

McArthur, David; Thorsen, Inge

**Conference Paper**

## Determinants of internal migration in Norway

50th Congress of the European Regional Science Association: "Sustainable Regional Growth and Development in the Creative Knowledge Economy", 19-23 August 2010, Jönköping, Sweden

**Provided in Cooperation with:**

European Regional Science Association (ERSA)

*Suggested Citation:* McArthur, David; Thorsen, Inge (2010) : Determinants of internal migration in Norway, 50th Congress of the European Regional Science Association: "Sustainable Regional Growth and Development in the Creative Knowledge Economy", 19-23 August 2010, Jönköping, Sweden, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/118913>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Determinants of internal migration in Norway <sup>\*</sup>

David Philip McArthur<sup>†</sup> and Inge Thorsen<sup>‡</sup>

## Abstract

This paper uses panel data from 2000-2008 on migration flows between municipalities in Norway. The relatively new method of fixed effect vector decomposition (fevd) is used to estimate parameters for time-invariant spatial structure variables as well as for labour market factors such as unemployment rates, income and industrial diversity. Results show that migrants tend to respond to ‘push’ factors in their home municipality rather than ‘pull’ factors at potential destinations. Several measures of spatial structure are experimented with, adding considerably to the explanatory power of the model, as expected.

*Keywords:* Internal migration, fixed effects vector decomposition, spatial structure, unemployment

*JEL codes:* J61, R23

---

<sup>\*</sup>This research is part of research programme 182639/V10 financially supported by the Research Council of Norway.

<sup>†</sup>Stord/Haugesund University College, Bjørnsonsgate 45, N-5528 Haugesund, Norway. e-mail: david.mcarthur@hsh.no

<sup>‡</sup>Stord/Haugesund University College, e-mail: inge.thorsen@hsh.no

# 1 Introduction

The aim of this paper is to explore the determinants of migration flows between municipalities in Norway. Understanding such flows is important in the formulation of both national and regional policy. For many years, successive Norwegian governments have actively pursued policies aimed at keeping the population spread out across the country. Despite this, there has been a trend of migration from the more peripheral areas to the cities. This suggests that the current policy approach is failing to address the root causes of the out migration.

The depopulation of the more peripheral areas of Norway can be decomposed into two components. The first is what can be termed ‘push’ factors, i.e. reasons for people to leave an area. The second can be termed ‘pull’ factors i.e. reasons for people to come to an area. If there are more reasons to leave than stay, an area will experience persistent net out-migration. Understanding the relative strength of such forces is important in order to design effective policy.

To account for the importance of ‘push’ and ‘pull’ factors, the flow of migrants between every possible origin and destination will be modelled as a function of the characteristics of that origin and destination. These characteristics will relate to the labour market conditions and the spatial structure characteristics of each origin and destination. Labour market conditions are measured using data on unemployment, average income and industrial diversity at the municipality level. Data on house prices is also included since the housing market interacts with the labour market. The effect of the presence of a higher education institution is also included in the model.

Measures of spatial structure take two forms in the analysis. Firstly, the centrality of a municipality is considered. An obvious hypothesis to test is whether more centrally located municipalities are more attractive to migrants than peripheral ones. Centrality measures are also used to test whether the response of migration flows to labour market variables is dependent on the centrality of the location. For instance, are migrants more responsive to changes in unemployment in peripheral regions than in centrally located regions? The underlying logic is that in a larger labour market the chances of re-employment will be higher than in a smaller one. The distance between each origin and destination is also included.

The responsiveness of migration to conditions in the labour market will provide important information about the responsiveness of the Norwegian economy to asymmetric regional shocks. For instance, for unemployment disparities to be eliminated, out-migration should be positively

related to unemployment while in-migration should be negatively related. If this is the case, then migration should act to smooth out differences over time. Asymmetries in the strength of the effects could lead to growing disparities over time. Information on this is an important input into the regional policy formation process.

The factors mentioned so far do not cover all possible influences on migration flows. For this reason, a fixed effects panel data approach is adopted to account for unobserved heterogeneity in origin-destination pairings. One problem with fixed effect models is that they are unable to estimate effects of time invariant regressors. Unfortunately, most spatial structure variables fall into this category e.g. distance. To deal with this problem, the fixed effect vector decomposition (fevd) method suggested by Plümper and Troeger (2007) is utilised. To the authors' knowledge, this is the first time this method has been employed in studies of internal migration.

The paper is structured as follows. Section 2 provides a brief outline of the literature relating to internal migration. This will primarily focus on the literature exploring the relationships between the labour market and aggregate flows of migration. Section 3 gives some details on the data used. Section 4 explains the methods adopted in the analysis while section 5 presents the results of the analysis. Section 6 provides a summary of the findings of the paper as well as looking at how these relate to policy.

## **2 Literature on Migration**

The literature covering the determinants of migration is vast and interdisciplinary in nature. Reviews are provided by Massey et al. (1993) and, more recently, by Cebula (2005). The main focus of this paper is the effect of the labour market and spatial structure on gross migration flows between an origin and destination. Of particular interest is the response of migration flows to regional disparities. Accordingly, the review presented here will focus on the literature relating to regional disparities.

Many of the studies of the relationship between labour market conditions and internal migration relates to the US and Europe. There are some differences in the findings. An important study of the functioning of regional labour markets was conducted by Blanchard and Katz (1992). The model is primarily concerned with disequilibrium disparities in unemployment rates between US states, and considers how adjustments in wages and the location decision of

households and firms act to close these disparities. Rather than directly modelling migration, it was assumed as a residual explanation. They found that it played an important role in closing regional disparities. However, subsequent work by a number of authors has questioned these findings.

Rowthorn and Glyn (2006) highlight the role played by measurements errors in time series approaches. They adjust for the likely magnitude for such errors in the Blanchard and Katz paper and find that the role of migration has most likely been overstated. An important assumption made in the paper was that employment rates were stationary. Partridge and Rickman (2006) test this assumption and fail to reject non-stationarity. Bartik (1993) examines the importance of the choice of functional form and also finds evidence of non-stationarity in employment rates.

While the role of migration in closing disparities in the US may have been overstated by Blanchard and Katz, it still seems that migration plays an important role. Bentivogli and Pagano (1999) and Decressin and Fátas (1995) found that migration flows were responsive to disparities in labour market conditions. They compare these results to Europe, where they conclude migration is much less responsive. Additional evidence relating to Europe is provided by Puhani (2001), who found that migration did not play a significant role when adjusting to a regional shock.

