Exploring Brexit with dynamic spatial panel models: some possible outcomes for employment across the EU regions

Bernard Fingleton[1] 📷

Abstract

Starting with a reduced form derived from standard urban economics theory, this paper estimates the possible job-shortfall across UK and EU regions using a time-space dynamic panel data model with a spatial moving average random effects structure of the disturbances. The paper provides a logical rational for the presence of spatial and temporal dependencies involving the endogenous variable, leading to estimates based on a state-of-the-art dynamic spatial generalized moments estimator proposed by Baltagi et al. (Reg Sci Urban Econ, 2018. https://doi.org/10.1016/j.regsciurbeco.2018.04.013). Given reliable interregional trade estimates, the simulations are based on a linear predictor which utilizes different regional interdependency matrices according to assumptions about interregional trade post-Brexit. The results indicate that heightened barriers to trade will evidently cause job-shortfalls both in the UK and across the EU, but it is also shown that there is a considerable amount of asymmetry in the outcomes across regions and sectors.

1 Introduction

The paper aims to simulate potential outcomes for employment across 255 EU regions as a result of the impending and, at the time of writing, probable departure of the UK from the EU, commonly known as Brexit. Most analysts agree that Brexit will have momentous consequences for the UK and (remaining) EU economies, but there is very minimal analysis at the regional scale, and analysis typically fails to account for interconnectivity at the regional level. Some regional impact studies have been carried out by Dhingra et al. (2017a, b), Los et al. (2017) and McCann (2018), and the current paper complements or contrasts with this research by applying a state-of-the-art dynamic spatial panel data model, in which a pan-European

---

✉ Bernard Fingleton
   bf100@cam.ac.uk

1   Department of Land Economy, University of Cambridge, Cambridge, UK

approach is adopted involving the majority of EU regions[1] and all UK regions. This modelling approach is ideally suited to capturing the impact of spatial interconnectivity of the European regions and projecting the long-run consequences of Brexit across EU and UK regions, thus enabling comparison of the impact on both sides of the Channel and the Irish Sea. To this end, we use the dynamic spatial panel data model and prediction equation recently introduced into the literature by Baltagi et al. (2018) and applied in different contexts by Fingleton et al. (2018) and Fingleton and Szumilo (2019). The model assumes that employment in a given region depends on the levels of production and investment within that region, as shown in the basic economic model which underpins the estimating equation, and it also depends on demand coming from all the regions of the EU and UK, as determined by interregional trade flows. Additionally, the model proposes that employment levels in any regional are closely linked to employment levels in the region in the previous period and on employment levels in trade-connected EU and UK regions in the previous period. Following this literature, a rational basis for the presence of spatial and temporal lags is introduced which more typically is ad hoc in the spatial econometrics literature. In addition, the model takes account of unobserved factors which also affect the level of employment. These are captured by region-specific random effects which are also spatially interdependent. An additional feature of the approach adopted is the way in which endogeneity is handled, with the application of internal instruments in the spirit of Arellano and Bond (1991), thus eliminating the often difficult search for valid external instruments.

The focus of analysis is the so-called job-shortfall which could arise due to Brexit. In other words, the intention is not to forecast what happens to the actual levels of employment in each region, which would be the predicted change in the number of jobs, but to simulate what the impact of Brexit would be assuming no consequential responses such as jobs created by new trade links formed post-Brexit, changes to the UK's competitivity and consequences for demand and employment due to changes in exchange rates and prices, changes to migration flows in and out of the UK, changes in the competitivity of firms if trade barriers are increased and regulations relaxed, and possible changes in levels of inward and outward investment and capital stock if capital relocates. Contemplating these and other possible consequences enhance uncertainty regarding what might be the actual change in the levels of employment in the UK regions, so in this paper the focus is on attempting to simulate the job-shortfall due to Brexit per se.

Stated more explicitly, the empirical analysis bases the spatial interdependence of levels of employment across different regions on how closely they are connected in terms of trade. We assume that employment levels partly reflect demand for a given region's good and services coming from the UK and EU regions. Naturally, since about 50% of the UK's trade in 2019 is with countries outside the EU, demand coming from these non-EU countries will also affect the levels of employment. For both UK and EU regions generally, we assume that the non-EU component of demand

---

[1] Although Switzerland and Norway are formally outside the EU, they are an integral part of the analysis. Their regions are referred to as EU regions for convenience.

is reflected by the levels of production and investment within each region. In this way, we have attempted to isolate the impact of reduced trade between UK and EU regions from the potential effects of changes in non-EU trade on employment. In the simulations, the non-EU trade flows and hence capital and output are assumed to remain at the same level as previously, thus leading to our focus on job-shortfall rather than job-loss or job-gain. Moreover, in an effort to make the simulations robust, simplifying assumptions are made regarding the impact of Brexit on trade flows between individual UK and EU regions, leading to a geography of the Brexit impact that is immune to changes in the actual level of assumed trade.

More specifically, estimates are made of employment levels across $N = 255$ EU regions both with and without Brexit. The explicit drivers of employment are output and capital, which are approximated by Gross Value Added (GVA) and a function of Gross Fixed Capital Formation (GFCF), respectively. Estimation is based on a viable data series over the period from 2001 to 2010. Data for 2011 and 2012 are not used in estimation but held for one- and two-step ahead prediction. Different assumptions can be made about post-2011 paths for GVA and GFCF, given that accessible data with the same geography are not available over the more recent period, although it has been found that these have relatively little effect on outcomes.

The structure of the paper is as follows. In Sect. 2, the model is outlined, and the data described in Sect. 3. Section 4 summarizes the estimator, and Sect. 5 the resulting estimates. Section 6 focuses on prediction methodology, and Sect. 7 gives details of the method for simulating the Brexit effect. Section 8 gives the simulation results, and Sect. 9 concludes. The "Appendices" section give the theoretical basis of the estimating equation, provide details of the outcomes from alternative estimators, and summarize the Chow–Lin approach to obtaining an interregional connectivity matrix. The last section lists the references.

2 The model

The reduced form used as a basis to simulate the Brexit effect assumes that employment partly depends on level of output, as measured by GVA (Gross Value Added), denoted by $\mathbf{q}_t$, and (a proxy for) the level of capital within the region, based on GFCF (Gross Fixed Capital Formation), which is denoted by $\mathbf{k}_t$. To show this, we start with the theoretical model given as Eq. (1), which is based on Eq. (30) given in "Appendix". The N by 1 vector $\mathbf{e}_t^d$ is the density of employment per unit area, and $\mathbf{a}_t$ is the level of efficiency of labour at time t, so that the product $\mathbf{e}_t^d \mathbf{a}_t$ is the number of labour efficiency units. This is related to $\widetilde{\mathbf{q}}_t$, which is a measure of output in the competitive final goods and services sector in each region at time t, via the constant parameters ϕ and $\widetilde{\gamma}$, thus

$$\widetilde{\mathbf{q}}_t = \phi(\mathbf{e}_t^d \mathbf{a}_t)^{\widetilde{\gamma}} \tag{1}$$

In order to obtain total output $\mathbf{q}_t$, it is assumed that $\widetilde{\mathbf{q}}_t = \boldsymbol{\pi} \mathbf{q}_t$, in which $\boldsymbol{\pi}$ is an N by 1 vector giving the share of total output in each region that is competitive final goods and services output. For simplicity of estimation, it is assumed that $\boldsymbol{\pi}$ is

constant over time. Also the employment levels are $\mathbf{e}_t = \mathbf{h}\mathbf{e}_t^d$ in which $\mathbf{h}$ is the area of land in each region. Taking logs gives

$$\ln \boldsymbol{\pi} + \ln \mathbf{q}_t = \ln \phi + \widetilde{\gamma} \ln \mathbf{e}_t + \widetilde{\gamma} \ln \mathbf{a}_t - \widetilde{\gamma} \ln \mathbf{h} \tag{2}$$

Rearranging (2) gives

$$\ln \mathbf{e}_t = \frac{1}{\widetilde{\gamma}} \left(\ln \boldsymbol{\pi} + \ln \mathbf{q}_t - \ln \phi \right) - \ln \mathbf{a}_t + \ln \mathbf{h}$$

To obtain (3), I assume that labour efficiency $\mathbf{a}_t = \frac{\mathbf{q}_t}{\mathbf{k}_t}$, with more efficient labour having a higher level of output per unit of capital $\widetilde{\mathbf{k}}_t$. As shown below in Eq. (14), an approximation to the log level of capital is $\ln \widetilde{\mathbf{k}}_t = -\ln \widetilde{a} + \widetilde{b} \ln \mathbf{k}_t$, hence $\ln \mathbf{a}_t = \ln \mathbf{q}_t + \ln \widetilde{a} - \widetilde{b} \ln \mathbf{k}_t$, and from this

$$\ln \mathbf{e}_t = \frac{1}{\widetilde{\gamma}} \left(\ln \boldsymbol{\pi} + \ln \mathbf{q}_t - \ln \phi \right) + \ln \mathbf{h} - \ln \mathbf{q}_t - \ln \widetilde{a} + \widetilde{b} \ln \mathbf{k}_t \tag{3}$$

Collecting together constants as c with $\boldsymbol{\iota}$ an N by 1 vector of ones, and reorganizing gives

$$\ln \mathbf{e}_t = c\boldsymbol{\iota} + \frac{1 - \widetilde{\gamma}}{\widetilde{\gamma}} \ln \mathbf{q}_t + \widetilde{b} \ln \mathbf{k}_t + \boldsymbol{\varepsilon}_t \tag{4}$$

in which the error term $\boldsymbol{\varepsilon}_t$ captures the time-invariant regional heterogeneity in land $\mathbf{h}$ and in shares $\boldsymbol{\pi}$, which are unobserved, as given in Eq. (12).

In the dynamic context, it is reasonable to assume that disparities in employment levels across locations will persist as an equilibrium outcome to unchanging and fundamental causes. We therefore proceed, following in parallel the exposition in Baltagi et al. (2018), to assume that (log) employment levels across regions, denoted by the N by 1 vector $\ln \mathbf{e}_t$ at time t will persist at dynamically stable levels so that $\ln \mathbf{e}_t = \ln \mathbf{e}_{t-1}$ unless there are changes in the levels of $\mathbf{q}_t$ or $\mathbf{k}_t$, or changes in common factors, interregional trade, or unobserved effects. If such a disturbance occurs at time t and is ephemeral, then $\ln \mathbf{e}_t \neq \ln \mathbf{e}_{t-1}$ but over a subsequent period of quiescence $t \rightarrow T$ then once again we expect employment levels to converge on a new equilibrium, at which $\ln \mathbf{e}_T = \ln \mathbf{e}_{T-1}$. Assume data are observed where $\ln \mathbf{e}_t \neq \ln \mathbf{e}_{t-1}$ but tending to converge, so that $\ln \mathbf{e}_t = f(\ln \mathbf{e}_{t-1})$, and an autoregressive process is assumed, hence

$$\ln \mathbf{e}_t = \boldsymbol{\varsigma} + \gamma \ln \mathbf{e}_{t-1} \tag{5}$$

in which $\boldsymbol{\varsigma}$ is an N by 1 vector and γ is a scalar parameter. In the long-run with $abs(\gamma) < 1$, and with no subsequent disturbances, the process converges to $\ln \mathbf{e}_T = \frac{\boldsymbol{\varsigma}}{(1-\gamma)}$.

Consider next connectivity between regions in the form of a matrix $\mathbf{W}_N^*$, which is a time-invariant N by N matrix, where N is the number of regions. Spatial interdependence between regions is a feature of many different situations, and can be modelled either via an autoregressive process involving the dependent variable, or via spatial interdependence of the errors, or by both as in this paper. The problem of how to model dependence between N regions is typically resolved by application of

an N by N matrix of constant quantitative values or weights assigned to the cells of $\mathbf{W}_N^*$ which indicate the existence and importance of a link between each pair of regions. In many spatial econometric applications, connectivity of the N regions will be some function of the distance between them, be it geographical distance or some measure of economic distance. In the paper, we proxy economic distance by the level of trade between each pair of regions, more trade equals shorter economic distance. Usefully, $\mathbf{W}_N^*$ provides a parsimonious parametrization for interdependence between, in this case, employment levels in different regions. As explained by LeSage and Pace (2009), once we allow for dependence relations between a set of N by N entities on a single variable, for example as represented by the N by 1 vector $\ln \mathbf{e}_t$, there are potentially $N^2 - N$ parameters that define individual interdependence, such as the relation between $\ln \mathbf{e}_{it}$ and $\ln \mathbf{e}_{jt}$, having excluded dependence of an observation on itself. This leads to an over-parametrization problem, which can be solved by imposing an a priori structure, or weights matrix $\mathbf{W}_N^*$, on the interdependence relations, thus reducing the number of parameters to be estimated from $N^2 - N$ to one, denoted here by ρ_1. For purposes of interpreting parameter estimates, we normalize by dividing $\mathbf{W}_N^*$ by the maximum eigenvalue of $\mathbf{W}_N^*$ to give[2] $\mathbf{W}_N$. Using this normalization, the maximum eigenvalue of $\mathbf{W}_N$ is 1, and the continuous range for which $\left(\mathbf{I}_N - \rho_1 \mathbf{W}_N\right)$ is non-singular is $\frac{1}{\min(eig)} < \rho_1 < \frac{1}{\max(eig)} = 1$, in which ρ_1 is a scalar spatial autoregressive parameter.

