant[13]. Given the reduction in sample size, and the small number of observations within each large neighbourhood, we were surprised that any of the small neighbourhood variables were close to being significant. The fixed effects logit results suggest that the trade qualification result is not robust. However, the small neighbourhood SES result is quite robust to a very general specification of the large neighbourhood effect. Living in an area where the immediate neighbourhood has low socioeconomic status has a negative effect on drop out propensities.

4. Conclusions

We have tested for the presence of both small and large neighbourhood effects on the drop out rate of Australian teenagers. Two neighbourhood effects appear to operate. The first works at the large neighbourhood level through the proportion of the adult population with vocational education. A high proportion of vocationally trained adults leads to a higher drop out rate. This would appear to be consistent with two possible mechanisms – one working through local labour market demand, the other through social networks. The results for the small neighbourhood variables suggest that the former is the most likely channel. A high proportion of adults with vocational qualifications in a small neighbourhood reduces drop out probability. This suggests that the high drop out rates associated with high concentrations of vocationally qualified adults reflect local labour market conditions. We have shown that the SES result for small neighbourhoods is quite robust to conditioning out all large neighbourhood effects.

We have also found that the socio–economic status of small neighbourhoods matters for drop out rate. The channels through which this variable might operate are presumably those identified by Wilson (1995), Akerlof (1997) and others. Such channels include effects on the assessment of returns to education, the importance of social networks and the influence of peer–group pressure.

Our results here do not allow us to separate out the channels through which small neighbourhood socio–economic status influences drop out rates. Information on average neighbourhood drop out rates might allow us to do this, but such information is not available and only a very poor proxy can be constructed from the data given the number

[13]It is actually significant at the 11% level.

of observations in each small neighbourhood. Our results are clearer on the channel through which the structure of large neighbourhoods impact on drop out rates.

The policy implications of these results are interesting. First, the fact that large neighbourhood effects seem to operate through the structure of local labour market demand rather than through some other neighbourhood mechanism suggests that high drop out rates may sometimes be a rational response to perceived local labour market conditions. This suggests that local employers of large numbers of unskilled workers may need to play an important role if governments wish to reduce drop out rates in certain neighbourhoods. Second, the fact that small neighbourhood effects exist, and seem to operate through the socio–economic status of the neighbourhood suggests that government policies placing small clusters of low SES families in better neighbourhoods may have little significant impact on drop out rates. 'Forced' mixing through government housing programs may need to ensure that low SES families are well dispersed throughout more affluent neighbourhoods, rather than concentrated in 'sink' estates. Refining the policy implications will involve separating out the mechanisms through which the effects operate. This identification is left to further work.

4

Unemployment clusters across European regions and countries

with Diego Puga

1. Introduction

When we think about differences in unemployment rates across Europe, we normally think of differences across countries as represented in Figure 4.1. This is a useful starting point that leads naturally to trying to understand, for instance, why the average unemployment rate of Spain is so much higher than that of Portugal[1]. However, the national averages represented in Figure 4.1 hide large differences in unemployment rates across regions *within* countries. The case of Italy is best known, with Campania having a 1996 unemployment rate 4.4 times as high as Valle d'Aosta. But large regional differences exist in all European countries. In the United Kingdom, Merseyside has an unemployment rate 3.2 times that of the Surrey-Sussex region; in Belgium, the unemployment rate of Hainut is 2.2 times that of Vlaams Brabant; in Spain, Andalucía has an unemployment rate 1.8 times that of La Rioja; in France, Languedoc-Roussillon has a rate twice that of Alsace; and so on.

The map at the top of Figure 4.2 plots regional unemployment rates for the contiguous European Community of 1986 (more details on the regional coverage are given below). While the map is drawn for 1986, the regional distribution would look very similar for earlier years[2]. In the decade up to the mid 1980s, the average European unemployment rate was rising. However, differences in unemployment rates across European regions were very stable, with regional labour forces adjusting just enough to offset ongoing changes in regional employment (see chapter 6 in Layard, Nickell, and Jackman, 1991). The map at the bottom of Figure 4.2 suggests that something has

[1]On this respect, see Blanchard and Jimeno (1995); Bover, García-Perea, and Portugal (1998); Castillo, Dolado, and Jimeno (1998*a*,*b*).

[2]Unfortunately, only a more limited regional coverage is available before 1986.

1986

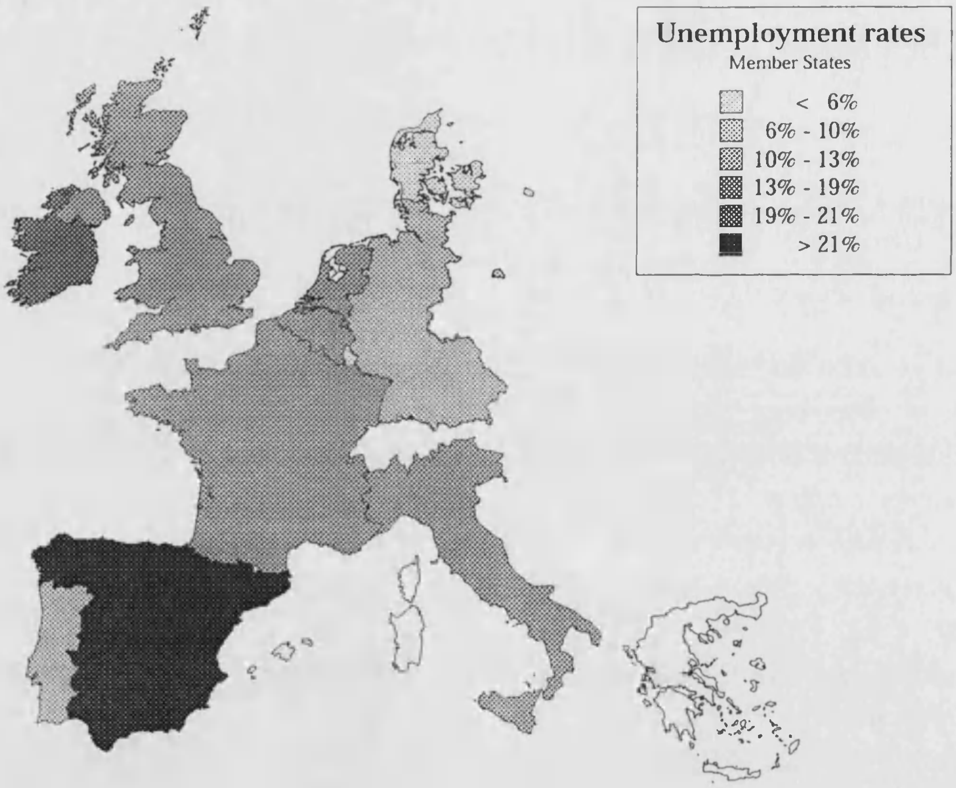


Figure 4.1. National unemployment rates in Europe

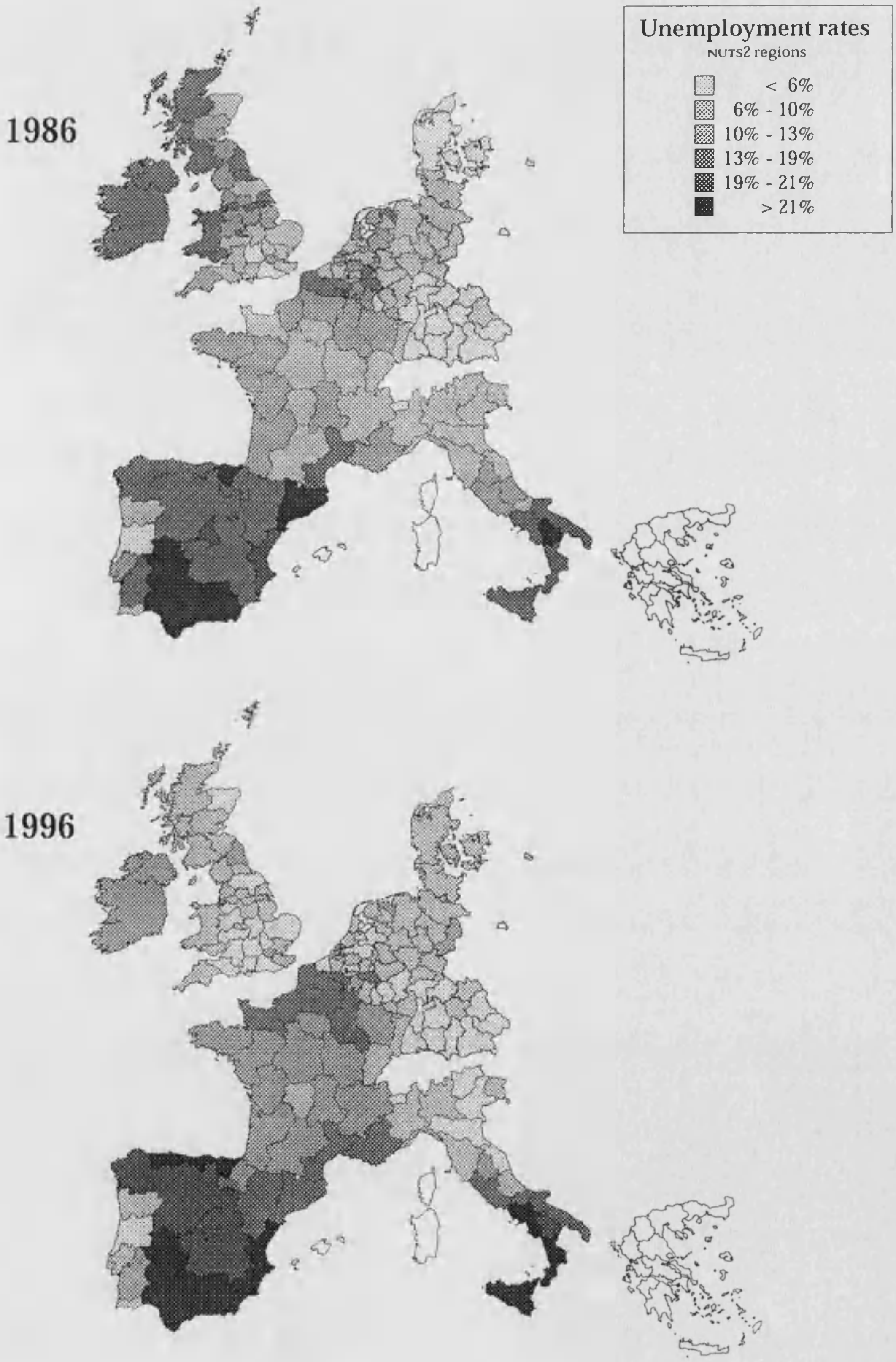


Figure 4.2. Regional unemployment rates in Europe

changed over the last decade, and that the stability described by Layard *et al.* (1991) up to the mid 1980s no longer holds. The average unemployment rate for regions in these maps was the same, 10.7%, in 1996 as in 1986, and the decade separating them could be thought of as covering a full cycle in unemployment rates[3]. Yet the map for 1996 looks different enough from that for 1986, that one starts to wonder what has happened to the distribution of European regional unemployment rates over this period. The answer to that question is the starting point of this chapter.

We begin by showing that, during the decade from 1986 to 1996, there has been a polarisation of unemployment rates across the regions of the European Union (EU). To go beyond the limited conclusions that can be drawn from comparing summary statistics over time, Section 2 looks at the evolution of the shape of the whole distribution of European unemployment rates. We also track the outcomes of individual regions. Regions that in 1986 had a low unemployment rate relative to the EU average still tended have a relatively low unemployment rate in 1996. Similarly, regions that in 1986 had a relatively high unemployment rate still tended have a relatively high unemployment rate in 1996. However, regions with intermediate initial unemployment rates had mixed fortunes. Some saw a marked fall in their relative unemployment rate, while others saw it rise, and still others saw it roughly unchanged.

We show that this process has been driven by changes in regional employment rather than by changes in demographic structure or labour market participation. There has been some labour force adjustment to regional employment changes. Regions with relatively low unemployment rates have typically experienced above average labour force growth, while regions with relatively high unemployment rates have generally experienced a below average increase, or a fall, in their labour force. However, this adjustment has been insufficient to prevent the polarisation of European unemployment rates.

What factors might be driving this polarisation? The simplest explanation would be that some countries have managed to sort out their unemployment problems, while others have not. However, other characteristics of regions may also matter. Regions differ in the sectoral composition of their employment; in the age, sex and skill structure of their populations; and in their geographical location within the EU. Regions initially specialised in agriculture or manufacturing may have seen their unemployment rates rise as the EU production structure moves away from those sectors. Similarly, regions with a high proportion of low skilled workers may have seen their unemployment rates rise as production shifts from low skilled to high skilled employment. Other changes to the EU production structure may be equally as important, but have received much less attention. Over the last decade, the Member States of the EU have pushed ahead with

[3]The average European unemployment rate in 1986 (for regions belonging to what was then the European Economic Community) was 10.7%, starting to come down from a peak of 10.8% one year before that. It kept coming steadily down to 8.1% in 1990, and then steadily up to a new peak of 11% in 1994, after which it fell back to its 1986 rate of 10.7% in 1996.

ever closer economic integration. Recent theoretical developments suggest that such a process can be associated with the emergence of spatial concentrations of employment, and that with falling barriers to trade these may extend across national borders. If regional labour forces do not fully adjust to such employment changes, then geographical location may be important in explaining the increased polarisation of unemployment rates.

We use two complementary techniques, one parametric, one nonparametric, to examine these alternative explanations. The nonparametric technique involves grouping regions by some common characteristic (like State Membership, or similar skill composition) and then examining the similarity of unemployment outcomes within groups. This technique has the distinct advantage that it allows for different regional characteristics to matter to different degrees for different parts of the distribution. Its main disadvantage is that it only allows one to consider a single factor at a time. To ensure that our results are robust in this respect, we finish with a more standard parametric analysis. This also allows us to consider the importance of cross border effects.

Both the parametric and nonparametric techniques show that regions' unemployment outcomes have closely followed those of neighbouring regions. This is only weakly explained by regions being part of the same Member State, having a similar skill composition, or broad sectoral specialisation. Remarkably, we find that neighbouring regions across national borders are as important as domestic neighbours in determining unemployment outcomes. The clusters of high and low unemployment that have emerged over the last decade show little respect for national borders.

2. The evolution of the distribution of unemployment rates

As the data to be studied we take *Europe relative* unemployment rates from 1986 to 1996. The Europe relative unemployment rate is defined as the ratio of the regional unemployment rate to the European wide average unemployment rate. Working with relative, as opposed to absolute unemployment rates, helps remove co-movements due to the European wide business cycle and trends in the average unemployment rate. As mentioned in the Introduction, the average European unemployment rate was the same in 1996 as in 1986, 10.7%, and the decade in between can be regarded as covering a full cycle.

The unemployment rate series are computed from the harmonised unemployment rates and labour force data contained in the Regio database produced by Eurostat (Eurostat, 1998). These data are based on the results of the Community Labour Force Survey, carried out in Spring each year.

The analysis focus on the contiguous European Community of 1986. That is, those regions of the EU that satisfy the following three criteria:

1. Have been part of the EU (European Economic Community before 1 November 1993) from 1986 to 1996.

2. Are in a Member State which has a land border with at least one other Member State containing at least one region satisfying (1).

3. Have a land border with at least one other region satisfying (1) and (2).

The definition of regions corresponds to level two of the Nomenclature of Territorial Units for Statistics (NUTS2), a hierarchical classification with three regional levels established by Eurostat to provide comparable regional breakdowns of EU Member States. There are 150 NUTS2 regions satisfying criteria (1) to (3) above. The average NUTS2 region in our data set had a land area of 13,800 square kilometres and a population of 2.1 million in 1996 (that is slightly larger than the US State of Connecticut and with two thirds of its population).

Data Appendix B gives full details of the regional coverage and data sources.

2.1 The shape of the distribution

What has happened to the distribution of regional unemployment rates over the decade beginning in 1986? One way to answer this question would be to compare summary statistics of the distribution of regional unemployment rates across time. For instance, the Theil index for the distribution of regional unemployment rates increased from 0.10 in 1986, to 0.13 in 1996. However, such an exercise gives at best limited conclusions (as a recent radio broadcast on behalf of Ontario's teachers put it 'averages, like promises, don't mean much'). Instead, we consider the evolution of the entire distribution. Figure 4.3 plots a sequence of kernel estimates of the density of Europe relative unemployment rates for four years: 1986, 1989, 1993, and 1996[4]. The density plots can be interpreted as the continuous equivalent of a histogram, in which the number of intervals has been let tend to infinity and then to the continuum. By definition of the data, 1 on the horizontal axis indicates the European average unemployment rate, 2 indicates twice the average, and so on.

Two features are particularly noticeable in Figure 4.3. First, as we move through the decade, the distribution of unemployment rates for a majority of regions becomes more concentrated below the European average: the peak of the distribution, close to the average in 1986, moves slightly leftwards and the mass becomes more narrowly concentrated around that peak. Second, there is a growing group of regions with unemployment rates above twice the European average: these regions produce the 'bulge' in the upper tail of the distribution — to see this most clearly, contrast the

[4] All densities are calculated nonparametrically using a Gaussian Kernel with bandwidth set as per Section 3.4.2 of Silverman (1986). The range is restricted to the positive interval using the reflection method proposed in Silverman (1986). Calculations were performed with Danny Quah's tSrF econometric shell (available from `http://econ.lse.ac.uk/~dquah/`).

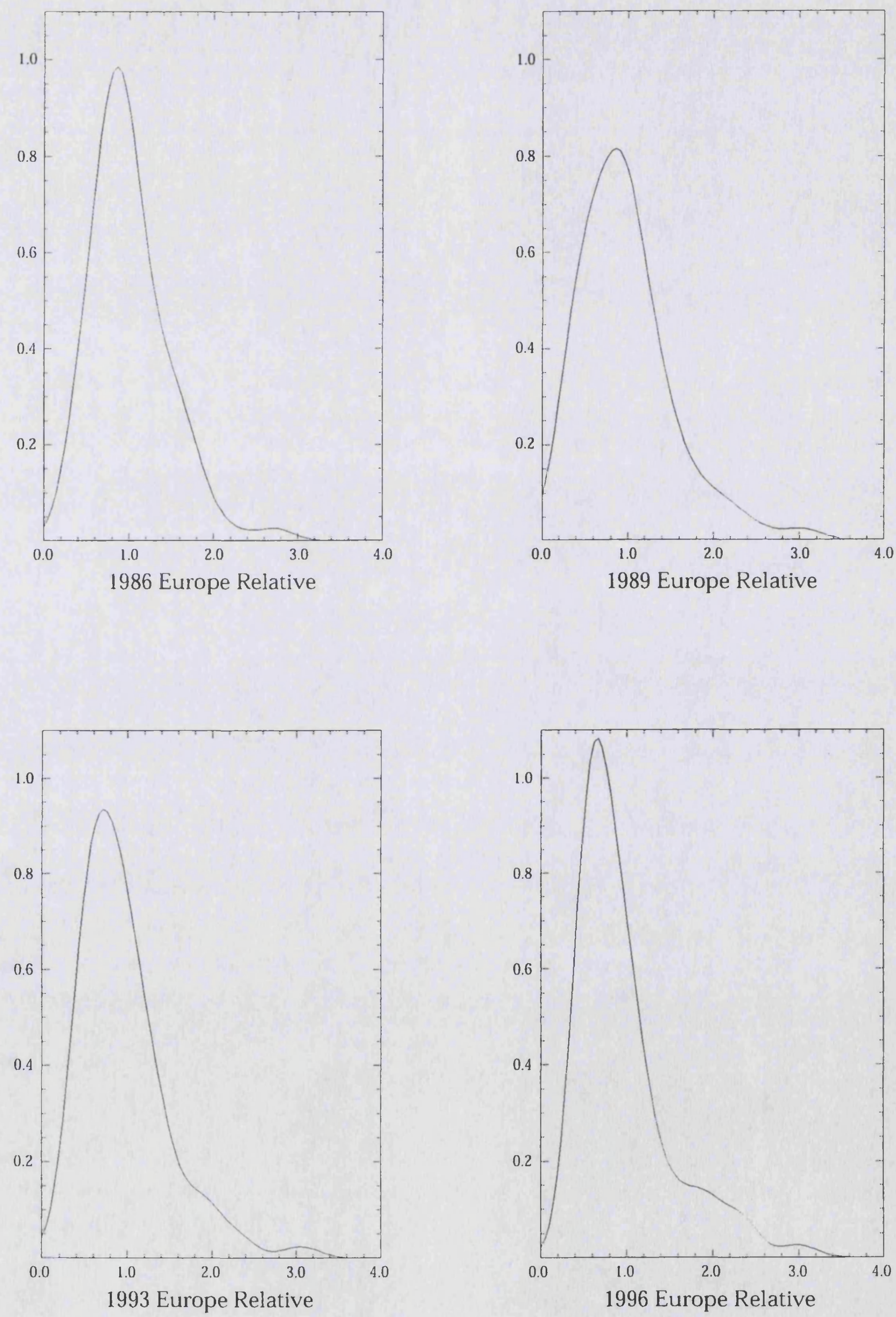


Figure 4.3. Densities of Europe relative unemployment rates

1986 Europe Relative		1996 Europe Relative				
n		[0-0.6)	[0.6-0.75)	[0.75-1)	[1-1.3)	[1.3-∞)
32	[1.3-∞)	0.00	0.00	0.16	0.22	0.62
32	[1-1.3)	0.06	0.22	0.34	0.19	0.19
42	[0.75-1)	0.24	0.29	0.26	0.21	0.00
23	[0.6-0.75)	0.52	0.26	0.09	0.09	0.04
21	[0-0.6)	0.81	0.19	0.00	0.00	0.00

Table 4.1. 1986 to 1996 Europe relative transition probability matrix

mass above twice the European average unemployment rate in 1986 and 1996. Looking through the four snapshots we see that these two features have slowly evolved over the decade. Therefore, over time more regions have unemployment rates below the European average, or above twice that average, and less regions have unemployment rates between the average and twice the average.

2.2 Mobility and persistence

The density plots are suggestive of a gradual polarisation of European regional unemployment rates. However, this interpretation cannot be supported by the density plots alone. The collection of densities tell us nothing about the identity of regions in the distribution of regional unemployment rates. Is it true that a group of low unemployment regions and a group of high unemployment regions has slowly emerged, while regions with intermediate unemployment rates have moved closer to the tails of the distribution? Certainly, more regions had low or high unemployment rates in 1996 than in 1986, but what was their relative position in previous years? Does this collection of snapshots actually just show churning of the unemployment rate distribution, the random ups and downs of regional fortunes, or are they the result of a more structured process?

The natural way to answer these questions is to track the evolution of each region's relative unemployment rate over time. An easy way to do this is to construct transition probability matrices. For a discrete stochastic process with an integral number of possible outcomes or states, each row of this matrix takes a given state and shows the probability of transiting to any other state. Constructing a transition probability matrix for a continuous variable requires a discretisation of the space of possible outcomes.

Table 4.1 does this with the space of relative unemployment rates, to construct the transition probability matrix between the 1986 and 1996 distributions of Europe

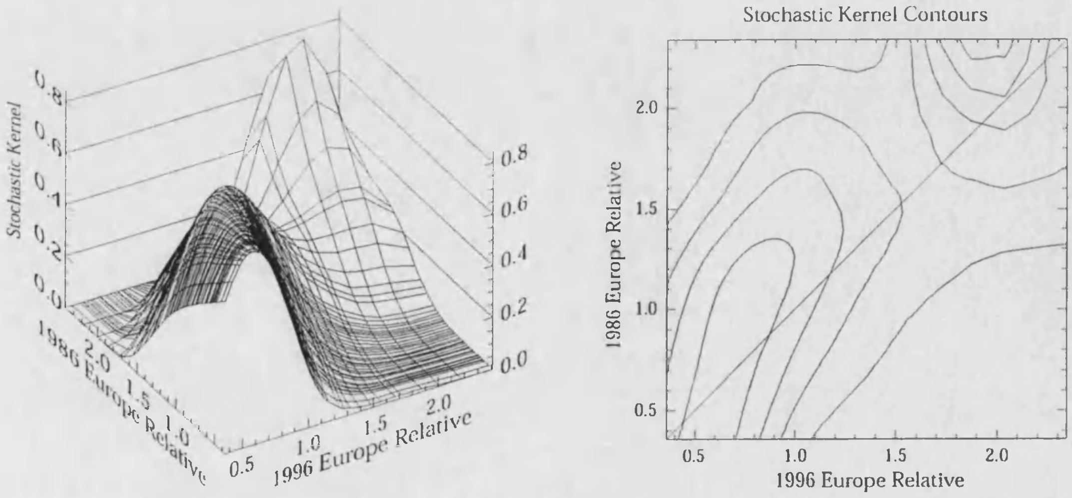


Figure 4.4. 1986 to 1996 Europe relative stochastic kernel

relative unemployment rates[5]. Reading along the bottom row of the matrix, we observe strong persistence for regions starting with an unemployment rate below 0.6 times the European average: by 1996, 81% remained below 0.6 times the European average, 19% had an unemployment rate between 0.6 and 0.75 times the average, and none had a relative unemployment rate higher than that. The next row up tells us that of those regions with an initial unemployment rate between 0.6 and 0.75 times the European average, 26% remained in that range, while 52% saw their unemployment rate fall below 0.6 times the average. Jumping to the top row we also see strong persistence amongst the regions with highest unemployment rates: of the regions with an initial unemployment rate above 1.3 times the European average, 61% remained above 1.3 times the average in 1996, while 23% moved to between the average and 1.3 times the average. However, regions with unemployment rates between 0.75 and 1.3 times the European average (third and fourth rows from the bottom) had experienced much greater mobility — regions with initial unemployment rates between 0.75 times the average and the average ended up almost equally distributed across the four intervals between 0 and 1.3 times the average.

Europe relative unemployment rates are, by nature, a continuous variable. There is a degree of arbitrariness involved in choosing a specific discretisation, and changing from one discretisation to another can easily distort the 'true' picture of transitions. In addition, many interesting details are lost as a result of the discretisation.

Figure 4.4 resolves these problems by avoiding any discretisation, and plotting the transition kernel from the 1986 distribution of Europe relative unemployment rates to the 1996 distribution of Europe relative unemployment rates. One can think of this kernel as the result of taking the transition probability matrix of Table 4.1 and letting

[5]The table gives two additional pieces of information. The first column gives n, the number of regions that begin their transitions in a give state. The second column gives the classes that divide up the state space.

the number of possible states tend to infinity and then to the continuum (see Technical Appendix C for a formal definition)[6]. The plot on the right hand side of the figure is a contour plot of the three dimensional kernel on the left. The contour plot works in exactly the same way as the more familiar contours on a standard geographical map. Lines on the contour plot connect points at the same height on the three dimensional kernel. An additional straight line is drawn in the contour plot to mark the diagonal, where all mass would be concentrated if there was complete persistence in the distribution.

Figure 4.4 confirms that there has been a polarisation of regional unemployment rates between 1986 and 1996, as suggested by the transition probability matrix[7]. Regions that in 1986 had a low unemployment rate relative to the European average tended to maintain or reduce their low relative unemployment rate over the next decade. Similarly, regions that in 1986 had a high unemployment rate relative to the European average in 1996 still tended to have a relatively high unemployment rate. However, regions with intermediate unemployment rates had mixed fortunes: some saw their relative unemployment rate fall, while others saw it rise. Still others saw it roughly unchanged.