As well as flows of migration responding to differences in unemployment rates, it seems reasonable to expect that they would also respond to differences in wages. In the Blanchard and Katz (1992) model, average relative wages were allowed to vary but it was assumed that relative wages were stationary. Under these assumptions it was found that wage flexibility reduced the impact of a regional labour market shock but that there was only a weak effect on job creation and firm relocation. Partridge and Rickman (2003) and Partridge and Rickman (2007) use structural vector autoregressive (SVAR) models to explore wage flexibility in the US and Canada respectively. Both studies find evidence of short-run wage flexibility.

The evidence in Europe is less encouraging. Jimeno and Bentolila (1998) examine the role of wages in closing unemployment disparities in Spain between 1976 and 1994. They found that wages were very unresponsive. Abraham (1996) examined regions of the EU and found little evidence of wage flexibility. In addition, migration was found to be much lower in the EU than in the US. Buettner (2007) compared the EU-15 to the accession countries and found that it

took the economies of the EU-15 around twice as long to adjust to shocks. Siebert (1997) also found that wages in the EU were unresponsive to local labour market conditions.

The discussion so far has largely focused on what can be termed disequilibrium disparities. Such disparities represent a move away from an equilibrium position which is corrected through flows of migration, commuters, wage and employment adjustments and firm relocations. However, regional disparities may not represent a disequilibrium situation if some kind of compensating differentials are present.

One way in which a regional differential can be equilibrium in nature is through the relationship between wages and unemployment. This theory was proposed by Harris and Todaro (1970) in the context of developing economies. The theory suggests that migration flows respond to expected wages rather than the overall level of wages or unemployment, i.e. the relationship between wages and unemployment is positive. Although the theory was initially used to explain migration flows in developing economies, it is potentially relevant for European economies. Mulhern and Watson (2009a) find that a Harris-Todaro effect is important in explaining migration patterns in Spain. Partridge and Rickman (1997b) investigate whether the wage curve (Blanchflower and Oswald, 1994), invalidates the Harris-Todaro hypothesis. According to the wage curve hypothesis, the relationship between unemployment and wages is negative. Partridge and Rickman (1997b) show that the result depends on the dataset used and the choice of model specification. However, in most cases they find a positive relationship between unemployment and wages, hence validating the Harris-Todaro model.

Another important contribution to the literature is provided by Marston (1985). In this study, workers are assumed to be compensated for higher unemployment rates by the presence of amenities. This is in contrast to the Harris-Todaro model where workers are compensated through higher wages. A common example given of an amenity is a favourable climate. In the context of Norway, amenities could represent the abundant natural amenities offered by the more rural areas or the urban amenities offered by the cities. One problem with the amenity explanation is that there is no obvious way to measure amenities. To complicate matters, preferences are likely to vary across individuals. For instance, some individuals may place a high value on natural amenities and a negative value on urban amenities while others may take an opposite view. The outcome in such a case will be determined by the distribution of

preferences across the population, and the supply of such amenities.

Despite the problems in measuring amenities, a number of studies have examined their role in determining migration flows and regional disparities. Tuck et al. (2009) study the effect of amenities on teacher salaries in the US. They find that school districts with a high level of amenities find it easier to recruit and retain staff. Berger et al. (2008) develop a quality of life index for Russian cities and find a positive correlation between quality of life and net in-migration. In their study of the Canadian economy, Partridge and Rickman (2007) found that amenities did not play a significant role in explaining the flows of migrants. In contrast, they found that amenities did play a role in the US (Partridge and Rickman, 2006). Additional empirical support for the hypothesis can be found in Marston (1985) and Partridge and Rickman (1997a).

Two papers particularly important to this current study are presented by Mulhern and Watson (2009b) and Andrienko and Guriev (2004). They use a panel data approach to model flows between a set of origin and destination regions. Panel data has a number of desirable properties. Perhaps chief among them is the ability to control for unobserved heterogeneity between regions. This is particularly useful if amenities exist, are stable over time but cannot be accurately measured. The approach also allows the effect of labour market variables such as unemployment and income to be measured. Section 4 will outline the methodology employed in this paper in more detail.

### 3 The geography and data

The data used in this paper are taken from the 430 municipalities which, in 2008, made Norway excluding the Arctic archipelago of Svalbard, Jan Mayen and dependent territories. The data were gathered from Statistics Norway<sup>1</sup>. Norway has population of just over 4.8 million who enjoy one of the highest standards of living in the world. GPD per capita is in excess of 65,000 EUR, spread fairly evenly by international standards. Unemployment over the study period was low, rising no higher than 4.5%. The municipalities (kommuner) represent the lowest level of government. The municipalities form 20 counties (fylker). These counties can be further grouped into five regions (landsdeler). Figure 1 shows the population development in these

---

<sup>1</sup>[www.ssb.no/english](http://www.ssb.no/english)

regions from 1986-2009. Most of the country is sparsely populated. The four largest cities are the capital Oslo (575,000) in the East, Bergen (250,000) and Stavanger (120,000) in the west and Trondheim (170,000) in Trøndelag.

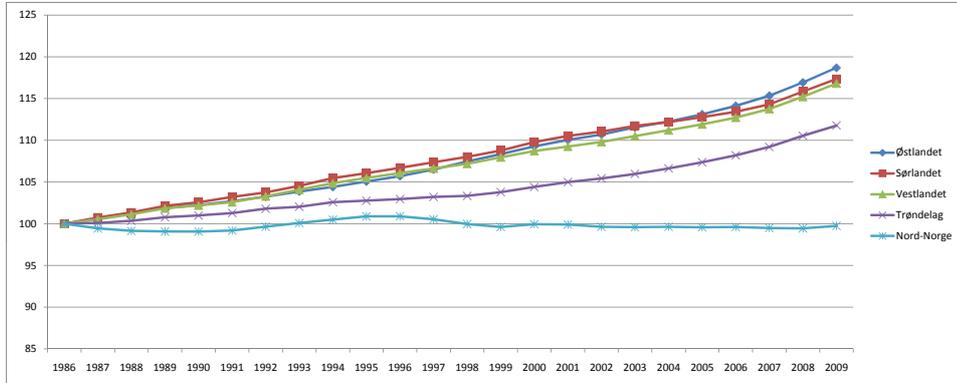


Figure 1: Population development in the 5 regions (landsdeler) from 1986-2009 where 1986=100. In 1986, the population in each region was as follows: Østlandet = 2,019,877; Sørlandet = 234,881; Vestlandet = 1,066,308; Trøndelag = 373,511 and Nord-Norge = 464,610.