Multiplying (5) by $\rho_1 \mathbf{W}_N$ gives

$$\rho_1 \mathbf{W}_N \ln \mathbf{e}_t = \rho_1 \mathbf{W}_N \varsigma + \rho_1 \mathbf{W}_N \gamma \ln \mathbf{e}_{t-1} \tag{6}$$

Subtracting (6) from (5) leads to another logically consistent expression, in which the spatial dependence implied by (6) can be seen in (7) as an explicit cause of variation in $\ln \mathbf{e}_t$. Thus,

$$\ln \mathbf{e}_t - \rho_1 \mathbf{W}_N \ln \mathbf{e}_t = \varsigma + \gamma \ln \mathbf{e}_{t-1} - \left(\rho_1 \mathbf{W}_N \varsigma + \rho_1 \mathbf{W}_N \gamma \ln \mathbf{e}_{t-1}\right)$$
$$\left(\mathbf{I}_N - \rho_1 \mathbf{W}_N\right) \ln \mathbf{e}_t = \left(\gamma \mathbf{I}_N - \rho_1 \gamma \mathbf{W}_N\right) \ln \mathbf{e}_{t-1} + \left(\mathbf{I}_N - \rho_1 \mathbf{W}_N\right)\varsigma$$

Writing $\theta = -\rho_1 \gamma$ gives

$$\ln \mathbf{e}_t = \mathbf{B}_N^{-1}\left[\mathbf{C}_N \ln \mathbf{e}_{t-1} + \mathbf{B}_N \varsigma\right] \tag{7}$$

in which $\mathbf{B}_N = \left(\mathbf{I}_N - \rho_1 \mathbf{W}_N\right)$, $\mathbf{C}_N = \left(\gamma \mathbf{I}_N + \theta \mathbf{W}_N\right)$ and $\mathbf{I}_N$ is an identity matrix of order N. In order to solve Eq. (7), given appropriate parameter restrictions, Eq. (7) converges to $\ln \mathbf{e}_T = \left(\mathbf{B}_N - \mathbf{C}_N\right)^{-1} \mathbf{B}_N \varsigma$.

Introducing additional covariates by writing $\mathbf{B}_N \varsigma = (c\boldsymbol{\iota} + \mathbf{x}\boldsymbol{\beta})$, in which $c\boldsymbol{\iota}$ is a constant N by 1 vector, $\mathbf{x}$ is an N by k matrix and $\boldsymbol{\beta}$ is a k by 1 vector, gives

$$\ln \mathbf{e}_t = \mathbf{B}_N^{-1}\left[\mathbf{C}_N \ln \mathbf{e}_{t-1} + c\boldsymbol{\iota} + \mathbf{x}\boldsymbol{\beta}\right]$$

---

[2] The matrix $\mathbf{W}_N$ retains (scaled) absolute levels rather than shares as the basis of interregional connectivity, and we make the standard assumptions for a weights matrix, that it comprises fixed (non-stochastic) non-negative values with zeros on the leading diagonal and its row and column sums are uniformly bounded in absolute value, and maintain the same assumption for $\mathbf{B}_N^{-1} = \left(\mathbf{I}_N - \rho_1 \mathbf{W}_N\right)^{-1}$ (Elhorst 2014, p. 99).

In order to maintain dynamically stable simulations, following Elhorst (2001, 2014, p. 98), Parent and LeSage (2011, p. 478; 2012, p. 731) and Debarsy et al. (2012, p. 162), the largest characteristic root $\left(e_{max}\right)$ of $\mathbf{B}_N^{-1}\mathbf{C}_N$ should be less than 1. This restriction ensures that employment converges to equilibrium levels

$$\ln \mathbf{e}_T = \left(\mathbf{B}_N - \mathbf{C}_N\right)^{-1}(c\boldsymbol{\iota} + \mathbf{x}\boldsymbol{\beta}) \tag{8}$$

Additional realism is introduced in three ways. First, the restriction that $\theta = -\rho_1\gamma$ is removed since this greatly simplifies estimation. However, we anticipate that $\widehat{\theta} \approx -\widehat{\rho}_1\widehat{\gamma}$. Second, taking account of the variables in Eq. (4), the time-invariant matrix $\mathbf{x}$ is replaced by time-varying matrix[3] $\mathbf{x}_t$. Third, spatially dependent unobservables are represented by the error term ε_t. Although the system may, depending on $\mathbf{B}_N^{-1}\mathbf{C}_N$, still tend towards equilibrium, equilibrium will be continuously disturbed and new equilibrium levels established as t varies. For simplicity of estimation, interregional connectivity is assumed to remain constant over the estimation period. These considerations lead to the model of employment levels[4] given in Eqs. (9,10,11,12), which is a time-space dynamic panel data model, thus

$$\ln \mathbf{e}_t = \mathbf{B}_N^{-1}\left[\mathbf{C}_N \ln \mathbf{e}_{t-1} + c\boldsymbol{\iota} + \mathbf{x}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t\right] \tag{9}$$

Given $\mathbf{x}_{1t} = \ln \mathbf{q}_t,\ \mathbf{x}_{2t} = \ln \mathbf{k}_t,\ \mathbf{x}_{3t} = \ln \overline{\mathbf{e}}_t, \mathbf{x}_t = \left[\mathbf{x}_{1t}\ \mathbf{x}_{2t}\ \mathbf{x}_{3t}\right]$ and $\boldsymbol{\beta} = \left[\beta_1\ \beta_2\ \beta_3\right]^T$, Eq. (9) can be stated more explicitly as

$$\begin{aligned} \ln \mathbf{e}_t = {}& c\boldsymbol{\iota} + \gamma \ln \mathbf{e}_{t-1} + \rho_1\mathbf{W}_N \ln \mathbf{e}_t + \beta_1 \ln \mathbf{q}_t \\ & + \cdots \beta_2 \ln \mathbf{k}_t + \beta_3 \ln \overline{\mathbf{e}}_t + \theta\mathbf{W}_N \ln \mathbf{e}_{t-1} + \boldsymbol{\varepsilon}_t \end{aligned} \tag{10}$$

$$\boldsymbol{\varepsilon}_t = \mathbf{u}_t - \rho_2\mathbf{M}_N\mathbf{u}_t \tag{11}$$

$$\begin{aligned} & u_{it} = \mu_i + v_{it} \quad i = 1, \ldots, N, t = 1, \ldots, T \\ & \mu_i \sim iid(0, \sigma_\mu^2) \\ & v_{it} \sim iid(0, \sigma_v^2) \end{aligned} \tag{12}$$

The presence of the district-invariant mean of the dependent variable $\overline{\mathbf{e}}_t$ attempts to allow for the presence of observed or unobserved common factors affecting all districts at each point in time. This approach is motivated by Pesaran (2015) who provides a major treatise on the different approaches to modelling dynamic spatial panel data with common factors, and by Bailey et al. (2016) who ask, 'to what extent are the observed dependencies between different spatial units due to common factors—for example, aggregate shocks—that affect different units rather than being the result of local interactions that generate spatial spillover effects?'. They propose the use of cross-unit averages to extract common factors, an approach that has also been

---

[3] We assume that the elements of $\mathbf{X}_t$ are uniformly bounded in absolute value.

[4] Changes in employment levels are driven by changes in the fundamental variables in the theoretical model, and could come about as a result of consequent changes in unemployment rates, levels of economic activity, and the number of UK and international inward and outward migrants.

applied by Fingleton et al. (2018) and Fingleton and Szumilo (2019). The introduction of common factors to spatial econometric models has also been considered by among others Vega and Elhorst (2016) and Ertur and Musolesi (2017).

The disturbances ε_t capture the effects of the spatially dependent unobserved variables, with a compound structure (12) comprising time-invariant unobserved unit-specific interregional heterogeneity represented by μ_i with $i = 1, \dots, N$ and unobserved idiosyncratic shocks represented by $v_{it}; i = 1, \dots, N, t = 1, \dots, T$. These are assumed to be independent of each other and are collectively represented by u_{it}. It is important to recognize that the $\mu_i s$ represent the net effect of unobserved variables which in the short run can be treated as time invariant.

Most usually, the assumption is that spatial dependence is an autoregressive (SAR-RE) process, such that $\varepsilon_t = \rho_2 \mathbf{M}_N \varepsilon_t + \mathbf{u}_t$. However, in this paper the assumption for the error process is a spatial moving average process (SMA-RE) as in Eq. (11), thus $\varepsilon_t = \mathbf{G}_N \mathbf{u}_t$, where $\mathbf{G}_N = \left(\mathbf{I}_N - \rho_2 \mathbf{M}_N \right)$. This means that the error process is such that a shock in a region affects only neighbouring regions as defined by a row-standardized interregional contiguity matrix[5] $\mathbf{M}_N$. In contrast, an SAR-RE process would entail shocks affecting all regions. There are two reasons for this. First, assuming SMA-RE rather than SAR-RE errors improves the predictive performance of the estimator, as described in Sect. 6. Second, SMA-RE errors might proxy for omitted spillovers, which otherwise might be captured by the spatial lags $\mathbf{W}_N \mathbf{x}_t$. This is pertinent since the presence of $\mathbf{W}_N \mathbf{x}_t$ on the right-hand side of (9) could adversely affect estimation. As explained by Fingleton et al. (2017) and Baltagi et al. (2018), an SMA-RE error specification 'mitigates against the problem for instrumental variable estimation identified by Pace et al. (2012)'. In two-stage least squares (2SLS) estimation, the instrument set should comprise the 'exogenous' variables ($\mathbf{x}_t$) and their spatial lags ($\mathbf{W}_N \mathbf{x}_t$), and kept to a low order to avoid linear dependence and retain full column rank for the matrix of instruments (1998, Kelejian and Prucha 1999). The performance of the estimation procedure could be suboptimal, as explained by Pace et al. (2012), by including $\mathbf{W}_N \mathbf{x}_t$ among the set of explanatory variables. This is because with spatial lags ($\mathbf{W}_N \mathbf{x}_t$) among the set of regressors, then spatial lags of the spatial lags ($\mathbf{W}_N^2 \mathbf{x}_t, \mathbf{W}_N^3 \mathbf{x}_t, \dots$) feature among the instruments, and this could lead to a weak instrument problem. To avoid this, SMA-RE errors are adopted as an alternative way to capture local spillovers.

3 Data

In estimating Eq. (10), data for employment ($\mathbf{e}_t$), output as measured by Gross Value Added (GVA, $\mathbf{q}_t$) and capital as proxied by a function of Gross Fixed Capital Formation (GFCF, $\mathbf{k}_t$), both denominated in €2005m, are taken from the Cambridge Econometrics European Regional Economic database, with observations over the

---

[5] The matrix $\mathbf{M}_N$ has $\widetilde{\widetilde{e}}_{max}^{-1} = 1$, where $\widetilde{\widetilde{e}}$ is the vector of purely real characteristic roots of $\mathbf{M}_N$. Often $\mathbf{M}_N = \mathbf{W}_N$, otherwise we assume that $\mathbf{M}_N$ has the same properties as $\mathbf{W}_N$, and with the restriction that $\widetilde{\widetilde{e}}_{min}^{-1} < \rho_2 < \widetilde{\widetilde{e}}_{max}^{-1} = 1$ one guarantees the invertibility of $\mathbf{G}_N$ as in Eq. (22).

10-year period from 2001 to 2010 used to estimate the model. Data are also available for 2011−2012, but are held back to allow out-of-sample prediction tests of the model and some rivals. $\mathbf{k}_t$ is used to reflect capital stock $\widetilde{\mathbf{k}}_t$, for which data are unavailable, on the basis of a simple relationship which is assumed to exist between the two variables. $\mathbf{k}_t$ measures gross net investment (acquisitions minus disposals of produced fixed assets) in fixed capital assets and so provides an indicator of changes to the stock of capital. The assumption is that $\mathbf{k}_t$ is a nonlinear function of a constant fraction $\widetilde{a}$ of $\widetilde{\mathbf{k}}_t$ so that

$$\mathbf{k}_t = \left(\widetilde{a} \widetilde{\mathbf{k}}_t \right)^{\frac{1}{\widetilde{b}}} \tag{13}$$

hence

$$\widetilde{\mathbf{k}}_t = \frac{1}{\widetilde{a}} \left(\mathbf{k}_t^{\widetilde{b}} \right) \tag{14}$$

As a test of the viability of this approximation, assume a standard model for the evolution of capital stock which is depreciating at a constant rate $\widetilde{d}$ so that

$$\widetilde{\mathbf{k}}_t = \mathbf{k}_t + (1 - \widetilde{d}) \widetilde{\mathbf{k}}_{t-1}; \quad t = 2, \dots, T \tag{15}$$

in which T is a large number. One problem with (15) is that it requires the initial capital stock at time $t = 1$, i.e. $\widetilde{\mathbf{k}}_1$. However, given arbitrary values for $\widetilde{\mathbf{k}}_1$ and $\widetilde{d}$, values for $\widetilde{a}$ and $\widetilde{b}$ can be found, whereby (14) provides a reasonable approximation to the outcome of iterations (15). A more realistic test is provided by the existence of both (albeit experimental estimates of) capital stock[6] (Derbyshire et al. 2010) and of well-founded GFCF data. Using the latest available data for both $\mathbf{k}_t$ and $\widetilde{\mathbf{k}}_t$, which is for the year, $t = 2008$, and taking logs of (14), leads to a loglinear regression of $\ln \widetilde{\mathbf{k}}_t$ on $\ln \mathbf{k}_t$ which gives OLS estimates of the constant $\ln \widetilde{a}^{-1} = 2.4546$ (t ratio = 13.5628) and slope $\widetilde{b} = 1.0195$ (t ratio = 50.8118), with $R^2 = 0.8888$. The plot of $\ln \mathbf{k}_t$ against $\ln \widetilde{\mathbf{k}}_t$ shows a significant linear relationship and no evidence of outliers or of heteroscedasticity It thus appears that the model given as Eq. (13) provides a good approximation. The estimated $\widetilde{a} = 0.0859$ suggests the approximate proportion of the capital stock that is invested, and, by comparison, $\sum \mathbf{k}_t / \sum \widetilde{\mathbf{k}}_t = 0.0686$.