2.3 Employment and labour force changes

By definition, unemployment rates equal one minus the ratio of employment to labour force. Thus the evolution of the distribution of regional unemployment rates can in principle reflect changes in regional demographic structure or labour market participation, as well as changes in regional employment. Has the recent polarisation of European regional unemployment rates been driven mainly by changes in regional employment? What role have changes in the regional distribution of labour force played? Or to put these questions in another way, how different would the distribution of regional unemployment rates have been in 1996, had the distribution of the European labour force across individual regions remained unchanged with respect to 1986[8]? Figure 4.5 provides the answer.

The plot on the left hand side of Figure 4.5 graphs the density of a 'counterfactual' distribution of 'unemployment rates'. These 'unemployment rates' are computed from actual values of regional employment in 1996, and hypothetical values of regional labour force constructed by disaggregating total European labour force in 1996 according

[6]The three dimensional stochastic kernel plots are drawn so that the density of lines reflects the underlying number of observations on which that part of the kernel is estimated. This procedure makes the pictures easier to read and more informative, but does not change the shape of the kernel.

[7]In fact, discrete intervals for the matrix were chosen to reflect accurately the 'true' continuous kernel equivalent.

[8]Our choice of unemployment rates rather than employment rates as the variable of interest is partly motivated by this analysis of labour force changes. Computation of both rates involves normalising the employment of regions of different sizes. However, normalising by labour force rather than working age population provides interesting additional insights. At the same time, it should be noted that our finding of a polarisation of the distribution of unemployment rates carries over to the distribution of employment rates.

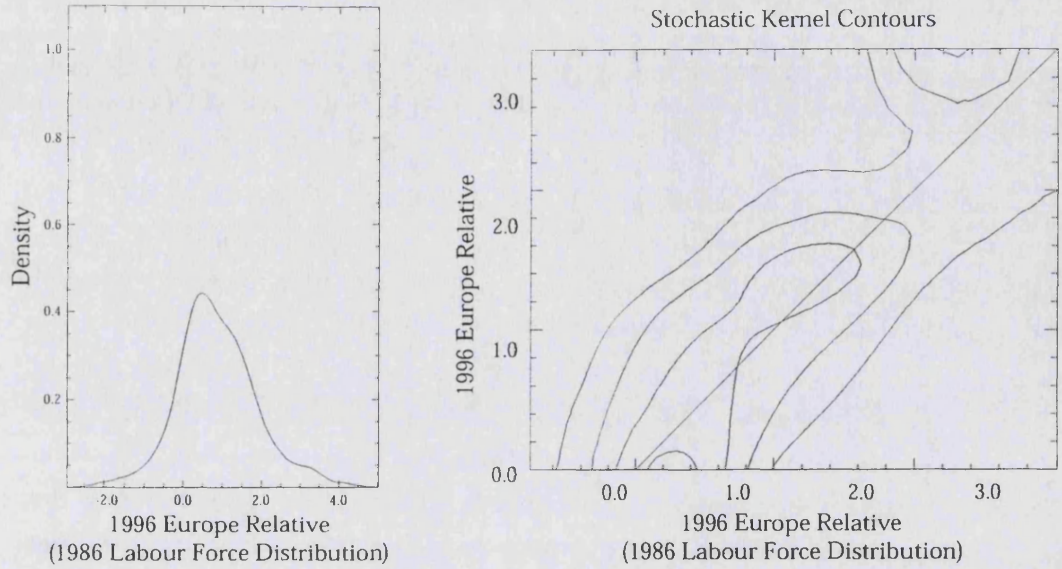


Figure 4.5. Counterfactual 1996 Europe relative (with 1986 labour force distribution)

to its 1986 distribution. This represents what the distribution of Europe relative unemployment rates would have looked like had there been no differences across regions in terms of labour force changes, but with employment still changing as it did in each region. The hypothetical nature of these rates is emphasised by the fact that, unlike actual unemployment rates, they are not bounded below by zero. This is because there are regions whose employment grew by more than the sum of their unemployed population in 1986 and the amount by which their labour force would have grown if it had grown at the same rate as total European labour force (6.3%). Comparing this density plot with the 'true' one (1996 plot in Figure 4.3), we see essentially the same features. However, there is a wider dispersion around the average (which, by construction, is the same in both cases) when the distribution of labour force is held constant. Changes in the regional distribution of the European labour force between 1986 and 1996 therefore made regional unemployment rates in 1996 less unequal than they would otherwise have been.

But have changes in the regional distribution of the European labour force significantly altered the relative position regions would otherwise have had in the distribution of Europe relative unemployment rates? The plot on the right hand side of Figure 4.5 shows that, in general, they have not. In Figure 4.4 we produced a stochastic kernel tracking regional positions in the distribution of Europe relative unemployment rates in 1996, given positions in the 1986 distribution. Similarly, in Figure 4.5 we produce the contour plot of a stochastic kernel tracking regional positions in the distribution of counterfactual 1996 Europe relative unemployment rates, given their positions in the distribution of actual 1996 Europe relative regional unemployment rates. Unlike the other kernels in the chapter, this one is not square, reflecting the fact that actual unemployment rates are bounded below by zero while the counterfactual ones are not.

The diagonal on the contour plot marks the position of regions with average labour force growth between 1986 and 1996. The concentration of mass close to the diagonal shows that the unemployment rates of individual regions would have been similar even without any differences in the evolution of their labour force. However, for all of the distribution there is some mass on both sides of the diagonal, showing that for all ranges of the unemployment rate distribution there have been regions with above average and below average labour force growth between 1986 and 1996.

The key is to identify whether, for a given interval on the vertical axis, there is more mass to the left of the diagonal (reflecting most regions in that range having above average labour force growth) or to its right (below average labour force growth). Starting from the top of the picture, regions with 1996 unemployment rates above 2.4 times the European average generally had above average labour force growth. However, from the 1996 plot in Figure 4.3 we see this part of the kernel is computed from very few regions (in fact only three). It is also almost entirely driven by the Spanish region Andalucía[9]. The rest of the distribution behaved pretty much as one would expect. Most of the regions with 1996 unemployment rates between 1.6 and 2.4 times the European average had either below average increases or decreases in their labour force (the exceptions were again a few Spanish regions with large increases in participation rates). Those with 1996 unemployment rates between the average and 1.6 times the average generally had above average increases in their labour force. These increases where even larger for most regions with below average unemployment rates. Thus, the distribution of labour force across European regions over this period tended to adjust to compensate, in part, for changes in the regional distribution of employment. Layard *et al.* (1991) explain that, between the 1960s and the late 1980s, regional labour force adjustment in Europe just offset changes in regional employment, leaving differences in unemployment rates and relative wages very stable[10].

In this section we have shown that, since 1986, labour force adjustment has no longer been able to keep up with employment changes, and has been clearly insufficient

[9]This region accounts for more than 50% of the labour force in this range of 1996 Europe relative unemployment rates (but for less than 2% of the total European labour force). Despite a 21.5% employment growth between 1986 and 1996, a 25.3% labour force growth over this decade kept Andalucía's unemployment rate as Europe's highest in 1996, at 32.4%. Andalucía's labour force growth resulted from a combination of demographic trends and changes in labour market participation. Natural population growth — helped by changes in age structure, but almost unaffected by tiny net immigration flows — resulted in a 9% increase in the population of working age between 1986 and 1996. At the same time, the increased participation of women in the labour force (42.5% of those between 15 and 64 years of age in 1996, up from 25.2% in 1986) more than offset the fall in male participation rates (from 75.9% in 1986 to 72.9% in 1996) and raised the total participation rate from 49.9% to 57.4%.

[10]Layard *et al.* (1991) focus on ongoing changes in employment sustaining persistent differences in unemployment rates. In contrast, Decressin and Fatàs (1995) study adjustment to one-off region-specific shocks, and show that the relative regional unemployment rate tends to come back to its trend within four years — a comparable time to the US, even though adjustment in Europe occurs mainly though changes in participation rates, while in the US adjustment takes place mainly through migration (see Blanchard and Katz, 1992).

to prevent a polarisation of European regional unemployment rates. We now turn to trying to understand the factors behind the markedly different unemployment outcomes of regions during this process.

3. Conditioning

How do we set about understanding the factors behind the features highlighted in Section 2? In this section, we consider a nonparametric approach which allows us to study the importance of these different factors in a simple way. In the next section we look at a parametric approach that provides complementary insights.

The nonparametric approach we develop here builds on a collection of tools proposed by Quah (1996, 1997*a*) for studying the dynamics of evolving distributions. These techniques are a first step in allowing us to understand the evolution of the entire cross section rather than the behaviour of a representative region. As will become clear, moving away from the standard representative region assumption gives us a number of interesting additional insights. Multiple equilibria and path dependency characterise a number of theories of regional development. Thus, regions with similar characteristics may have different development paths. Interactions between regions may further distort the link between individual regional characteristics and development paths. A proper understanding of the evolution of the distribution of unemployment rates may therefore involve more than understanding the evolution of a single representative region as in standard regression analysis.

The underlying idea is to look at how closely the evolution of each region's unemployment rate has followed that of some group of regions which we would expect to behave similarly. To do this we establish a mapping from a region's unemployment rate relative to the European average to the same region's unemployment rate relative to the group average. We group regions by a number of different criteria. Specifically, these groups of regions will be regions in the same Member State, regions that are geographical neighbours, regions with similar sectoral composition, and regions with similar proportions of low skilled.

These mappings are an extension of the transition kernels used in Section 2. Those kernels characterise the transitions across a decade. They are a mapping from the 1986 distribution of unemployment rates to the 1996 distribution. Technical Appendix C shows that this interpretation can be formalised using basic definitions and results from measure theory. That Appendix also shows that a similar construction can be used to explain the mapping between any two distributions, not just distributions of the same variable at different points in time.

We study the evolution of the distribution of unemployment rates in levels, not the pattern of changes in these unemployment rates. To see why this is more informative, imagine two situations, one where unemployment rates are converging, the other where unemployment rates are diverging. The distribution of changes in unemployment rates

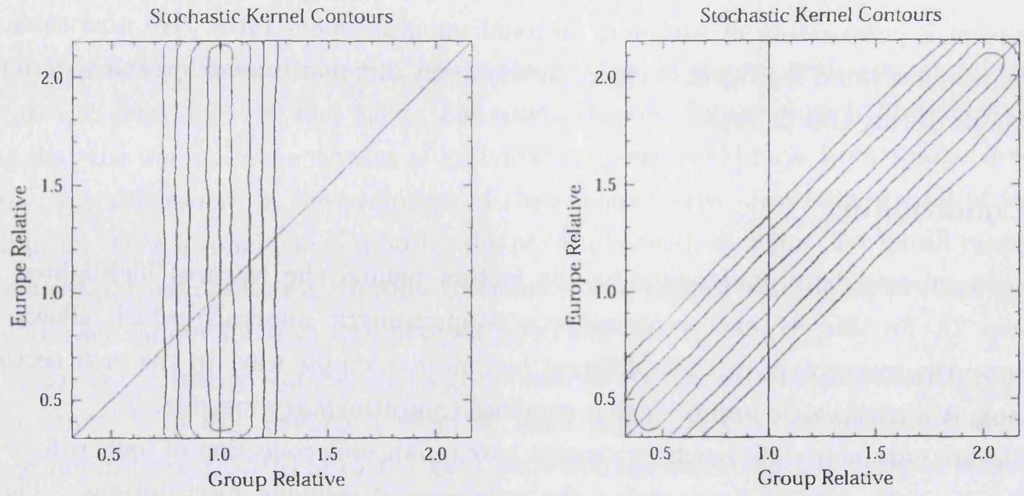


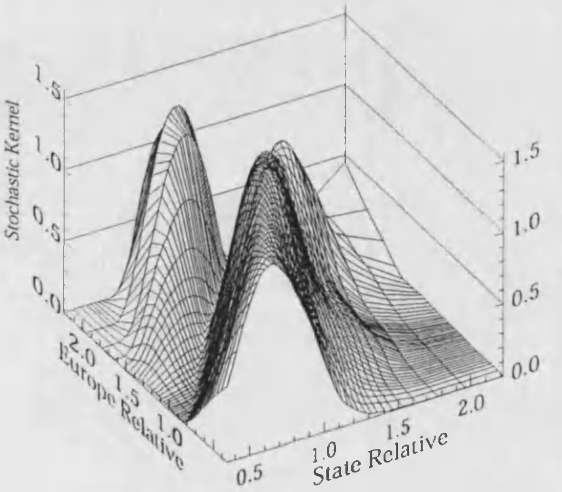
Figure 4.6. Benchmark stochastic kernels

across regions could be identical for both cases — some regions with positive changes, some with negative changes. However, studying the evolution in levels allows the two situations to be clearly distinguished: convergence shows up as a collapsing of the distribution, divergence as a spreading out. For similar reasons, conditioning in terms of levels is more informative than conditioning in terms of changes. However, the main reason for working with levels rather than changes is to exploit one of the most useful features of our approach: the ability to identify the same factor as having a different degree of relevance for different ranges of the original distribution. This is only possible if the distribution is specified in terms of a variable where similar values correspond to similar experiences. In our case, that implies working with unemployment rates rather than with changes in unemployment rates.

3.1 Conditioning on Member State

Possibly the simplest explanation for the polarisation of unemployment rates is that over this decade some EU Member States have managed to sort out their unemployment problems, while others have not.

An extreme version of this argument would have all regions within each State with almost identical unemployment rates throughout the decade. In that case, any differences in regional unemployment rates would be due to regions being in States with different national unemployment rates, and the polarisation of unemployment rates would have arisen as countries with intermediate rates drifted apart. In this extreme benchmark case, regardless of a region's *Europe relative* unemployment rate, its unemployment relative to the average for other regions in the same Member State (*State relative*) will be close to one. The stochastic kernel mapping Europe relative to State relative unemployment rates would then have almost all mass on the vertical line centered at one. The contour plot on the left of Figure 4.6 illustrates this benchmark.



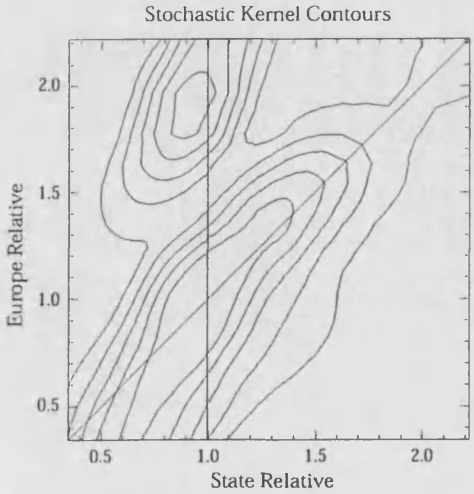


Figure 4.7. Europe relative to State relative stochastic kernel

The opposite extreme would have a similar regional distribution within each State, and almost identical State averages throughout the decade. In that case, the polarisation of unemployment rates could have arisen from mean preserving spreads of the regional distribution within Member States. In the corresponding benchmark, each region's State relative unemployment rate would be very close to its Europe relative unemployment rate. The stochastic kernel mapping Europe relative to State relative unemployment rates would then have almost all mass concentrated on the diagonal. The contour plot on the right of Figure 4.6 illustrates this benchmark.

As we move through the kernels in the remainder of the chapter, it will be useful to keep these two benchmarks in mind. When looking for criteria by which to group regions, our objective will be to find one that produces a kernel as close as possible to the benchmark on the left of 4.6, and as different as possible from the benchmark on the right.

In reality we see neither of these extremes. Figure 4.7 shows the actual Europe relative to State relative stochastic kernel. The kernel is calculated using data for all eleven years. For unemployment rates below 1.5 times the European average, the kernel is concentrated close to the diagonal, showing that each region's position with respect to the European average is not dissimilar from its position with respect to its State average.

Further, regions do not even tend to move strongly with their State over time. If a region followed changes in its State average, there would be a wide vertical spread of mass, which is not present in Figure 4.7. This is because the Europe relative unemployment would change over time with changes in the State average, but the State relative unemployment rate would remain constant. This is consistent with other

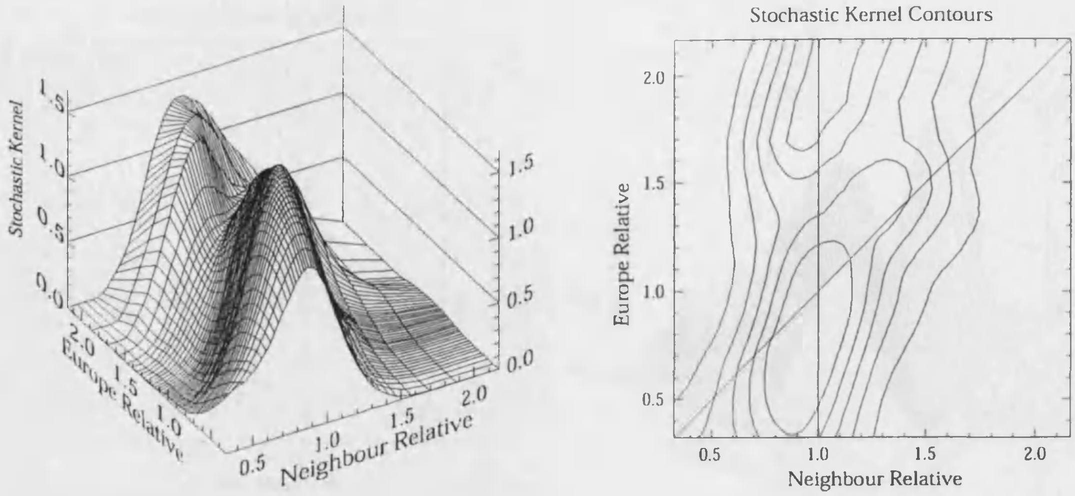


Figure 4.8. Europe relative to neighbour relative stochastic kernel

evidence about the diminishing economic significance of national borders in Europe[11].

The range above 1.5 times the European average stands out from the rest. Some high Europe relative unemployment outcomes correspond to high State outcomes. The spike at around the European average in this range corresponds to approximately the one half of Spanish regions with unemployment rates close to the Spanish average, plus Ireland (which is classified as a single NUTS2 region, so by construction its unemployment rate is the State average) prior to 1994. However, there are also regions in this range whose outcome differs as much from their State average as from the European average, leading to a wide spread of mass above one and close to the diagonal. This was a small group of regions in 1986, formed by Basilicata and Campania in Southern Italy, Northern Ireland, and five regions in the North of England and the South of Scotland. Over the next decade the British regions dropped from this group as their unemployment rates came closer to those of their Southern neighbours. At the same time, this group expanded to include regions on both sides of the French-Belgian border, all of Southern Italy, and the regions on France's Mediterranean Coast.

3.2 Conditioning on geographical neighbours

We have suggested in the previous subsection that ongoing European integration may mean that national borders are becoming less important in determining regional outcomes. Geographical location may still matter however, though perhaps at levels below the nation state. Could the evolution of European unemployment disparities be

[11] For instance, Fatàs (1997, p.759) finds that during the period '1966–1992, the correlation [of employment growth rates] of regions across national borders has been increasing over time while, at the same time, the cross-regional correlation within countries has decreased. [...] For example, in the post-EMS [European Monetary System] period, northern Italian regions display higher correlations with German regions than with southern Italian regions.'

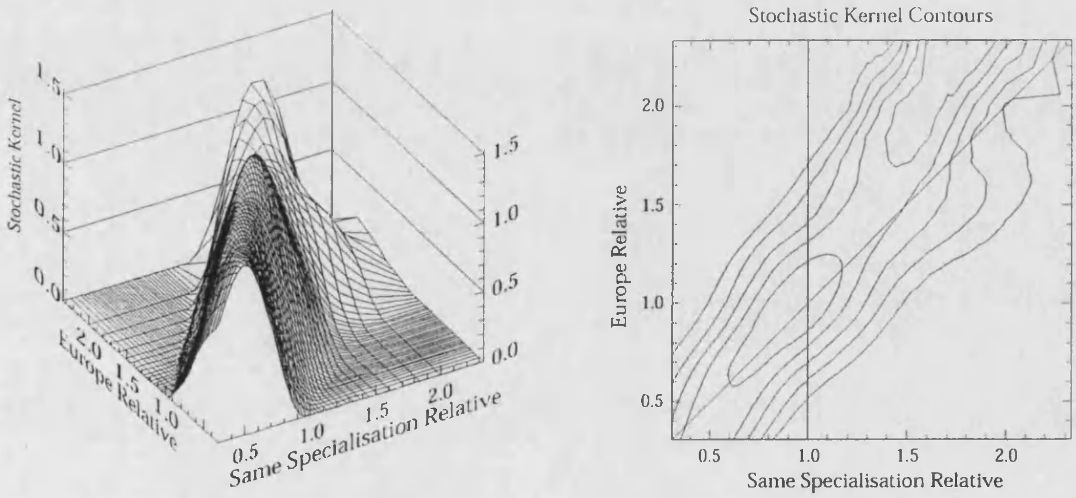


Figure 4.9. Europe relative to same specialisation relative stochastic kernel

understood in terms of the evolution of groups of neighbouring regions with similar outcomes that transcend national boundaries?

To answer this question we construct a kernel mapping Europe relative to *neighbour relative* unemployment rates, defined as each region's unemployment rate divided by the labour force weighted average of the unemployment rates of contiguous regions (not including the region itself).

Comparison of Figure 4.8 with Figure 4.7 shows that regional outcomes are much closer to outcomes of neighbours than to those of regions in the same Member State, except for the highest range of unemployment rates. Although the neighbour relative kernel still twists towards the diagonal for the middle unemployment regions, it is far more concentrated around the vertical line on one for regions with low and middle rates. This shows that while regions have followed very different evolutions relative to the European average, they have had very similar outcomes to those of their neighbours. This is particularly clear when one contrasts Figures 4.7 and 4.8, in the 'twist' of the bottom peak and the 'depth' of the valley between the two peaks in the three dimensional plot. Alternatively, one can count up the number of lines from the 'bottom' of the contour plot in Figures 4.7 and 4.8 (they are plotted at the same heights). Both the lower peak and the valley between the peaks in the neighbour relative kernel incorporate far more mass than the corresponding areas in the State relative kernel. The fact that the valley in the neighbour relative kernel is not as deep is particularly relevant, because it is in this intermediate range of unemployment rates that regions with similar starting positions have had very different evolutions. Also, note that a regions' domestic neighbours are part of the groups used to construct either kernel. In Figure 4.8, however, other regions in the same State are included. In Figure 4.7 they are not, but foreign neighbours are. Foreign neighbours are therefore much more closely related to a region in terms of unemployment outcomes than regions in the same State that are not contiguous.

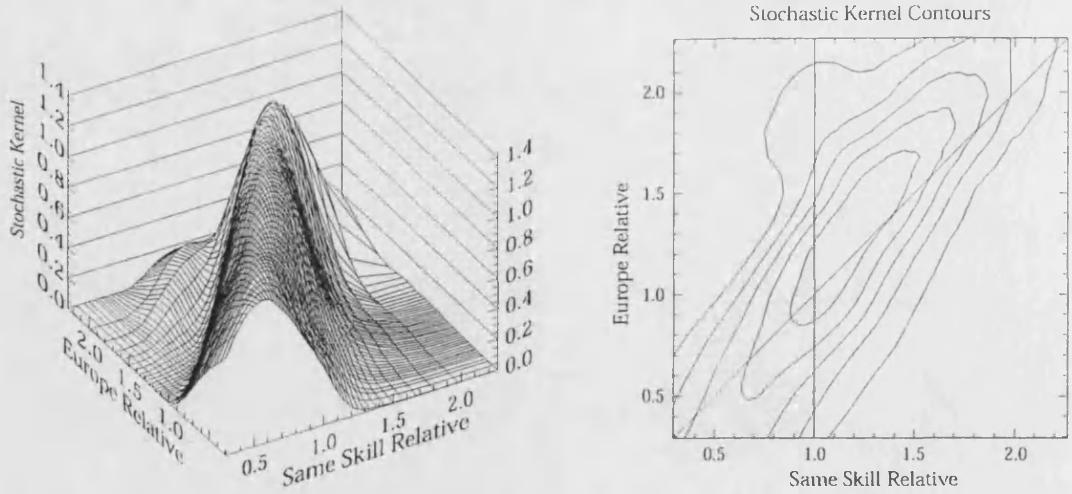


Figure 4.10. Europe relative to same skill relative stochastic kernel

In Section 4 we show that, in fact, foreign neighbours are as important as domestic neighbours.

The similarity of outcomes across neighbours could simply be driven by neighbouring regions having similar characteristics that are important determinants of unemployment rates. We now turn to two such determinants which have received particular attention.

3.3 Conditioning on same broad sectoral specialisation

The period 1986 to 1996 saw the continuation of an ongoing shift of European employment from agriculture, mining, and industry into services. If, as we have seen, labour force adjustment is slow, then regions with high initial specialisation in declining sectors may have seen their unemployment rates rise and not recover. Could this be driving the polarisation of unemployment rates across Europe? And can the importance of neighbours be justified by those regions with heavy industrial or primary employment being contiguous? Figure 4.9 suggests that the answer to both questions is no. This figure provides the stochastic kernel mapping Europe relative unemployment rates to *same specialisation relative* unemployment rates. This conditioning groups regions by the sector (agriculture and other primary sectors, manufacturing, or services) in which the initial share of regional employment was highest, relative to the average European share.

The concentration of mass on the diagonal of Figure 4.9 suggests that regions with similar initial specialisation have seen very different outcomes. This is probably due to the fact that the largest drop in agricultural and manufacturing employment had already taken place before the beginning of the period we consider. In the 15 years between 1971 and 1986 the share of manufacturing in European employment fell from 41% to 33%, while the share of services rose from 45% to 59%. In the next ten years to 1996, the share of manufacturing only fell by another three percentage points to 30%, while that of services rose to 65%. Spatial concentrations of declining sectors are not the key component driving the neighbours effect.

3.4 Conditioning on similar skill composition

There has been some discussion as to whether changes in the patterns of relative labour demand and supply in Europe have resulted in a rise in unemployment rates for the low skilled relative to unemployment rates for the high skilled (see, for instance, Krugman, 1994; Manacorda and Petrongolo, 1998; Nickell and Bell, 1995). One possible implication of this is that the evolution of regional unemployment rates may reflect the underlying skill composition of regional labour forces. Have regions with a large proportion of workers with low skills seen their unemployment rate rise, while regions with a small proportion of workers with low skills have seen their unemployment fall?

Figure 4.10 plots the stochastic kernel mapping Europe relative to *same skill relative* unemployment rates. We construct the kernel using nine groups of regions that have a similar percentage of adult population with less than upper secondary education (divided into equally spaced intervals between 0% and 90%). The concentration of mass on the diagonal reflects that the distribution of unemployment rates across each of our nine groups of regions with similar labour force skill composition is not dissimilar from the distribution of unemployment rates across all European regions. A region's skill composition tells us very little about the evolution of its unemployment rate since 1986. This is clearly not the key component driving the neighbours effect either.

3.5 Discretisation

In order to check the visual ranking of the kernels, we discretise the state space of relative unemployment rates and calculate the transition matrices that are the discrete versions of the continuous stochastic kernels. These discretisations, presented in Table 4.2, allow us estimate the relative mass in different areas of the kernels without having to integrate explicitly.