Figure 1 shows that the most populous region, by far, is Østlandet. This region contains the capital Oslo. It has shown a fairly large increase in population over the period. The second most populous is Vestlandet. This prosperous region contains the cities of Bergen and Stavanger and is home to Norway's thriving petro-maritime industry. Vestlandet has shown robust population growth over the period. The remaining three regions are much smaller. Sørlandet has shown significant growth over the period, as has Trøndelag albeit not quite as strong. The population of Nord-Norge has stagnated over the time period shown. It is worth noting that these regions are large and hide many of the centralisation tendencies. For instance, the population level of Vestlandet has significantly increased over the period, while at the same time there has been a significant move from the peripheral areas of the region to the centres of Bergen and Stavanger.

The most important data collected were the total number of migrants between each pair of municipalities for the years 2000 to 2008 inclusive. The use of municipalities gives a more refined description of migration patterns than is possible in many other countries. It is possible in Norway because the national population register records where every citizen is living. This makes it possible to see where people move from and to in any given year. In total, 184,900 observations on flows between municipalities were recorded. Figure 2 shows Norway divided into its 430 municipalities and 5 regions.

Data on unemployment and income were also collected from Statistics Norway. When exper-

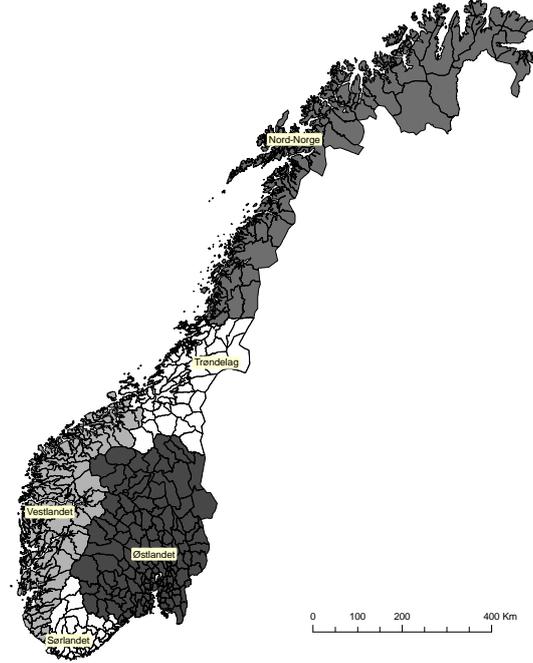


Figure 2: Norway, its 430 municipalities and 5 regions (landsdeler).

imenting with model specification, it was found that average male income gave more significant results than overall average income. If women are more likely to engage in part-time work, then the overall figure for average income may give skewed results. However, it is also possible that the outcome for a male is more important in determining migration decisions than those for women. There is evidence to support this hypothesis. For example, Taylor (2007) studies tied migration in the UK and finds that the employment concerns of the husband tends to dominate.

Another variable which is included is a measure of industrial diversity. The diversity measure attempts to capture the nature of the employment opportunities available in a municipality. It is based on the relative diversity index (RDI) suggested by Duranton and Puga (2000) and presented in McCann (2001 p. 82):

$$RDI_r = \frac{1}{\sum_i \left| \frac{E_{ir}}{E_r} - \frac{E_{in}}{E_n} \right|} \quad (1)$$

Here,  $RDI_r$  is the relative diversity index of municipality  $r$ ,  $E_{ir}$  is employment in industry  $i$  in municipality  $r$ ,  $E_r$  is the total employment in municipality  $r$ ,  $E_{in}$  is employment in industry  $i$  at the national level and  $E_n$  is total national employment. The index therefore measures how diverse a municipalities's labour market is relative to the national labour market.

A summary of the data is given in Table 3. One period lags of these variables are used in

the models to avoid problems of endogeneity, as in Mulhern and Watson (2009b).

Variable	Mean	Std. Dev.	n	Years
Population (Age 15-66)	5,438	16,407	430	2000-2008
Unemployment (%)	2.86%	1.64%	430	2000-2008
Male Income (NOK)	311,650	61,811	430	2000-2008
House prices (NOK/m <sup>2</sup> )	11,268	3,575	430	2002-2008

Table 1: Some summary statistics on the data.

The average house price data come with a caveat attached. Figures were available for most municipalities, but not all. When a municipality had no data, the county average was used instead. This means that care must be taken when interpreting results regarding house prices.

## 4 Methodology

The data consist of repeated observations on the same cross sectional units over time. This means that they constitute panel data. There are a number of ways in which such data can be handled. The treatment given here follows Greene (2008), Wooldridge (2002) and Johnston and DiNardo (1997). The most simple form of estimation would be to ignore the panel structure and estimate a pooled OLS model.

$$y_{ijt} = \alpha + \mathbf{x}'_{ijt}\beta + \mathbf{z}'_{ij}\gamma + \epsilon_{ijt} \quad (2)$$

Here,  $y_{ijt}$  is the flow of migrants from origin  $i$  to destination  $j$  in time  $t$  as a percentage of the population at  $i$ ,  $\alpha$  is a constant term,  $\mathbf{x}_{ijt}$  are variables describing the origin and destination,  $\mathbf{z}_{ij}$  are variables describing spatial structure which do not vary over time,  $\epsilon_{ijt}$  is an error term and  $\beta$  and  $\gamma$  are parameters to be estimated. In this model there are  $ijt$  observations from  $ij$  units. There are reasons to expect that observations from a particular unit will be more alike than observations from another. This can be captured by decomposing the error term:

$$\epsilon_{ijt} = u_{ij} + \eta_{ijt} \quad (3)$$

where  $u_{ij}$  is a origin/destination specific effect which varies across origin/destination pairs but is constant over time, and  $\eta_{ijt}$  is an error term uncorrelated with  $x$  and  $z$ . Consistent estimates of  $\beta$

and  $\gamma$  can be obtained using OLS if a number of assumptions hold. The first assumption is that  $E[\eta_{ijt}|x_{ij1}, x_{ij2}, \dots, z_{ij}] = 0$ , i.e. strict exogeneity. The second is that  $E[u_{ij}|x_{ij1}, x_{ij2}, \dots, z_{ij}] = u_{ij}$  i.e. mean independence.

While OLS will give consistent parameter estimates under these assumption, the standard errors will be incorrect since the information gained from  $ij$  units observed  $t$  times is not the same as the amount gained from  $ijt$  observations. The estimation will also be inefficient since it fails to incorporate information about the structure of the error term.

More efficient estimation can be achieved by using the random effects model. In order to do this, the additional orthogonality assumption is made that  $E[\eta_{ijt}|x_{ij1}, x_{ij2}, \dots, z_{ij}, u_{ij}] = 0$ . The random effects model views the group specific components,  $u_{ij}$ , as random variables being drawn from the same distribution. This is the reason why such strict orthogonality conditions are required. The estimation proceeds by utilising information about the variance-covariance structure in a feasible generalised least squares estimation.