The matrix $\mathbf{W}_N$ is based on estimated bilateral trade flows between EU NUTS2 regions. The data come from the PBL (the Netherlands Environmental Assessment Agency)[7] who developed a new methodology which is close to that of Simini et al. (2012). Details of the methodology are given in Thissen et al. (2013, 2013a, b), see also Gianelle (2014). The method follows a top-down approach and therefore is consistent with the national accounts of the different countries. Given the total international exports and imports on the country level, interregional trade flows are derived using data on business travel (services) and on freight transport (goods).

---

[6] I am grateful to Cambridge Econometrics for providing these data.

[7] We are grateful to Mark Thiessen, who kindly provided the data. The data can be visualized at http://themasites.pbl.nl/eu-trade/index2.html?vis=net-scores.

Additionally, exports that went to EU destination countries' final demand[8] were also included. Trade flows involving regions of non-EU countries such as Switzerland and Norway were obtained on the basis of interregional trade flows estimated by the best linear disaggregation method of Chow and Lin (1971), which was initially used to break down annual time series into quarterly series (see Abeysinghe and Lee 1998; Doran and Fingleton 2014). In this, commencing with aggregate trade values[9] between 21 EU counties, these were allocated to the NUTS2 regions. A parallel approach has been used by Polasek et al. (2010), Vidoli and Mazziotta (2010), and Fingleton et al. (2015). More details of the method are provided in "Appendix". Finally, OLS regression of the log PBL trade flows on log Chow–Lin trade flows produced parameters used to predict the missing PBL regional trade flows for Switzerland and Norway using the values for these regions obtained via the Chow–Lin approach. For estimation, the start-of-period trade flows for the year 2000 are used. This year is chosen because it is the earliest available, so it is treated as exogenous to $\mathbf{e}_t$, $\mathbf{q}_t$ and $\mathbf{k}_t$, for $t = 2001$ to 2010. Prediction is based on the 2010 trade flows supplemented in the same way by Chow–Lin data. Estimates are also given in "Appendix" Table A3 based on a $\mathbf{W}_N$ matrix constructed entirely from the Chow–Lin trade flows. These simply use great circle distances and year 2000 GVA levels, and so are also assumed to be exogenous. The comparative predictive performance of each set of estimates is discussed in Sect. 6.

4 Estimator for the time-space dynamic panel data model

Comprehensive overviews of spatial panel econometrics are given by Pesaran (2015, Chapters 29 and 30) and Baltagi (2013, Chapter 13) which highlight its growing importance for the applied econometrician. The estimator used in this paper, introduced by Baltagi et al. (2018), adds to the available methodology by allowing a wider range of spatial interaction effects which include the spatial lag of the temporal lag of the dependent variable $W_N \ln \mathbf{e}_{t-1}$, thus avoiding bias due to constraints necessary for dynamic stability and stationarity, and also by allowing spatial moving average compound error dependence rather than the usual autoregressive compound error process found in the majority of spatial econometric models. The estimator, which is applied to Eq. (9), is based on the earlier paper by Baltagi et al. (2014), which extends the approach of Arellano and Bond (1991) by the introduction of extra moments in line with the presence and availability of spatial lags (see also Bouayad-Agha and Védrine 2010). Since the estimator is described elsewhere, a simple outline sketch is provided here focussing on the treatment of regressors

---

[8] 1. Final consumption expenditure by households and non-profit organizations

  2. Final consumption expenditure by government

  3. Net capital formation

  4. Inventory adjustment

[9] They are downloadable from http://cid.econ.ucdavis.edu/data/undata/undata.html, see also Feenstra et al. (2005).

as predetermined rather than exogenous.[10] Hence in Eq. (10), $\ln \mathbf{q}_t$ and $\ln \mathbf{k}_t$ are considered to be predetermined alongside endogenous right-hand side variables $\ln \mathbf{e}_{t-1}, \mathbf{W}_N \ln \mathbf{e}_{t-1}$ and $\mathbf{W}_N \ln \mathbf{e}_t$ and $\ln \bar{\mathbf{e}}_t$.

Focussing on the endogenous dependent variable $\ln \mathbf{e}_t$, the instruments include $\ln \mathbf{e}_t$ lagged by two periods, and its spatial lag $\mathbf{W}_N \ln \mathbf{e}_t$ also lagged by two periods, so that the moments equations (16 and 17) hold assuming v_{it} is serially uncorrelated and $E(\Delta v_{it}, \Delta v_{it-2}) = 0$. Thus, following Baltagi et al. (2018), with, we have

$$E\left(\ln e_{il} \Delta v_{it}\right) = 0 \quad \forall i, l = 1, 2, \ldots, T-2; t = 3, 4, \ldots T \tag{16}$$

$$E\left(\sum_{i \neq j} w_{ij} \ln e_{il} \Delta v_{it}\right) = 0 \quad \forall i, l = 1, 2, \ldots, T-2; t = 3, 4, \ldots T \tag{17}$$

in which E denotes the expectation. Also, if we were to assume exogenous rather than predetermined regressors $(\mathbf{x}_1, \mathbf{x}_2)$ this leads to (18)

$$\mathbf{Z}_t = \begin{pmatrix} \ln \mathbf{e}_1, \ldots, \ln \mathbf{e}_{t-2}, \mathbf{W}_N \ln \mathbf{e}_1, \ldots, \mathbf{W}_N \ln \mathbf{e}_{t-2}, \mathbf{x}_{11}, \ldots, \mathbf{x}_{1T}, \\ \mathbf{x}_{21}, \ldots, \mathbf{x}_{2T}, \mathbf{W}_N \mathbf{x}_{11}, \ldots, \mathbf{W}_N \mathbf{x}_{1T}, \mathbf{W}_N \mathbf{x}_{21}, \ldots, \mathbf{W}_N \mathbf{x}_{2T}, \\ \mathbf{x}_{31}, \ldots, \mathbf{x}_{3t-2}, \mathbf{W} \mathbf{x}_{31}, \ldots, \mathbf{W}_N \mathbf{x}_{3t-2,} \end{pmatrix} \tag{18}$$

for $t = 3, \ldots, T$. Given that in (18) the regressors $(\mathbf{x}_1, \mathbf{x}_2)$ are exogenous, the moments equations are satisfied including the entire set
$\mathbf{x}_{11}, \ldots, \mathbf{x}_{1T}, \mathbf{x}_{21}, \ldots, \mathbf{x}_{2T}, \mathbf{W}_N \mathbf{x}_{11}, \ldots, \mathbf{W}_N \mathbf{x}_{1T}$ and $\mathbf{W}_N \mathbf{x}_{21}, \ldots, \mathbf{W}_N \mathbf{x}_{2T}$ regardless of time t. As explained in Baltagi et al. (2018), additional instruments can be generated via the matrix $\mathbf{W}_N^2$, but for simplicity these are omitted from the estimators used in the current paper.

Strict exogeneity rules out any feedback from past shocks to current values of the variable, and the need to accommodate feedback leads to the preferred estimator based on predetermined regressors (see Bond 2002; Pesaran 2015). Predetermined regressors are contemporaneously uncorrelated, so that $corr(\mathbf{x}_t, \mathbf{v}_t) = 0$, but do depend on earlier shocks so that, for example, $corr(\mathbf{x}_t, \mathbf{v}_{t-1}) \neq 0$. This means that an adjustment to $\ln \mathbf{e}$, which embodies $\mathbf{v}$, at time t does not have an instantaneous effect on output and capital investment time t but takes effect at $t + 1$ and later. This allows an extension to the set of instruments (compared with assuming endogeneity, where all endogenous variables are lagged by two periods), by the inclusion of $\mathbf{x}_{1t-1}, \mathbf{x}_{2t-1}, \mathbf{W}_N \mathbf{x}_{1t-1}$, and $\mathbf{W}_N \mathbf{x}_{2t-1}$ so that

$$\mathbf{Z}_t = \begin{pmatrix} \ln \mathbf{e}_1, \ldots, \ln \mathbf{e}_{t-2}, \mathbf{W}_N \ln \mathbf{e}_1, \ldots, \mathbf{W}_N \ln \mathbf{e}_{t-2}, \\ \mathbf{x}_{11}, \ldots \mathbf{x}_{1t-2}, \mathbf{x}_{1t-1}, \mathbf{x}_{21}, \ldots \mathbf{x}_{2t-2}, \mathbf{x}_{2t-1}, \mathbf{x}_{31}, \ldots \mathbf{x}_{3t-2}, \\ \mathbf{W}_N \mathbf{x}_{11}, \ldots \mathbf{W}_N \mathbf{x}_{1t-2}, \mathbf{W}_N \mathbf{x}_{1t-1}, \mathbf{W}_N \mathbf{x}_{21}, \ldots \mathbf{W}_N \mathbf{x}_{2t-2}, \mathbf{W}_N \mathbf{x}_{2t-1}, \\ \mathbf{W}_N \mathbf{x}_{31}, \ldots \mathbf{W}_N \mathbf{x}_{3t-2} \end{pmatrix} \tag{19}$$

Given the set of instruments as in Eq. (19), these are used to obtain initial estimates of $\gamma, \rho_1, \theta, \beta_1, \beta_2$ and β_3, having first differenced the data to eliminate the time-invariant

individual effects μ which are correlated with the time and space-lagged dependent variables. The resulting estimates are then used to give estimated errors which lead to estimates of the parameters of the spatial moving average error process, namely ρ_2, σ_μ^2 and σ_ν^2 using moments equations given in Fingleton (2008). Given these, preliminary one-stage consistent spatial GM estimates are obtained, followed by the two-stage Spatial GM estimates of γ, ρ_2, θ and β based on a robust version of the variance-covariance matrix.

5 Estimates

Table 1 shows that the θ estimate for the spatial lag of the temporal lag ($\mathbf{W}_N \ln \mathbf{e}_{t-1}$) is not dissimilar to $-\widehat{\gamma}\widehat{\rho}_1$, in line with expectation stemming from an equilibrium

Table 1 Estimates of Eq. 9

Parameter	Param. est.	St. error	t-ratio
γ	0.6765	0.002948	229.4
ρ_1	0.5566	0.01037	53.67
β_1	0.1216	0.002062	58.94
β_2	0.03274	0.0007239	45.22
β_3	− 0.0201	0.006553	− 3.067
θ	− 0.4009	0.008314	− 48.22
ρ_2	− 0.7646	0.047227	− 16.27
σ_μ^2	0.1347		
σ_ν^2	0.0003		

ρ_2 t-ratio based on null bootstrap distribution of ρ_2
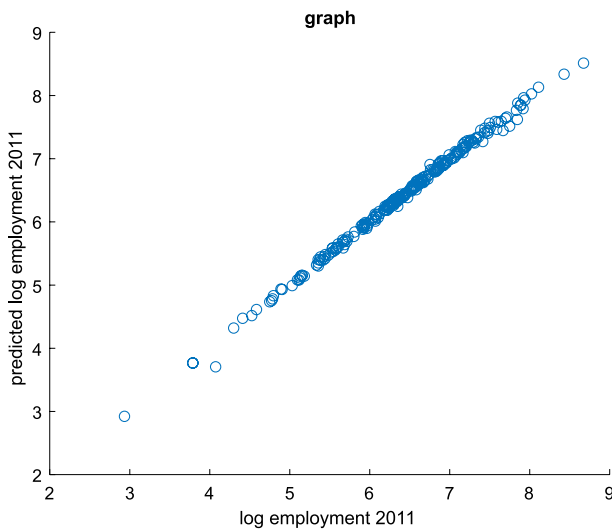

Fig. 1 Out-of-sample predictions for 2011

Table 2 London—doubling trade reduction and regressor levels

Trade reduction	2%	4%	8%	16%			
	Job-shortfall				Ratios		
2011 levels	− 2.56	− 5.11	− 10.20	− 20.35	1.996	1.996	1.995
2 × 2011 levels	− 2.70	− 5.40	− 10.79	− 21.52	2.000	1.998	1.994
Ratios	1.055	1.055	1.056	1.057			

process. Also, Table 1 estimates are stationary and dynamically stable, as shown by the largest characteristic root of $\mathbf{B}_N^{-1}\mathbf{C}_N$ which is equal to 0.6874, and the stationary bounds for ρ_2 are $\widetilde{e}_{min}^{-1} = -1.1239 < \rho_2 < \widetilde{e}_{max}^{-1} = 1$. Observe that the negative values of $\widehat{\rho}_2$ imply positive spatial dependence among the errors. Among the instrument set, we have several endogenous variables, one is the dependent variable lagged by two periods, $\ln \mathbf{e}_{t-2}$, its spatial lag lagged by two periods $\mathbf{W}_N \ln \mathbf{e}_{t-2}$, $\ln \overline{\mathbf{e}}_{t-2}$ and $\mathbf{W}_N \ln \overline{\mathbf{e}}_{t-2}$. To satisfy the orthogonality conditions and moments equations for these instruments, we require a lack of serial correlation in the v_{it}, in other words we need to satisfy the assumption that $E\big(\Delta v_{it}, \Delta v_{it-2}\big) = 0$. Arellano and Bond (1991) give a