To interpret these matrices it is useful to compare them with the same benchmarks we used to interpret the corresponding stochastic kernel: large numbers on the column for the interval containing one, versus large numbers on the diagonal. We see that the Europe relative to neighbour relative matrix has all diagonal elements smaller than those of the other three kernels. At the same time, all other elements in the central column are larger in the Europe relative to neighbour relative matrix.

This confirms our earlier conclusion, that the unemployment outcomes of individual regions have closely followed those of their neighbours, much more so than the average outcomes of other regions within the same Member State, or other European regions with the same sectoral specialisation, or skill composition. That suggests that there is a truly spatial component to the neighbours effect. To be reasonably sure, however, we have to check that the neighbours effect remains strong, even after controlling for

State Relative

n	Europe Relative					
240	[1.45-∞)	0.00	0.06	0.43	0.13	0.38
201	[1.15-1.45)	0.00	0.13	0.16	0.52	0.18
577	[0.75-1.15)	0.00	0.05	0.68	0.18	0.09
330	[0.55-0.75)	0.06	0.30	0.50	0.12	0.02
302	[0-.55)	0.30	0.36	0.32	0.01	0.00
		[0-0.55)	[0.55-0.75)	[0.75-1.15)	[1.15-1.45)	[1.45-∞)

Neighbour Relative

n	Europe Relative					
240	[1.45-∞)	0.00	0.04	0.45	0.30	0.21
201	[1.15-1.45)	0.00	0.06	0.40	0.37	0.16
577	[0.75-1.15)	0.01	0.07	0.63	0.21	0.07
330	[0.55-0.75)	0.04	0.16	0.72	0.07	0.01
302	[0-.55)	0.22	0.17	0.55	0.06	0.00
		[0-0.55)	[0.55-0.75)	[0.75-1.15)	[1.15-1.45)	[1.45-∞)

Same Specialisation Relative

n	Europe Relative					
240	[1.45-∞)	0.00	0.00	0.01	0.25	0.74
201	[1.15-1.45)	0.00	0.00	0.35	0.36	0.29
577	[0.75-1.15)	0.00	0.15	0.68	0.16	0.01
330	[0.55-0.75)	0.23	0.45	0.32	0.00	0.00
302	[0-.55)	0.68	0.32	0.00	0.00	0.00
		[0-0.55)	[0.55-0.75)	[0.75-1.15)	[1.15-1.45)	[1.45-∞)

Same Skill Relative

n	Europe Relative					
240	[1.45-∞)	0.00	0.00	0.15	0.24	0.61
201	[1.15-1.45)	0.00	0.01	0.28	0.56	0.14
577	[0.75-1.15)	0.00	0.11	0.64	0.18	0.07
330	[0.55-0.75)	0.17	0.33	0.41	0.07	0.02
302	[0-.55)	0.48	0.25	0.26	0.00	0.00
		[0-0.55)	[0.55-0.75)	[0.75-1.15)	[1.15-1.45)	[1.45-∞)

Table 4.2. Europe relative to group relative transition probability matrices

similarities in regional characteristics. With that purpose, we now move to parametric techniques that will also complement the kernel results in other respects.

4. Regression results

The stochastic kernels of the previous section are attractive for a number of reasons. Grouping regions by common characteristics can be a useful way of thinking about which interactions between them help the most in understanding individual outcomes. The kernels also have a distinct advantage over parametric specifications, in that they make it easy to identify different behaviour at different parts of the distribution. For example, we have seen that many of the regions with very high unemployment rates have similar, large, fractions of their population with low education; at the same time, skill composition does not appear to be important for discriminating between unemployment outcomes for any other than these very high unemployment regions. Their main disadvantage over parametric approaches is that the kernels only allow us to consider one factor at a time.

In this section, we complement the stochastic kernel results with a number of parametric specifications. These regression results confirm the robustness of the kernel results. Even after controlling for a variety of other important factors, geographical neighbours remain key in explaining the evolution of regional unemployment. The parametric specification also allows us to separate out neighbours in the same Member State from neighbours in different States.

To keep the parametric specification simple, we examine the crosssection of changes in regional unemployment rates as a function of State, regional and neighbour characteristics[12].

Heuristically, we can divide changes in a region's unemployment rate into two components. They can be seen as being partly the result of a regions' initial structure — initial sectoral specialisation, skill composition, age and sex structure of population, and national differences in labour market structure and institutions, have all been identified as important explanatory factors for unemployment outcomes. This suggests that variables describing those initial characteristics should be an important element of our regressions. At the same time, there is a more endogenous component to the evolution of unemployment rates, related to the movement of firms and workers in to, and out of, regions. Further, this correlation in movements is interesting in its own right — especially if those flows seem to be correlated across national borders. Information on such flows is not readily available, and finding suitable instruments to incorporate them into empirical work is not easy. Even for sectoral structure of

[12] The closest counterpart to the stochastic kernel analysis would probably be a suitably defined panel specification. Unfortunately, the lack of reasonable exogenous time varying instruments makes it unfeasible to estimate such a panel, while allowing for the endogeneity of right hand side variables and the (auto)correlation structure of the regional residuals.

employment, there is no time series for the regions covered. However, we can capture some of this endogenous effect if, as suggested by the location argument outlined above, firm and worker movements are correlated across neighbouring regions. We do this by using the unemployment rate of surrounding regions[13].

Table 4.3, column 1, shows ordinary least squares results for our first empirical specification. The dependent variable is the (logarithm of the) change in the unemployment rate of region i between 1986 and 1996. We consider a number of different explanatory variables. Two variables capture the initial structure of employment in the region — percentage of regional employment in agriculture, mining, forestry, and fishing, and percentage of regional employment in manufacturing. Two variables capture the skill composition of the the region — the percentage of adult population with low skills (less than upper secondary education), and the percentage with medium skills (completed upper secondary education). The change in neighbours' unemployment rate is constructed from the average unemployment rate of each regions' geographical neighbours, as in Section 3. All explanatory variables are expressed in logarithms. Country dummies are included, but not reported, in this and all other specifications. We exclude from the regressions Member States classified as a single NUTS2 region (Denmark, Ireland, and Luxembourg). Further details on data definitions and sources are given in Data Appendix B. Heteroscedastic robust standard errors are reported in parenthesis under each estimate.

We can see that the coefficient on the percentage of adult population with low skills is positive, large, and significant, as would be expected. After conditioning on the other variables, a high proportion of population with low skills is associated with an increase or less of a decrease in regional unemployment. The coefficient on medium skills, however, is not significantly different from zero. This suggests that it is the lower end of the skill distribution that most markedly affects regional labour market outcomes.

Regarding the initial sectoral composition of employment, the coefficient on the percentage of employment in agriculture and other primary sectors is not significantly different from zero. However, the percentage of employment in industry at the beginning of the period has a negative effect on unemployment rate changes. This somewhat surprising result can be explained by noting that, for most of the Northern and Central European regions traditionally specialised in heavy industry, the worst part of the adjustment was over by the mid 1980s. Since then many of these regions have in fact seen their unemployment rate fall. Something that distinguishes these regions from heavy industrial regions in Southern Europe, where adjustment has taken place later, is the different proportions of population with low skills. It is therefore not unreasonable that, after controlling for skills, the effect of manufacturing specialisation on unemployment changes comes out to be negative.

[13]Of course, there are other reasons why a region's unemployment rate may be related to that of its neighbours. In particular, functional labour markets might extend across the administrative boundaries that define our regions. We return to this issue below.

Dependent variable Δ unemployment 1986–1996	1 OLS	2 IV	3 OLS	4 IV	5 OLS	6 IV	7 OLS	8 IV	9 OLS	10 IV
% low skill	0.241**	0.239**	0.231**	0.233**	0.225**	0.230**	0.235**	0.222**	0.222**	0.228*
	(0.094)	(0.932)	(0.101)	(0.099)	(0.103)	(0.098)	(0.091)	(0.080)	(0.096)	(0.083)
% medium skill	−0.109	−0.058	−0.115	−0.081	−0.063	−0.005	−0.104	0.013	−0.050	0.049
	(0.152)	(0.189)	(0.157)	(0.195)	(0.166)	(0.191)	(0.152)	(0.192)	(0.165)	(0.192)
initial % agriculture	−0.024	−0.029	−0.024	−0.028	−0.032*	−0.038*	−0.025	−0.038*	−0.034*	−0.045**
	(0.021)	(0.020)	(0.021)	(0.020)	(0.019)	(0.020)	(0.020)	(0.021)	(0.019)	(0.020)
initial % manufacturing	−0.220**	−0.204**	−0.219**	−0.207**	−0.258**	−0.231**	−0.216**	−0.176**	−0.252**	−0.207**
	(0.086)	(0.085)	(0.089)	(0.092)	(0.090)	(0.095)	(0.088)	(0.086)	(0.091)	(0.092)
initial female participation			0.002	−0.016	−0.208	−0.222			−0.221	0.253
			(0.133)	(0.149)	(0.157)	(0.165)			(0.158)	(0.171)
% young			0.090	0.066	0.331	0.266			0.290	0.157
			(0.205)	(0.213)	(0.204)	(0.217)			(0.210)	(0.234)
initial unemployment					−0.182**	−0.165**			−0.187**	−0.165**
					(0.070)	(0.071)			(0.072)	(0.077)
Δ neighbours' unemployment	0.589**	0.735**	0.584**	0.674**	0.524**	0.693**				
	(0.106)	(0.272)	(0.109)	(0.256)	(0.114)	(0.259)				
Δ domestic neighbours							0.606**	0.935**	0.551**	0.828**
							(0.115)	(0.252)	(0.120)	(0.240)
Δ foreign neighbours							0.514**	0.670**	0.415**	0.581*
							(0.174)	(0.325)	(0.169)	(0.342)
Adjusted R^2	0.72	0.72	0.72	0.72	0.73	0.73	0.72	0.72	0.72	0.72
No. observations	147	147	147	147	147	147	147	147	147	147

Country dummies included, but not reported, in all specifications
Heteroscedastic robust standard errors in parenthesis
** denotes coefficient significantly different from zero with 5% confidence level, * with 10% level

Table 4.3. Regression results

The most remarkable aspect of these results, however, is that the evolution of the unemployment rate in neighbours has a very strong and significant effect, even after controlling for regional industrial structure and skill composition. To understand the evolution of a region's unemployment we therefore need to consider its geographical position in addition to regional specific characteristics. We return to the interpretation of this result below. Before that, let us discuss a number of econometric issues.

We have chosen to capture the linkages between neighbouring regions through the incorporation of a labour force weighted unemployment rate variable, rather than through covariance assumptions on the error structure. We think that in the present context this specification is preferable. We would expect that predictable increases in neighbouring unemployment should feed through to regional unemployment through a number of mechanisms. Such expected increases are, by definition, orthogonal to the error, and thus best captured through the inclusion of a 'spatially lagged' dependent variable[14]. Introduction of a spatially lagged dependent variable is problematic, however, as the variable is correlated with the error (a region's unemployment effects its neighbour's unemployment, which in turn effects the region's unemployment, and so on). To solve this problem, we instrument for the spatially lagged dependent variable.

Our earlier discussion suggests that neighbour's initial sectoral employment shares, and the skill, age and sex composition of their workforces are all possible instruments for the spatially lagged unemployment rates. These variables should pick up the exogenous impact that we outlined above. We would also like to instrument for the endogenous effect of the movement of firms and workers across regions. Recent location theories suggest that such movements will be related to some measure of 'market potential'[15]. To do this, we construct a market potential variable, defined as the inverse of distance weighted sum of European regional Gross Domestic Products[16]. Instrumental variables (IV) results using this set of instruments are presented in Table 4.3, column 2. The table shows that instrumenting does not change our initial results. The proportions of low educated and initial industrial employment remain significant. The effect of neighbours' unemployment remains strong and significant[17].

Our second specification introduces two additional variables. As youth unemployment rates are high and rising, and regions differ in the age structure of their population, we control for the percentage of population that reached working age during the period (those aged between 15 and 25 in 1996). Additionally, in the mid-1980s regions female participation rates differed widely across European regions. Some regions, in Spain, had

[14]See Anselin (1988) for further discussion.

[15]See Fujita and Krugman (1995) for theoretical foundations, and Hanson (1998) for an empirical implementation.

[16]Thus, for region i, market potential is defined as $mp_i = \sum_{j \neq i} \text{GDP}_j / d_{i,j}$, where $d_{i,j}$ is the great circle distance between region i and region j, and GDP_j is the GDP of region j, and the sum is over all regions in the European Union excluding region i itself.

[17]In this, and all subsequent specifications we cannot reject the validity of our instrument set at the 5% confidence level using the test proposed by Davidson and MacKinnon (1993).

participation rates as low as 18%, while others, in the UK, had rates above 50%. Over the decade, female participation rates have significantly converged across European regions. This has resulted in huge increases in labour force, not always matched by comparable increases in employment[18]. We therefore control for the initial female participation rate in each region. OLS results are in column 3. Both coefficients have the expected sign, but are insignificant. Further investigation reveals that the percentage young becomes significant if we drop percentage low skilled and female participation. Female participation remains (just) insignificant when we drop out percentage young and low skilled. This occurs because all three variables are highly correlated — although percentage low skilled appears to matter most. Column 4 shows that instrumenting does not change the results.

Column 5 shows the OLS results when we introduce the initial unemployment rate. The only change here is that the agriculture variable becomes significant, but only at the 10% level. Column 6 shows that, once again, instrumenting doesn't change these results.

We have seen that neighbours are important. In Section 3 we argued that foreign neighbours mattered more than regions in the same State that are not contiguous. We now take this one step further and ask how important foreign neighbours are relative to domestic neighbours. The surprising answer is that they are equally as important. This is shown in Columns 7–10, where we split the neighbours variable for border regions into two components, that due to domestic neighbours and that due to foreign neighbours[19]. There are 51 such border regions (around a third of the sample)[20]. Column 7 provides OLS results for the basic specification. We see that foreign neighbours have a significant effect on border regions. Further, we are unable to reject the hypothesis that the coefficients on both domestic and foreign neighbours are identical (the test has a value of 0.9 and is distributed $\chi^2(1)$).

Again, both neighbours effects are possibly endogenous. To correct for this we instrument for both domestic and foreign neighbours. The results are reported in column 8. We see that foreign neighbours continue to have a significant effect on border regions. Again, we are unable to reject the hypothesis that the coefficients on both domestic and foreign neighbours are identical. Next we introduce the additional variables considered before. This specification is presented in columns 9 (OLS) and 10 (IV). We see that the results are consistent with previous ones, although the significance of foreign neighbours drops slightly. However, we still cannot reject the hypothesis that the coefficients on both domestic and foreign neighbours are identical.

---

[18] See Wasmer (1998) for an exposition of this argument.

[19] For the domestic and foreign neighbours variables, the labour force weights are those used when constructing our original neighbourhood variable. This ensures, that the sum of the two variables is the original neighbourhood variable, and that the coefficients are directly comparable.

[20] If we drop out the UK's 35 regions, which include only one border region, then border regions make up nearly half the sample. The results do not change for this restricted sample.

We have also tried a number of alternative specifications, not reported in the table. For instance, we have tried including the average change for unemployment for regions with a similar initial sectoral specialisation, a similar skill composition of adult population, and so on. The results are still remarkably robust.

There are a number of possible interpretations for the importance of geographical neighbours. First, the results could be driven by neighbouring regions having in common important determinants of unemployment rates. However, we have already taken this into account by controlling for the State to which regions belong, as well as for important regional characteristics[21].

A second, rather mechanical, explanation is that functional labour markets extend across our geographical units. That is, neighbouring NUTS2 regions may actually form one labour market with substantial commuting flows between regions. Although relevant for smaller regions, this is not so important for NUTS2 regions, with the known exceptions of the Netherlands and areas surrounding London, Paris, Brussels, Bremen and Hamburg (see Cheshire and Carbonaro, 1996, for further discussion). Further, neighbourhood effects remain equally strong across national borders, and we know that cross border commuting flows are tiny. Cross border flows represented only 0.2% of the total European labour force in 1990 (de Falleur and Vandeville, 1996). Of these 316,000 cross-border workers, roughly 50% are workers commuting to Switzerland (not included in our sample). A further 40,000 represent flows into Luxembourg (which is excluded from our regressions). Thus, there are only approximately 100,000 cross-border commuting flows for the border regions in our sample. Even on the German-French border, where commuting flows are strongest, the total flows are 43,970, which is less than 0.8% of the combined border region labour force of 5,300,000.

A third explanation is that the location and relocation decisions of workers and firms, in combination with weak labour force adjustment, have resulted in clusters of high and low unemployment larger than our geographical units and crossing national boundaries. However, we know that net migration flows across European regions are tiny, and not very responsive to differences in wages or unemployment rates (see, for instance, Eichengreen, 1993). This is particularly marked for cross-country migration flows, to the extent that only 1.5% of EU workers have a job in a Member State different from that in which they were born (`http://citizens.eu.int/en/en/newsitem-2.htm`)

All of this suggests that the spatial spillover results could reflect firm relocations that are occurring on the basis of geographical areas somewhat larger than NUTS2, but somewhat smaller than Member States. Looking with hindsight at the maps in Figure 4.2, we can in fact see the emerging clusters of high and low unemployment. These clusters do not conform to a standard core-periphery gradient. Instead high and low unemployment clusters have appeared in both the core and the periphery of the EU,

[21] The fact that we are estimating in changes rather than levels should largely take care of fixed effects as well.

often extending across national borders.

5. An example of two border regions in Belgium

In 1986 the Belgian region of Limburg had an unemployment rate 1.2 times the Belgian average and 1.3 times the European Union average. By 1996 its unemployment rate had fallen below both the Belgian and EU averages. Just across the border from Limburg (Belgium), two Dutch regions had similar experiences. The unemployment rates of Limburg (Netherlands) and Noord-Brabant fell relative to both the Dutch and EU averages.

Back in Belgium, 90 kilometres South-West of Limburg, the region of Hainaut started with a similar unemployment rate in 1986. However, instead of falling as it did in Limburg, this rate rose both in absolute terms and relative to both the Belgian and EU averages. Just across the border from Hainaut, the French region of Nord-Pas de Calais also saw its unemployment rate increase in both absolute and relative terms.

The different fortunes of these two Belgian regions were not driven by changes in demographic structure or labour market participation. Both regions had growing labour forces, but Limburg's actually grew more than twice as fast. The reason for Limburg's success is that its employment grew even faster than its labour force, and over four times faster than Hainaut's. A similar process occurred in the two Dutch neighbours of the Belgian Limburg. These regions that did relatively well had large and growing labour forces. But they also had a rate of employment growth that more than matched their labour force growth, and that brought their unemployment rates down. By contrast Nord-Pas de Calais, the French neighbour of Hainaut that did relatively badly, lost employment while its labour force was rising.

The drop in Limburg's unemployment rate versus Hainaut's rise cannot be put down to differences in the skill composition of their labour force. Both these Belgian regions had a similar percentage of their population with less than upper secondary education. And the French region of Nord-Pas de Calais, despite having a smaller fraction of people with less than upper secondary education than either of the Belgian regions, had a worse unemployment outcome.

Further, the evolution of these regions was not due to their different initial sectoral composition. Admittedly in 1986 Nord-Pas de Calais was a predominantly industrial region. But Hainaut also saw its unemployment rate rise and in 1986 was concentrated in services. In contrast, the Belgian success story Limburg was concentrated in industry and of its two neighbours, one was mainly industrial (Noord-Brabant), the other service based (Limburg). No simple story of sectoral changes explains the relative performance of these regions[22].

Given the small flows of workers across these borders, both in terms of commuting and permanent moves, one can hardly argue that there are functional labour markets extending across these regions. However, firms do seem to find it attractive to exploit other advantages of location close to these borders, such as the ability to use suppliers from different countries. The areas on the borders between Belgium and France and Belgium and the Netherlands have provided traditional locations for industry. However, in recent years these two borders have experienced very different evolutions. The most publicised case came in 1997 as Renault announced the closure of its Vilvoorde plant on the Belgian border with France. This raised protests at the loss of 3,100 jobs, at a time when Renault was planning to expand operations in other parts of Europe. At about the same time in Limburg (Netherlands), Volvo introduced a three-shift working schedule in its Nedcar joint plant with Mitsubishi, to double production over the following three years, drawing on suppliers from both sides of the Belgian-Dutch border. And on the Belgian side of this border, General Motors was also expanding production at its Antwerp plant.

Starting from similar conditions, the Belgian regions Limburg and Hainaut saw very different evolutions in their unemployment rates, but in each case these were very similar to those of their foreign neighbours. In this chapter we have shown that this story is not unique, but representative of a broader pattern that has developed across Europe.

6. Concluding comments

This chapter has shown that European regions have experienced a polarisation in their unemployment rates between 1986 and 1996. Regions with low rates in 1986 had low rates in 1996, regions with high rates in 1986 had high rates in 1996, while regions with intermediate rates in 1986 have tended to move towards the extremes of the distribution. This process has been driven by changes in regional employment rather than by changes in demographic structure or labour market participation. While there has been some labour force adjustment to regional employment changes, this has been insufficient to prevent the polarisation of European unemployment rates. Further, the outcomes of individual regions have closely followed those of their geographical neighbours.

This neighbours result could be driven by neighbouring regions having similar characteristics. For example, neighbouring regions often have similar employment structures, or similarly skilled labour forces. However, we have shown that the importance of neighbours' outcomes is only weakly driven by skill composition and broad

[22]Possible differences between the Flemish and French speaking regions of Belgium cannot explain these changes either. Contiguous to both the Flemish speaking Belgian Limburg and to the Dutch Limburg is the French speaking Belgian region of Liège, which also experienced a reduction in its unemployment rate.

sectoral specialisation. The same is true with respect to other regional characteristics, such as the sex and age structure of population.

Alternatively, the neighbours result could be driven by the fact that different European Union Member States have had different unemployment experiences, and regions within the same Member State tend to move together. However, we have also shown that regional outcomes only follow average Member State outcomes to a small extent. Further, the outcome of both own state and foreign neighbours matters equally for regional outcomes.

So, what is driving this emerging pattern of cross border unemployment clusters? We think it may be the result of firm location and relocation decisions, reflected in agglomerations of activity over geographical areas somewhat larger than NUTS2, but somewhat smaller than nation states. Worker relocations could also matter, but we know net flows of workers between European regions are small.

The EU has experienced a period of rapid and deep integration over the last decade. Portugal and Spain became Member States in 1986. Customs formalities for shipments of goods across the internal borders of the EU disappeared 1 January 1993. Border controls for movements of people across the Member States signing the Schengen agreement (Belgium, France, Germany, Luxembourg, Netherlands, Portugal, and Spain) disappeared 26 March 1995. Transport infrastructure has also been greatly improved — for instance, the number of kilometres of motorways in the European Economic Community of 1986 increased by a third between 1986 and 1994, and in Portugal and Spain it more than tripled.

Over this same period, there has been a revival of interest by economists in location issues. Recent models of trade and location formalise cumulative causation mechanisms, to show that regions which are similar, or even identical, in underlying structure can end up having very different development paths. Many of those models focus on how the propensity of firms and workers to agglomerate in space changes as regions become more integrated (see Ottaviano and Puga, 1998, for a survey). With little worker mobility, and institutional constraints on regional wage disparities, the conclusion is that closer economic integration will result in increasing concentration of economic activities across space (Puga, 1999).

Where would we expect to see agglomeration reflected? Looking at Mexico and the United States, Hanson (1997*a*,*b*, 1998) and Ciccone (1997) point to wages. However, the weak responsiveness of European regional wages to local economic conditions suggests that in Europe agglomeration will be reflected instead in employment. The aforementioned models of location do not incorporate unemployment explicitly. However, with limited labour force adjustment to regional employment changes (as found in Section 2), we can expect changes in employment to be largely translated into changes in unemployment. The distinguishing feature of this story is that regions with similar characteristics may have very different outcomes. At the same time, if clusters of activity are of a size larger than the regions considered, neighbouring regions will tend to experience similar outcomes, even if they are in different Member States.

The fact that unemployment outcomes are so much more homogenous across neighbours, foreign and domestic, than across regions in the same Member State also tells us something about the spatial dimensions of the emerging clusters of high and low unemployment in Europe. The average Member State has 13.6 regions, while the average neighbourhood has 5.6 regions. Hence these are clusters of typically less than one half of the size of the average Member State of the European Union, but often extend across national borders and include regions from more than one Member State. This is similar to the geographical dimensions of agglomerations that Hanson (1998) finds looking at regional wages in the United States (US).

That also has important implications for policy. European regional policy has traditionally targeted mainly regional differences in income per capita, but is increasingly shifting its focus towards tackling regional differences in unemployment rates. Contrasting our results with those of Quah (1997*b*) shows the empirical reality underlying this change in emphasis — in contrast to the divergence of unemployment rates across European regions, Quah shows that differences in regional incomes per capita are narrowing. But there is one important additional difference. While inequalities in incomes per capita exhibited a core-periphery gradient (Keeble, Offord, and Walker, 1988), unemployment clusters are more localised and emerging in both the core and the periphery of the EU. There is strong political opposition to tackling these growing unemployment rate differences through increased labour mobility. Recent location theories suggest that the self-reinforcing nature of agglomerations will make these hard to break once they become established. However, given that the unemployment clusters we find are of not very large size and scattered across Europe, it may be politically viable as well as more efficient to implement policies that accept some clustering and larger mobility within a neighbourhood.

5

The cross–sectional evolution of the US city size distribution

with Yannis Ioannides

1. Introduction

Empirical studies of the distribution of city sizes have a long and distinguished history. At least 80 years ago, it was observed that the distribution of cities within an urban system is often remarkably well approximated by a Pareto distribution. This observation has generated a vast body of empirical work aimed at testing this and related propositions. Much of this work has concentrated on testing the rank–size rule first proposed by Zipf (1949)[1]. This large empirical literature has, in turn, lead to the development of a number of theoretical models which attempt to generate this apparent regularity. This collection of models are essentially stochastic – they seek to generate, rather than explain the regularity. To do this, they abstract from underlying economic or social processes that drive the evolution of city sizes.

The importance of the rank–size rule in framing the discussion about the distribution of city sizes has had two important implications for the literature on the development of the urban system. First, it has led to the acceptance of simplistic models that downplay important economic and social forces but that are capable of replicating the regularity. Second, it has relegated work on other aspects of the distribution to a distant second place. This chapter is primarily concerned with these other aspects of the distribution.

With respect to the first issue, recent work by a number of theorists, who developed the so–called new economic geography, highlight the problems that the rank size rule has presented for theoretical work. In common with an older theoretical literature, these authors have emphasised the interplay of agglomeration and dispersion forces as key in determining city sizes. However, they have also emphasised the fact that 'when

[1]The rank–size rule (or Zipf's law) states that the city size distribution follows a Pareto distribution with exponent one. See Section 2.

it comes to the size distribution of cities, [...] the problem we face is that the data offer a stunningly neat picture, one that is hard to reproduce in any plausible (or even implausible) theoretical model'[2]. For a discussion of the issues see Fujita, Krugman, and Venables (1999*b*). Simon (1955), Krugman (1996), and Gabaix (1997) all propose models capable of generating regularities in the distribution of city sizes.

The second issue has received very little direct attention. The empirical work on the rank size rule is essentially involved with one particular characteristic of the distribution of city sizes – the shape of that distribution. In contrast, this chapter considers the importance of intra–distribution dynamics. It asks questions about how cities develop relative to the rest of the urban system, both in terms of rankings and relative sizes. We propose a number of techniques for characterising this intra–distribution mobility.