One of the main problems with the random effects model is that its strict assumptions are rarely met in practice. It is often the case that the group specific effects are correlated with the included regressors. Take as an example Marston (1985). The theory states that regions with a high level of amenities will experience higher levels of unemployment. Many amenities, e.g. average climate, are fixed over time. This would lead to a correlation between the group specific effects and the regressors, and hence a violation of the assumptions of both the pooled OLS and random effects models.

The fixed effects estimator is an alternative which allows for correlation between  $u_{ij}$  and  $\mathbf{x}_{ijt}$  and  $\mathbf{z}_{ij}$ . There are two ways to estimate the fixed effect model. One way involves the use of dummy variables. However, this can lead to computational intractability when the number of observation is large. This is the case with the data in this paper. One way to avoid this problem is to subtract the mean of each of the variables over time. The deviation of  $m_{ijt}$  from its mean  $\bar{m}_{ij}$  is regressed on the deviation of  $\mathbf{x}_{ijt}$  and  $\mathbf{z}_{ij}$  from their means,  $\bar{\mathbf{x}}_{ij}$  and  $\bar{\mathbf{z}}_{ij}$ . At this stage, another problem emerges. Note that the  $\mathbf{z}$  variables (measures of spatial structure) do not have a time subscript. As a result, subtracting the time mean simply gives zero. The model to be estimated is given by:

$$y_{ijt} = \alpha + \mathbf{x}'_{ijt}\beta + \mathbf{z}'_{ij}\gamma + u_{ij} + \eta_{ijt} \quad (4)$$

However, when the model is time demeaned, the fixed effect  $u_{ij}$  is indistinguishable from the effect of time invariant variables  $\mathbf{z}_{it}$ , and both effects are removed from the equation. This could be termed a new combined fixed effect  $u_{ij}^*$ . If it is only the parameters attached to the time varying regressors,  $\beta$ , which are of interest, then this is not a problem. However, if the effect of the time invariant regressors is of interest, this method cannot be used.

A number of options are available at this stage. Hausman and Taylor (1981) suggest a consistent and efficient method using instrumental variables. The drawback, as with all instrumental variables approaches, is that its success in estimating consistent coefficients lies in the suitability of the instruments. Finding appropriate instruments is not an easy task. In a recent paper, Plümper and Troeger (2007) propose a new method which they call fixed effects vector decomposition (fevd). The method works by decomposing the combined fixed effects,  $u_{it}^*$ , into a component explained by the variables in  $\mathbf{z}$  and an unexplained part representing the unexplained fixed effect  $\zeta$ .

$$u_{ij}^* = u_{ij} + \mathbf{z}'\gamma \quad (5)$$

Estimates for  $\gamma$  can be obtained through OLS. Once  $\hat{\mathbf{u}}_{ij}$  has been obtained, the following model can be estimated:

$$y_{ijt} = \alpha + \mathbf{x}'\beta + \mathbf{z}'\gamma + \hat{u}_{ij} + \eta_{ijt} \quad (6)$$

Plümper and Troeger (2007) show, using Monte Carlo evidence, that the fevd model performs better than the pooled OLS model, the random effects model and the Hausman-Taylor model in estimating coefficients for time invariant regressors. The superior performance of this model is due to the fact that it uses information gained from cross-unit variation while controlling for fixed effects and time variable regressors.

## 5 Results

The fevd method has three distinct stages. The first is the estimation of a standard fixed effects panel model using all of the time-varying variables. Once a satisfactory model has been fitted, the vector of fixed effect is saved. The second stage is to decompose this vector into a part explained by time invariant regressors, e.g. distance, and an error component. The third stage combines the results of the previous two stages and gives a complete model of how both time varying and time invariant regressors affect the dependent variable.

### 5.1 The fixed effects model

Table 2 presents two model specifications. M1 shows a model with all of the variables under consideration included. M2 presents M1 with the insignificant variables removed.

	Description	M1	M2
lnUnempO	Log unemployment rate at origin	0.1015 (0.0056)	0.0979 (0.0051)
lnUnempD	Log unemployment rate at destination	-0.0059 (0.0049)	- -
lnUnempO*CENT	lnUnempO * centrality dummy	-0.0491 (0.0066)	-0.0487 (0.0066)
lnUnempD*CENT	lnUnempD * centrality dummy	-0.0132 (0.0054)	-0.0166 (0.0047)
lnmIncO	Log average male income at origin	-0.0819 (0.0200)	-0.0766 (0.0170)
lnmIncD	Log average male income at destination	-0.0005 (0.0172)	- -
lnHPO	Log average price per $m^2$ at origin	0.0280 (0.0082)	0.0302 (0.0081)
lnHPD	Log average price per $m^2$ at destination	0.0060 (0.0074)	- -
lnRDIO	Log regional diversity index at origin	0.0456 (0.0235)	0.0476 (0.0233)
lnRDID	Log regional diversity index at destination	0.0973 (0.0233)	0.1029 (0.0231)
CONS	Constant	-10.0614 (0.2025)	-10.1006 (0.1855)
Hausman test	Test for random effects	0.0000	0.0000
F-Test that $u_i = 0 \forall i$	Test that fixed effects are jointly zero	0.0000	0.0000

Table 2: Results of fixed effect panel estimation. Robust standard errors are given in parentheses.

The results given by M1 are broadly in line with expectations. Unemployment in a municipality is associated with an increase in out migration. This is to be expected and indicates that migration flows act to reduce unemployment within a municipality. Unemployment at a destination does not appear to act as a significant deterrent to in-migration. More information is provided about the effect of unemployment by the two interaction terms. These measure

the effect of unemployment in the most central areas. The most central areas were defined by Statistics Norway, who have classified all Norwegian municipalities into four levels of centrality. More will be said on this in Section 5.2 The coefficient on  $\ln\text{UnempO}*\text{CENT}$  is negative. When combined with the effect of  $\ln\text{UnempO}$ , this shows that unemployment acts to increase out-migration from the centrally located municipalities, but that the effect is not as strong as in less central areas.

This difference with respect to the effect of unemployment is in line with expectations. In a municipality with high accessibility to surrounding areas and with a thick labour market, there is less of a need to move out in response to a temporary negative employment shock. The second interaction term shows that unemployment acts as a slight deterrent to potential migrants. The magnitude of the effect is somewhat low. This is unsurprising given that overall unemployment is low, with no large regional disparities.

One interesting point to note is the asymmetry between the push and pull effects of unemployment. Unemployment has a stronger effect on out-migrants than in-migrants. This asymmetry suggests that, *ceteris paribus*, an adverse employment shock to a region will stimulate out-migration but that a fall in unemployment would not see a significant rise in in-migration. In such a case, repeated negative employment shocks, even if temporary, would lead to a fall in the population of a region. This is particularly true in less central municipalities.