Table 3 Paris—doubling trade reduction and regressor levels

Trade reduction	2%	4%	8%	16%			
	Job-shortfall				Ratios		
2011 levels	− 1.30	− 2.59	− 5.16	− 10.27	1.992	1.992	1.990
2 × 2011 levels	− 1.37	− 2.74	− 5.47	− 10.88	2.000	1.996	1.989
Ratios	1.054	1.058	1.060	1.059			
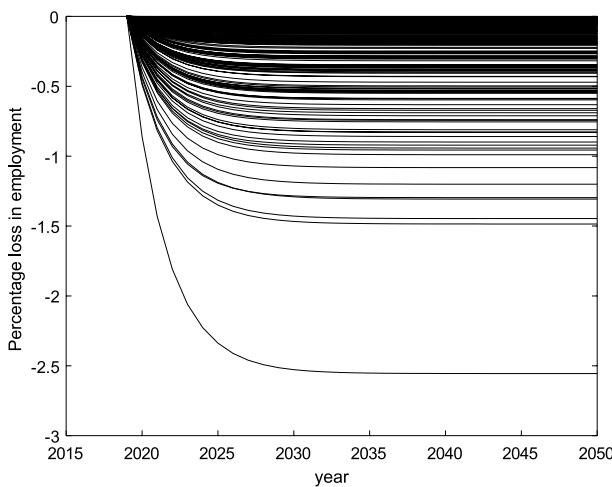
Fig. 2 Dynamic paths for % employment shortfall
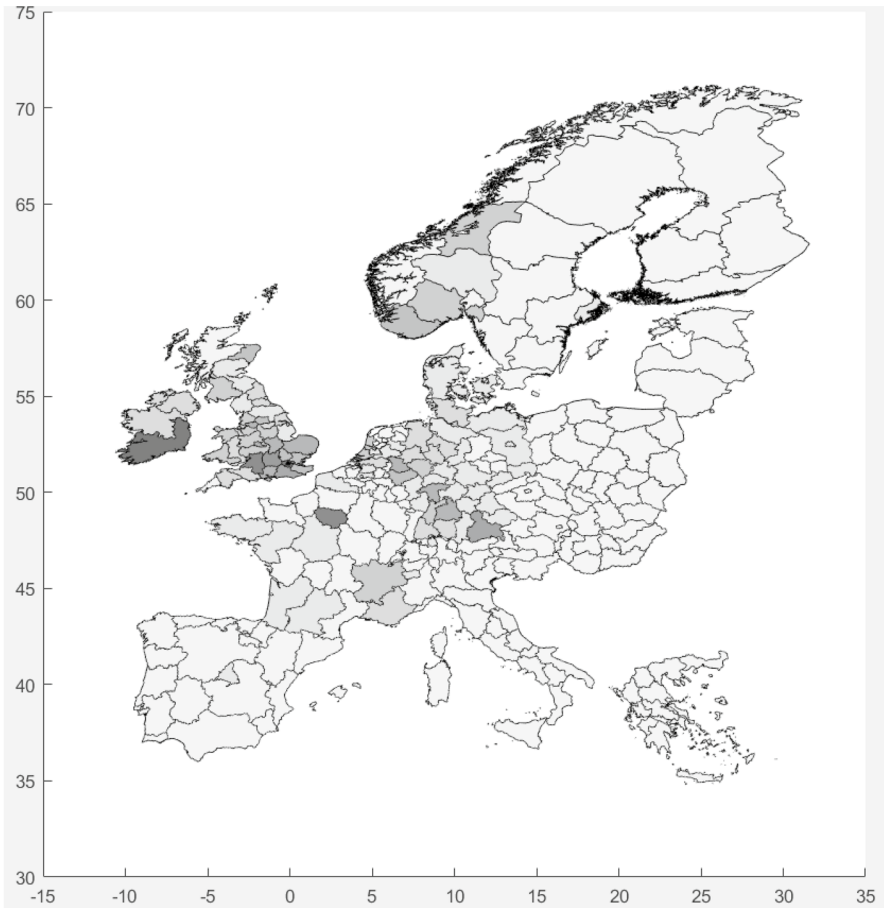
Fig. 3 % employment shortfall across 255 EU regions

test $m_2 = cov\left(\Delta v_{it}, \Delta v_{it-2}\right)/s.e.$ which is asymptotically $N(0, 1)$ under the null of no serial correlation. In our case $m_2 = -0.9389$ with two-tailed p value equal to 0.3478. Thus, we assume that there is an absence of serial correlation as required. Note also that $m_1 = cov\left(\Delta v_{it}, \Delta v_{it-1}\right)/s.e = -6.22$ indicating significant first-order serial correlation as one would expect, since if the v_{it} are serially uncorrelated, Δv_{it} has first order moving average serial correlation. A second complementary approach to testing the validity of the instrument set is via the application of the Sargan–Hansen test of over-identifying restrictions, which is equal to 253.3. This is insignificant when referred to the χ^2_{314} distribution, and while this evidently supports the moments conditions implied by our dynamic spatial panel model, one should be cautious because it may have low power, given the presence of many moments conditions (Bowsher 2002; Pesaran 2015).

"Appendix" Table 4 gives the estimates of some rival estimators, including one with SMA-RE errors but assuming exogenous regressors (Table A1), and with
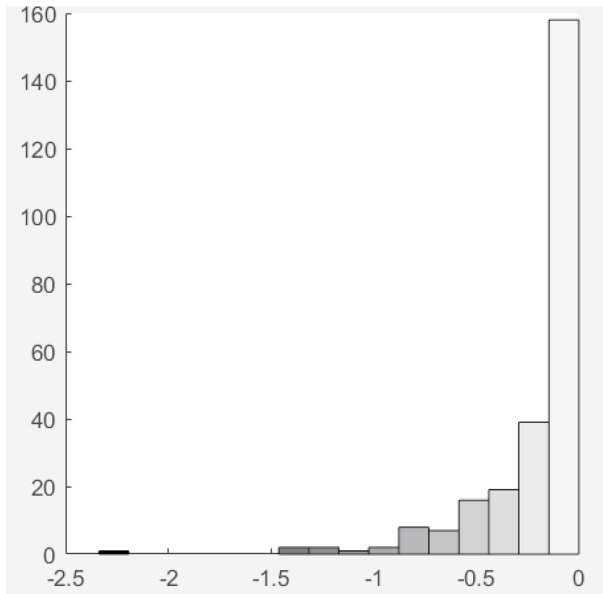
Fig. 4 % employment shortfall across 255 regions

SAR-RE errors assuming predetermined regressors (Table A2). As noted in Sect. 6, the predictive ability of these rivals is not as good as obtained via the preferred estimates summarized in Table 1.

6 Prediction

In order to support the preferred model summarized by Table 1, a cross-validation strategy is employed to assess the performance of competing estimators 'by comparing their predictive ability on data which have not been used in model estimation' (Anselin 1988). Out-of-sample predictions of the level of employment across regions are obtained for the years 2011 and 2012 using 2011 and 2012 data combined with the parameter estimates obtained for data over the estimation period from 2001 to 2010.

Following Chamberlain (1984), Sevestre and Trognon (1996), and Baltagi et al. (2014, 2018), the linear predictor is

$$E\left[\ln \mathbf{e}_t\right] = \mathbf{B}_N^{-1}\left[\mathbf{C}_N E\left[\ln \mathbf{e}_{t-1}\right] + \mathbf{x}_t \boldsymbol{\beta} + c\boldsymbol{\iota} + \mathbf{G}_N E\left[\mathbf{u}_t\right]\right] \tag{20}$$

in which $E[.]$ denotes the expectation, so this can be seen to be identical to Eq. (9) but with expectations. With regard to the estimate of the time-invariant component of the error term $\boldsymbol{\mu}$, assuming a spatial moving average error process gives Eq. (9) rewritten thus
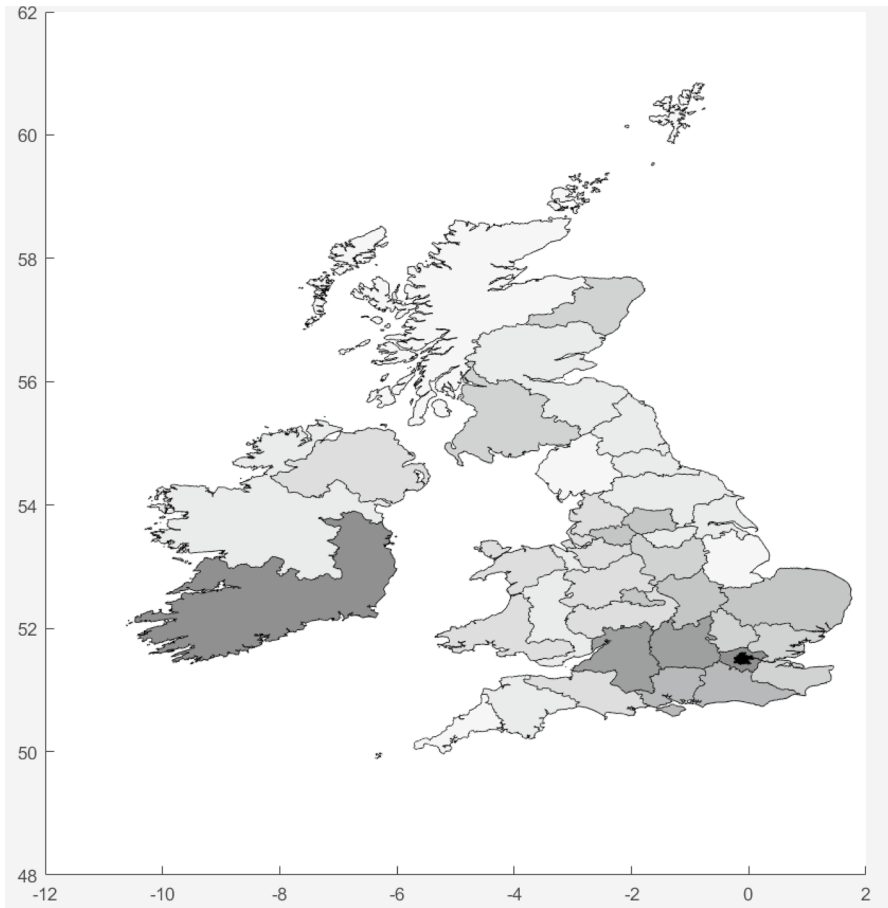
Fig. 5 % employment shortfall UK and Ireland

$$\boldsymbol{\varepsilon}_t = \mathbf{B}_N \ln \mathbf{e}_t - \mathbf{C}_N \ln \mathbf{e}_{t-1} - \mathbf{x}_t \boldsymbol{\beta} - c\boldsymbol{\iota}$$
$$\mathbf{G}_N \mathbf{u}_t = \mathbf{B}_N \ln \mathbf{e}_t - \mathbf{C}_N \ln \mathbf{e}_{t-1} - \mathbf{x}_t \boldsymbol{\beta} - c\boldsymbol{\iota} \tag{21}$$

$$\mathbf{u}_t = \boldsymbol{\mu} + \boldsymbol{v}_t$$
$$\boldsymbol{\mu}^{(t)} = \mathbf{G}_N^{-1} \left(\mathbf{B}_N \ln \mathbf{e}_t - \mathbf{C}_N \ln \mathbf{e}_{t-1} - \mathbf{x}_t \boldsymbol{\beta} - c\boldsymbol{\iota} \right) - \boldsymbol{v}_t \tag{22}$$

In order to obtain estimates $\widehat{\boldsymbol{\mu}}^{(t)}$ estimates $\widehat{\mathbf{G}}_N = \left(\mathbf{I}_N - \widehat{\rho}_2 \mathbf{M}_N \right), \widehat{\mathbf{B}}_N = \left(\mathbf{I}_N - \widehat{\rho}_1 \mathbf{W}_N \right),$ $\widehat{\mathbf{C}}_N = \left(\widehat{\gamma} + \widehat{\theta} \mathbf{W}_N \right)$ and $\widehat{c}$ and $\widehat{\boldsymbol{\beta}}$ are used along with random draws from $\boldsymbol{v}_t \sim N(0, \widehat{\sigma}_v^2)$. We then take the mean over time of the $\widehat{\boldsymbol{\mu}}^{(t)} s$ for $t = 2, \dots, T$, subsequently scaling so that it has variance equal to $\widehat{\sigma}_\mu^2$, thus giving the estimate $\widehat{\boldsymbol{\mu}}$ of the time-invariant error component. The outcome is the prediction Eq. (23) for $T + 1 = 2011$, in which $\mathbf{x}_{1T+1} = \ln \mathbf{q}_{T+1}, \mathbf{x}_{2T+1} = \ln \mathbf{k}_{T+1},$ and $\mathbf{x}_{3T+1} = \ln \bar{\mathbf{e}}_{T+1}, t = 1, \dots, T.$
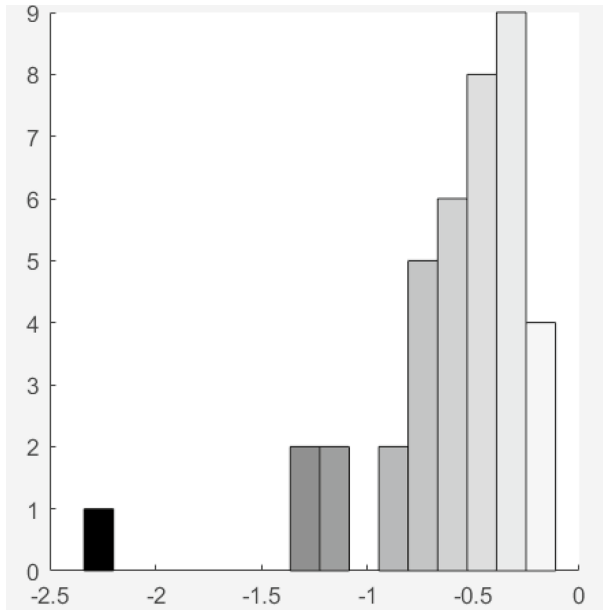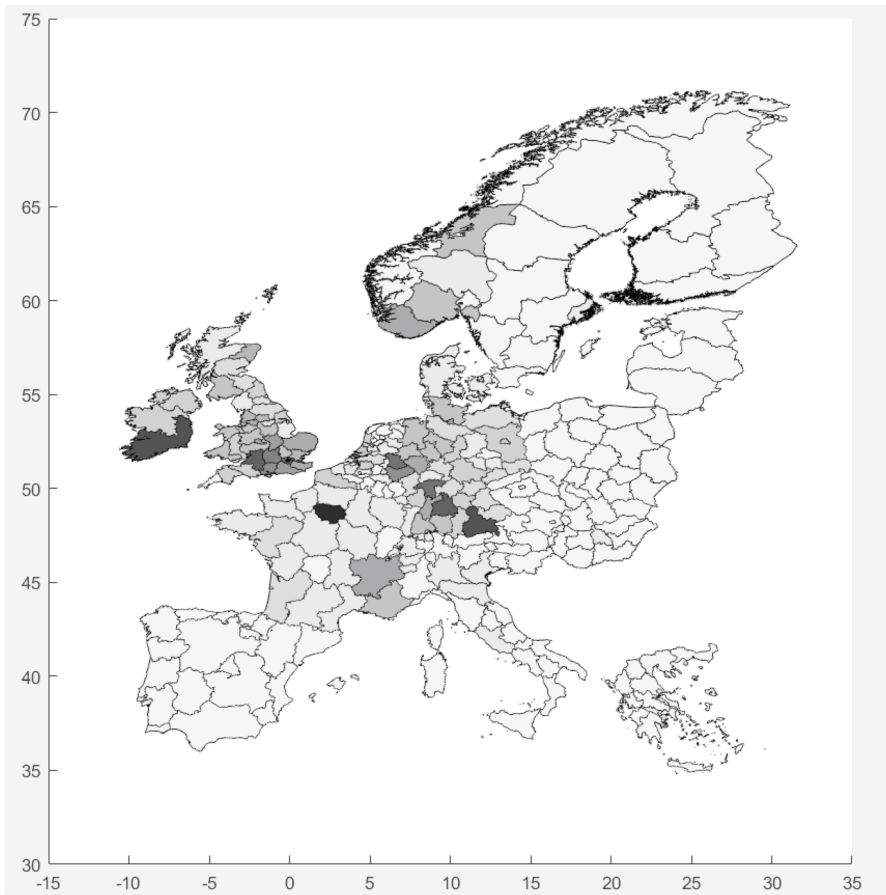
Fig. 6 % employment shortfall