We do not see characterising this intra–distribution mobility as a substitue for direct tests of either the economic or the stochastic models of the development of the urban system. Economic models are not typically asked to predict the shape of the distribution of endogenous variables, so there is no reason to be unduly demanding with regard to the dynamics of the distribution of city sizes. To the extent that economic models help us understand the economic forces that might promote agglomeration or drive dispersion, failure to match empirical regularities on city sizes should not lead to an outright rejection of those models. However, given that the aim of stochastic models is to help us understand the nature of the process that might produce the rank–size rule, it would seem important that these models also deliver on other aspects of the city size distribution. Stochastic models which generate the shape of the distribution, but only at the expense of unrealistic intra–distribution dynamics, may well be uninformative about the processes at work.

This chapter proceeds as follows. Section 2 reviews some of the related empirical and theoretical literature. Section 3 briefly describes the data. Section 4 develops a number of empirical tools which can be used to analyse intra–distribution dynamics. We use these tools to examine the evolution of the US city size distribution from 1900 to 1990. Section 5 concludes.

2. Related literature

There is a vast empirical literature on the distribution of city sizes. A very selective account follows, which seeks to highlight the main issues and those most closely related to the empirical work in this chapter. A number of extensive surveys exist: Carroll (1982) covers earlier work in some detail; Cheshire (1999) provides a survey of more recent work.

At least as early as Auerbach (1913) it had been proposed that the city size distribution could be closely approximated by a Pareto distribution. Thus, if we rank

[2]See (Fujita, Krugman, and Venables, 1999*b*, chapter 12).

cities from the largest (rank 1) to the smallest (rank N) to get the rank r(p) for a city of size p, then:

$$r(p) = Ap^{-\alpha}. \tag{5.1}$$

Or, taking logs

$$\log r(p) = \log A - \alpha \log p. \tag{5.2}$$

Zipf (1949) went further. He proposed that, not only did the distribution of city sizes follow a Pareto distribution, but that it took a special form of the distribution where $\alpha = 1$. This expression of the regularity is known as the 'rank–size' rule (or Zipf's rule) and has formed the starting point for much of the empirical literature. It implies that the second largest city is half the size of the largest, the third largest is a third the size of the largest, etc.

Rosen and Resnick (1980) brought together the questions from a large body of literature developed from the 1950s to the 1970s.[3] They highlighted the importance of the definition of the lower threshold size for cities[4] and considered how the urban system might be best defined. We will return to this second issue briefly in Section 4.2 below.

A further two decades of work has followed with two key conclusions. The first, less controversial, is that the city size distribution is reasonably well approximated by a Pareto distribution, at least for the largest cities. The second, far more controversial, is that the exponent α in Equation 5.1 is close to one. Some authors, notably Krugman (1996), have argued that the combined evidence suggests that the rank size rule holds for a number of different samples over a number of different time periods. Others, such as Alperovich (1993), reject this stronger conclusion, but accept the first. The debate still rages. Dobkins and Ioannides (1999), in common with other work, find that the exponent (α) is around one for a sub–sample of the largest cities, but below one for the whole sample. However, when they compare the fit of the Pareto Law with a nonparametric one, they find that the fit is poor around the tails of the distribution, thus raising doubts about the validity of the strict rank–size rule. Black and Henderson (1999) use similar (though not identical) data, and reject Equation 5.1 due to the inclusion of a significant quadratic term.

These last two papers also consider a number of issues related to the intra–distribution mobility characteristics of the city size distribution. Both build on Eaton and Eckstein (1997) who use transition probability matrices to characterise the evolution of the French and Japanese urban systems and find that both systems are characterised by parallel growth. Cities tend to grow at the same rate, maintaining their place in the relative distribution and consequently showing little intra–distribution mobility. In contrast, Dobkins and Ioannides (1999) find that the US system is characterised by the

[3] Key contributions included Allen (1954), Madden (1956) and Berry (1961).

[4] A recurring theme in the urban systems literature. See Black and Henderson (1999) and Dobkins and Ioannides (1999) for a recent discussion.

entry of new cities and a higher degree of mobility. Black and Henderson (1999) confirm this result. They show that new entry means that cities tend to be more mobile up the distribution, but less mobile down the distribution. The expected transition time from lowest state to highest is around 500 years. Movement in the opposite direction takes, on average, 5500 years. This chapter builds on these three papers to provide a more detailed characterisation of the intra-distribution mobility of cities within the US city size distribution.

3. Data

There are a variety of ways to define cities[5]. In this chapter we use contemporaneous Census Bureau definitions of metropolitan areas, with adaptations for availability. From 1900 to 1950, we have metropolitan areas defined by the 1950 census. That is, for years previous to 1950, we use reconstructions from Bogue (1953) of what populations would have been in each metropolitan area in each year if the cities had been defined as they were in 1950. For each decennial year from 1950 to 1980, we use the metropolitan area definitions that were in effect for those years. Between 1980 and 1990, the Census Bureau redefined metropolitan areas in such a way that the largest US cities would seem to have taken a huge jump in size, and several major cities would have been lost. While this might be appropriate for some uses of the data, it would introduce 'artificial' intra-distribution mobility for the 1980–1990 period. Therefore, Dobkins and Ioannides reconstructed the metro areas for 1990, based on the 1980 definitions, much as Bogue did earlier. We believe that this gives us the most consistent definitions of US cities (metropolitan areas) that we are likely to find.

The method raises a question as to which cities, as defined or reconstructed, should be included. In the years from 1950 to 1980, we use the Census Bureau's listing of metropolitan areas. Although the wording of the definitions of metropolitan areas has changed slightly over the years, the number 50,000 is a minimum requirement for the core area within the metropolitan area. Therefore, we used 50,000 as the cutoff for including metropolitan areas as defined by Bogue prior to 1950. Consequently we have a changing number of cities over time, from 112 in 1900 to 334 in 1990. While it is often difficult to deal with an increasing number of cities econometrically, we think that this is a key aspect of the US system of cities, and is worthy of being factored into our studies.

We also have data on earnings in all cities in the sample for all years, drawn from Census reports, although the data set is not ideal because the Census Bureau changed the categories it reported over the years. We have data on schooling in each city over the century, reported as the percentage of the population in the 15 to 20 year old category who are in school. We also have data on regional location according to the Census

[5]This section draws extensively from Dobkins and Ioannides (1999).

Bureau division of the country into nine regions. We recombine these regions into five regions, when necessary. Table 5.4 provides summary statistics of the data for each census year. Table 5.5 provides additional statistics for the whole sample in 1990.

There are two important distinctions between our data and the data used by Black and Henderson (1999) for the same time period. First, they define the geographical area of a city as the collection of counties that form that city in 1990. They then use the urban population of each of these counties to give city size in each census year from 1900-1990. This gets round problems relating to changing definitions of metro areas between 1950 and 1990 that apply to our data. However, it introduces an additional source of mismeasurement relating to the use of contemporaneous definitions of urban population that may change throughout the period. It also means that collections of small towns in areas that will become cities are treated identically to genuine metro areas of a similar size. Second, they use a relative cut–off point to define when a city enters in to the sample, whereas we use an absolute cut–off point. Their use of a relative cut–off point in combination with metro areas defined on the basis of urban populations means that their sample will tend to overstate the number of functional metro–areas in any given sample period. In contrast our approach based on an absolute cut–off point will tend to understate the number of functional metro–areas. Black and Henderson (1999) show that estimates of intra–distribution mobility are sensitive to the choice of an absolute versus relative cut–off. However, a–priori there is no reason to prefer one definition over enough.

4. Intra–distribution dynamics

As Quah (1996, 1997*a*,*b*) has forcefully argued, typical cross–section, or panel data techniques, do not allow inference about patterns in the intertemporal evolution of the entire cross-sectional *distribution*. They do not allow us to consider the impact over time of one part of the distribution upon another, i.e., of the development of large cities as a group upon smaller cities. Making such inferences requires that one model and estimate directly the full dynamics of the entire distribution of cities. In contrast, typical panel data analyses involve efficient and consistent estimation of models where the error consists of components reflecting individual effects (random or fixed), time effects and purely random factors. In addition, the evolution of urbanization and suburbanization may affect individual cities so drastically as to render conventional methods of accounting for attrition totally inappropriate. Particularly, as smaller urban units may fuse to create larger units. Given this, and the small number of time series observations, non–parametric or semi–parametric distributional approaches such as the one proposed here may be the only appropriate ones.

We examine intra–distribution dynamics by first considering nonparametrically the long run transition patterns in the US city size distribution. Next we introduce measures of intra–distribution mobility in the form of suitably defined statistics of dispersion and serial correlation in changes in rankings. Finally, we examine patterns in the intra–distribution dynamics within different groupings of cities, that is, in terms of geographical regions and hierarchical tiers.

4.1 Intra–distribution mobility

We will consider two inter-related types of intra–distribution mobility: changes in the rankings of cities and changes in their relative sizes. Previous studies of intra–distribution mobility have studied both types of mobility without clearly distinguishing between the two different concepts. Eaton and Eckstein (1997), Black and Henderson (1999) and Dobkins and Ioannides (1999) consider the size of cities relative to the mean city size. They then discretize the state space of relative city sizes by defining discrete intervals[6] and calculate the transition matrices corresponding to this discretization. Only Dobkins and Ioannides (1999) consider mobility in terms of the rankings of cities by discretizing the state space in each period on the basis of quantiles (bottom 10%, second 10%, ..., top 10%). They argue that this gives a more detailed insight in to the intra–distribution dynamics, without making it clear that the mobility that they are studying is subtly different.

To see why the distinction is important, we need to think about what the two types of exercise tell us. Considering the first type of mobility, in city sizes relative to the mean, allows us to answer a number of interesting questions about the long run city size distribution. Thus, we can examine whether the distribution has a tendency to become uniform (flatten out), to collapse to a single point (all cities converge to the same size), or to, say, become bimodal. To do this, after discretising the state space and calculating the transition probability matrix, we calculate the ergodic distribution of that markov process (assuming that it exists). All three of the above mentioned papers do exactly that. Black and Henderson (1999) emphasize that the long run ergodic distribution is remarkably close to the current distribution.

Notice that all of these are questions about what mobility implies for the overall shape of the distribution. These exercises also tell us about changes in the rankings of cities. Take any two neighbouring discrete states. If some cities move up from the lower state to the higher state, while others move from the higher state to the lower state, then the rankings of those cities must have changed. There are, however, two problems with this method of characterising the changing rankings of cities. First, the discretisation of the state space means that we do not observe what happens within each discrete state. Only when cities move between states do we get information on mobility. Second, observing the movement of an individual city between states does not necessarily imply a change in rankings. This is where the second approach of Dobkins and Ioannides (1999) based on quantiles differs from the approach based on an (arbitrary) fixed discretisation. For

[6]For example Dobkins and Ioannides (1999) divide the state space in to .30, .50, .75, 1.00, 2.00 and 20.00 times the contemporaneous mean.

this second approach, movement of one city up a discrete state, *must* be accompanied by a corresponding move down a state by another city and vice–versa. Thus all movements between states correspond to changes in rankings. However, this second approach still suffers from the fact that we do not observe mobility and changes of rankings within cells.

The problem of movements within cells arises because we discretize a continuous state space in order to calculate transition probability matrices. Previous attempts to characterise the intra–distribution dynamics of the city size distribution also face two other related problems. The first, is that there are a group of very large cities whose mobility characteristics may be different from the rest of the system. Including these cities may over–state the degree of persistence in the distribution. A simple solution would be to exclude those cities from the sample and recalculate the transition matrices. However, this brings us to the second problem, that the number of cities is such that we can only discretize the state space in to relatively few discrete states. For example, Black and Henderson report results for a five–state markov process, but the top state is occupied by the very immobile largest cities, leaving four states to capture the dynamics of the remaining cities. Such a limited number of states may lead us to underestimate the degree of mobility. Dobkins and Ioannides (1999) allow for ten discrete quantile states. However, that number of states leaves very few cities in each state, and mobility may be overstated due to the movements of a very few cities. Finally a large number of states for a small number of cities means that small changes in the discretization may lead to large changes in our estimates of the degree of persistence or mobility.

Quah (1996, 1997*a*,*b*) has proposed a set of tools for analysing evolving distributions which avoid the need to discretize the state space. Instead, he suggests calculating a non–parametric estimate of the underlying continuous transition kernel. Let f_t denote the density function of P_{it} at time t. Let us assume that the intertemporal evolution of f_t may be described in general by

$$f_{t+1} = \mathcal{M}(f_t, \varepsilon_{t+1}), \tag{5.3}$$

where $\mathcal{M}$ is an operator that maps (f_t, ε_{t+1}) to a probability measure, and ε_{t+1} is a appropriately defined stochastic function representing random shocks. E.g., the random growth model in Simon (1955) may be considered as a special case of processes consistent with specification (5.3). We may estimate the above law of motion of the evolution of city sizes in the form of estimating the probability distribution function of city i population in time $t+1$, conditional on its population at time t, $f(P_{i,t+1}|P_{i,t})$. Appendix C, presents technical issues necessary to establish that stochastic kernel estimation techniques may be used to estimate transition, and more generally mapping, probability functions.

Figure 5.1 presents a nonparametrically estimated transition kernel, along the lines of Equation 5.3 showing the extent of mobility for the periods 1910–1920 and 1980–1990.[7]

[7]All transition kernel and cross-profile plot calculations were performed with Danny Quah's tSrF econometric shell (available from `http://econ.lse.ac.uk/~dquah/`)

The underlying data at time t are the (logarithm of) population of each city relative to the mean city size at time t. these result illustrate the extent of mobility as we discuss shortly below.

We estimate the stochastic kernel as follows. First, we derive a non-parametric estimate of the joint distribution $f(P_{i,t}, P_{i,t+1})$, where $P_{i,s}$ is city population at time s. We then numerically integrate under this joint distribution with respect to $P_{i,t+1}$ to get the marginal distribution of population at time t - $f(P_{i,t})$[8]. Next we estimate the marginal distribution of market potential conditional on population size by dividing through $f(P_{i,t}, P_{i,t+1})$ by $f(P_{i,t})$. Thus we estimate $f(P_{i,t+1}|P_{i,t})$ by $\hat{f}(P_{i,t+1}|P_{i,t}) = \hat{f}(P_{i,t}, P_{i,t+1})/\hat{f}(P_{i,t})$. Under regularity conditions, this gives us a consistent estimator for the conditional distribution. See Rosenblatt (1971) and Yakowitz (1985) for details.

Figure 5.1 gives a selection of our results, for the periods 1910–1920 and 1980–1990. To interpret the diagram, take a cross–section from any point on the 1910 axis. The cross–section is the distribution of relative city sizes in 1920 conditional on the city size in 1910. Figure 5.1 shows that there is almost the same degree of mobility at both the beginning and end of the sample. We cannot directly test for the stationarity of the underlying markov process, although evidence from Black and Henderson (1998) suggest that the process is stationary. If so, we can pool across time periods to get a better estimate of the underlying transition process. Figure 5.2 shows two such pooled transition kernels. One for the entire US, and one for pooled data for regions, where each city's population is taken relative to the regional mean.

These stochastic kernels give a pictoral representation which allow us to compare mobility across samples and time periods. They suffer from two problems, however. First, when estimated for smaller samples, the degree of precision is reduced, giving the appearance of more mobility. Second, they do not give us statistics with which to compare mobility across different samples. We could discretize the state space, estimate transition matrices and calculate the standard mobility indices – but then we are back to the problems with the previous literature that we have highlighted above.

Instead, we proceed by reporting in Figure 5.3 the cross profile plots proposed by Dolado, González-Páramo, and Roldán (1994) and Quah (1997*b*). The top left hand corner of Figure 5.3 shows such a cross–profile plot for (the logarithm of) relative city sizes for the cities that 'exist' in 1920. Reading upward from the bottom of the figure the curves show the situation in 1920, 1940, 1960 and 1980 respectively. For 1920, cities have been ranked in order of increasing size, and the horizontal axis marks this ranking in a linear fashion. In 1920, the cross–profile plot is monotone rising. We then maintain this ordering for each of the subsequent years. Thus, the extent of choppiness (or jaggedness) depends on the degree of intra–distribution mobility. The shape of the plots gives us information on both types of mobility that we discussed above. If the

[8]We could also estimate the marginal distribution $f(P_{i,t})$ using a univariate kernel estimate. The asymptotic statistical properties of both estimators are identical, and in practice tend to produce very similar estimates.

cross–profile plots were always monotone rising. but the slope increased over time, then city population ranks are invariant, but the spread of the distribution is increasing. If the cross profile plot becomes jagged, then cities are changing rankings over time.

We can calculate a number of statistics which capture the features of the distribution that we have described above. We report these statistics in Table 5.1. The first measure, *Slope*, gives the (OLS estimated) slope of the (resorted) cross–profile plot at each point in time. This gives an idea of the degree of inequality in the city size distribution. When *Slope* has the value x, then being 10 cities larger means having a population e^{10x} times higher. For the whole sample, this measure decreases slightly over time, but stabilizes towards the end of the sample.

Additional insight in to the degree of intra–distribution mobility is provided by the measures *SerCorr* and *Variation* which capture the changing choppiness of the cross–profile plot. These measures are defined as follows. Let (r) denote the ranking in 1920 when the cities are ordered in terms of increasing size. Thus, $r = 1$ for the smallest city; $r = 2$ for the second smallest city, etc. Then, for each period, *SerCorr* is the first–order serial correlation coefficient of sequential changes across this ordering. If $p_{(r)}$ is the relative population of the city with rank r, then sequential changes are defined as the difference in relative sizes of the cities with two successive rankings:

$$\Delta^* p_{(r)} = p_{(r)} - p_{(r-1)}. \tag{5.4}$$

Then, *SerCorr* is defined as the 'serial' (along the ranking in 1920) correlation coefficient:

$$SerCorr = \frac{(\sum_r (\Delta^* p_{(r)} - E\Delta^* p)(\Delta^* p_{(r-1)} - E\Delta^* p)}{(\sum_r (\Delta^* p(r) - E\Delta^* p)}, \tag{5.5}$$

where $E\Delta^* p$ is the average of $\Delta^* p_{(r)}$. As with all correlation coefficients, the definition of *SerCorr* ensures that it lies between -1 and $+1$. If the cross–profile plot is a straight line, then *SerCorr* is zero regardless of the slope of that cross–profile plot. However, because it measures the rank–serial correlation in changes, it may be negative even when the cross–profile plot is monotonically rising. Usually, *SerCorr* differs from zero, because relative sizes do not differ uniformly across rankings. It becomes more and more negative when the choppiness of the cross–profile plot increases. It is positive when the cross–profile plot is monotone increasing, with with increasing slope, in a sense becoming more convex.

The other measure of intra–distribution mobility is *Variation* – the mean–square variation across successively ranked cities. If N is the number of cities, then:

$$Variation = (N-2)^{-1} \sum_r (\Delta^* p_{(r)} - \Delta^* p_{(r-1)})^2. \tag{5.6}$$

Variation is non–negative and becomes larger with increasing variance of the cross–profile plot. As with *SerCorr*, *Variation* is zero when the cross–profile is a straight line, regardless of the slope of the profile. However, it will be positive for all other cross–profile plots, regardless of whether they are monotonically increasing.

	Slope	*SerCorr*	*Variation*
1920	0.020	0.391	0.063
1940	0.019	−0.479	0.513
1960	0.017	−0.612	1.324
1980	0.015	−0.619	1.708
1990	0.015	−0.609	1.924

Table 5.1. Whole sample cross profile statistics

Table 5.1 provides these summary statistics for the cross–profile of all cities existing in 1920. The intra–distribution dynamics for the whole sample settle down rapidly: *SerCorr* has value -0.612, -0.619 and -0.609 in 1960, 1980 and 1990 respectively. One can see from the cross–profile plot that this does not mean that the profile is actually frozen in time. Rather, the ongoing churning of the distribution has characteristics that are stable. This is consistent with our earlier observation on the stationarity of the Markov process for city transitions. However, now we are directly examining the mobility properties of the entire distribution. Notice that *Variation* shows ongoing increase over time, capturing the increasing amplitudes of changes in relative sizes, as evidenced by the cross–profile plots in the upper left hand corner of Figure 5.3.

Our results for the cross–profile of cities that exist in 1920 suggests that the churning characteristics of the distribution are relatively stable over time and parameterised by a value of *SerCorr* around -0.6. Both these statistics and the stochastic kernels indicate the degree of mobility that characterises the evolution of the US city size distribution. Models that seek to explain the evolution of that distribution could use these figures as upper bound bench marks[9]. These tools can also be used to compare the mobility patterns of different groupings of cities. It is to this issue that we now turn.

4.2 Regional urban subsystems

A key issue, seldom addressed in the rank–size literature, is the appropriate definition of the urban system. Our approach allows us to characterise the degree of mobility within different urban subsystems. Here, we demonstrate the technique by considering the evolution of nine subsystems defined by the Census Bureau's regional divisions[10].

Figure 5.2 shows the stochastic kernel for the evolution of city size relative to the average city size of cities in the same region. This stochastic kernel is estimated assuming that the transition process is stationary over time and identical across regions. This

[9]We would argue upperbound, as the actual urban system is hit by shocks that presumably increase mobility relative to the underlying economic mechanisms captured by current theoretical models.

[10]The nine regions are New England (ned); Middle Atlantic (mad); South Atlantic (sad), East South Central (escd); East North Central (encd), West North Central (wncd); West South Central (wscd); Mountain (mtd); Pacific (pad) These regions may not correspond exactly to functional urban subsystems. However, they provide a convenient division that allows us to demonstrate the general approach.

	Slope	*SerCorr*	*Variation*
encd			
1920	0.084	0.279	0.169
1940	0.087	0.117	0.276
1960	0.085	−0.355	0.589
1980	0.083	−0.529	0.788
1990	0.085	−0.544	0.867
escd			
1920	0.213	−0.304	0.185
1940	0.203	−0.276	0.209
1960	0.190	−0.879	0.468
1980	0.148	−0.296	0.396
1990	0.142	−0.397	0.449
mad			
1920	0.159	0.448	0.250
1940	0.154	0.240	0.388
1960	0.146	−0.013	0.612
1980	0.134	−0.355	0.945
1990	0.128	−0.272	0.983
ned			
1920	0.211	−0.025	0.497
1940	0.212	0.019	0.549
1960	0.241	−0.346	1.206
1980	0.219	−0.352	1.210
1990	0.226	−0.400	1.332
pcd			
1920	0.242	−0.427	0.581
1940	0.222	−0.461	1.117
1960	0.162	−0.243	1.318
1980	0.120	−0.216	1.395
1990	0.095	−0.237	1.489
sad			
1920	0.090	0.376	0.148
1940	0.082	0.048	0.329
1960	0.079	−0.254	0.882
1980	0.073	−0.275	1.265
1990	0.074	−0.304	1.549
wncd			
1920	0.198	−0.071	0.308
1940	0.193	−0.296	0.484
1960	0.188	−0.672	1.120
1980	0.184	−0.622	1.255
1990	0.188	−0.671	1.579
wscd			
1920	0.120	−0.154	0.222
1940	0.134	−0.857	0.781
1960	0.148	−0.821	1.032
1980	0.164	−0.797	1.539
1990	0.172	−0.713	1.601

Table 5.2. Regional systems cross profile statistics

allows us to pool observations across both dimensions. It appears that the pattern of mobility of cities within their regional subsystems is not much different from the pattern of mobility relative to the US-wide average city size. However, remember that this result is conditional on the assumption that we can pool observations across both regional subsystems and across time. The results of the cross–profile plots suggest that this is not a valid assumption. In fact, there may actually be substantial differences between regional subsystems.

Figure 5.3 shows the cross–profile plots for the nine regions. The cross–profile plots are for the years 1920, 1940, 1960 and 1980 as before. Because of the varying numbers of cities in each region the plots are hard to compare visually. However, some stark differences do immediately jump out. For example, compare the cross–profile plots for the South Atlantic and Mid Atlantic regions. Both regions have similar numbers of cities[11], but the transition dynamics appear very different. The measures in Table 5.2 allow a more direct comparison. For example, we see that the visual impression that the Mid Atlantic region shows more churning than the South Atlantic, is actually driven by higher variation, rather than increased churning. Thus *SerCorr* has similar values for the two regions, but *Variation* is much higher for the South Atlantic. To take another example, we see that the West South Central and West North Central districts show a higher level of churning than all the other regions. The cross–profile plots suggest some evidence that there are differences in intra–distribution mobility within the different regional subsystems. Some areas of the US have urban systems that are characterised by far higher intra–distribution mobility.

4.3 City tiers

Classical hierarchical theories of cities divide cities in to tiers, depending on the functions of each city. More recent theoretical work has incorporated insights from this older literature in to the new economic geography literature. These theoretical analyses suggest that the highest tier cities, which are most diversified, may display different patterns of evolution from lower tier cities. See Fujita *et al.* (1999*b*) for details. In this section, we examine whether the intra–distribution dynamics do appear to differ substantially among tiers.

In order to construct the tiers, we took as our basic classification a listing of US cities by 'function' (nodal centers) from Knox (1994). We amended the top tier slightly to include Atlanta, Chicago, Denver, Houston, Los Angeles, New York City, Miami, San Francisco, Seattle and Washington D.C.. The next classification is the regional nodal centres, which includes fourteen large cities: Baltimore, Boston, Cincinnati, Cleveland, Columbus OH, Dallas, Indianapolis, Kansas City MO, Minneapolis, New Orleans, Philadelphia, Phoenix, Portland OR and St. Louis MO. The third classification is the sub–regional nodal centres. This comprises nineteen cities: Birmingham, Charlotte,

[11] 24 and 21 respectively.

	Slope	*SerCorr*	*Variation*
top tier			
1920	0.431	0.050	0.560
1940	0.394	−0.353	1.059
1960	0.302	−0.441	1.077
1980	0.191	−0.475	1.040
1990	0.169	−0.546	1.094
second tier			
1920	0.212	0.032	0.286
1940	0.181	−0.247	0.306
1960	0.123	−0.204	0.346
1980	0.073	−0.162	0.404
1990	0.055	−0.191	0.497
third tier			
1920	0.150	0.360	0.243
1940	0.132	−0.025	0.465
1960	0.109	−0.232	0.782
1980	0.082	−0.426	1.073
1990	0.073	−0.462	1.163
fourth tier			
1920	0.019	−0.334	0.036
1940	0.017	−0.625	0.491
1960	0.012	−0.649	1.279
1980	0.010	−0.636	1.597
1990	0.010	−0.632	1.828

Table 5.3. Functional tiers cross profile statistics

Des Moines, Detroit, Hartford, Jackson MS, Little Rock, Memphis, Milwaukee, Mobile, Nashville, Oklahoma City, Omaha, Pittsburgh, Richmond, Salt Lake City, Shreveport, Syracuse and Tampa. The remaining cities are placed in the lowest tier.

From the classification, it is clear that the number of cities in the different tiers differs substantially between tiers. Thus the top tier comprises ten cities, the second tier fourteen cities, the third tier nineteen cities and the lowest tier the remaining 291 cities. With such small numbers of cities within tiers, it makes no sense to calculate stochastic kernels for each tier. Instead, in Figure 5.4 we show the cross–profile plots for each of the four tiers. Table 5.3 gives the corresponding measures.