Average male income is significant and signed as expected at the origin but not at the destination. Rising income in a municipality causes a fall in out-migration. It does not seem to encourage any in-migration. This could reflect differences in earnings between residents and new migrants. If migrants are engaged in speculative moving then it would seem reasonable to expect wages to act as a pull factor. However, if migrants secure employment before moving, they will be less concerned with the average wage at their destination.

House prices are signed as expected and significant at the origin but not at the destination. As house prices rise, it encourages out migration. This may reflect the fact that residents can release equity from their property by moving to a municipality with lower house prices. There appears to be no deterrence effect for in-migrants of rising house prices. This may be a real effect or simply the fact that the data on housing are not of sufficiently good quality. One municipality may contain several sub-markets, hence the average may not give much useful information. In

addition, the missing data for some municipalities may have had an effect.

It is not clear, *a priori*, what sign to expect for the regional diversity index (RDI) at the origin. In M1, there is a slight positive effect at the origin. A positive sign should be expected if more diverse labour markets experience higher turnover, i.e. more people moving in and out. At the destination, the effect of the RDI on in-migration is positive and significant. This is as expected since a thicker labour market can be expected to attract workers.

Specification M2 is a re-estimation of M1 with insignificant variables removed. This is the model which will be used in the next stage of the estimation. The vector of fixed effects from this regression,  $\mathbf{u}_i$ , is saved and will be regressed on time invariant regressors which could not be included in the fixed effect models presented in Table 2.

## 5.2 Time invariant regressors

One important time-invariant regressor is likely to be distance. Indeed, this is the basis of the gravity modelling tradition (Sen and Smith, 1995). There are a number of modelling decisions to be made regarding the inclusion of distance in the model. The first is on how to measure distance. The dataset used in this paper has three measures: Euclidean distances, driving distances and driving time. Euclidian distances are measured from the geometric centroid of each municipality with the other two measured using population weighted centroids. The second decision relates to the functional form of the distance measure. A number of different ways of including distance were tested. The exponential form presented in Equation 7 gave the most reasonable results.

$$D(d_{ij}) = \exp(-\rho d_{ij}) \quad (7)$$

The choice of distance measure gave only minor differences in results. The driving time was ultimately chosen as the best measure since it has an easier interpretation and is also better at capturing the effect of ferries. An additional benefit is that driving time is a potential policy instrument. This is particularly true in areas of western Norway where ferry connections have been substituted with bridges and tunnels, and where further such projects are currently being evaluated.

The distance between an origin and destination is not the only spatial structure variable

which is likely to be relevant in explaining the flow of migrants. The centrality or accessibility of a given location is likely to be important. Centrally located municipalities will have access to labour markets in several other municipalities or urban centres. This is likely to reduce the need for out migration. At the same time, a central location is likely to be a strong pull factor in attracting migrants. Access to a wider labour market is likely to improve the the job matching process and reduce the risk of prolonged periods of unemployment for a given individual.

To capture this effect, a centrality measure developed by Statistics Norway <sup>2</sup> is utilised. All municipalities in Norway were classified into 4 categories: least central, less central, quite central and central <sup>3</sup>. These data were used to create 16 possible origin-destination combinations e.g. a less central municipality to a central municipality. This allows trends in centralisation or decentralisation to be measured. In addition to measures of centrality, regional dummies are included for the five regions (landsdeler) of Norway: Nord-Norge, Trøndelag, Vestlandet, Østlandet and Sørlandet. These 5 broad regions were used to create dummies for the origin and destination of each migration flow. Trøndelag was omitted to avoid perfect multicollinearity.

As one further measure of labour market diversity a dummy indicating the presence of a university and one indicating the presence of a university college (Høgskole) in a municipality were included in one specification of the model. Another specification was estimated with a dummy variable indicating the presence of a university and/or university college. It is particularly useful to include such dummies since it is clear that a university or university college attracts students, some of whom will take up residence in the institution's municipality. All three model specifications are presented in Table 3.

Model M3 in Table 3 shows the results with all spatial structure variables and excluding any dummies representing universities or university colleges. The first point to note is that all of the variables included are highly significant. For most of the spatial structure dummies it is hard to know what signs to expect. The parameters attached to distance are at least correct, with the distance deterrence parameter,  $\rho$ , having a negative sign showing that migration falls with distance. The effect is also highly significant.

The centrality measures show some interesting trends. All of these dummies are positively signed. This is since they are measured relative to migration from the least central municipalities

---

<sup>2</sup>[www.ssb.no/english](http://www.ssb.no/english)

<sup>3</sup>Incidentally, this classification was the one used to create the centrality dummy CENT in Table 2.

	Description	M3	M4	M5
Constant	Constant	-0.5776 (0.0042)	-0.6130 (0.0036)	-0.6271 (0.0038)
Least-Less	Centrality O-D dummy	0.1618 (0.0049)	0.0848 (0.0043)	0.0390 (0.0045)
Least-Quite	Centrality O-D dummy	0.2770 (0.0043)	0.1468 (0.0038)	0.0953 (0.0039)
Least-Central	Centrality O-D dummy	0.4518 (0.0036)	0.2500 (0.0032)	0.3101 (0.0033)
Less-Least	Centrality O-D dummy	0.1832 (0.0049)	0.1592 (0.0043)	0.1523 (0.0045)
Less-Less	Centrality O-D dummy	0.3817 (0.0076)	0.2834 (0.0067)	0.2306 (0.0069)
Less-Quite	Centrality O-D dummy	0.5228 (0.0063)	0.3701 (0.0056)	0.3116 (0.0058)
Less-Central	Centrality O-D dummy	0.7063 (0.005)	0.4820 (0.0044)	0.5353 (0.0045)
Quite-Least	Centrality O-D dummy	0.3018 (0.0043)	0.2639 (0.0038)	0.2561 (0.0039)
Quite-Less	Centrality O-D dummy	0.4894 (0.0064)	0.3761 (0.0056)	0.3224 (0.0058)
Quite-Quite	Centrality O-D dummy	0.7076 (0.0054)	0.5426 (0.0048)	0.4831 (0.005)
Quite-Central	Centrality O-D dummy	0.9062 (0.0044)	0.6671 (0.0039)	0.7200 (0.004)
Central-Least	Centrality O-D dummy	0.2458 (0.0036)	0.2017 (0.0032)	0.2101 (0.0033)
Central-Less	Centrality O-D dummy	0.3795 (0.005)	0.2599 (0.0044)	0.2226 (0.0045)
Central-Quite	Centrality O-D dummy	0.5081 (0.0044)	0.3343 (0.0039)	0.2919 (0.004)
Central-Central	Centrality O-D dummy	0.5664 (0.0038)	0.3247 (0.0034)	0.3931 (0.0035)
Nord-Norge O	Region dummy (origin)	0.0978 (0.0033)	0.0929 (0.0029)	0.0941 (0.003)
Østlandet O	Region dummy (origin)	-0.0872 (0.0031)	-0.0832 (0.0027)	-0.0871 (0.0028)
Sørlandet O	Region dummy (origin)	-0.0950 (0.0042)	-0.0688 (0.0037)	-0.0626 (0.0038)
Vestlandet O	Region dummy (origin)	0.1254 (0.0031)	0.1330 (0.0027)	0.1350 (0.0028)
Nord-Norge D	Region dummy (destination)	0.1217 (0.0033)	0.0951 (0.0029)	0.1084 (0.003)
Østlandet D	Region dummy (destination)	-0.1144 (0.0031)	-0.0834 (0.0027)	-0.1121 (0.0028)
Sørlandet D	Region dummy (destination)	-0.2674 (0.0042)	-0.1747 (0.0037)	-0.1319 (0.0038)
Vestlandet D	Region dummy (destination)	-0.0862 (0.0031)	-0.0580 (0.0027)	-0.0430 (0.0028)
UniO	University at origin dummy	-	0.5277 (0.0061)	-
UniColO	Uni College at origin dummy	-	0.2365 (0.0028)	-
UniD	University at destination dummy	-	2.7945 (0.0061)	-
UniColD	Uni College at destination dummy	-	0.8071 (0.0028)	-
HEO	Uni or UC at origin dummy	-	-	0.2953 (0.0026)
HED	Uni or UC at destination dummy	-	-	1.2084 (0.0026)
Distance	Driving Time	0.0000 (0.0151)	6.5708 (0.0131)	6.6036 (0.0137)
$\rho$	Exponential distance parameter	-0.0132 (0.0000)	-0.0130 (0.0000)	-0.0131 (0.0000)
$R^2$		0.38	0.52	0.49