$$\ln \widehat{\mathbf{e}}_{T+1} = \widehat{\mathbf{B}}_N^{-1} \left[\widehat{\mathbf{C}}_N \ln \widehat{\mathbf{e}}_T + \mathbf{x}_{T+1} \widehat{\boldsymbol{\beta}} + \widehat{c}\boldsymbol{\iota} + \widehat{\mathbf{G}}_N \widehat{\boldsymbol{\mu}} \right] \tag{23}$$

For two-step ahead,[11] $\mathbf{x}_{1T+2} = \ln \mathbf{q}_{T+2}$, $\mathbf{x}_{2T+2} = \ln \mathbf{k}_{T+2}$ and $\mathbf{x}_{3T+2} = \ln \overline{\mathbf{e}}_{T+2}$. Figure 1 shows a close correlation between predicted log employment $\ln \widehat{\mathbf{e}}_{T+1}$ and observed log employment, suggesting that the preferred estimator giving Table 1 estimates would be a good basis for simulating the impact on employment following Brexit.

The preference for Table 1 estimates is based on the mean of the RMSE = $\sqrt{\sum_{i=1}^{N} \left(\ln e_{i,T+s} - \ln \widehat{e}_{i,T+s} \right)^2 / N}$ for $s = 1, 2$, denoted by $\overline{\text{RMSE}}$. In the case of Table 1, $\overline{\text{RMSE}} = 0.0781$. Rival estimators ("Appendix" Table 4) give less accurate one- and two-step ahead predictions. In the case of assuming SMA-RE errors and exogenous regressors, $\overline{\text{RMSE}} = 0.1791$. Assuming SAR-RE errors with predetermined regressors gives $\overline{\text{RMSE}} = 0.2890$. Note that in the case of SAR-RE errors, $\widehat{\mathbf{G}}_N = \left(\mathbf{I}_N - \widehat{\rho}_2 \mathbf{M}_N \right)^{-1}$ in Eqs. (20, 21, 23). Table 4 also gives estimates relating to SMA-RE errors and predetermined regressors, but are based on $\mathbf{W}_N$ derived using the Chow–Lin approach. In this case $\overline{\text{RMSE}} = 0.2529$, providing support for the choice of $\mathbf{W}_N$ based on the PBL trade data. Table 4 also gives estimates based

---

[11] Data limitations mean that for 2012, $\mathbf{k}$ in each region is estimated using each region's previous growth rate.

Fig. 7 % employment shortfall due to industry

on SMA-RE errors and predetermined (and exogenous) regressors, but with the additional variables $\mathbf{W}_N\mathbf{x}_1$, the spatial lag of $\ln\mathbf{q}$, and $\mathbf{W}_N\mathbf{x}_2$, the spatial lag of $\ln k$, with $\mathbf{W}_N$ given by the PBL trade data. This is thus a form of spatial Durbin specification, but with regressors $\mathbf{x}_t = (\mathbf{x}_{1t}, \mathbf{x}_{2t}, \mathbf{W}_N\mathbf{x}_{1t}, \mathbf{W}_N\mathbf{x}_{2t})$ the additional covariates evidently cause a problem of weak instruments, giving dynamically unstable nonstationary estimates, as reflected by the largest characteristic root of $\mathbf{B}_N^{-1}\mathbf{C}_N$ equal to 1.0663 (1.9041) and, with $\mathbf{x}$ in Eqs. (22) and (23) $\overline{\mathrm{RMSE}} = 7.4403$ (3.0918). The same spatial Durbin specification again assuming predetermined regressors but with ρ_2 restricted to zero gives a largest characteristic root equal to 1.1127 and $\overline{\mathrm{RMSE}} = 3.3017$. The same spatial Durbin specification assuming exogenous regressors and with a spatial autoregressive (SAR) error process gives a largest characteristic root equal to 2.489 and $\overline{\mathrm{RMSE}} = 20.9333$. These results point to the viability of Table 1 estimates for prediction purposes.
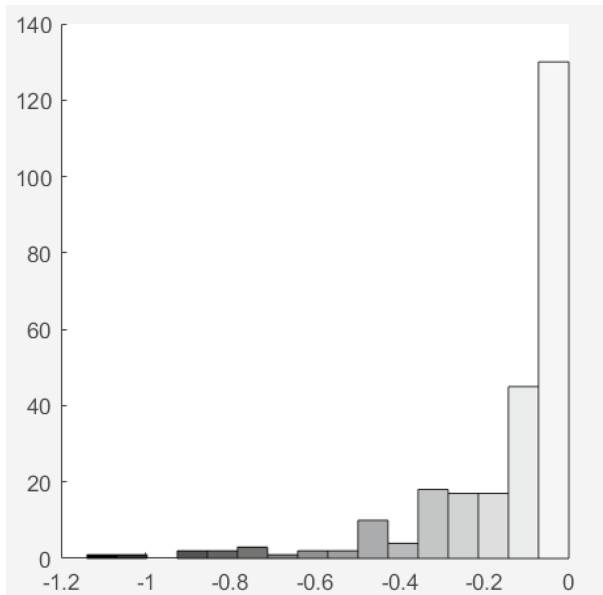
Fig. 8 Frequency distribution from Fig. 7

7 Simulating the Brexit effect

The approach adopted is to use the parameters estimates in Table 1 to predict the impact on employment of presumably reduced trade between the UK and the remaining EU regions in the year 2020 and beyond. Attention is focussed on 2020 and later, given that the UK's formal exit from the EU is scheduled for the first half of 2019, so 2020 will be the first full year outside the EU. Given a lack of appropriate and accessible data, for instance with the same geography as up to 2011, beyond 2011 employment could be predicted on the basis of assumptions about the level of $\mathbf{q}, \mathbf{k}$ and $\bar{\mathbf{e}}$ in 2020.

From $\tau = 2020$ onwards, there are two scenarios, one based on the trade flows assuming no-Brexit effect, and the other assuming a Brexit effect on trade flows, and the difference between them is taken as the Brexit effect. Regarding the no-Brexit effect scenario, this applies matrix $\mathbf{W}_N$, which is based on the latest available trade flows pertaining to the year 2010. The prediction is then given by the solution to Eq. (24) with $\widehat{\mathbf{B}}_N = \left(\mathbf{I}_N - \widehat{\rho}_1 \mathbf{W}_N\right), \widehat{\mathbf{C}}_N = \left(\widehat{\gamma} + \widehat{\theta}\mathbf{W}_N\right)$ and $\widehat{\mathbf{G}}_N = \left(\mathbf{I}_N - \widehat{\rho}_2 \mathbf{M}_N\right)$. Also $\mathbf{x}_\tau$ is an (N by 3) matrix containing the forward projections $\ln \mathbf{q}_\tau, \ln \mathbf{k}_\tau$ and $\ln \bar{\mathbf{e}}$, thus

$$\ln \widehat{\mathbf{e}}_\tau = \widehat{\mathbf{B}}_N^{-1}\left[\widehat{\mathbf{C}}_N \ln \widehat{\mathbf{e}}_{\tau-1} + \mathbf{x}_\tau \widehat{\boldsymbol{\beta}} + \widehat{c}\boldsymbol{\iota} + \widehat{\mathbf{G}}_N \widehat{\boldsymbol{\mu}}\right] \tag{24}$$

The second scenario is to assume that bilateral trade between the UK regions and the (remaining) EU regions is, for example, 2% lower than it would otherwise be. Thus, of the $N = 255$ UK plus EU regions, there are $N^2 - N = 64,770$ bilateral
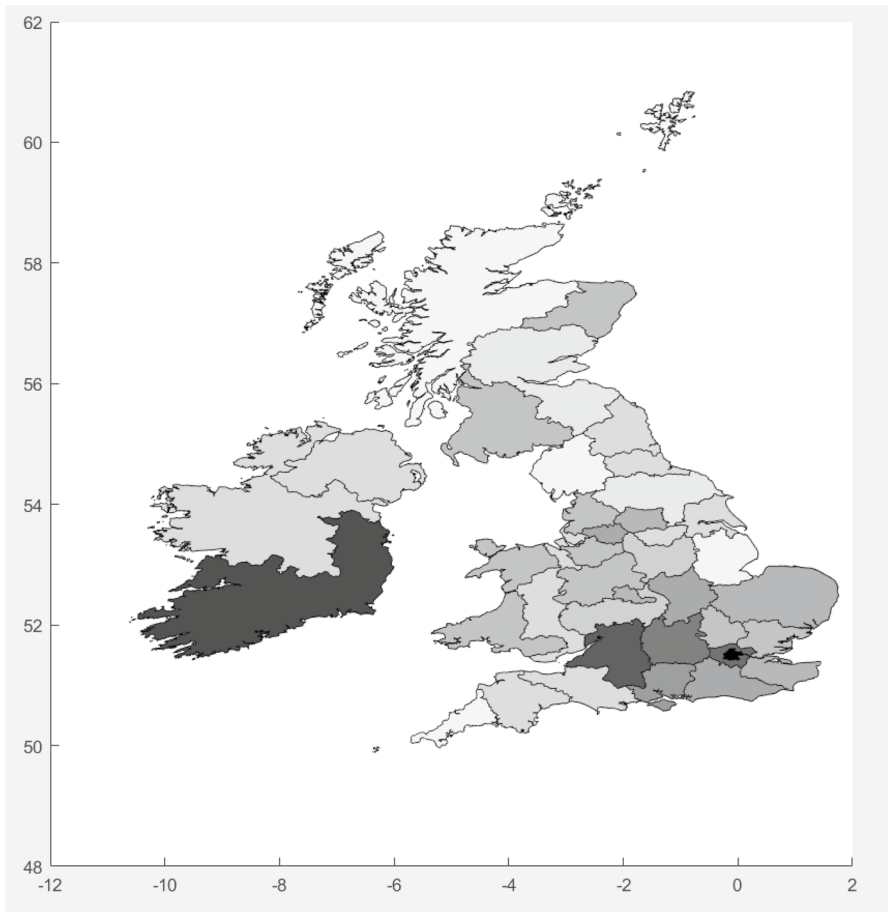
Fig. 9 % employment shortfall UK and Ireland due to industry

trade flows in any one year involving the regions. With 37 UK regions and 218 EU regions $(2 \times 37 \times 218) = 16{,}132$ interregional trade flows are assumed to be 2% smaller than under an assumption of no-Brexit effect. This Brexit-affected trade flow matrix is denoted by $\widetilde{\mathbf{W}}_N$ which leads to $\widetilde{\mathbf{B}}_N = \left(\mathbf{I}_N - \widehat{\rho}_1 \widetilde{\mathbf{W}}_N \right), \widetilde{\mathbf{C}}_N = \left(\widehat{\gamma} + \widehat{\theta} \widetilde{\mathbf{W}}_N \right)$ and the prediction equation

$$\ln \widetilde{\mathbf{e}}_\tau = \widetilde{\mathbf{B}}_N^{-1} \left[\widetilde{\mathbf{C}}_N \ln \widetilde{\mathbf{e}}_{\tau-1} + \mathbf{x}_\tau \widehat{\boldsymbol{\beta}} + \widehat{c}\boldsymbol{\iota} + \widehat{\mathbf{G}}_N \widehat{\boldsymbol{\mu}} \right]$$

Thus, the % job-shortfall at time τ is $\ln \widetilde{\mathbf{e}}_\tau - \ln \widehat{\mathbf{e}}_\tau$.