The table shows that the top tier actually shows a surprising degree of mobility. Looking at the cross–profile plot suggests that this mobility is mainly due to changes in the relative sizes and rankings of cities at the lower end of the tier. By 1940, the rankings of the top four cities appears set, although they still display mobility with respect to relative sizes[12]. For the bottom five cities, there is a surprising degree of mobility both in terms of rankings and relative sizes. Results for the second tier are again surprising.

[12]Relative sizes are now defined with respect to the average city size for cities in the same tier.

It is actually this second tier of cities that show remarkable stability, both in terms of relative size and rankings. The measures and the shape of the cross–profile plot show that this is easily the most stable subsystem that we have studied. Mobility patterns for the third tier lie somewhere between the first and second. Finally, the fourth tier shows the highest degree of mobility. In standard analysis using transition probability matrices, nearly all the action for the top three tiers would be disguised by the fact that they all fall in the top discrete state. Our results here suggests that there are interesting differences in mobility for subsystems of cities that usually fall within this highest state.

5. Conclusions

This chapter has studied a number of aspects of intra–distribution mobility for the US city size distribution. Characterising the nature of such intra–distribution mobility should help guide the two different theoretical strands that seek to explain the evolution of urban systems. For the literature that attempts to generate urban systems that obey the rank size rule, these results provide benchmarks for the upper level of intra–distribution mobility that would ensure these models are consistent with real world intra–distribution dynamics. The results on regional and tier subsystems also throws up questions for the literature that tries to model the economic mechanisms that may govern the evolution of urban systems. Are their economic forces that can explain the apparent differences in the nature of intra–distribution mobility between different regional subsystems? More interestingly, what explains the apparent stability of the second tier of cities relative to the other two higher tiers?

Table 5.4. Descriptive statistics: decennial data 1900–1990

1	2	3	4	5	6	7	8
Year	US Pop. (000)	US Urban Pop. (000)	Mean Size	Median Size	GNP billion $	Distance miles	Nearest miles
1900	75,995	29,215	259952	121830	71.2	802.5	70.9
1910	91,972	39,944	286861	121900	107.5	863.8	68.3
1920	105,711	50,444	338954	144130	135.9	864.0	66.2
1930	122,775	64,586	411641	167140	184.8	876.9	64.8
1940	131,669	70,149	432911	181490	229.2	884.9	64.4
1950	150,697	85,572	526422	234720	354.9	890.8	65.3
1960	179,323	112,593	534936	238340	497.0	940.4	56.9
1970	203,302	139,419	574628	259919	747.6	981.3	52.5
1980	226,542	169,429	526997	232000	963.0	998.7	45.9
1990	248,710	192,512	577359	243000	1277.8	1005.3	45.5

All figures are taken from *Historical Statistics of the United States from Colonial Times to 1970*, Volumes 1 and 2, and *Statistical Abstract of the United States, 1993*. Column 6: GNP adjusted by the implicit price deflator, constructed from sources above; 1958=100. Column 7 gives the average distance to all other cities. Column 8 gives the average distance to the nearest city. Distances are calculated as great circle distances based on latitudes and longitudes from the Times Atlas 1999 edition (See Appendix D).

Table 5.5. Descriptive statistics for all cities – 1990

Variable	Mean	Std. Dev.	Skewness	Kurtosis	Min	Max
Population (000)	479.5	1001.5	6.6	58.8	50.7	9,372.0
Log(Population)	12.4028	0.9895	1.0	4.1	10.8343	16.374
Growth Rate (%)	10.62	41.98	-1.1	5.8	-.999	1.8752
Education (%)	57.1085	20.9284	-0.4	1.8	11.80	92.73
Real Wage ($)	3197.92	1132.37	0.2	2.3	1020.00	7311.00
New England	8.8%					
Mid Atlantic	12.8%					
South Atlantic	16.7%					
East North Central	20.3%					
East South Central	6.6%					
West North Central	9.1%					
West South Central	12.2%					
Mountain	4.6%					
Pacific	8.8%					

Data on education and real wage are taken from *Historical Statistics of the United States from Colonial Times to 1970, Vol. 1 and 2, and* Statistical Abstract of the United States, 1993. Educational percentage refers to the mean percent of 15 to 20 age cohort in school. Mean real annual earnings, by city proper or metro area, are in dollars, deflated by the Consumer Price Index, 1967 = 100.

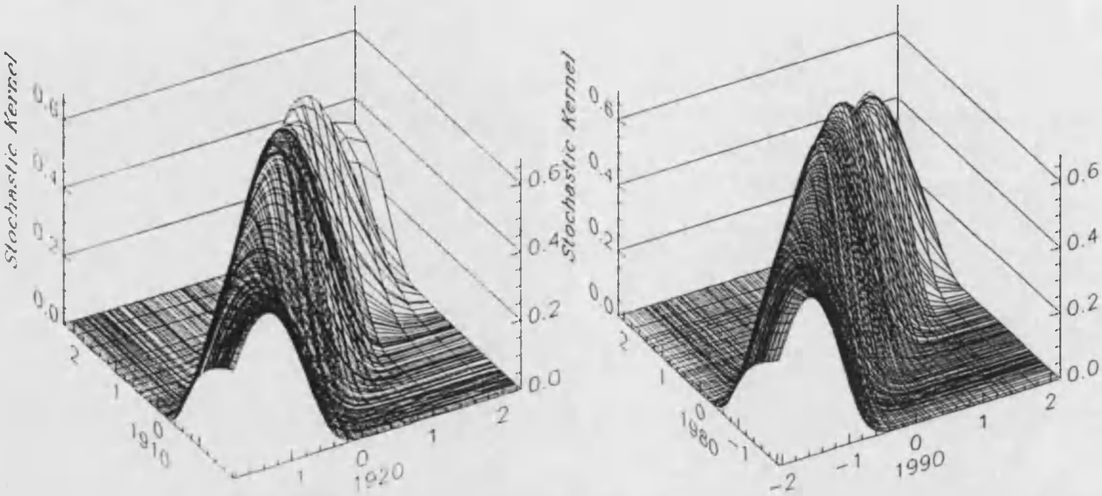


Figure 5.1. Selected decades transition kernels

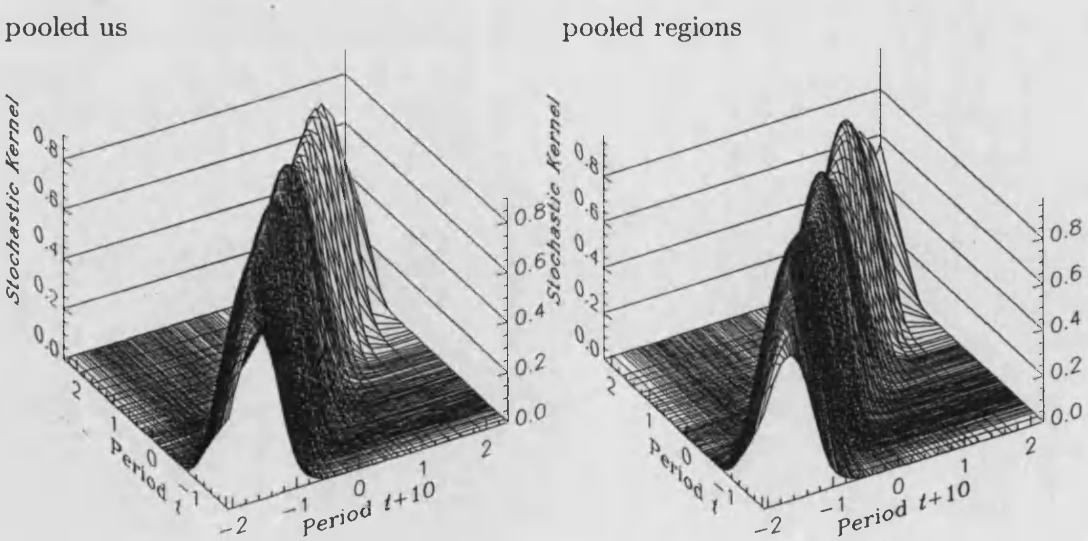
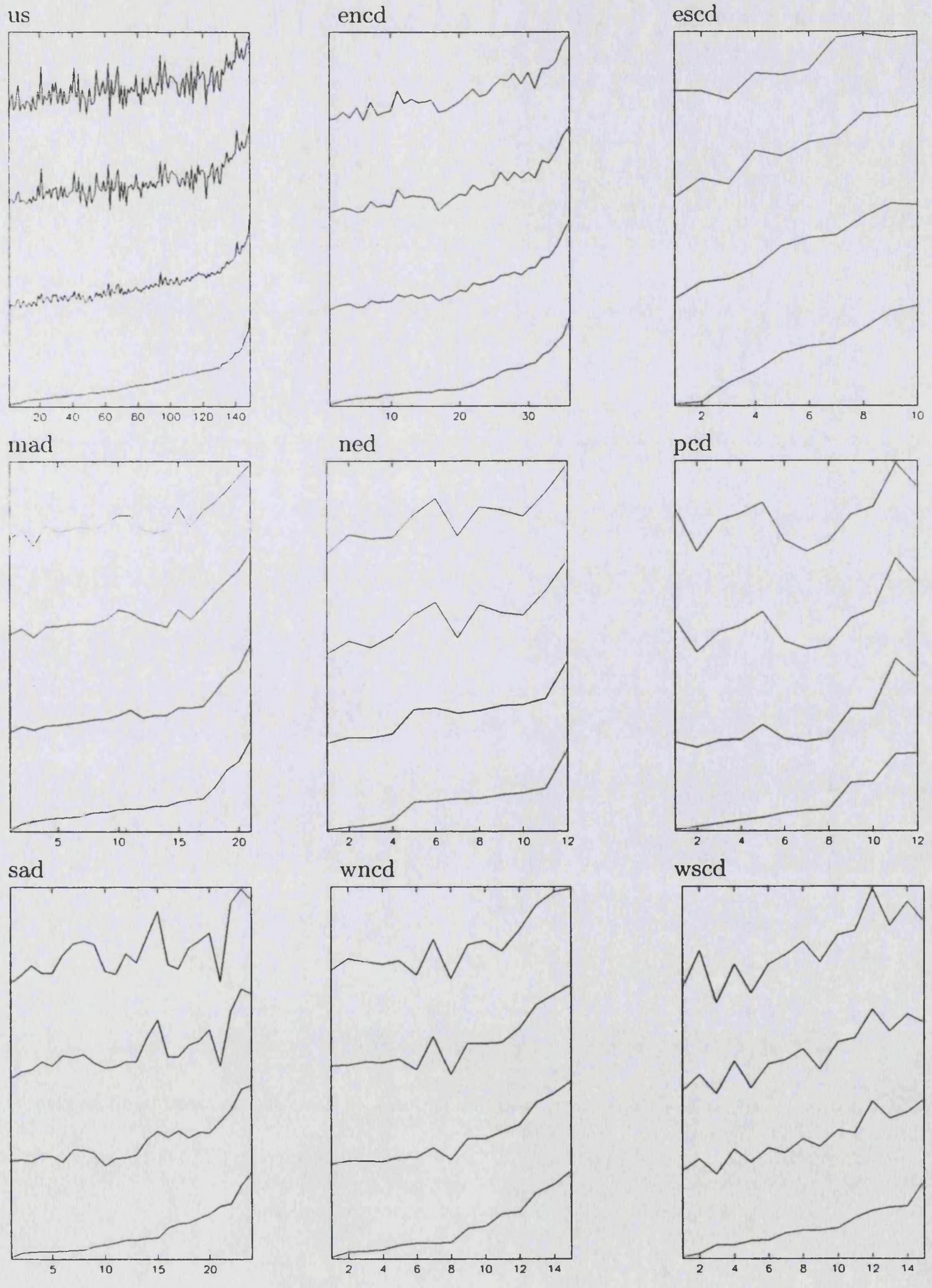
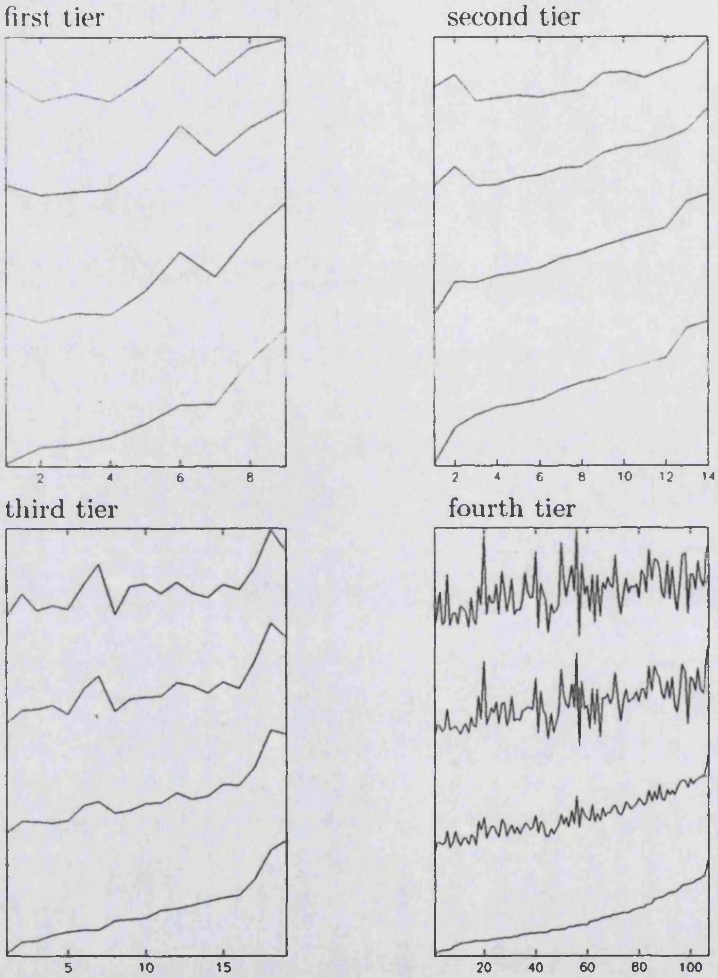


Figure 5.2. Pooled US and regions transition kernels



For each diagram, the lowest line gives the cross-profile for 1920, then moving upwards, the lines give cross-profiles for 1940, 1960 and 1980, respectively.

Figure 5.3. US and regional cross profile plots



For each diagram, the lowest line gives the cross-profile for 1920, then moving upwards, the lines give cross-profiles for 1940, 1960 and 1980, respectively.

Figure 5.4. Functional tiers cross profile plots

6

The spatial evolution of the US urban system

with Yannis Ioannides

1. Introduction

Questions pertaining to the location of economic activity, to the relative sizes of cities in different countries, and to changing roles for different geographical areas in the process of economic growth have attracted considerable interest recently. Work by several theorists, who developed the so-called new economic geography, including recent contributions by several researchers, but in particular by Masahisa Fujita, Paul Krugman and Anthony Venables [Fujita *et al.* (1999*b*)] have added important new spatial insight to the established system of cities literature, represented most notably by the research of Henderson (1974, 1988). The system of cities approach features powerful models of the intra-metropolitan spatial structure, but lacks an explicit model of inter-metropolitan spatial structure. In contrast, certain aspects of the inter-metropolitan spatial structure have played a key role in the new economic geography literature, as, for example, in Krugman (1991). Krugman (1998) provides an excellent overview of this literature. Tabuchi (1998) proposes a step towards a synthesis of the older system of cities literature with the newer economic geography based theories, by incorporating intra-metropolitan commuting costs in addition to inter-metropolitan transport costs.

This chapter attempts to examine empirically some of the spatial aspects of the evolution of the US system of cities. The evidence that we consider falls in to two broad categories. First, we consider spatial features of the system about which theory is relatively silent. This includes, for example, evidence on the co-evolution of the distribution of market potentials and relative city sizes. Second, we consider spatial features of the system with which the theory deals more directly. This includes, for example, evidence on the relationship between relative growth rates and relative market potentials. This second set of results could be characterised as 'tests' of the new economic geography. Care is needed here, however, as the results that we get are also consistent with other models of the evolution of the urban system.

This chapter builds on work by Dobkins and Ioannides (1998) and Black and Henderson (1999). Dobkins and Ioannides (1998) examine the basic dynamics of spatial interactions among US cities and its impact on their populations. They use a data set on US metro areas, which spans this century from 1900 to 1990, to look at patterns of city growth and the distribution of city sizes as new cities enter the distribution. They emphasise that entry of new cities is an important characteristic of the United States system of cities. The key spatial characteristics they consider are the presence of neighboring cities, regional influence, and distance between a city and the nearest one in a higher tier. The present chapter takes a broader view of temporal *cum* spatial interactions by estimating models of joint dynamic and spatial interdependence that do not restrict intercity interactions through notions of adjacency. Black and Henderson (1999) consider the importance of both first and second nature geography in explaining the growth rates of cities. First nature characteristics are those that are intrinsic to a site. Second nature characteristics are a result of the spatial structure of the economic system. See Section 4 for more discussion. They find that both factors are important in explaining city growth. Our results in Section 4 build directly on their analysis. However, we consider issues relating to second nature geography in far more detail, as this is our primary interest in this chapter.

The chapter is organized as follows. Section 2 briefly describes the data set. Section 3 characterises some initial spatial features of the US urban system. Section 4 outlines the theoretical predictions from the new economic geography that relate directly to the spatial evolution of the urban system. Empirical models are developed, and we find that the data are consistent with some key implications of the new economic geography models. Section 5 concludes.

2. Data

There are a variety of ways to define cities. In this chapter we use contemporaneous Census Bureau definitions of metropolitan areas, with adaptations for availability. From 1900 to 1950, we have metropolitan areas defined by the 1950 census. That is, for years previous to 1950, we use Bogue (1953) reconstructions of what populations would have been in each metropolitan area in each year if the cities had been defined as they were in 1950. For each decennial year from 1950 to 1980, we use the metropolitan area definitions that were in effect for those years. Between 1980 and 1990, the Census Bureau redefined metropolitan areas in such a way that the largest US cities would seem to have taken a huge jump in size, and several major cities would have been lost. While this might be appropriate for some uses of the data, it would introduce 'artificial' differences in growth patterns for the 1980–1990 period. Therefore, we reconstructed the metro areas for 1990, based on the 1980 definitions, much as Bogue did earlier. We believe that this gives us the most consistent definitions of US cities (metropolitan areas) that we are likely to find.

The method raises a question as to which cities, as defined or reconstructed, should be included. In the years from 1950 to 1980, we use the Census Bureau's listing of metropolitan areas. Although the wording of the definitions of metropolitan areas has changed slightly over the years, the number 50,000 is minimum requirement for a core area within the metropolitan area. Therefore, we used 50,000 as the cutoff for including metropolitan areas as defined by Bogue prior to 1950. Consequently we have a changing number of cities over time, from 112 in 1900 to 334 in 1990. While it is often difficult to deal with an increasing number of cities econometrically, we think that this is a key aspect of the US system of cities, and is worthy of being factored into our studies.

We also have data on earnings in all cities in the sample for all years, drawn from Census reports, although the data set is not ideal because the Census Bureau changed the categories it reported over the years. We have data on schooling in each city over the century, reported as the percentage of the population in the 15 to 20 year old category who are in school. We also have data on regional location according to the Census Bureau division of the country into nine regions. See Dobkins and Ioannides (1999) and Chapter 5 for more details.

3. Spatial features of the US urban system

The key empirical implication common to the newer theoretical frameworks is a prediction that the dynamic evolution of wages and population reflects spatial considerations[1]. In this section, we seek to characterise certain spatial characteristics of the evolution of the US system. Theory suggests that there are complex interactions between spatially dispersed economic agents, with those interactions partly governed by distances between the location of those agents. We use tools developed by Quah (1996, 1997*a*,*b*) to characterise some key aspects of that evolution. We start with the relationship between the distribution of city sizes and market potential.

Most explicitly spatial models of the urban system predict a relationship between city size and market potential. Market potential is supposed to capture the importance of demand from other cities or regions while allowing for the 'friction of distance'. The models suggest that market potential should be a function of city incomes, distances between cities and the city price indices for manufactured goods. Making such a definition operational is difficult given the data that we have available. Some authors have estimated models with market potential measures as explanatory variables. For an early example, see Harris (1954). Krugman (1992) showed that his own economic geography model may be interpreted as justifying the market potential concept. Hanson

[1]See, for example, chapters eight to thirteen in Fujita *et al.* (1999*b*).

(1998) is the only paper that estimates a Krugman–type model as a structural form and compares it with market potential model, using data for counties for much later time periods.

Due to data availability we are forced to move to a simpler definition of market potential. We use three different definitions of market potential for city i based on the following formula:

$$mp_{it} = \sum_{j \neq i} \frac{P_{jt}}{D_{ij}}.$$

The first two measures differ depending on whether the summation is across all cities or all counties in the US. In words, city i's market potential is the sum over all other cities (counties) j of population in city (county) j $[P_{jt}]$, weighted by the distance between i and j $[D_{ij}]$. Various weighting functions are possible. In this chapter we present results where each city's (county's) weight is inversely proportional to it's distance from city i. When the summation is across all cities we will refer to this as city based market potential and when across counties as county based market potential[2]. Taking different definitions is interesting for two reasons. First, it allows us to see whether spatial interactions between cities differs from general spatial interactions between cities and other (non–city) locations in the US. Second, we have wage data for cities back till 1900. We do not have similar information for counties. This data allows us to construct a third measure of market potential where cities are weighted by average wages as well as distance. This measure may better capture the importance of demand from other cities and regions than the measures that only consider population.

At present, we have to make a somewhat arbitrary choice on the importance of distance. That is, whether distance should enter linearly, or whether it is more important (squared – say), or less important. Results do not appear to be too sensitive to these assumptions. For example, the GMM results that we report in Section 4 are not markedly different if we weight by the square root of distance – although the degree of variation in market potential is substantially reduced and we tend to see higher standard errors. It would also be possible to allow for the effect of distance to decrease through time. However, the changing composition of consumption from manufacturing to services, means that, at an aggregate level it is not clear whether general transport costs have risen or fallen over time. Thus, Hanson (1998) finds that transport costs actually appear to have increased between 1970 and 1980. Without further data on actual transport costs, we have chosen to take the "neutral" viewpoint that general transport costs are unchanged over the sample period. Further, in common with many authors, we assume that transport costs are directly related to the distance between cities without any consideration of actual transport networks and costs. Again, without any further

[2] For the county based market potential measure, note that the sum is over all counties that are not part of that metropolitan area in 1990.

information on transport costs over the period, it is unclear what alternative assumption would be better.

As the urban system develops, the distribution of population and the distribution of market potential co–evolve. We are interested in examining the nature of this co–evolution. To do this we estimate a series of stochastic kernels which give us the distribution of one of the variables conditional on the other variable.

We actually report results for five sets of kernels – city size conditional on city and county based market potential; city and county based market potential conditional on city size; and relative wages conditional on city market potential. To understand the construction of the stochastic kernel, consider the kernel showing the distribution of city size conditional on market potential. To estimate that stochastic kernel, we first derive a non–parametric estimate of the joint distribution $f(x,y)$, where y is city population and x is market potential. We then numerically integrate under this joint distribution with respect to y to get the marginal distribution of market potentials $f(x)$[3]. Next we estimate the marginal distribution of city size conditional on population size by dividing through $f(x,y)$ by $f(x)$. Thus we estimate $f(y|x)$ by $\hat{f}(y|x) = \hat{f}(x,y)|\hat{f}(x)$. Under regularity conditions, this gives us a consistent estimator for the conditional distribution for any market potential x. The stochastic kernels plot this conditional distribution for all values of x.

All variables are relative. That is, they are normalised by sample means as follows:

$$RPOP_{i,t} = pop_{i,t}/\overline{pop}_t$$

$$RMP_{i,t} = mp_{i,t}/\overline{mp}_t$$

Where $\overline{pop}_t$ is the mean population in time t; $\overline{mp}_t$ is the mean market potential in time t[4].

Relative city sizes vary dramatically across the US. At points in the sample period, New York is up to 25 times the mean city size (1930). Including these very large cities is conceptually simple, but technically problematic. Even though we consider log relative city sizes, very large outliers automatically drive up the optimal bandwidth that we use to nonparametrically calculate the stochastic kernels[5] When this happens, the detail in the lower end of the distribution (comprising the main body of cities) is obscured, as the estimates are over–smoothed. We have tried two different solutions to this problem. First, we restricted the sample range to all cities below a certain size. Second, we restricted the cities that we consider on the basis of a functional urban hierarchy classification. We used such a classification in Chapter 5 and showed that

[3] We could also estimate the marginal distribution $f(x)$ using a univariate kernel estimate. The asymptotic statistical properties of both estimators are identical, and in practice tend to produce very similar estimates.

[4] We also normalise the wage weighted market potential variable.

[5] The optimal bandwidth is based on Silverman (1986) and is a function of the range or the variance whichever is the larger.

there were some differences in intra–distribution mobility across different tiers in the urban system. In fact, it turns out that the two methods deliver very similar results. The results that we present in this section use the restriction implied by the urban hierarchy classification.

reference

In order to construct the tiers, we took as our basic classification a listing of US cities by "function" (nodal centers) from Knox (1994). We amended the top tier slightly to include Atlanta, Chicago, Denver, Houston, Los Angeles, New York City, Miami, San Francisco, Seattle and Washington D.C.. The next classification is the regional nodal centres, which includes fourteen large cities: Baltimore, Boston, Cincinnati, Cleveland, Columbus OH, Dallas, Indianapolis, Kansas City MO, Minneapolis, New Orleans, Philadelphia, Phoenix, Portland OR and St. Louis MO. The third classification is the sub–regional nodal centres. This comprises nineteen cities: Birmingham, Charlotte, Des Moines, Detroit, Hartford, Jackson MS, Little Rock, Memphis, Milwaukee, Mobile, Nashville, Oklahoma City, Omaha, Pittsburgh, Richmond, Salt Lake City, Shreveport, Syracuse and Tampa. The remaining cities are classed in the lowest tier. See Chapter 5 for more details.

We represent the resulting estimates graphically. We start with the results for the distribution of city sizes conditional on market potential. Figure 6.1.1 is the stochastic kernel for the bottom tier showing the distribution of relative city size conditional on county based market potential for 1910[6]. The stochastic kernel for the top three tiers looks similar but is not reported. The way to interpret this stochastic kernel is as follows. If you take a point on the relative market potential axis, say 0.0, and cut across the stochastic kernel parallel to the relative city size axis, this gives the conditional distribution of relative city sizes for cities with mean county based market potential. The stochastic kernel plots these conditional distributions for all values of market potential.

From Figure 6.1.1 we see that the 1910 kernel is somewhat skewed towards the diagonal. At the start of the period, the smallest cities tend to have smaller market potentials and larger relative city size is associated with larger relative market potential. To see this, observe that for 1910, there is a peak in the stochastic kernel, centred in the lower southwest corner, which contains most of the mass for the smaller cities. In contrast the conditional distribution for the largest cities is relatively flat. Figures 6.1.2 to 6.1.5 present a series of snapshots showing the same stochastic kernel for the years 1930, 1950, 1970 and 1990 respectively. The sequence of pictures clearly shows the stochastic kernel slowly twisting back until, by 1990, relative city sizes appear to have become virtually independent of the relative market potential. The peak becomes less and less pronounced, as the distribution of city sizes conditional on low market potential shows greater variance. By 1990, the conditional distributions of relative city sizes are almost identical across all values of relative market potential. Only for the very largest

[6]The picture looks similar for 1900, but city entry means that in 1910 we estimate on 98 cities rather than 74.

cities is city size positively related to market potential. Again, the pictures for the upper three tiers are not included, but show a similar evolution. Simple regressions of population on market potential (not reported) find a similar relationship for $E(y|x)$ although endogeneity means that the coefficients are hard to interpret. This is an interesting finding in that it suggests, at least from the non-parametric picture, that the distribution of city sizes conditional on market potential is nearly independent of relative market potential.