Table 3: Non-linear least squares regression of fixed effects,  $\mathbf{u}$ , on time invariant regressors.

to other least central municipalities. It is reasonable to assume that the least migration would occur in such a combination. The ordering of these dummies shows a clear tendency for migration to be higher when the destination is a more central municipality.

The values of the regional dummies from Model M4 are shown in Figure 3. These dummies measure the average fixed effect for each of the regions of Norway. They potentially capture a number of different kinds of effects, making their interpretation quite a complex matter. They are all, however, significantly different from zero. The first dummy shows that the number of migrants with an origin or destination in Nord-Norge is higher than would be expected compared to the other regions. The most likely explanation for this is that the North of Norway is remote. There are fewer municipalities and the distances between them and the rest of Norway is high. Therefore, if someone wants to migrate either in, out or within this region, they would have to tolerate a higher distance between their origin and destination.

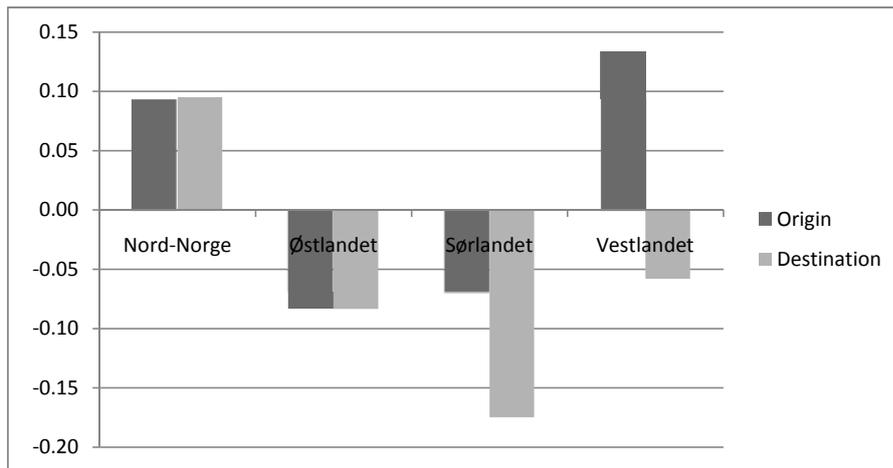


Figure 3: Regional dummies from Model M4.

The opposite kind of spatial structure exists in Østlandet, where there are more municipalities with shorter distances between them. Given this structure, it might be expected that there would be more migration. However, both the origin and destination dummies show that the number of migrants is actually lower than what would otherwise be expected. This may reflect greater commuting possibilities in this region than in the rest of the country. If people are able to access several labour markets from one location then they have less need to migrate.

Sørlandet experiences a significantly lower number of migrants when a destination lies there rather than in another region. This could reflect the fact that despite many of the municipal-

ities in this area having a reasonably high number of neighbours, the economy of the area has performed less well than the adjacent regions of Vestlandet and Østlandet.

The coefficients attached to the Vestlandet dummies show that flows originating in this region tend to have a slightly higher number of migrants while the opposite is true for flows terminating there. This would seem to suggest some degree of depopulation, when ignoring other factors. However, this is a large region with fairly diverse municipalities ranging from cities such as Bergen and Stavanger to some of the most rural of the municipalities in Norway. As a result, while parts of the region are shrinking, other parts, such as the cities, are growing.

Model M4 includes four dummies representing the presence of a university at the origin and destination and of a university college at the origin or destination. All four of these dummy variables are significant and signed as expected. The presence of one of these institutions at an origin increases the outward flow of migrants. This is to be expected since a portion of the students graduating at a given institution will leave. Also in line with expectations, a university has a larger effect than a university college. A similar story applies at the destination. A university acts as a significant pull factor, with a university college having a weaker but still significant effect. University colleges tend to draw a large proportion of their student population from the local area, therefore it is unsurprising that they have a smaller effect on migration flows than universities.

One possible problem with this specification is that Norway has only 14 institutions with university status. These are found exclusively in city areas. It is also worth noting that they are not spread across 14 municipalities e.g. Bergen has 2 universities. The correlation between ‘city effects’ and the effect of a university may lead to an inflated coefficient on the university dummy. The problem is less pronounced with the university colleges because these are spread from the cities to some very rural municipalities. As some of the university colleges are the result of mergers, dummies are included for the separate campuses. This further increases the geographic spread.

To examine the sensitivity of the coefficients to the correlation between universities and urban areas, a dummy was created to indicate the presence of a higher education institution with no distinction made between the type. The results of this are shown in Model M5. The two new dummies were significant and signed as expected. Some clues regarding the correlation between

the university dummy from Model M4 and ‘city effects’ can be seen in the coefficients relating to the centrality of the municipality. The coefficients on the centrality measures with a destination in a central municipality are highest in Model M3. They are lowest in Model M4 suggesting that some of their explanatory power has been stolen by the inclusion of four dummies. However it is unclear which is the best model since in M3 the centrality measure is correlated with the fact that central municipalities have access to either a university or a university college.