Using the equilibrium solution of (8), but also taking into account $\mathbf{x}$ at time $\tau = T$ and $\widehat{\mathbf{G}}_N \overline{\boldsymbol{\mu}}$, employment converges to

$$\ln \widehat{\mathbf{e}}_T = \left(\widehat{\mathbf{B}}_N - \widehat{\mathbf{C}}_N \right)^{-1} \left[\mathbf{x}_T \widehat{\boldsymbol{\beta}} + \widehat{c}\boldsymbol{\iota} + \widehat{\mathbf{G}}_N \overline{\boldsymbol{\mu}} \right]$$
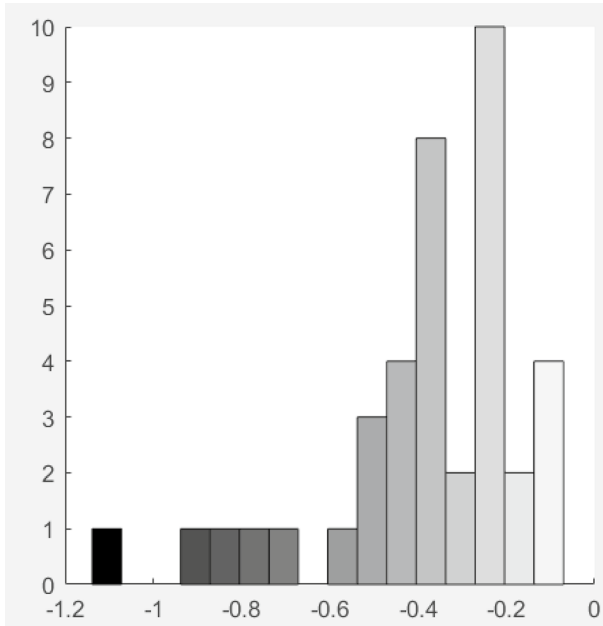
Fig. 10 Frequency distribution from Fig. 9

Similarly

$$\ln \widetilde{\mathbf{e}}_T = \left(\widetilde{\mathbf{B}}_N - \widetilde{\mathbf{C}}_N \right)^{-1} \left[\mathbf{x}_T \widehat{\boldsymbol{\beta}} + \widehat{c} \boldsymbol{\iota} + \widehat{\mathbf{G}}_N \overline{\boldsymbol{\mu}} \right]$$

Thus, the % job-shortfall with long-run convergence at time T is $\ln \widehat{\mathbf{e}}_T - \ln \widetilde{\mathbf{e}}_T$, hence

$$\ln \widehat{\mathbf{e}}_T - \ln \widetilde{\mathbf{e}}_T = \left[\left(\widehat{\mathbf{B}}_N - \widehat{\mathbf{C}}_N \right)^{-1} \right] \left(\mathbf{x}_T \widehat{\boldsymbol{\beta}} + \widehat{c} \boldsymbol{\iota} + \widehat{\mathbf{G}}_N \overline{\boldsymbol{\mu}} \right)$$
$$- \left[\left(\widetilde{\mathbf{B}}_N - \widetilde{\mathbf{C}}_N \right)^{-1} \right] \left(\mathbf{x}_T \widehat{\boldsymbol{\beta}} + \widehat{c} \boldsymbol{\iota} + \widehat{\mathbf{G}}_N \overline{\boldsymbol{\mu}} \right)$$

One assumption might be that $\mathbf{q}, \mathbf{k}$ and $\overline{\mathbf{e}}$ in 2020 are at the same level as observed in each region in 2011. An alternative assumption could be that from 2011 onwards they grow at their historical rates, taken over the period from 1991 to 2011 in each region. On this basis on average the level of $\mathbf{q}$ and $\mathbf{k}$ in 2025 is approximately 25% more than the 2011 levels. However, Table 2 gives simulation outcomes for the examples of Inner London and Paris which illustrate the relative insensitivity of
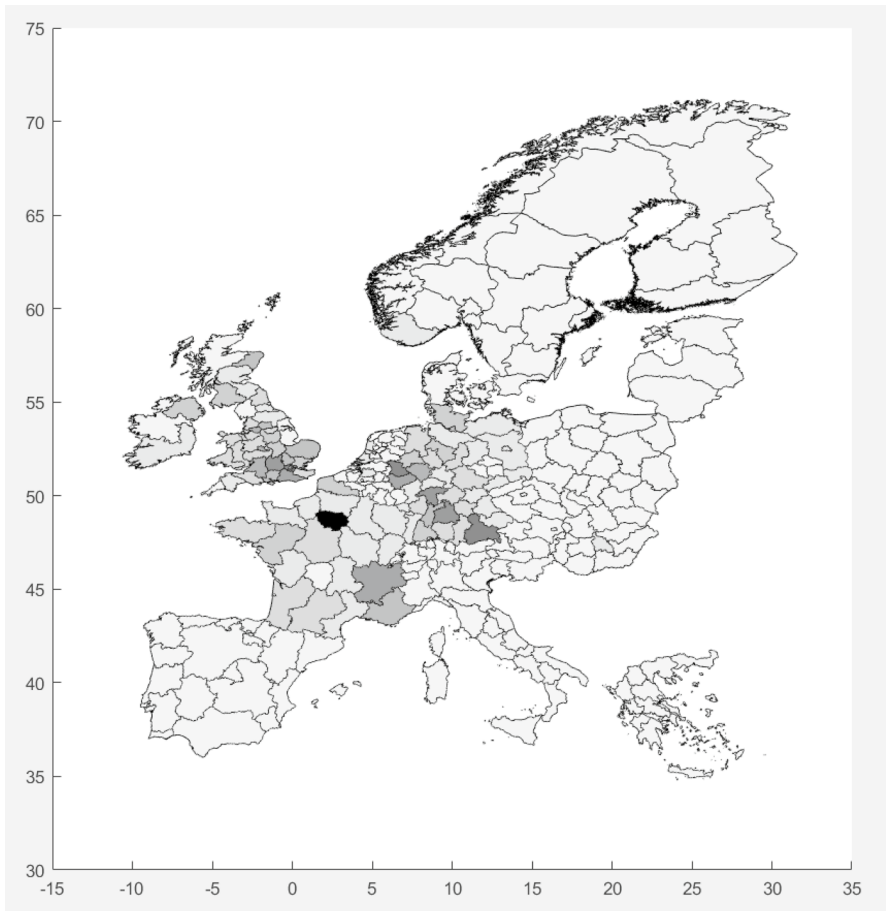
Fig. 11 % employment shortfall due to transport equipment, etc.

job-shortfall to assumed regressor levels. The table shows that doubling the level of **q** and **k** gives equilibrium job-shortfalls that are only about 6% higher, as a result of the same increases applying to both Brexit and no-Brexit outcomes. Table 2 also shows that doubling the trade reduction in each region in effect doubles the job-shortfall. It shows that doubling the % trade reduction, say from 2 to 4%, has the effect of doubling the job-shortfall in each region. Increasing the % reduction by a factor of 8, going from 2 to 16%, increases each region's job-shortfall by a factor of 8. This means that ratio of Inner London to Paris remains stable (which for this pair of regions is approximately 1.98) regardless of what is assumed for % trade reduction. This stability of the outcome ratio exists for any pair of regions so that maps of job-shortfalls would be in a sense identical—identifying the same regions with large or small levels giving constant outcome ratios—irrespective of the assumed % trade reduction. This geographical stability is a result of the assumptions made within the
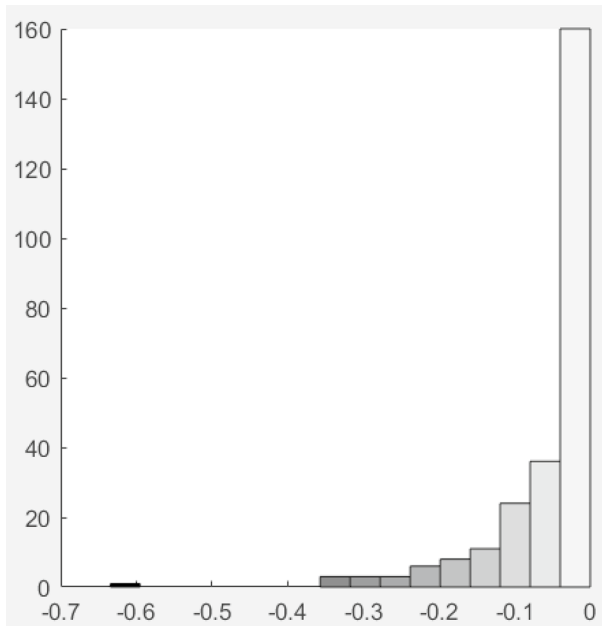
Fig. 12 Frequency distribution from Fig. 11

simulation exercise, with trade for all UK to EU trade flows reduced by the same %. This means that the subsequent map patterns are immune to the assumed reduction in trade, although the scales would differ, where we focus on a trade reduction other than 2%. So in this way, we see we have an element of robustness in our simulations. Of course in an ideal world, one might wish to make changes to trade on an individual region by region and sector by sector basis rather than assume that trade reduces by the same amount across all regions and all sectors. However, this is very much the unknown, although some sectorally specific estimates are given elsewhere.

Sectorally specific Brexit impacts are obtained by assuming that trade in specific sectors alone is restricted. While this is unrealistic, it is likely that there will be sectorally differentiated impacts but it is difficult to know by how much trade in manufactures, for example, will be reduced compared with trade in services.[12] A simple approach is therefore to assume that a specific sector is impacted by Brexit, but that there is zero impact on other sectors. This highlights the geography of the sector-specific trade impacts, because the sectoral trade patterns have different geographies and therefore the impacts have different geographical distributions to the outcomes assuming a global reduction across all sectors (Table 3).

---

[12] Excluding non-market services.
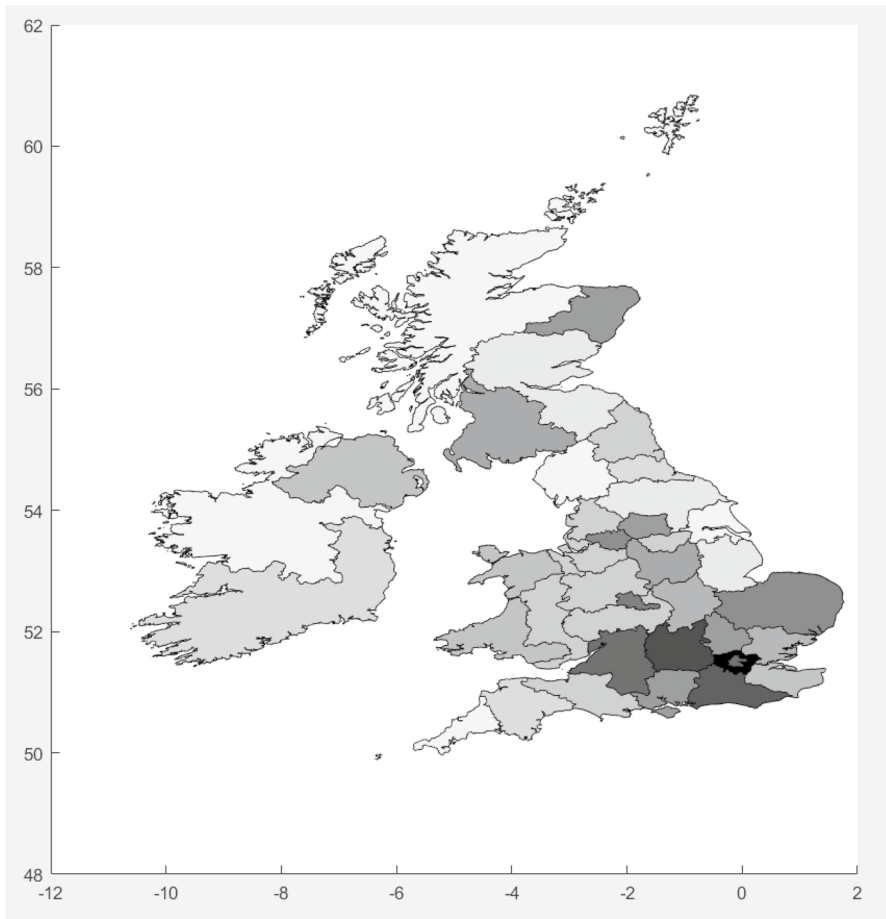
Fig. 13 % shortfall UK and Ireland due to transport equipment, etc.