Similar results hold for the distribution of city size conditional on city based market potential. Figure 6.2.1 and 6.2.5 show that for both 1910 and 1990 the conditional distribution of city size is virtually independent of relative market potential. Again, the only exception is for the very largest cities in the lowest tier, where market potential is positively related to relative city size.

We can also consider the distribution of market potential conditional on city size for the city based market potential. Figure 6.3.1 is the stochastic kernel for the bottom tier showing the distribution of relative city based market potential conditional on relative population size for 1910. The sequence of pictures clearly shows the stochastic kernel slowly twisting back until it is virtually parallel with the relative population axis. By 1990, market potential appears to be "independently" distributed with respect to relative population. The results in Figure 6.4 suggests that this result is not driven by our restriction of market potential. These kernels show the distribution of relative county based market potential conditional on city size. The result for county based market potential is actually somewhat stronger: the only noticeable difference is that in 1910 cities that were relatively smaller showed a larger variance in relative market potential, than cities that were relatively large.

Finally, we briefly consider the co–evolution of market potential and the distribution of wages across cities. We use the stochastic kernel approach to look at the evolution of relative wages conditional on relative market potential. Figure 6.5 shows the resulting stochastic kernels for 1910 and 1990. Economic geography models suggest that the relative wage distribution should be skewed towards the diagonal. Cities with high relative market potential should have high relative wages. This prediction is not confirmed by the 1910 data. Wages are relatively high for cities with low market potential. This is more clearly seen from the contour plot in Figure 6.6[7]. As the system develops the relationship changes. The stochastic kernel is slowly twisted towards the diagonal with higher wages associated with larger market potential. This finding agrees with a backward linkages interpretation of the Krugman model, as in Fujita *et al.* (1999*b*) (or Krugman (1993))) namely that the value of labor is higher in locations which are 'closer' in terms of transport costs to areas with high consumer demand.

We note, however, that the weakly positive relationship implied by our finding is

[7]The contours work exactly like the more standard contours on a map. Any one contour connects all the points on the stochastic kernel at a certain height.

actually consistent with the broad implications of what Krugman calls the "no black-hole" condition: increasing returns, which are responsible for the backward linkages effect, must not be too strong, or else all economic activity would concentrate in one location. In fact, Hanson's results are broadly consistent with this condition. Second, urban congestion, too, can dampen the benefits from agglomeration.

Our initial results suggest that there is no simple relationship between market potential and city size. Indeed, in this section, we have shown that the co–evolution of the city size distribution and market potential may actually conflict with traditional views on the forces driving the evolution of the city size distribution[8]. We now consider parametric formulations which allow us to examine the same relationship.

4. Growth and the spatial structure of the urban system.

Given the previous results on the evolution of the distribution of city sizes, we look here at the relationship between city growth rates and the spatial structure of the urban system. The basic economic geography story suggests that cities with the highest market potential should grow fastest. Newer versions of this story suggest that the effects of high market potential on city growth are not necessarily monotonic. A city that is very close to a big city will have high market potential, but may fall within the agglomeration shadow of the bigger city [See Fujita *et al.* (1999*b*)]. Thus, a–priori we cannot say whether higher market potential is good or bad for growth[9].

Both first nature and second nature characteristics of city locations are presumably important for understanding the relative growth rate of cities. First nature characteristics are those that are intrinsic to a site. For example, good climate, good access to raw materials and a natural harbour are all first nature characteristics. Second nature characteristics are a result of the spatial structure of the economic system. For example, the distribution of market potential, the distribution of wages and the positioning of neighbours are all second nature characteristics. Our main interest is in the importance of second nature variables.

To avoid constructing information on first nature variables we use the fact that we have a panel of cities and absorb all first nature variables in the fixed effect for any given city. Thus, we are assuming that the effect of a "good" site on growth rates is constant over the entire time period. After absorbing first nature factors into the fixed effects, we are left with a group of time–varying second nature variables that we think may influence the growth rate of cities.

[8]See, for example, Harris (1954).

[9]We have not yet found a satisfactory solution to this problem. Black and Henderson (1999) using a quadratic form in a similar specification find that there appears to be a negative relationship between growth and market potential at the very top of the market potential distribution. This result is suggestive, but does not get around the problem that trade models predict that the coefficient on market potential will vary as a function of the distance from the cities casting agglomeration shadows. Thus high and low growth rates are consistent with high market potential.

The first type of variables are the different normalized market potential measures. Again, as in Section 3, we may want to consider market potential calculated on the basis of either cities or counties, with or without weighting by wages.

The second type of variable is a dummy variable for entry of a neighbouring city. As the urban system grows, new cities reach the threshold size of 50000 which is necessary for inclusion in our sample. Thus, our sample is characterized by "entry" of new cities. So, for example, in 1900 we have 112 cities, by 1990 there are 337 cities. City entry occurs in all census years although, more cities enter towards the end of the period. This is hardly surprising for two reasons. First, is our choice of an absolute cut–off point for city definitions. In some senses this is a "higher" hurdle at the beginning of the period. Second, is that we would expect the growing rate of urbanization towards the end of the sample to result in a faster rate of city creation. It is interesting to examine the effect of city entry on the growth rates of the surrounding cities. Some versions of the new economic geography models predict bifurcation of the city system as the system grows [See Fujita, Krugman, and Mori (1999*a*)] . When a new city enters, these models predict that the population size of it's nearest neighbour will decline. As absolute population declines are rare in the data we do not test for this strict result. Instead, we consider a "growth equivalent". It may be possible that when a city enters close to an existing city, that the existing city does not grow as fast as we would predict given the levels of the other explanatory variables. The entry dummy tries to capture this effect. It is defined as follows:

$\mathrm{ED}_{it} = 1,$ if city i is the nearest neighbour to a newly entering city at time t;

$\mathrm{ED}_{it} = 0,$ otherwise.

The third type of variable that we consider is the lagged population size of a city. Again, a–priori it is hard to predict the impact of lagged population size on city growth. Convergence type reasoning would suggest that lagged population size should be negatively related to growth. However, if we think of own city size as a proxy for "self–potential", then we would expect lagged population size to be non–negatively related to growth. This would then take account of the fact that the size of the home market is excluded from our calculation of market potential.

Finally, we consider the interaction between own city size and market potential. Some new economic geography models suggest that it is actually the ratio of city size to market potential that is important for city growth. Cities enter the urban system at sites where market potential reaches some threshold. That threshold is established relative to the high market potential of existing cities. Thus when cities enter, they will be small relative to the high value market potential at the site where they enter. When cities are small relative to the market potential of their site, they grow quickly. In the theory this fast growth takes the form of a bifurcation of the urban system. Small cities grow very (infinitely) fast at the cost of larger cities that loose population. We discussed this above with reference to the entry variable. Pushing this theoretical proposition somewhat, we would expect to see fast city growth when market potential is large relative to current city size.

Before turning to the results, we briefly summarise our discussion above:

- City growth should be a function of market potential. Traditionally, models predicted that market potential should have an unambiguous, positive, effect on growth. New economic geography models suggest that large cities may cast agglomeration shadows, which make the effects of market potential on growth ambiguous.

- City growth should be effected by the entry of other cities. In traditional models city entry should have a positive effect on growth, working through increases in market potential for the existing city. New economic geography models suggest that entry should have a negative effect on the growth rate of nearby cities. Strictly, city entry represents a bifurcation of the urban system and should lead to absolute population decline in nearby cities.

- Own lagged city size has an ambiguous effect on growth. Models that predict convergence of city size predict a negative impact of own lagged city size on growth [as do some new economic geography models]. Models that emphasises intra as well as inter–metropolitan distance also may predict a negative effect of own lagged city size on growth. This reflects congestion forces internal to the city that may reduce growth rates. Finally, some models predict a positive impact of lagged city size on own growth. This positive impact may reflect the fact that own lagged city size is a measure of self–potential and thus should have a positive impact on growth

- New economic geography models that consider the spatial evolution of the urban system allowing for endogenous entry make clearer predictions about the ratio of own city size and market potential, than they do about the effect of either variable separately. A city should grow fast when it is small relative to it's market potential.

4.1 The distribution of growth rates

In Section 3 we used stochastic kernels to look at the co–evolution of the distribution of city sizes and various measures of market potential. We can also use this approach to look at the relationship between the distribution of growth rates, and the distribution of measures of market potential. The discussion above suggested that we may want to condition out first nature variables that may make some cities grow faster than others independent of second nature geography. To do this, we consider the difference between this periods relative growth rates and the (time) average of growth rates for that city. We also do the same for relative market potential. Figure 6.7 shows a stochastic kernel for the distribution of relative growth rates conditional on the distribution of relative market potentials. This figure is for 1920 – showing the growth rate in the decade between 1910 and 1920 (relative to the long run growth rate) conditioned on the market potential of the city in 1910 (again, relative to the long run average market potential). The picture on the right hand side gives the contour plot corresponding to the 3D stochastic kernel plot on the left hand side. The complex nature of the stochastic kernels mapping relative growth rates to relative market potentials means that these contour plots are actually more informative than the corresponding 3D plots. The remaining pictures in Figure 6.7 show the contour plots for 1940, 1960 and 1980. These plots suggest that there is no simple stable relationship between the distribution of relative growth rates and the distribution of relative market potentials. This suggests why the results that we get in the following section tend to be fragile. In the parametric specifications that follow, market potential tends to have a weak impact on relative growth rates. This is, perhaps, unsurprising when we observe the degree of instability in the relationship over time.

4.2 Parametric results

We begin with the relationship between market potential and city growth. The equation that we estimate is thus:

$$GRPOP_{it} = \alpha_i + \gamma_t + \beta MP_{it} + \varepsilon_{it}. \tag{6.1}$$

$GRPOP_{it}$ is the growth rate of city i between time period t and $t+1$; MP_{it} is the market potential of city i at time t.

Table 6.1 (column 1) gives results for fixed effects (FE) estimates on the unbalanced panel for the time period 1900 to 1990. For consistency with later results, the time period is restricted to 1930 to 1990. Only cities that have entered the urban system by 1950 are included in the sample. However, the whole urban system is used when calculating the value of market potential.

The fact that market potential is a function of the whole urban system introduces a significant complication. Standard fixed effects estimates assume strict exogeneity, but market potential is endogenous to the system. A high value of the error for city i this period, drives up the growth rate of city i. But higher growth rate of city i changes the market potential, and hence growth rates, of all the other cities in the system. This, in turn, feeds back in to future values of market potential for city i. To allow for this we switch to a GMM formulation. We first difference Equation (6.1) to eliminate the fixed effects. As instruments, we use predetermined values of market potential and lagged values of the city size. For efficient estimation, we allow the number of instruments exploited to vary across time periods[10]. For year t, time varying instruments are thus market potential and lagged city size for time period $t-s$ where $s > 2$. After differencing operations and construction of instruments, we are left with an unbalanced panel with

[10] For details see, for example, Arrelano and Bond (1991).

seven years of data. Results for GMM estimation of Equation (6.1) are reported in column 2 of Table 6.1[11].

The results reinforce our earlier results from the stochastic kernel showing the mapping from population to market potential. City growth and market potential tend to be negatively related. This is true even when we allow for the growth of the south–west (pulled out by the fixed effects) which we know could not be driven by market potential.

Next we consider the importance of neighbouring city entry for growth. The fixed effects results show that both market potential and entry are negatively related to growth. The coefficient on market potential is lower, suggesting that some of the negative result may be due to the fact that cities with high market potential tend to see neighbouring (competing?) cities enter. The results are reported in column 3 of Table 6.1. The GMM results are somewhat disappointing. Allowing for entry of a neighbour has a negative effect on growth rates, but the coefficient is (just) insignificant at the 10% level if we allow for heteroscedasticity. The results are reported in column 4. We suspect that this lack of significance reflects the lack of good instruments for the entry variable. We have to instrument entry, because entry may not be exogenous with respect to neighbour size[12]. However, lagged city size and market potential may not be good instruments for the entry of a neighbouring city. We experienced similar problems with other specifications.

Next, we allow for the introduction of lagged own city size. The results here are somewhat surprising. If we account for lagged own city size, the effect of the market potential variable becomes insignificant with fixed effects, but significantly *positive* in the GMM specification. The effect of entry is now insignificant in both specifications. Lagged own city size has a large negative effect on growth rates. See columns 5 and 6.

As outlined above, new economic geography models actually suggest that what matters for city growth is the size of the city relative to it's market potential. New cities should enter when market potential at a site is above the market potential of existing cities. Thus cities will grow fastest when they are small relative to the market potential at the site. This suggests that we should actually enter population and market potential in ratio form. The results for entering them individually are consistent with this – we cannot reject the hypothesis that the coefficients are equal but opposite in size. Columns 7 and 8 show that when we enter the variable in ratio form, the effect is significant and negative.

As for the stochastic kernel specifications, we have recalculated market potential weighting each city by wage. The results in terms of parameter signs and significance

[11]For both fixed effects and GMM we report one–step estimates with robust standard errors. See Arrelano and Bond (1991) for why this is preferable to either non–robust errors or two–step estimators with robust standard errors.

[12]Dobkins and Ioannides (1998) provide compelling evidence which suggests that entry *is* driven by spatial features of the urban system.

are identical using this alternative market potential variable. Results are reported in Table 6.2.

The results that we have reported so far use city based market potential (with and without weighting by wage). Table 6.3 shows that these results are not robust to the use of county based versus city based market potential. The major difference between these sets of results is that market potential is insignificant when entered market potential and lagged own city size are entered separately in levels. However, the results for population relative to market potential are the same for all three types of market potential.[13]

To summarize:

- Market potential has a negative effect on growth rates if we do not take in to account own lagged city size. This result is robust to the use of the three different definitions of market potential.

- Entry has a weak negative impact on the growth rates of neighbouring cities. This result is not very robust. However, this may reflect the lack of good instruments for the entry variable.

- Own lagged city size has a robust negative effect on growth rates. When both own lagged city size and market potential are entered in levels, market potential has a positive effect on city growth. The results are not very robust to the definition of market potential.

- The ratio of own lagged city size to market potential has a robust negative impact on city growth. Cities grow fastest when they are small relative to their market potential.

5. Conclusions

This chapter has used a number of different approaches to analyse the spatial evolution of the US urban system over the period 1900 to 1990. The results confirm some theoretical insights, but also throw up a number of puzzles.

The first group of findings concern the relationship between city size and market potential. Our results in Section 3 suggest that there is no simple positive relationship between the distribution of city sizes and the distribution of market potentials. Indeed this relationship appears to change substantially over time. There is some evidence of a

[13]How do we reconcile these results with those of Black and Henderson (1999)? The stochastic kernels in Figure 6.7 suggest one possible solution. As discussed above our definition of cities uses an absolute cut–off point of 50000, whereas Black and Henderson use a relative cut–off point. One of the implications of this choice of cut–off is that cities enter the sample later in our data set. However Figure 6.7 shows that the positive relationship between city growth rates and market potential is stronger at the start of the century. Thus, one explanation of the difference between our results is that our estimations place *less* weight on the period when the positive relationship between growth rates and market potential is strongest. This factor is re–enforced by the fact that Black and Henderson use a balanced panel of cities that existed in 1930, whereas we use an unbalanced panel which allows for entry.

positive relationship between city sizes and market potential at the start of the century. That relationship is much weaker at the end of the century, apparently only holding for the largest cities.

Our second group of findings concern the relationship between city growth rates and market potential. Again, our non–parametric results show that this is a complex relationship which appears to have evolved over time. Parametric specifications appear to be quite fragile, presumably as a result of this evolution in the relationship over time. Initial parametric results suggest that there is a negative relationship between city size and market potential if we do not take in to account own lagged city size. Once we allow for own lagged city size, there is a positive relationship between market potential and city growth. Own lagged city size has a negative effect. These results are not robust to the definition of market potential.

By far the most robust parametric result relates to the ratio of lagged own city size to market potential. When cities are small relative to their market potential they grow faster. This result is consistent with a theoretical models advanced as part of the new economic geography. However, if the results are driven by the own lagged city size variable, then these results may also be consistent with theoretical models that emphasise congestion effects within cities. Separating out these two hypothesis is left to further work.

	FE 1	GMM 2	FE 3	GMM 4	FE 5	GMM 6	FE 7	GMM 8
ℓn market potential	-0.137** (.052)	-0.082** (0.042)	-0.104** (0.046)	-0.082* (0.048)	-0.09 (0.06)	0.200** (0.101)		
entry			-0.036* (0.021)	-0.032 (0.021)	-0.02 (0.019)	-0.021 (0.020)		
ℓn pop					-0.126 (0.02)	-0.174** (0.05)		
ratio							-0.146** (0.020)	-0.127** (0.035)

* Significant at 10% level. ** Significant at 5% level.
Due to city entry, we use an unbalanced panel. Total sample comprises 160 cities (112 cities with 10 obs., 27 cities with 9 obs., 10 cities with 8 obs., 8 cities with 7 obs., 3 cities with 6 obs.).

Table 6.1. City growth rates – city based market potential

	FE 1	GMM 2	FE 3	GMM 4	FE 5	GMM 6	FE 7	GMM 8
ℓn market potential	-0.064** (0.030)	-0.110** (0.040)	-0.065** (0.030)	-0.112 ** (0.043)	-0.03 (0.03)	0.08* (0.04)		
entry			-0.030* (0.02)	-0.034 (0.021)	-0.029 (0.020)	-0.020 (0.020)		
ℓn population					-0.116 (0.019)	-0.100** (0.03)		
ratio							-0.113** (0.017)	-0.091** (0.031)

* Significant at 10% level. ** Significant at 5% level.
Sample as for Table 6.1.

Table 6.2. City growth rates – wage weighted market potential

	FE 1	GMM 2	FE 3	GMM 4	FE 5	GMM 6	FE 7	GMM 8
ℓn market potential	-0.181** (0.065)	-0.178* (0.094)	-0.175** (0.065)	0.182** (0.083)	0.055 (0.109)	0.216 (0.162)		
entry			-0.034 (0.021)	0.036 (0.038)	-0.030 (0.019)	0.048 (0.045)		
ℓn population					-0.111** (0.020)	-0.140** (0.045)		
ratio							-0.124** (0.018)	-0.086** (0.029)

* Significant at 10% level. ** Significant at 5% level.
Sample as for Table 6.1.

Table 6.3. City growth rates – county based market potential

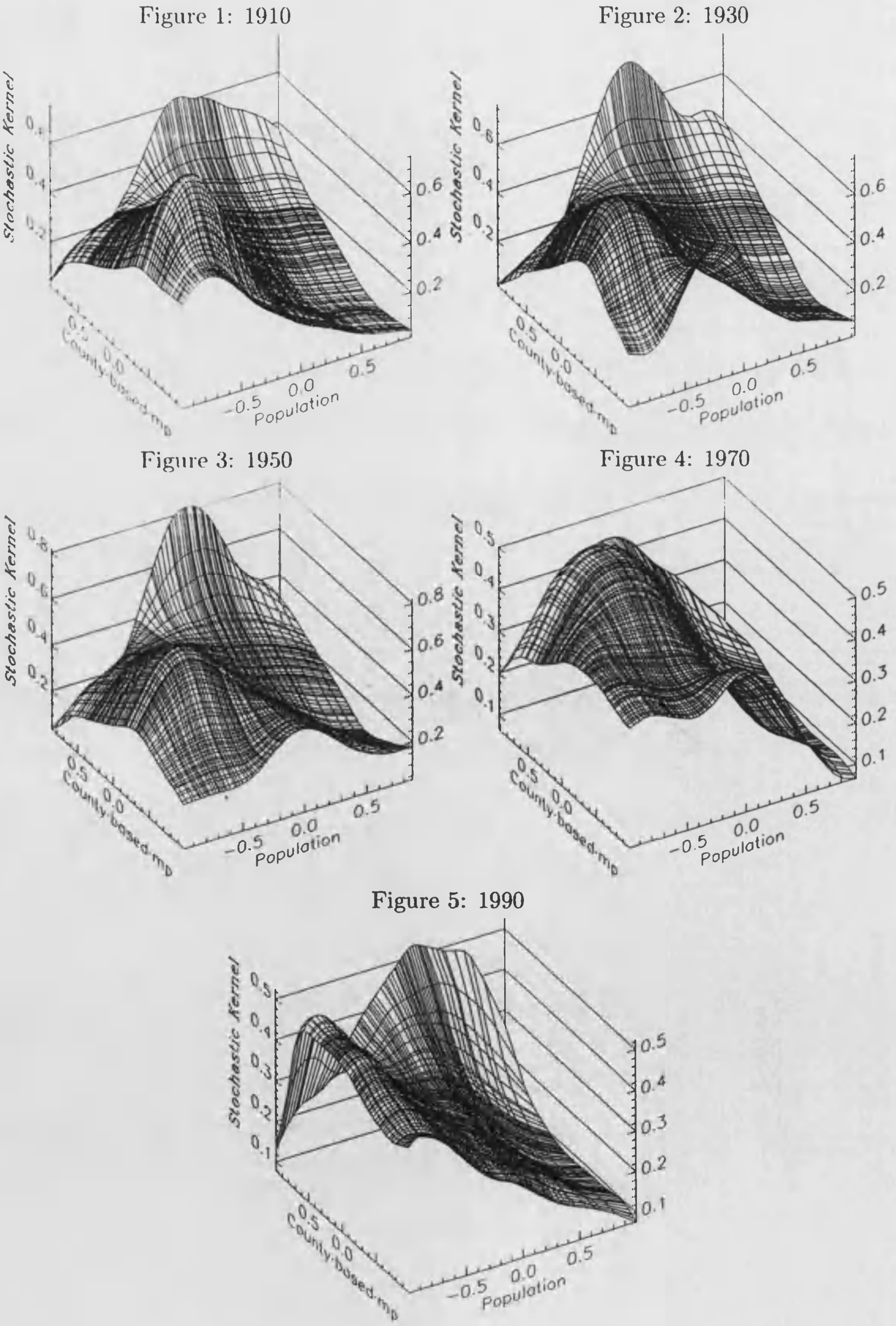


Figure 6.1. Stochastic kernels - Population conditional on county based market potential

Figure 1: 1910

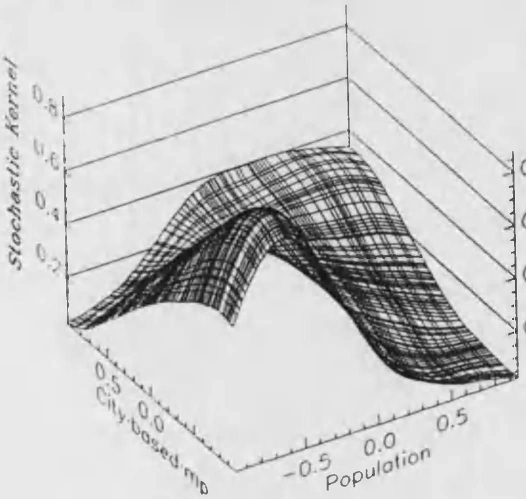


Figure 2: 1930

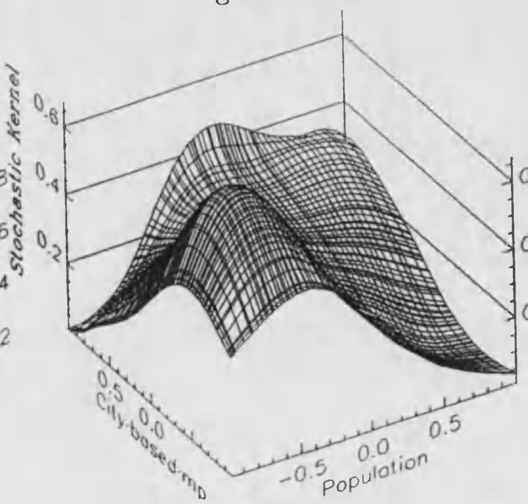


Figure 3: 1950

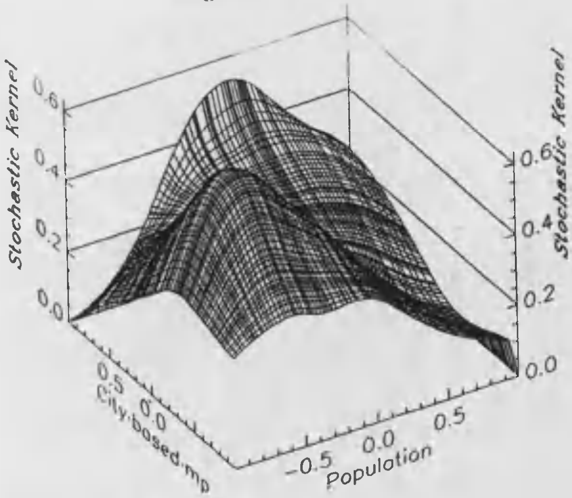


Figure 4: 1970

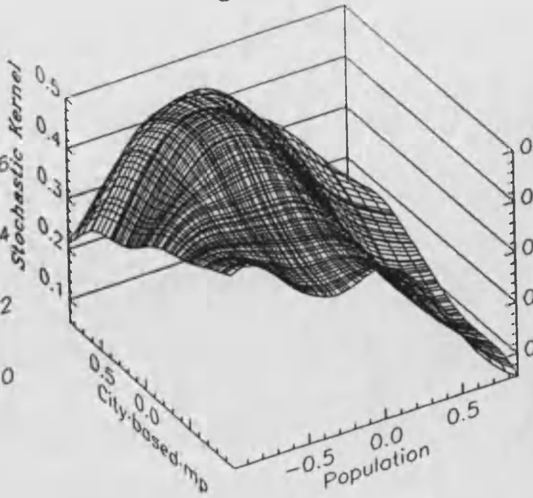


Figure 5: 1990

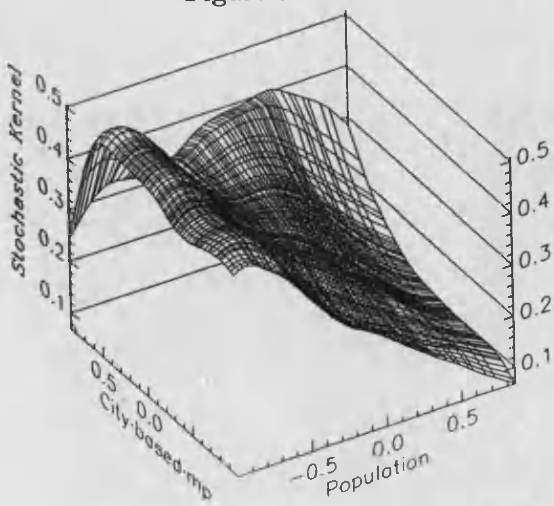


Figure 6.2. Stochastic kernel - Population conditional on city based market potential