Clearly the determination of the precise effect of a higher education institution on migration flows lies beyond the scope of this paper and is a subject for future research. However, since the effect measured in M4 is consistent with prior expectations and increases the explanatory power of the model, this model is the one which will be carried forward to the final stage in the estimation process. Before moving on, it is worth commenting on the explanatory power of the model. The variables included in M4 explain over half of the variation in the fixed effects vector. This is respectable given the large number of omitted variables.

### 5.3 The final model

The final stage in the estimation process is to combine the two regressions shown in M2 (Table 2) and M4 (Table 3 ). This allows the coefficients for the time varying and invariant factors to be presented in the same model. To achieve this, the residuals from M4 in Table 3 are saved into vector  $\zeta$ . This vector now represents the fixed effects,  $\mathbf{u}$ , which are unexplained by the time invariant regressors. These are then included in an OLS regression of  $\ln\text{Migration}$  on all of the regressors. Results are presented in Table 4.

The coefficients in this model are identical to those presented in the two previous regressions, although the standard errors are different. The vector of unexplained fixed effects,  $\zeta$ , is included and has a coefficient of 1. This is due to the way it is constructed. Its inclusion allows the use of OLS estimation rather than panel data because it accounts for the fact that the model is built using repeated observations on the same units.

The adjusted  $R^2$  of this model is 0.61. There are a few points to note. Firstly, it is rather high given that the dataset is constructed at the relatively geographically disaggregated level of municipalities. Given this, it is only to be expected that there will be a substantial amount of noise in the system. Secondly, this explanatory power relies on the fact that the vector  $\zeta$  is

	Description	Coefficient	Std. Err.
Constant	Constant	-10.7136	0.0912
lnUnempO	Log unemployment rate at origin	0.0978	0.0026
lnUnempO*CENT	lnUnempO * centrality dummy	-0.0487	0.0031
lnUnempD*CENT	lnUnempD* centrality dummy	-0.0166	0.0020
lnmIncO	Log average male income at origin	-0.0766	0.0082
lnHPO2	Log average price per $m^2$ at origin	0.0302	0.0039
lnRDID	Log regional diversity index at origin	0.0476	0.0029
lnRDID	Log regional diversity index at destination	0.1029	0.0028
Least-Less	Centrality O-D dummy	0.0929	0.0034
Least-Quite	Centrality O-D dummy	-0.0832	0.0031
Least-Central	Centrality O-D dummy	-0.0688	0.0040
Less-Least	Centrality O-D dummy	0.1330	0.0032
Less-Less	Centrality O-D dummy	0.0951	0.0032
Less-Quite	Centrality O-D dummy	-0.0834	0.0033
Less-Central	Centrality O-D dummy	-0.1747	0.0041
Quite-Least	Centrality O-D dummy	-0.0580	0.0030
Quite-Less	Centrality O-D dummy	0.0848	0.0087
Quite-Quite	Centrality O-D dummy	0.1468	0.0087
Quite-Central	Centrality O-D dummy	0.2500	0.0085
Central-Least	Centrality O-D dummy	0.1592	0.0118
Central-Less	Centrality O-D dummy	0.2834	0.0141
Central-Quite	Centrality O-D dummy	0.3701	0.0133
Central-Central	Centrality O-D dummy	0.4820	0.0129
Nord-Norge O	Region dummy (origin)	0.2639	0.0118
Østlandet O	Region dummy (origin)	0.3761	0.0132
Sørlandet O	Region dummy (origin)	0.5426	0.0131
Vestlandet O	Region dummy (origin)	0.6671	0.0127
Nord-Norge D	Region dummy (destination)	0.2017	0.0116
Østlandet D	Region dummy (destination)	0.2599	0.0125
Sørlandet D	Region dummy (destination)	0.3343	0.0126
Vestlandet D	Region dummy (destination)	0.3247	0.0124
UniO	University at origin dummy	0.5277	0.0052
UniColO	Uni College at origin dummy	2.7945	0.0109
UniD	University at destination dummy	0.8071	0.0048
UniColD	Uni College at destination dummy	0.2365	0.0034
Distance	Driving Time (exponential)	6.5708	0.0131
$\zeta$	Unexplained fixed effect vector	1	0.0018
$R^2$		0.61	

Table 4: OLS regression of lnMigration on time varying and invariant regressors including unexplained fixed effect vector,  $\zeta$ .

present in the list of regressors. In a sense, this does not explain anything as it has no clear interpretation. The model does, however, allow the effects of the other regressors to be measured and predictions to be made about what might happen if the exogenous variables change.

## 6 Conclusion

The conclusions of this paper can be split into two components: the first relating to the method of fevd and the second relating to the results obtained. The fevd method proved very successful. It allowed the estimation of coefficients for time variant and invariant regressors within a panel-data framework. This is important for any model hoping to explain some kind of spatial interaction. In the analysis of migration flows presented in this paper, the measures of spatial structure employed (distance, measures of centrality and regional dummies) accounted for over a third of the variation in the fixed effects of the original panel model. Adding dummies representing the presence of a university or university college increased the explanatory power to over 50%. The ability to attribute this variation to a set of variables is very useful.

The results obtained were also of interest. With regard to regional disparities, the results were mixed. The model showed that migration acts to close unemployment disparities, with higher unemployment encouraging out-migration. However it only acted as a deterrent in the centrally located municipalities. Changes in the unemployment rate had a much stronger effect on the current residents of a municipality than on potential in-migrants, irrespective of centrality. Increases in the average price per square meter for housing tended to cause a rise in the number of migrants leaving a municipality but did not seem to discourage new ones from moving in. An increase in the average level of male income is associated with a fall in out-migration but did not appear to attract new migrants.

There were also evidence in the data of increasing divergence between regions. Rising industrial diversity acted as a strong incentive for new migrants. It is likely that such an effect would be self re-enforcing. If a thick labour market attracts more labour, particularly highly skilled labour, then in the next period industrial diversity may further increase. This could be caused by greater specialisation and/or human capital spillover effects. The dummies included in the model to capture any centralisation tendencies also presented evidence of divergence. There is a clear trend for migrants to move from the least central municipalities to the more central

ones. Such a trend could also re-enforce itself. As rural areas depopulate, it becomes harder to provide the services necessary to sustain a community. If this happens, it could accelerate any depopulation and give even stronger incentives for further centralisation.