8 Results

The initial outcomes relate to a reduction in EU-UK trade of 2% across all sectors. The predicted % changes in employment across the EU and UK regions assuming the 2011 levels for $\mathbf{q}$, $\mathbf{k}$ and $\bar{\mathbf{e}}$ are shown by Fig. 2. This shows the dynamic paths for each region to 2050, with convergence to steady state occurring after 2030. From this, it is evident that the maximum equilibrium job-shortfall is -2.56%, in the case of Inner London, with most other regions falling below 1%. Figure 3 shows the geographical pattern of the Brexit impact equal to $\ln \widetilde{\mathbf{e}}_{\tau} - \ln \widehat{\mathbf{e}}_{\tau}$ for $\tau = 2025$, indicating a maximum shortfall by 2025 of -2.34% (Inner London). The picture which emerges from the simulation is that the negative Brexit impact is diverse across regions and bilateral, with both UK regions and EU regions likely to see a job-shortfall. Figure 3 shows larger negative impacts in regions with strong trading links to the UK,
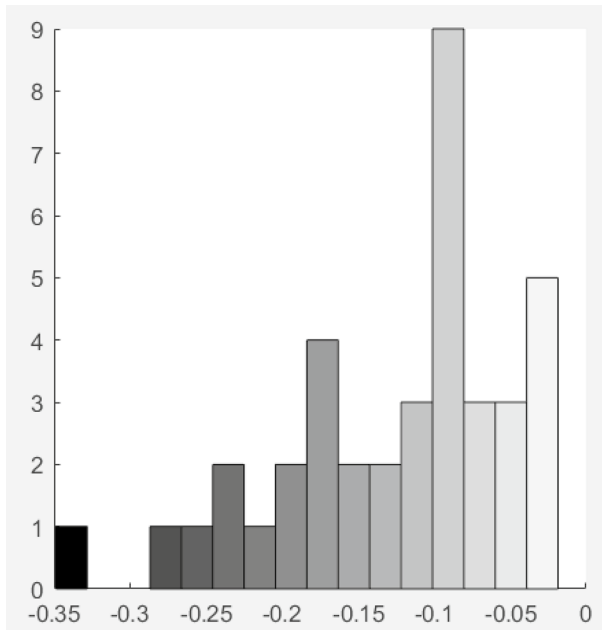
Fig. 14 Frequency distribution from Fig. 13

most notably in the Ile de France (Paris) region (−1.19%), the Southern and Eastern region of Ireland (−1.35%), and the Oberbayern region centred on Munich(− 0.99%). Figure 4 gives the frequency distribution of Fig. 3 data, highlighting the fact that despite some large impacts, for about 160 of the 255 regions, Brexit is likely to have close to zero effect on employment. Figure 5 shows that within the UK, Inner and Outer London (−1.32%) are expected to have the biggest % shortfall by 2025, with impacts generally higher along the Thames valley in Berkshire, Bucks and Oxford-shire (−1.09%) towards Gloucestershire, Wiltshire and North Somerset (−1.19%). Generally, %s are higher around the Greater South East and in some of the large conurbations (Birmingham − 0.78%, Manchester − 0.76%, West Yorkshire − 0.67%) than in more rural and peripheral regions. Figure 6 gives the frequency distribution of Fig. 5 data, emphasizing the Inner London outlier, with many regions having a job-shortfall of less than − 0.5%. As noted above, if one were to assume different reductions in trade other than 2%, the outcomes for employment would be different, but proportional to the 2% impact, so that the ratio of impacts in different regions and the geographical pattern would be identical.

Next, consider the separate impacts on employment of restricted trade in the manufacturing sector, defined as the production of food, beverages and tobacco, tex-tiles and leather, coke, refined petroleum, nuclear fuel and chemicals, electrical and optical equipment and transport equipment and other manufacturing. Simulating on the same basis, region paths converge to equilibrium levels as with Fig. 2, although
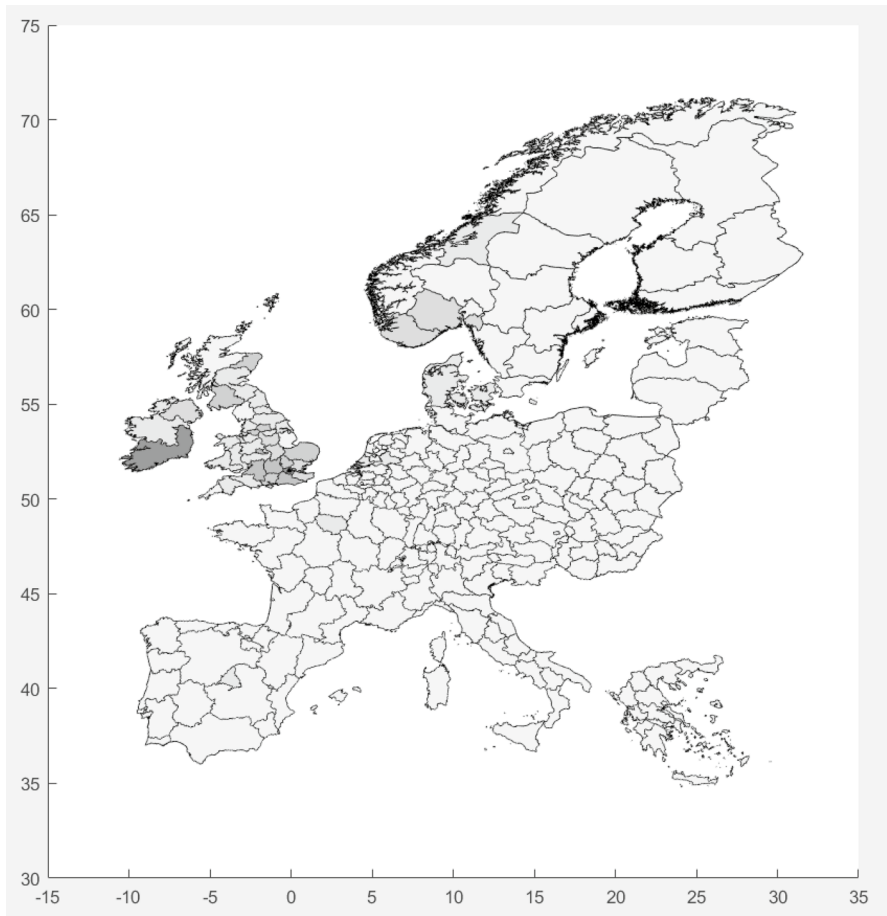
Fig. 15 Services impact: % employment shortfall across 255 EU regions

the equilibrium levels differ from those of Fig. 2. We take a snapshot across the dynamic paths in Figs. 7, 8, 9 and 10, showing the % shortfall in employment in manufactures by region for the year 2025. Figure 7 shows that the geography of the impact due to 2% less industry trade is very similar to the overall pattern shown in Fig. 3. However, a comparison of Figs. 4 and 8 emphasizes the differences in the levels of impact, with the maximum level in Inner London. Figures 9 and 10 show the % job-shortfall in the UK and Ireland. Again the impact in the South and East of Britain and, especially, London is clear, and the effect on Ireland remains pronounced.

We also estimate the impact of reduced trade within the manufacturing sector. Of particular interest is the group of industries defined for trade purposes as comprising 'electrical and optical equipment and transport equipment', which includes
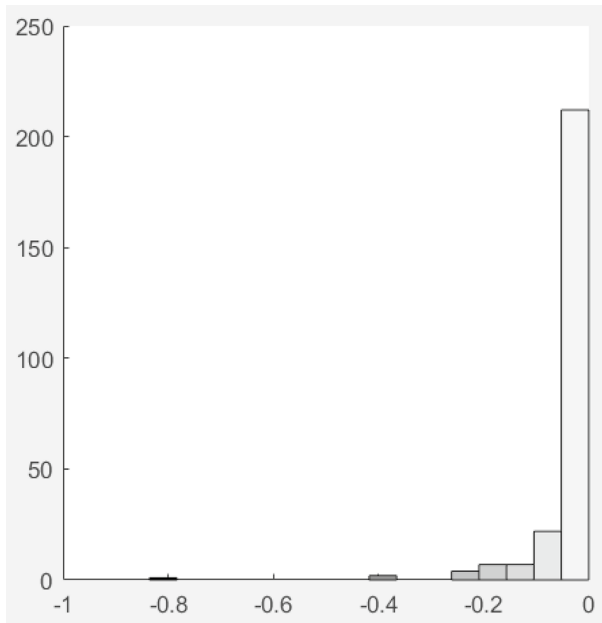
Fig. 16 Frequency distribution from Fig. 15

the all-important production of vehicles. With technological development, we have seen the geographical fragmentation of production processes involved in vehicle manufacture, with the development of spatially dispersed value chains as different elements of the production processes optimally located in different regions or countries. Also just in time processes means that quick and easy access to parts and components used in manufacturing vehicles is important, so interregional connectivity is important, and any disruption of it due to increased barriers to trade will have a significant impact. The impacts of a 2% trade reduction are summarized in Figs. 11 and 12, which picks out some of the production hot spots. In Figs. 13 and 14, the integrated nature of production and dispersed knock-on effects of reduced trade are evident from the relatively even spread of impacts across regions.

Compared with industry, the impact of reduced trade in services[13] is much less symmetrical, with the bulk of the job-shortfall occurring in Britain and Ireland. This is clear from Figs. 15 and 16. Inner London clearly stands out with the most significant projected job-shortfall compared with all other EU regions, and apart from the South and Eastern region of Ireland, almost all nonzero job-shortfalls occur in Great Britain. Figures 17 and 18 emphasize the polarized effect of service trade reduction, with Inner London standing out as an outlier. Outer London and Southern and Eastern Ireland see comparable effects, ahead of all the other UK regions. Focussing

---

[13] Distribution, Hotels and restaurants, Transport storage and communication, Financial intermediation and Real estate renting and business activities
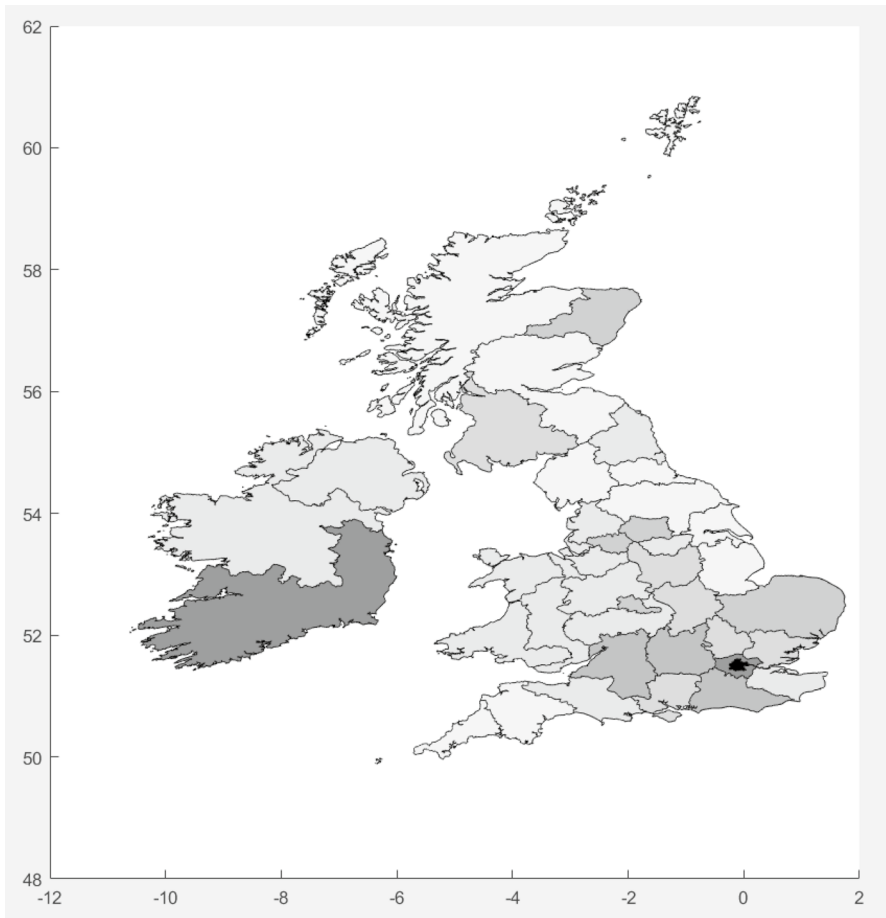
Fig. 17 % employment shortfall UK and Ireland due to services

on the highly important Financial Intermediation sector, Figs. 19 and 20 emphasize even more strongly the asymmetric impact on Brexit, with more than 200 non-UK–EU regions having almost zero job-shortfall. The most affected non-UK–EU region is the South and East of Ireland, followed by Luxembourg. Figures 21 and 22 illustrate the role of Inner London in particular as a centre for financial intermediation, but overall for the UK increased trade barriers for this element of services seems to have a less profound impact on the job-shortfall than the more geographically widespread and deeper impact of reduced trade in transport equipment.
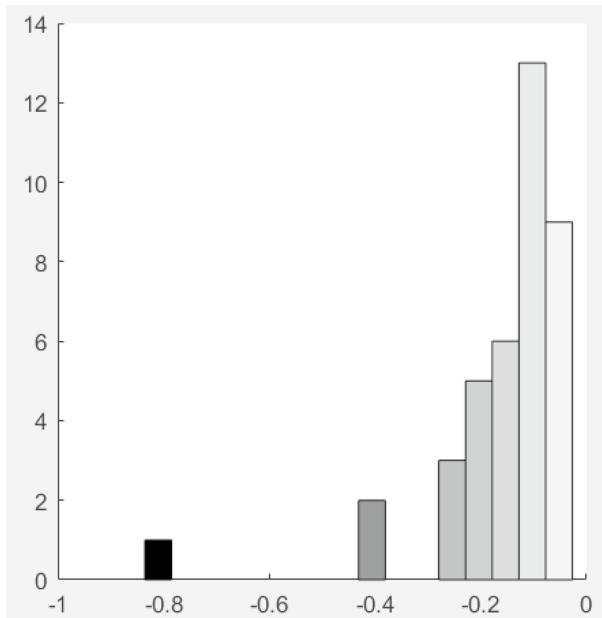
Fig. 18 Frequency distribution from Fig. 17

8.1 Conclusion

The paper shows negative Brexit-induced impacts on employment which affect not only the UK regions but also employment levels in EU regions, especially those which are close trading partners. This pan-European interregional interdependency is captured in the state-of-the-art model by spatial and temporal interactions based on the best available trade flow estimates which determine the strength of interdependence. This means that employment within a region not only depends on the levels of output and capital within the region, but also on demand coming from other regions which are trading partners. The impacts will be multi-way, what happens to employment in London depends partly on what happens in Paris, which depends on what happens in Munich, which depends on what happens in London, etc. The approach adopted has been to assume a reduction in trade between EU and UK regions which gives a corresponding reduction in demand for jobs. The predicted job-shortfalls depend on the assumed global % reduction in trade between UK and EU regions, but the modelling assumptions ensured stability in the geography of Brexit impact.