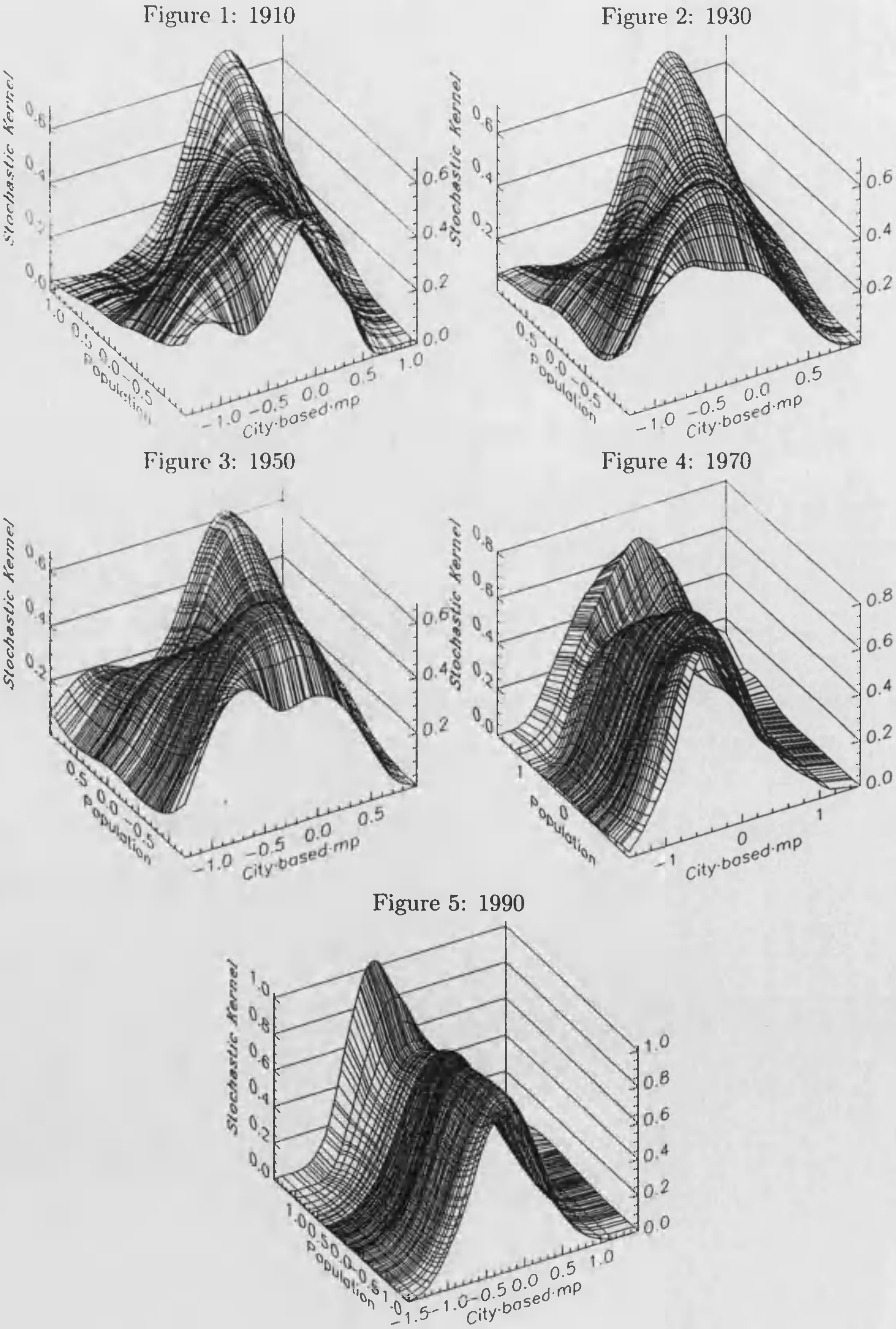


Figure 6.3. Stochastic kernels - City based market potential conditional on population

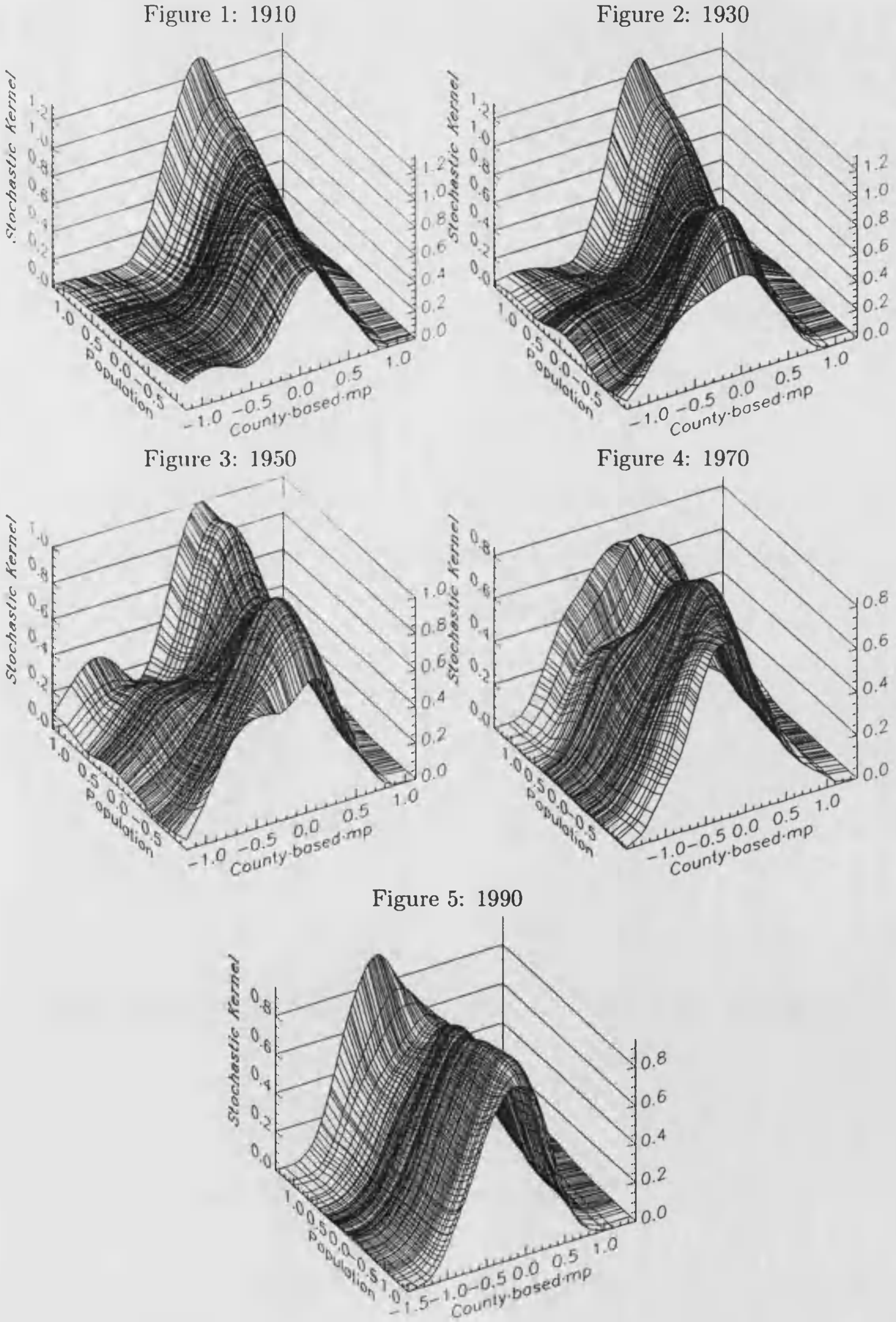


Figure 6.4. Stochastic kernels - County based market potential conditional on population

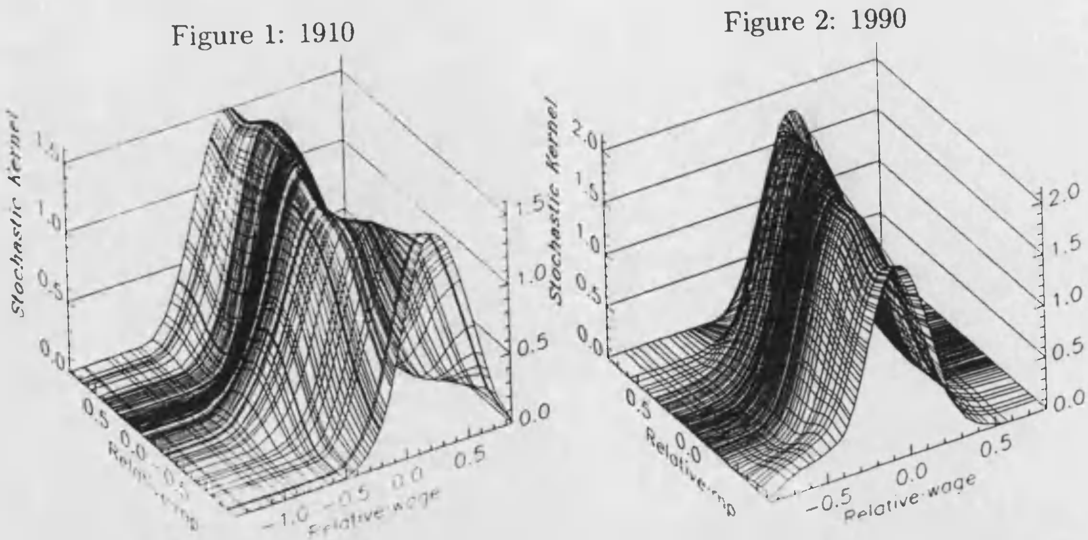


Figure 6.5. Stochastic kernels - Wage conditional on market potential

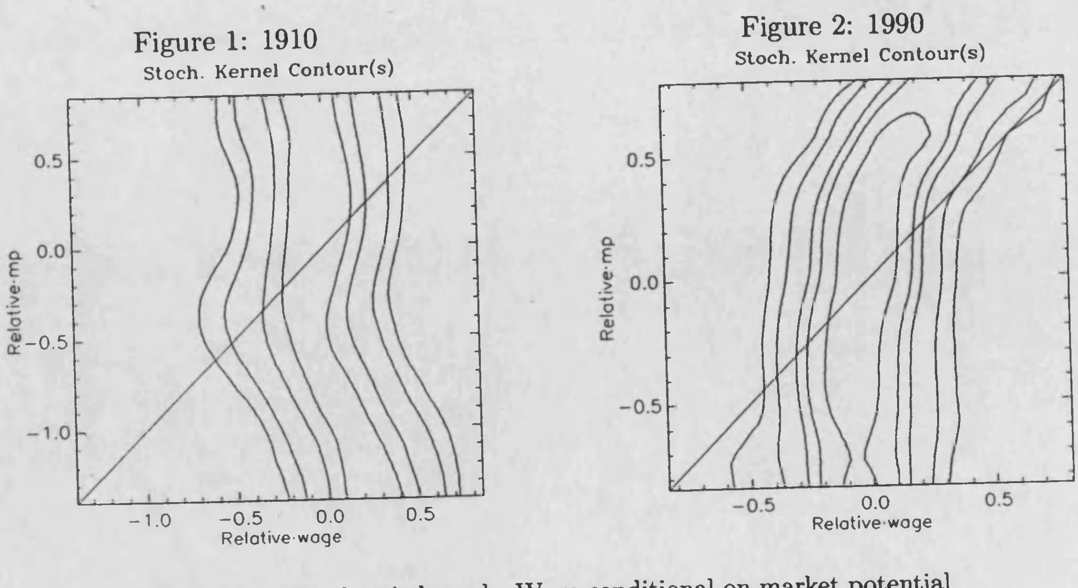


Figure 6.6. Stochastic kernel - Wage conditional on market potential

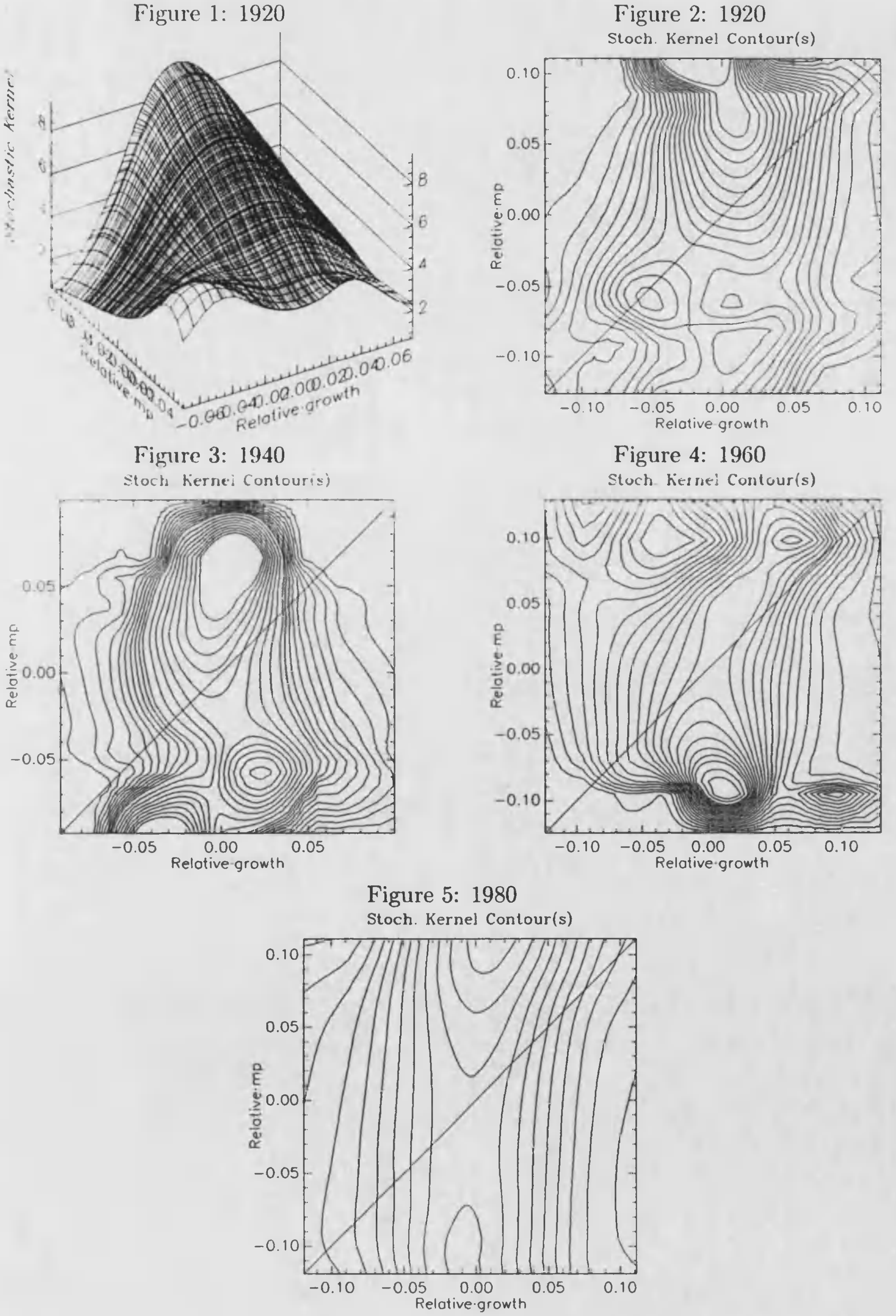


Figure 6.7. Stochastic kernels - City growth conditional on relative market potential

Appendices

A. AYS data appendix

The following variables take the value 1 when the characteristic is present and 0 otherwise;

- Personal characteristics: male, born overseas;
- Parent's characteristics: parent not employed when the respondent was 14, parent not present in the household when the respondent was 14;
- Parent's education: has a degree, has a trade qualification, has other post school qualifications (omitted category: parent has completed high school or less);
- Section of State: other city, rural area, country town (omitted category: capital city); and
- State: Victoria, South Australia, Western Australia, Queensland, Tasmania, ACT (omitted category: New South Wales; Northern Territory dropped due to too few observations).

The following variables are count variables:

- Age, number of siblings.

The language proficiency variables are defined as:

- 'English good' takes the value one if the respondent does not have English as a first language, but regards their proficiency as 'very good' or 'good'; and
- 'English poor' takes the value one if the respondent does not have English as a first language and regards their proficiency as 'fair', 'poor' or 'very poor'.
- The omitted category contains respondents who speak English as their first language.

The 'School' variable is defined as:

- 'Catholic' takes the value one if the respondent is currently studying at a Catholic high school, or whose last school was a Catholic high school;

- 'Other non-government' takes the value one if the respondent is currently studying at a non-government, non-Catholic high school, or whose last school was an 'other non-government' school.

- The omitted category contains respondents whose current or last school was a government high school.

Parent's occupational status is measured as the socio-economic status of the respondent's parent when the respondent was 14. If the parent was not present in the household or was not employed the index is set to zero.

The neighbourhood variables are defined as:

- 'average personal income' is the average personal income of the respondent's postcode or CD;

- 'proportion with grad qual' is the proportion of the respondent's postcode or CD who recorded having a higher degree, a degree, a graduate diploma, or an undergraduate diploma.

- 'proportion with trade qual' is the proportion of the respondent's postcode or CD who recorded having skilled vocational training or basic vocational training;

- 'unemployment rate' is the unemployment rate of the respondent's postcode or CD.

- the SES variable is a socio-economic status indicator based on factor-component analysis using a variety of socio-economic variables. It has a mean of 1.00 and a range of 0.714 to 1.42. See Hunter (1996) for details of the construction.

B. Unemployment clusters data appendix

Our definition of regions corresponds to level two of the Nomenclature of Territorial Units for Statistics (NUTS), 1995 version (Eurostat, 1995). The NUTS was established by Eurostat to provide comparable regional breakdowns of the Member States of the European Union. It is a hierarchical classification with three regional levels: each Member State is partitioned into an integral number of NUTS1 regions, each of which is in turn partitioned into an integral number of NUTS2 regions, each of which is in turn partitioned into an integral number of NUTS3 regions. (There are two additional sub-regional or local levels, NUTS4 and NUTS5, of which only the latter, consisting of Communes or their equivalent, is defined for all Member States). In 1996 the EU had 77 NUTS1 regions, 206 NUTS2 regions, and 1,031 NUTS3 regions. Eurostat (1995) also calls NUTS2 regions 'Basic Regions', and describes these as the appropriate level for analysing regional-national problems; it is also the level at which both national and Community regional policies are generally implemented.

NUTS2 regions correspond to national administrative units in Austria (Bundesländer), Belgium (Provinces), Finland (Suuralueet), Germany (Regierungsbezirke), Greece (Development Regions), Italy (Regioni), Netherlands (Provincies), Portugal (Comissaoes de Coordenaçao Regional), and Sweden (Riksområden). NUTS2 regions also correspond to national administrative units, but with exceptions, in France (Régions, plus the four Departements d'Outre Mer), and Spain (Comunidades Autónomas, plus Ceuta y Melilla). Three Member States are classified as a single NUTS2 region: Denmark, Ireland, and Luxembourg. In the United Kingdom, Groups of Counties have been introduced as an intermediate (NUTS2) level between NUTS1 (Standard Regions) and NUTS3 (a combination of Counties and Local Authority Regions) units.

The data set includes (with a single exception, documented below) all the NUTS2 regions of the EU that satisfy the following three criteria:

1. Have been part of the EU (European Economic Community before 1 November 1993) from 1986 to 1996.
2. Are in a Member State which has a land border with at least one other Member State containing at least one region satisfying (1).
3. Have a land border with at least one other NUTS2 region satisfying (1) and (2).

We include as land borders water borders less than five kilometres wide. This leads us to consider as geographical neighbours regions separated by a river (such as Zeeland and Zuid-Holland in Netherlands). It also leads to the inclusion of Sicilia (Italy), which, although an island, is only separated from Calabria (Italy) by the 3,300 metres-wide Strait of Messina — soon to be joined by a single span suspension bridge (see `http://www.strettodimessina.it/`).

From the 206 NUTS2 regions that formed the EU in 1996, 30 are excluded from the analysis because they were not part of the European Economic Community in 1986:

the nine NUTS2 regions of Austria, the six NUTS2 regions of Finland, and the eight NUTS2 regions of Sweden, all of which became part of the EU with the accession of these three Member States in 1995; and the seven NUTS2 regions of Germany that were part of the former Democratic Republic of Germany (Brandenburg, Mecklenburg-Vorpommern, Sachsen, Dessau, Halle, Magdeburg, and Thüringen), which only became part of the EU with German reunification in 1990.

Greece has no land border with any other Member State, so its 13 NUTS2 regions are also excluded.

Finally, another 12 NUTS2 regions are excluded because they have no land border with any other NUTS2 region satisfying criteria (1) and (2): Baleares, Ceuta y Melilla, and Canarias (Spain), Corse, Guadeloupe, Martinique, Guyane, and Réunion (France), Sardegna (Italy), Açores, and Madeira (Portugal), are all entirely surrounded by water and/or by territories which are not part of the EU; Berlin (Germany) is entirely surrounded by NUTS2 regions which were part of the former Democratic Republic of Germany.

Flevoland (Netherlands) is the only region that satisfies criteria (1)-(3) above but has been excluded due to lack of data: there is no labour force or unemployment data for Flevoland for 1986, even from national sources (see Centraal Bureau Voor de Statistiek, 1987). Flevoland was created as a separate administrative unit (Provincie) in 1986 from the union of the Noordoost, Oostelijk Flevoland, and Zuidelijk Flevoland polders, reclaimed from the IJsselllake (a lake that used to be part of Zuiderzee, a former inlet of the North Sea), and in 1996 accounted for 1.8% of the population and 5.8% of the land area of Netherlands.

The 150 NUTS2 regions used are:

Belgium (11)	Brussels, Antwerpen, Limburg (Belgium), Oost-Vlaanderen, Vlaams Brabant, West-Vlaanderen, Brabant Wallon, Hainaut, Liége, Luxembourg (Belgium), Namur.
Denmark (1)	
France (21)	Ile-de-France, Champagne-Ardenne, Picardie, Haute-Normandie, Centre, Basse-Normandie, Bourgogne, Nord-Pas-de-Calais, Lorraine, Alsace, Franche-Comté, Pays de la Loire, Bretagne, Poitou-Charentes, Aquitaine, Midi-Pyrénées, Limousin, Rhône-Alpes, Auvergne, Languedoc-Roussillon, Provence-Alpes-Côte d'Azur.
Germany (30)	Stuttgart, Karlsruhe, Freiburg, Tübingen, Oberbayern, Niederbayern, Oberpfalz, Oberfranken, Mittelfranken, Unterfranken, Schwaben, Bremen, Hamburg, Darmstadt, Giessen, Kassel, Braunschweig, Hannover, Lüneburg, Weser-Ems, Düsseldorf, Köln, Münster, Detmold, Arnsberg, Koblenz, Trier, Rheinhessen-Pfalz, Saarland, Schleswig-Holstein.

Ireland (1)	
Italy (19)	Piemonte, Valle d'Aosta, Liguria, Lombardia, Trentino-Alto Adige, Veneto, Friuli-Venezia Giulia, Emilia-Romagna, Toscana, Umbria, Marche, Lazio, Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria, Sicilia.
Luxembourg (1)	
Netherlands (11)	Groningen, Friesland, Drenthe, Overijssel, Gelderland, Utrecht, Noord-Holland, Zuid-Holland, Zeeland, Noord-Brabant, Limburg (Netherlands).
Portugal (5)	Norte, Centro (Portugal), Lisboa e Vale do Tejo, Alentejo, Algarve.
Spain (15)	Galicia, Asturias, Cantabria, País Vasco, Navarra, Rioja, Aragón, Madrid, Castilla-León, Castilla-La Mancha, Extremadura, Cataluña, Comunidad Valenciana, Andalucía, Región de Murcia.
United Kingdom (35)	Cleveland-Durham, Cumbria, Northumberland-Tyne and Wear, Humberside, North Yorkshire, South Yorkshire, West Yorkshire, Derbyshire-Nottinghamshire, Leicestershire-Northamptonshire, Lincolnshire, East Anglia, Bedfordshire-Hertfordshire, Berkshire-Buckinghamshire-Oxfordshire, Surrey-East-West Sussex, Essex, Greater London, Hampshire-Isle of Wight, Kent, Avon-Gloucestershire-Wiltshire, Cornwall-Devon, Dorset-Somerset, Hereford-Worcestershire-Warwickshire, Shropshire-Staffordshire, West Midlands (County), Cheshire, Greater Manchester, Lancashire, Merseyside, Clwyd-Dyfed-Gwynedd-Powys, Gwent-Mid-South-West Glamorgan, Borders-Central-Fife-Lothians-Tayside, Dumfries-Galloway-Strathclyde, Highlands-Islands, Grampian, Northern Ireland.

Regional unemployment rates and labour force from 1986 to 1996 are taken from the harmonised unemployment rates (table `regio/unemp/un3rt`) and labour force (table `regio/unemp/un3wpop`) in the May 1998 version of the Regio database published by Eurostat (Eurostat, 1998).

These data are based on the results of the Community Labour Force Survey (LFS). The Community LFS is carried out in Spring each year and for each Member State provides the number of the unemployed (in accordance with the definition of the International Labour Office), and the labour force (labelled 'working population') for April. The national unemployment data are subsequently regionalised to NUTS2 level on the basis of the number of persons registered at unemployment offices in April of the reference year (with the exceptions of Greece, Spain, Italy, Portugal, Finland, and Sweden, where the regional unemployment structures are taken from the Community LFS). The national labour force data are regionalised to NUTS2 level according to the results of the Community LFS. The regional unemployment rates are then obtained by dividing the number of the unemployed by the labour force.

The Regio database has no data on unemployment rates or labour force for two years, 1986 and 1987, for 13 of the targeted regions: all the NUTS2 regions of Netherlands, and Algarve (Portugal). For all of them (except the Dutch region of Flevoland, as documented above) comparable data has been obtained as follows. For the NUTS2 regions of the Netherlands in 1986 and 1987, the total number of the unemployed in the Netherlands in table `/regio/unemp/un3pers` of the Regio database has been regionally disaggregated to NUTS2 level, on the basis of the number of the unemployed in each region from table II.4 of Eurostat (1989), which are also derived from the Community LFS. Similarly, the total labour force of the Netherlands in table `/regio/unemp/un3wpop` of the Regio database has been regionally disaggregated to NUTS2 level, on the basis of regional labour force figures from table II.2 of Eurostat (1990) (for 1986), and of regional labour force figures computed by dividing the number of the unemployed by the corresponding unemployment rates in table II.4 of Eurostat (1989) (for 1987). Regional unemployment rates have then been calculated by dividing the number of the unemployed by the labour force. For Algarve (Portugal) in 1986 and 1987, employment and unemployment figures have been privately obtained from national sources (Portugal's Instituto Nacional de Estatística for employment, and Direcçao de Serviços de Estudos de Mercado de Emprego for unemployment), and corrected for the factor by which each of these sources underestimates the corresponding Community LFS data for all the other NUTS2 regions that, together with Algarve, constitute the NUTS1 region Continente (Norte, Centro, Lisboa e Vale do Tejo, and Alentejo). Labour force has been calculated as the sum of the employed and the unemployed, and the unemployment rate by dividing the number of the unemployed by the labour force.

Regional unemployment rates and labour force are used to construct five series of relative unemployment rates: unemployment rates relative to the European average (*Europe relative* for brevity), unemployment rates relative to the average for other regions in the same Member State (*State relative*), unemployment rates relative to the average for contiguous regions (*neighbour relative*), unemployment rates relative to the average for other regions with the same broad sectoral specialisation (*same specialisation relative*), and unemployment rates relative to the average for other regions with a similar split of low/high educational attainment (*same skill relative*). In all cases averages used to construct the relative series refer only to regions included in the analysis. The information on State membership and contiguity is taken off the paper maps in Eurostat (1995).