It is likely that centralisation tendencies will have other feedback mechanisms too. It is much easier for a more densely populated area to support a diverse labour market than a rural area. Given the importance of such labour markets in attracting migrants, centralisation at a particular point in time is likely to lead to more in the future. The same is true of universities and university colleges which were also found to play a role in attracting migrants. These self-perpetuating factors could make it very difficult to sustain a geographically dispersed population in the long run. If this is the aim of Norwegian policy makers, then a more proactive approach is needed since it appears the current trend is running in the opposite direction.

## References

ABRAHAM, F., 1996. Regional adjustment and wage flexibility in the European Union. *Regional Science and Urban Economics*, **26**(1), 51-75.

ANDRIENKO, Y. and GURIEV, S.M., 2004. Determinants of interregional mobility in Russia. *Economics of Transition*, **12**(1), 1-27.

BARTIK, T., 1993. Who benefits from local job growth: migrants or the original residents? *Regional Studies*, **27**(4), 297-311.

BENTIVOGLI, C. and PAGANO, P., 1999. Regional Disparities and Labour Mobility: the Euro-11 versus the USA. *Labour*, **13**(3), 737-760.

BERGER, M.C., BLOMQUIST, G.C. and SABIRIANOVA PETER, K., 2008. Compensating differentials in emerging labor and housing markets: Estimates of quality of life in Russian cities. *Journal of Urban Economics*, **63**(1), 25-55.

BLANCHARD, O.J. and KATZ, L.F., 1992. Regional Evolutions. *Brookings Papers on Economic Activity*, **1**(Macroeconomics), 1-75.

BLANCHFLOWER, D.G. and OSWALD, A.J., 1994. The Wage Curve. First edn. MIT Press.

BUETTNER, T., 2007. Unemployment disparities and regional wage flexibility: comparing EU members and EU-accession countries. *Empirica*, **34**(4), 287-297.

CEBULA, R.J., 2005. Internal migration determinants: Recent evidence. *International Advances in Economic Research*, **11**(3), 267-274.

CEBULA, R.J., 2005. Internal Migration Determinants: Recent Evidence. *International Advances in Economic Research*, **11**(3), 267-274.

DECRESSIN, J. and FATS, A., 1995. Regional labor market dynamics in Europe. *European Economic Review*, **39**(9), 1627-1655.

DURANTON, G. and PUGA, D., 2000. Diversity and specialisation in cities: why, where and when does it matter? *Urban Studies*, **37**(3), 533-555.

FIDRMUC, J., 2004. Migration and regional adjustment to asymmetric shocks in transition economies. *Journal of Comparative Economics*, **32**(2), 230-247.

GREENE, W.H., 2008. Econometric analysis. Sixth edn. Upper Saddle River, NJ: Pearson: Prentice Hall.

HARRIS, J.R. and TODARO, M.P., 1970. Migration, Unemployment and Development: A Two-Sector Analysis. *The American Economic Review*, **60**(1), 126-142.

HAUSMAN, J.A. and TAYLOR, W.E., 1981. Panel Data and Unobservable Individual Effects. *Econometrica*, **49**(6), 1377-1398.

JIMENO, J.F. and BENTOLILA, S., 1998. Regional unemployment persistence (Spain, 1976–1994). *Labour Economics*, **5**(1), 25-51.

JOHNSTON, J. and DINARDO, J., 1997. Econometric Methods. Fourth edn. Singapore: McGraw Hill.

LPEZ-BAZO, E., BARRIO, T.D. and ARTS, M., 2005. Geographical distribution of unemployment in Spain. *Regional Studies*, **39**(3), 305.

MARSTON, S.T., 1985. Two Views of the Geographic Distribution of Unemployment. *The Quarterly Journal of Economics*, **100**(1), 57-79.

MASSEY, D.S., HUGO, G., KOUAOUCI, A., PELLEGRINO, A. and TAYLOR, J.E., 1993. Theories of international migration: a review and appraisal. *Population and development review*, **19**(3), 431-466.

MCCANN, P., 2001. Urban and Regional Economics. 1st edn. Oxford: Oxford University Press.

MULHERN, A. and WATSON, J.G., 2009. Spanish Internal Migration: Is there Anything

New to Say? *Spatial Economic Analysis*, **4**(1), 103-120.

MULHERN, A. and WATSON, J., 2009. Spanish inter-regional migration: an enigma resolved. *Applied Economics Letters*, **99999**(1), 1-5.

PARTRIDGE, M.D. and RICKMAN, D.S., 1997. The Dispersion of US State Unemployment Rates: The Role of Market and Non-market Equilibrium Factors. *Regional Studies: The Journal of the Regional Studies Association*, **31**(6), 593-606.

PARTRIDGE, M.D. and RICKMAN, D.S., 2006. Fluctuations in aggregate US migration flows and regional labor market flexibility. *Southern Economic Journal*, **72**(4), 958-980.

PARTRIDGE, M.D. and RICKMAN, D.S., 2003. The waxing and waning of regional economies: the chicken-egg question of jobs versus people. *Journal of Urban Economics*, **53**(1), 76-97.

PARTRIDGE, M.D. and RICKMAN, D.S., 1997. Has the wage curve nullified the Harris-Todaro model? Further US evidence. *Economics Letters*, **54**(3), 277-282.

PARTRIDGE, M.D. and RICKMAN, D.S., 2007. Canadian regional labour market evolutions: a long-run restrictions SVAR analysis. *Applied Economics*, **99999**(1), 1-17.

PLÜMPER, T. and TROEGER, V.E., 2007. Efficient estimation of time-invariant and rarely changing variables in finite sample panel analyses with unit fixed effects. *Political Analysis*, **15**(2), 124-139.

PUHANI, P.A., 2001. Labour Mobility: An Adjustment Mechanism in Euroland? Empirical Evidence for Western Germany, France and Italy. *German Economic Review*, **2**(2), 127-140.

ROWTHORN, R. and GLYN, A., 2006. Convergence and stability in US employment rates. *Contributions to Macroeconomics*, **6**(1), 1-42.

SEN, A. and SMITH, T., 1995. **Gravity** Models of Spatial Interaction Behavior. First edn. Berlin: Springer.

SIEBERT, H., 1997. Labor Market Rigidities: At the Root of Unemployment in Europe. *The Journal of Economic Perspectives*, **11**(3), 37-54.

TAYLOR, M. and PARK, W., 2007. Tied migration and subsequent employment: Evidence from couples in Britain. *Oxford Bulletin of Economics and Statistics*, **69**(6), 795-818.

TUCK, B., BERMAN, M. and HILL, A., 2009. Local amenities, unobserved quality, and market clearing: Adjusting teacher compensation to provide equal education opportunities.

*Economics of Education Review*, **28**(1), 58-66.

WOOLDRIDGE, J.M., 2002. *Econometric analysis of cross section and panel data*. First edn. Cambridge MA: The MIT press.