To obtain groupings by broad sectoral specialisation, regions are classified according to the sector in the NACE-CLIO R3 classification (agricultural, forestry and fishery products; manufactured products; and market services) in which their share of total employment was highest relative to the EU average in 1988. The basis for these calculations are the total employment data by NACE-CLIO R3 sector (table /regio/lfs-r/lf2emp) in Eurostat (1998). These data are available for the 150 regions we are interested in only for 1988, but this is close enough to the beginning of the time frame considered to describe early specialisation.

To obtain groupings by low/high educational attainment, regions are classified according to the percentage of their population aged 25 to 59 in 1995 with less than upper secondary education — less than level 3 of the International Standard Classification of Education (ISCED) classification (UNESCO, 1976). These data are from table E14 in Eurostat (1997). These data are not ideal in that they refer to the adult population and not to the labour force, and they are only available for the 150 regions we are interested in for a single year, 1995. However, they are the best available at this level of regional disaggregation. We use them to construct nine groups of regions: regions where less than 10% of 25 to 59 year olds have less than upper secondary education, regions with more than 10% but less than 20%, and so on in ten percentage points intervals until regions where more than 80% but less than 90% of 25 to 59 year olds have less than upper secondary education.

The regression analysis of Section 4 uses the same data sources as the non parametric section. For the purpose of splitting population by skill there, low skill is taken to be an educational attainment of less than upper secondary education (below level 3 of the ISCED classification). Medium skill is an educational attainment of upper secondary education (level 3 of the ISCED classification). High skill is an educational attainment of higher education (levels 5, 6, and 7 of the ISCED classification). To calculate the percentage of young population, the young are taken to be those that reached working age during the sample period (those aged between 15 and 25 in 1996). These data are obtained from table /regio/lfs-r/lf2emp) in Eurostat (1998). Initial female participation rates are those for 1986 from table /regio/lfs-r/lf2actrt) in Eurostat (1998), completed with Eurostat (1989). For the calculation of the measure of initial market potential, used as one of the instruments in the instrumental variable estimations of Section 4, 1986 regional GDP levels are from table /regioecon-r/egdp/e2gdp) in Eurostat (1998). The distance between each pair of NUTS 2 regions is the great circle distance between their geographical centres, the coordinates of which have been obtained from http://shiva.pub.getty.edu/tgn_browser/.

C. Unemployment clusters technical appendix

More familiar applications of stochastic kernels use observations on random draws from a Markov process to estimate the underlying transition characteristics of that process. In contrast, in this paper we are interested in mappings from one distribution to another distribution. For example, this may be a mapping from the distribution of Europe relative unemployment rates at one point in time to the distribution of Europe relative unemployment rates at another point in time, or it may be the mapping from the distribution of Europe relative unemployment rates to the distribution of neighbour relative unemployment rates. In this Technical Appendix, we show that standard stochastic kernels can still be used to characterise the mappings between any two distributions, providing that we are careful about the space on which we define those stochastic kernels.

Let the two distributions of interest be γ and λ. Then we seek a mapping $\mathcal{T}^*$ such that $\lambda = \mathcal{T}^*(\gamma)$. Our underlying state space is the pair $(\mathrm{I}, \mathfrak{R}_{\mathrm{I}})$, where I is the unit interval and $\mathfrak{R}_{\mathrm{I}}$ is the collection of Borel sets of the real line that are subsets of the unit interval. However, we define stochastic kernels on the more general state space $(\mathbb{R}, \mathfrak{R})$, where $\mathbb{R}$ is the real line and $\mathfrak{R}$ the collection of its Borel sets. We do so with the understanding that these definitions are valid for restrictions of the general state space to the specific unit interval state space.

Consider the most familiar case first, where we are interested in transitions over time and the distributions of interest are λ_t and λ_{t-1}. Recall the standard definition of a transition function.

Transition function definition. Let $(Z, \mathfrak{Z})$ be a measurable space. A transition function is a function $Q : (Z, \mathfrak{Z}) \rightarrow [0, 1]$ that satisfies two conditions:

(i) For each $z \in Z$, $Q(z, .)$ is a probability measure on $(Z, \mathfrak{Z})$.

(ii) For each $A \in \mathfrak{Z}$, $Q(., A)$ is a $\mathfrak{Z}$-measurable function.

The standard interpretation is that $Q(a, A)$ is the probability that next periods realisation lies in the set A, given that this period's realisation is a. There are two useful functions associated with the standard transition function.

Two useful functions.

1. For any $\mathfrak{Z}$-measurable function f, define $\mathcal{C}f$ by $(\mathcal{C}f)(z) = \int f(z')Q(z, \mathrm{d}z')$, for all $z \in Z$.

2. For any probability measure λ on $(Z, \mathfrak{Z})$ define $\mathcal{C}^*\lambda$ by $(\mathcal{C}^*\lambda)(A) = \int Q(z, A)\lambda(\mathrm{d}z)$, for all $A \in \mathfrak{Z}$.

The interpretation is as follows. $(\mathcal{C}f)(z)$ is the expected value of the function next period, given that the current state is z. $\mathcal{C}$ maps the space of bounded functions to the space of bounded functions and is known as the Markov operator associated with Q. $(C^*\lambda)(A)$ is the probability that the state next period lies in the set A if the current state is drawn according to the probability measure λ. C^* maps the space of probability measures to the space of probability measures and is known as the adjoint of C. Thus $\lambda_t = C^*(\lambda_{t-1})$.

This C^* is closely related to the mapping $\mathcal{T}^*$ that we are interested in estimating. However two extensions are necessary. First, we want to allow for mappings between any two distributions, not just sequential distributions. Second, for empirical applications, we want to allow for generalised disturbances that may affect the mapping between distributions.[14] The extension to any two distributions is achieved through the use of the standard stochastic kernel definition.

Stochastic kernel definition. Let $(X, \mathfrak{X})$ and $(Y, \mathfrak{Y})$ be measurable spaces. Let ϕ be a probability measure on $(X, \mathfrak{X})$ and ψ be a probability measure on $(Y, \mathfrak{Y})$. A stochastic kernel relating ϕ to ψ is a mapping $M_{\phi,\ } : (X, \mathfrak{Y}) \rightarrow [0, 1]$ that satisfies three conditions:

(i) For all $y \in X$ the restriction $M_{(\phi,\ )}(y, .)$ is a probability measure.

(ii) For all $A \in \mathfrak{Y}$ the restriction $M_{(\phi,\ )}$ is $\mathfrak{X}$-measurable.

(iii) For all $A \in \mathfrak{Y}$ we have $\phi(A) = \int M_{(\phi,\ )}(y, A) \mathrm{d}\psi(y)$.

Consider (iii). In the initial distribution, for given y, there is some fraction $\mathrm{d}\psi(y)$ of regions with unemployment rates close to y. Count up all regions in that group whose unemployment rate subsequently fall in a given $\mathfrak{Y}$-measurable subset $A \subseteq \mathbb{R}$ of the second (later/conditional) distribution. When normalised by the fraction of the total number of regions this count is precisely $M_{(\phi,\ )}(y, A)$. Thus $M_{(\phi,\ )}(y, A)$ is the probability that a region's realisation in the later/conditional distribution lies in the set A, given that the initial realisation is y. Evaluate the integral $\int M_{(\phi,\ )}(y, A) \mathrm{d}\phi(y)$. This gives the fraction of regions that end up in state A regardless of their initial position. If this equals $\phi(A)$ for all measurable sets A, then ϕ must be the measure associated with the subsequent distribution of unemployment rates. Conditions (i) and (ii) just ensure that this interpretation is valid. In particular, (ii) ensures that the right hand side of (iii) is a well defined Lebesgue integral, while (i) ensures that the right hand side of (iii) is a weighted average of probability measures and thus is itself a probability measure. It is easy to see that a transition kernel is a stochastic kernel for which the two spaces $(X, \mathfrak{X})$ and $(Y, \mathfrak{Y})$ are the same.

To allow for generalised disturbances we need to be able to model random elements drawn from a collection of probability measures. Following Quah (1997a) we proceed as follows. First we define a Banach space that contains all possible probability measures. We then use this Banach space and suitably defined open sets on that space to define

[14] We have implicitly absorbed this generalised error in to our definition of $\mathcal{T}^*$.

a measurable space which we can, in turn, use to model random elements drawn from collections of probability measures.

Let $\mathbf{B}(\mathbb{R}, \mathfrak{R})$ be the Banach space of bounded finitely additive set functions on the measurable space $(\mathbb{R}, \mathfrak{R})$ with total variation norm

$$\text{for all } \phi \text{ in } \mathbf{B}(\mathbb{R}, \mathfrak{R}) : \|\phi\| = \sup \sum_j |\phi(A_j)|,$$

where the supremum is taken over all $\{A_j : j = 1, 2, \ldots n\}$ finite measurable partitions of $\mathbb{R}$.

Empirical distributions on $\mathbb{R}$ are identified with probability measures on $(\mathbb{R}, \mathfrak{R})$. Probability measures are elements of $\mathbf{B}(\mathbb{R}, \mathfrak{R})$ that are countably additive and assign value one to the entire space $\mathbb{R}$. We use the set of bounded finitely additive set functions, because a collection of probability measures can never form a linear space. The set of boundedly-additive set functions includes probability measures and does form a linear space. We can then use the total variation norm to make this space Banach. Once probability measures are embedded in a Banach space, it makes sense to talk about two probability measures (and the associated distributions) getting closer to one another. Further, if we define a measure of distance, we can define open sets of probability measures (relative to this distance measure) and use these open sets to generate (Borel) σ-algebras on the Banach space. Given such a σ-algebra, we can model random elements drawn from collections of probability measures. This is the data of interest when we are modelling the dynamics of distributions.

Let $\mathfrak{B}$ denote the σ-algebra generated by the open sub-sets (relative to the total variation norm topology) of $\mathbf{B}(\mathbb{R}, \mathfrak{R})$. Then $(\mathbf{B}, \mathfrak{B})$ is another measurable space. By construction, each ϕ_i associated with an observed (or derived) empirical cross sectional distribution F_i is a member of $(\mathbf{B}, \mathfrak{B})$. If $(\Omega, \mathfrak{F}, \Pr)$ is the underlying probability space, then ϕ_i is the value of an $\mathfrak{F}/\mathfrak{B}$-measurable map $\Phi(\Omega, \mathfrak{F}) \to (\mathbf{B}, \mathfrak{B})$. We can define probability measures on $(\mathbf{B}, \mathfrak{B})$ that will allow us to deal with the generalised disturbances that affect the mapping between distributions.

Now, let $\mathbf{b}(\mathbb{R}, \mathfrak{R})$ be the Banach space under sup norm of bounded measurable function on $(\mathbb{R}, \mathfrak{R})$. Fix a stochastic kernel M and construct an operator $\mathcal{T}$ (similar to $\mathcal{C}$) that maps the space of bounded measurable functions on to itself:

$$\text{for any } f \in \mathbf{b}(\mathbb{R}, \mathfrak{R}) \text{ define } \mathcal{T}f \text{ by} (\mathcal{T}f)(z) = \int f(z')M(z, \mathrm{d}z'), \text{ for all } z \in \mathbb{R}.$$

This mapping has the same interpretation as $\mathcal{C}$ in the (useful) function 1 above. Now we can denote the adjoint of $\mathcal{T}$ by $\mathcal{T}^*$. Thus:

$$\text{for any probability measure } \lambda \text{ on } (\mathbb{R}, \mathfrak{R}) \text{ define } \mathcal{T}^*\lambda \text{ by } (\mathcal{T}^*\lambda)(A) = \int M(z, A)\lambda(\mathrm{d}z), \text{ for all} A \in \mathfrak{R}.$$

From the Riesz Representation Theorem, the dual space of $\mathbf{b}(\mathbb{R}, \mathfrak{R})$ is $\mathbf{B}(\mathbb{R}, \mathfrak{R})$, the collection of bounded finitely additive set functions. Thus $\mathcal{T}^*$ maps the collection of bounded finitely additive set functions on to itself. It is also precisely the mapping (iii) in the stochastic kernel definition. In our empirical analysis, we estimate $M(\phi, \psi)(y, .)$ (the probability distribution of a region's realisation in the later/conditional distribution given that the initial realisation is y) for a whole range of y values. Here, we have shown that this does indeed allow us to trace out $\mathcal{T}^*$, the generalised mapping between any two distributions.

D. Distance calculations

For any two locations A and B, we can calculate the angle formed by a ray joining the two points A and B and a ray joining A to the centre of the earth as follows:

$$\text{angle} = (\sin(\text{latA}) \times \sin(\text{latB})) + (\cos(\text{latA}) \times \cos(\text{latB}) \times \cos(\text{longA} - \text{longB}))$$

where latA and longA are the latitude and longitude of location A measured in radians. Similarly for latB and longB. For regions (in Chapter 4) the latitude and longitude are from `http://shiva.pub.getty.edu/tgn_browser/`. For cities (in Chapters 5 and 6), they are the latitudes and longitudes given in the 1999 Times World Atlas. For counties (in Chapters 5 and 6) they are the latitudes and longitudes of the largest human settlement.

The distance is then

$$\text{distance} = 3954 \times \text{acos}(\text{angle})$$

acos(angle) gives us the approximate distance if the two points were located on a circle of radius one. We then need to multiply by the radius of (a circular) earth (3954 miles) to get an estimate of the distance.

Bibliography

Akerlof, George. 1997. Social distance and social decisions. *Econometrica* 65(5):1005–1027.

Allen, G.R.P. 1954. The 'courbe des populations': a further analysis. *Bulletin of the Oxford University Institute of Statistics* 16:179–189.

Alperovich, Gershon. 1993. An explanatory model of city-size distribution: Evidence from cross-country data. *Urban Studies* 30(9):1591–1601.

Anselin, Luc. 1988. *Spatial Econometrics: Methods and Models.* Dordrecht: Kluwer Academic Publishers.

Arrelano, Manuel and Stephen Bond. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economics and Statistics* 58:277–297.

Auerbach, F. 1913. Das gesetz der bevolkerungskoncentration. *Petermanns Geographische Mitteilungen* 59:74–76.

Banerjee, Abhijit and Timothy Besley. 1990. Peer group externalities and learning incentives: A theory of nerd behaviour. Technical Report 68, John M.Olin Program for the Study of Economic Organisation and Public Policy.

Becker, Gary S. 1993. *Human capital: A theoretical and empirical analysis with special reference to education.* Chicago: The University of Chicago Press.

Berry, B.J.L. 1961. City size distributions and economic development. *Economic Development and Cultural Change* 9:573–587.

Black, Duncan and Vernon Henderson. 1999. Urban evolution in the USA. Processed, London School of Economics.

Blanchard, Olivier Jean and Juan Francisco Jimeno. 1995. Structural unemployment: Spain versus Portugal. *American Economic Review* 85(2):212–218.

Blanchard, Olivier Jean and Lawrence F. Katz. 1992. Regional evolutions. *Brooking Papers on Economic Activity* (1):1–61.

Bogue, Donald. 1953. Population growth in standard metropolitan areas 1900-1950. Technical report, Scripps Foundation in Research in Population Problems.

Borjas, G. 1992. Ethnic capital and intergenerational income mobility. *Quarterly Journal of Economics* 57:123–150.

Borjas, G. 1995. Ethnicity, neighbourhoods, and human–capital externalities. *American Economic Review* 85:365–390.

Bover, Olimpia, Pilar García-Perea, and Pedro Portugal. 1998. A comparative study of the Portuguese and Spanish labour markets. Working Paper 9807, Banco de España.

Carroll, G. 1982. National city size distributions: What do we know after 67 years of research. *Progress in Human Geography* 6:1–43.

Case, Anne and Lawrence Katz. 1991. The company you keep: The effects of family and neighbourhood on disadvantaged youth. Discussion Paper 3705, National Bureau for Economic Research.

Castillo, Sonsoles, Juan José Dolado, and Juan Francisco Jimeno. 1998*a*. The fall in consumption from being unemployed in Spain and Portugal. Discussion Paper 1953, Centre for Economic Policy Research.

Castillo, Sonsoles, Juan José Dolado, and Juan Francisco Jimeno. 1998*b*. A tale of two neighbour economies: Labour market dynamics in Spain and Portugal. Discussion Paper 1954, Centre for Economic Policy Research.

Centraal Bureau Voor de Statistiek. 1987. *Regionaal Statistich Zakboek 1986.* The Hague: Centraal Bureau Voor de Statistiek.

Chamberlain, G. 1980. Analysis of covariance with qualitative data. *Review of Economic Studies* 47:225–38.

Cheshire, Paul. 1999. Trends in sizes and structures of urban areas. In Paul Cheshire and E Mills (eds.), *Handbook of Regional and Urban Economics.* Elsevier Science B.V.

Cheshire, Paul and Gianni Carbonaro. 1996. Urban economic growth in Europe: Testing theory and policy prescriptions. *Urban Studies* 33(7):1111–1128.

Ciccone, Antonio. 1997. Dynamic externalities and the spatial distribution of wages in the US. Processed, University of California, Berkeley.

Cleveland, William S. 1993. *Visualising data.* New Jersey: Hobart Press.

Cooper, Russell and Andrew John. 1988. Coordinating coordination failures in Keynesian models. *Quarterly Journal of Economics* 103(3):441–463.

Corcoran, Mary, Roger Gordan, Deborah Laren, and Gary Solon. 1992. The association between men's economic status and their family and community origins. *The Journal of Human Resources* 27(4):575–601.

Crane, Jonathan. 1991. Effects of neighbourhoods on dropping out of school and teenage child-bearing. In Christopher Jencks and P.E. Peterson (eds.), *The Urban Underclass.* Washington D.C.: Brookings Institution.

Crane, Jonathan. 1997. The epidemic theory of ghettos and neighbourhood effects on dropping out and teenage childbearing. *American Journal of Sociology* 96(5):1226–1259.

Datcher, Linda. 1982. Effects of community and family background on achievement. *Review of Economics and Statistics* 64(1):32–41.

Davidson, Russell and James G. MacKinnon. 1993. *Estimation and Inference in Econometrics.* New York: Oxford University Press.

de Falleur, M. and V. Vandeville. 1996. Cross-border flows of workers in Europe: Facts and determinants. In Harvey Armstrong and R. W. Vickerman (eds.), *Convergence and Divergence Among European Regions.* London: Pion.

Dearden, Lorraine and Alex Heath. 1996. Income support and staying in school: What we can learn from australia's austudy experiment? *Fiscal Studies* 17(4):1–30.

Decressin, Jörg and Antonio Fatàs. 1995. Regional labor market dynamics in Europe. *European Economic Review* 39(9):1627–1655.

Dobkins, Linda Harris and Yannis Ioannides. 1998. Spatial interactions among US cities. Processed, Tufts University.

Dobkins, Linda Harris and Yannis Ioannides. 1999. Dynamic evolution of the US city size distribution. In J.M. Huriot and Jacques Thisse (eds.), *The Economics of Cities.* Cambridge University Press.

Dockery, A.M. and K. Norris. 1996. The 'rewards' for apprenticeship training in Australia. *Australian Bulletin of Labour* 22(2):109–125.

Dolado, Juan J., Jose Manuel Gonzélez-Páramo, and José Mariá Roldán. 1994. Convergencia económica entre las provincias Españolas: Evidencia empírica (1955-1989). *Moneda y Crédito* 198:81–119.

Eaton, Jonathan and Zvi Eckstein. 1997. Cities and growth: Theory and evidence from France and Japan. *Regional Science and Urban Economics* 27(4–5):443–474.

Eichengreen, Barry. 1993. Labor markets and European monetary unification. In Paul R. Masson and Mark P. Taylor (eds.), *Policy Issues in the Operation of Currency Unions.* Cambridge: Cambridge University Press, 130–162.

Eurostat. 1989. *Regions: Statistical Yearbook 1988.* Luxembourg: Office for Official Publications of the European Communities.

Eurostat. 1990. *Regions: Statistical Yearbook 1989.* Luxembourg: Office for Official Publications of the European Communities.

Eurostat. 1995. *Regions: Nomenclature of Territorial Units for Statistics (*NUTS*).* Luxembourg: Office for Official Publications of the European Communities.

Eurostat. 1997. *Education Across the European Union: Statistics and Indicators 1996.* Luxembourg: Office for Official Publications of the European Communities.

Eurostat. 1998. *Regio Database.* Luxembourg: Eurostat. May 1998 CD-ROM version.

Fatàs, Antonio. 1997. EMU: Countries or regions? Lessons from the EMS experience. *European Economic Review* 41(3–5):743–751.

Fernandez, Raquel and Richard Rogerson. 1988. Income distribution, communities and the quality of education. *Quarterly Journal of Economics* 111(1):135–164.

Fujita, Masahisa, Paul Krugman, and Tomoya Mori. 1999*a*. On the evolution of hierarchical urban systems. *European Economic Review* 43(2).

Fujita, Masahisa, Paul Krugman, and Anthony Venables. 1999*b*. *The Spatial Economy: Cities, Regions and International Trade.* Cambridge MA: MIT Press.

Fujita, Masahisa and Paul R. Krugman. 1995. When is the economy monocentric? Von Thünen and Chamberlin unified. *Regional Science and Urban Economics* 25(4):508–528.

Gabaix, X. 1997. Zipf's law for cities. Processed, University of Harvard.

Greene, William H. 1997. *Econometric Analysis.* New Jersey: Prentice-Hall.

Gregory, Robert and B. Hunter. 1995. The macro economy and the growth of ghettos in urban poverty in Australia. Discussion Paper 325, Australian National Univesity.

Gregory, Robert and F. Vella. 1996. Selection bias and human capital investment: Estimating the rates of return to education of young males. *Labor Economics* 3(2).

Hanson, Gordon H. 1997*a*. Increasing returns, trade, and the regional structure of wages. *Economic Journal* 107(440):113–133.

Hanson, Gordon H. 1997*b*. Localization economies, vertical organisation, and trade. *American Economic Review* 86(5):1266–1278.

Hanson, Gordon H. 1998. Market potential, increasing returns, and geographic concentration. Working Paper 6429, National Bureau of Economic Research. URL `http://www.nber.org/`.

Harris, C. D. 1954. The market as a factor in the localization of industry in the united states. *Annals of the Association of American Geographers* 44:315–348.

Heath, Alex. 1998. Job search methods, neighbourhood effects and the youth labour market. Processed, Reserve Bank of Australia.

Henderson, Vernon. 1974. The sizes and types of cities. *American Economic Review* 64:640–656.

Henderson, Vernon. 1988. *Urban development: Theory, fact and illusion.* Oxford: Oxford University Press.

Hunter, B. 1996. Changes in the geographic dispersion of urban employment in Australia: 1976-1991. Unpublished PhD manuscript, Australian National University.

Jencks, Christopher and Susan Mayer. 1990. The social consequences of growing up in a poor neighbourhood. In L.E. Lynne Jr. and M.G.H. McGeary (eds.), *Inner City Poverty in the United States.* Washington D.C.: National Academy Press, Washington D.C., 111–186.

Keeble, David, John Offord, and Sheila Walker. 1988. *Peripheral Regions in a Community of 12 Member States.* Luxembourg: Office for Official Publications of the European Communities.

Krugman, Paul R. 1991. Increasing returns and economic geography. *Journal of Political Economy* 99:483–499.

Krugman, Paul R. 1993. First nature, second nature, and metropolitan location. *Journal of Regional Science* 33(2):129–144.

Krugman, Paul R. 1994. Past and prospective causes of high unemployment. *Federal Reserve Bank of Kansas City Economic Review* 79(4):23–43.

Krugman, Paul R. 1996. *The Self-Organizing Society.* Blackwell Publishers.

Krugman, Paul R. 1998. Space: The final frontier. *Journal of Economic Perspectives* 12:161–174.

Layard, Richard, Stephen Nickell, and Richard Jackman. 1991. *Unemployment: Macroeconomic Performance and the Labour Market.* Oxford: Oxford University Press.

Madden, C.J. 1956. Some indicators of stability in the growth of cities in the United States. *Economic Development and Cultural Change* 4:436–452.

Manacorda, Marco and Barbara Petrongolo. 1998. Skill mismatch and unemployment in OECD countries. Discussion Paper 307, Centre for Economic Performance, London School of Economics.

Manski, Charles. 1993. Identification of endogenous social effects: The reflection problem. *Review of Economic Studies* 60(3):531–542.

Manski, Charles. 1995. *Identification problems in the social sciences.* Cambridge: Harvard University Press.

Meghir, Costas. 1997. The effect of school quality on educational attainment and wages. Processed, University College London.

Miller, Paul and Paul Volker. 1987. The youth labour market in Australia. *Economic Record* 63(182):203–19.

Montgomery, James. 1991. Social networks and labour market outcomes: Towards an economic analysis. *American Economic Review* 81(5):1407–18.

Nickell, Stephen and Brian Bell. 1995. The collapse in demand for the unskilled and unemployment across the OECD. *Oxford Review of Economic Policy* 11(1):40–62.

Ottaviano, Gianmarco I. P. and Diego Puga. 1998. Agglomeration in the global economy: A survey of the 'new economic geography'. *The World Economy* 21(6):707–731.

Puga, Diego. 1999. The rise and fall of regional inequalities. *European Economic Review* 43(2):303–334.

Quah, Danny T. 1996. Regional convergence clusters across Europe. *European Economic Review* 40(3–5):951–958.

Quah, Danny Tyson. 1997*a*. Empirics for growth and distribution: Stratification, polarization and convergence clubs. *Journal of Economic Growth* 2:27–59.

Quah, Danny Tyson. 1997*b*. Regional cohesion from local isolated actions: I. historical outcomes. Discussion paper 378, Centre for Economic Policy Research.

Robertson, Donald and Jim Symons. 1996. Do peer groups matter? Peer group versus schooling effects on academic achievement. Discussion Paper 311, Centre for Economic Performance.

Rosen, K. and M. Resnick. 1980. The size distribution of cities: An examination of the Pareto law and primacy. *Journal of Urban Economics* 8:165–186.

Rosenblatt, M. 1971. Curve estimates. *Annals of Mathematical Statistics* 42:1815–1842.

Silverman, B. W. 1986. *Density Estimation for Statistics and Data Analysis.* New York: Chapman and Hall.

Simon, H. 1955. On a class of skew distribution functions. *Biometrika* 42:425–440.

Struefert, W. 1991. The effect of underclass isolation on schooling choice. Processed, University of Wisconsin.

Tabuchi, Takatoshi. 1998. Urban agglomeration and dispersion: A synthesis of Alonso and Krugman. *Journal of Urban Economics* 44:333–351.

UNESCO. 1976. *International Standard Classification of Education (*ISCED*).* Paris: United Nations Educational, Scientific and Cultural Organization.

Wasmer, Etienne. 1998. Can labour supply explain the rise of unemployment and inequality in OECD countries? Discussion Paper 410, Centre for Economic Performance, London School of Economics. URL `http://www.ecare.ulb.ac.be/ecare/etienne/home.htm`.

Wilson, William J. 1995. *The truly disadvantaged : the inner city, the underclass, and public policy.* Chicago: Chicago University Press.

Yakowitz, Sidney. 1985. Nonparametric density estimation, prediction, and regression for Markov sequences. *Journal of the American Statistical Association* 80:215–21.

Zipf, G. 1949. *Human Behaviour and the Principle of Least Effort.* Reading MA: Addison-Wesley.