In the paper, the impact of Brexit is measured in terms of the 2025 job-shortfall, which is the reduction in the number of jobs in each region due to Brexit assuming no alternative sources of employment are put in place. This of course might be a false assumption, as the pro-Brexit lobby has consistently emphasized the potential stimulus of new trade deals with other non-EU countries. Therefore, the Brexit impact as reflected in the maps of job-shortfall indicates those regions which
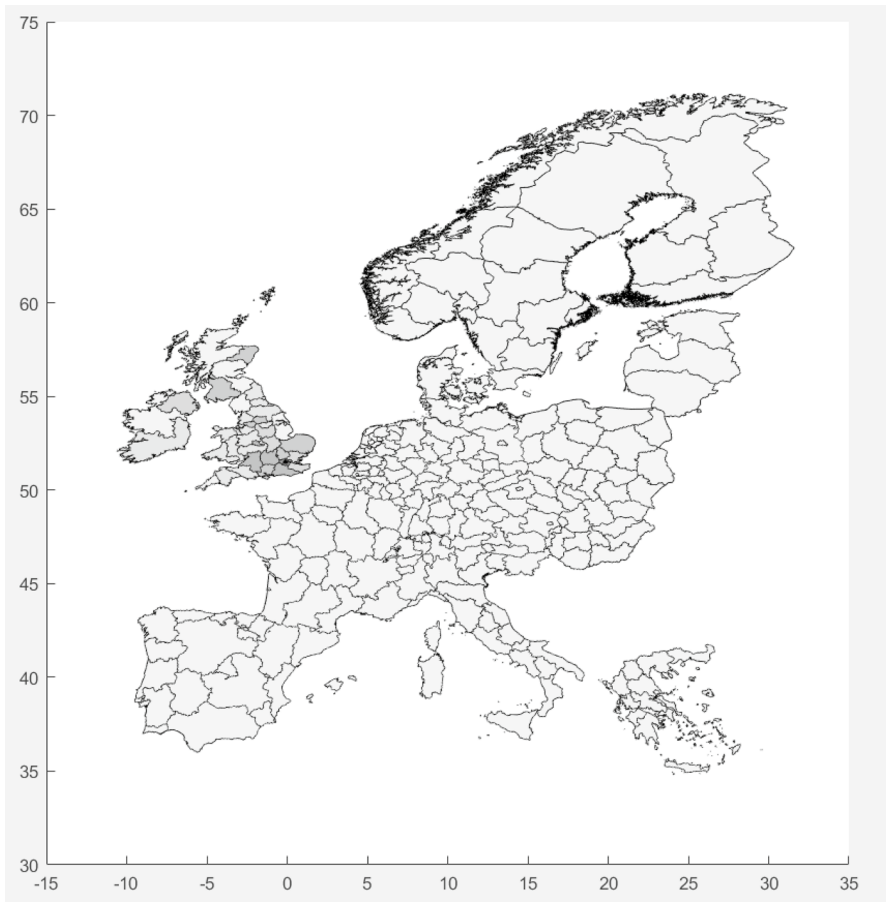
Fig. 19 % employment shortfall due to financial intermediation

could be in the greatest need of alternative compensating sources of employment. Thus, the paper is not predicting a job-loss per se, simply a potential job-loss without successful alternative trade arrangements post-Brexit. Additional employment due to trade diversion effects due to higher UK–EU trade barriers (Ortiz Valverde and Latorre 2018, Dhingra et al. 2017a, Krueger 1999) could possibly be captured within the current modelling set-up via changes to the levels of output and capital in each region, but these would be difficult to estimate and there is some empirical evidence that they might be quite small (Krueger 1999; Magee 2008).

Key outcomes are as follows. First, both UK and EU regions are negatively affected, this is a lose–lose scenario. Second, the deepest most concentrated impact is for the UK, many EU regions are barely affected, but the South and East regions of Ireland are the worst affected EU region, with impacts on a par with the worst affected UK regions. In addition the Ile de France, Oberbayern, Stuggart and
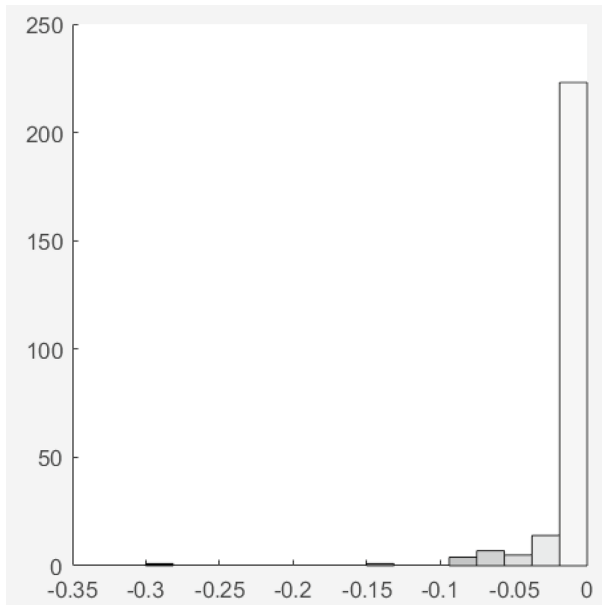
Fig. 20 Frequency distribution from Fig. 19

Dusseldorf stand out as regions likely to see significant job-shortfalls. Third, in the UK, Southern regions, especially in and around London, and big cities, are expected to see the largest job-shortfall. Fourth, the effect of manufacturing trade reduction will be evidently larger than for services, but the service impact is more asymmetric, with the bulk of the job-shortfall focussed on UK regions, especially London. This is even more the case when comparing the effect of trade reductions in vehicles and financial intermediation.

Overall, the simulations suggest that the biggest Brexit impact on UK regions will occur in the richer South East and urban areas, which is in line with work from LSE based on GVA, which shows that 'areas in the South of England, and urban areas, are harder hit by Brexit...the areas that were most likely to vote remain are those that are predicted to be most negatively impacted by Brexit' (Dhingra et al. 2017b). This interpretation is in direct contrast to other work which maintains that 'the regions which voted Leave also tended to be more dependent on Europe for their prosperity than the regions which voted Remain' (Los et al. 2017).

Clearly, Brexit is a complex phenomenon leading to diverse interpretations of outcomes, as evident in the special issue of Papers in Regional Science (McCann 2018). The outcomes presented in this paper are based on model assumptions, but it is argued that the main driver of the results is the data, not imposed assumptions. Nevertheless, great caution is needed in interpreting the validity and value of any 'prediction' effort. It is worth recalling the words of Box and Draper (1986), 'Essentially, all models are wrong but some are useful'. David Spiegelhalter, Professor of the Public Understanding of Risk at the University of Cambridge, refers to Donald Rumsfeld as the patron saint of Risk Analysis, who will be remembered for famously
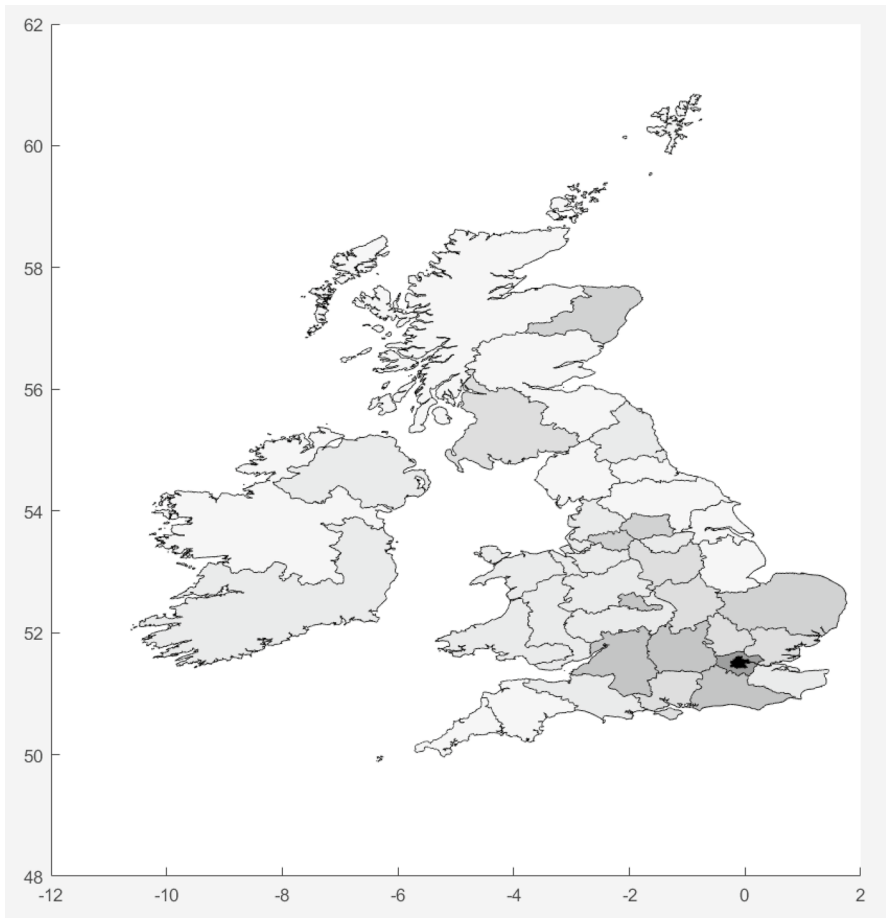
Fig. 21 % employment shortfall UK and Ireland due to financial intermediation

saying that 'but there are also unknown unknowns. There are things we do not know we don't know'. We should therefore put forward predictions with all due humility, but clearly and without fear, because we don't want to come across as 'dithering scientists'. In defence of the approach adopted, there is support from the words of Pesaran (1990), who points out that 'Econometric models are important tools for forecasting and policy analysis, and it is unlikely that they will be discarded in the future. The challenge is to recognize their limitations and to work towards turning them into more reliable and effective tools. There seem to be no viable alternatives'.
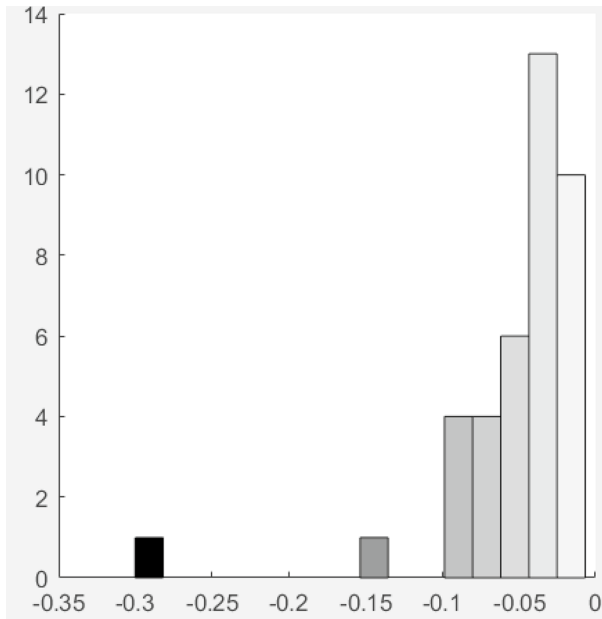
Fig. 22 Frequency distribution from Fig. 21

Appendices

See Table 4.

Table 4 Estimates of Eq. 9

Parameters	A1	A2	A3	A4	A5
γ	0.5634 (0.001425) (395.2)	0.6908 (0.003387) (204.0)	0.7102 (0.005737) (123.8)	0.3357 (0.00209) (160.6)	0.5673 (0.002129) (266.4)
ρ_1	0.5909 (0.006908) (85.54)	0.7355 (0.02388) (30.8)	0.6539 (0.01081) (60.48)	1.516 (0.01156) (131.2)	1.1990 (0.01358) (88.3)
β_1	0.1787 (0.0006129) (291.5)	0.1493 (0.001614) (92.46)	0.1577 (0.002976) (52.98)	0.3604 (0.001307) (275.6)	0.1897 (0.001245) (152.4)
β_2	0.01966 (0.0001589) (123.7)	0.002392 (0.0008091) (2.956)	0.01858 (0.001353) (13.73)	-0.03702 (0.0007463) (-49.61)	0.01581 (0.0002156) (73.34)
$\widetilde{\beta}_1$	…	…	…	0.2910 (0.005024) (57.91)	0.3576 (0.006044) (59.18)
$\widetilde{\beta}_2$	…	…	…	-0.3930 (0.003494) (-112.5)	-0.3025 (0.002825) (-107.1)
θ	-0.3807 (0.005913) (-64.39)	-0.6739 (0.02267) (-29.73)	-0.7551 (0.01528) (-49.42)	-0.8863 (0.008452) (-104.9)	-0.9465 (0.009802) (-96.56)
ρ_2	-0.6545	0.7119	-0.9456	-0.3079	-0.5852
σ_μ^2	0.2786	0.0539	0.1851	0.0097	0.4663
σ_ν^2	0.0003	0.0692	0.0005	0.0006	0.0003
Max char. root	0.5845	0.7595	0.8665	1.0663	1.9041
Forecasting RMSE					
2011	0.1413	0.2248	0.2343	7.4094	3.0746
2012	0.2168	0.3532	0.2715	7.4711	3.1091
Average	0.1791	0.2890	0.2529	7.4403	3.0918

A1: assuming exogenous regressors, SMA-RE errors

A2: assuming predetermined regressors, SAR-RE errors

A3: assuming predetermined regressors, SMA-RE errors, Chow–Lin $\mathbf{W}_N$

A4: Spatial Durbin, predetermined regressors, SMA-RE errors

A5: Spatial Durbin, exogenous regressors, SMA-RE errors

Theoretical background

The theoretical background to the model specification commencing with Eq. (1) is derived from standard urban economics as given by Abdel-Rahman and Fujita (1990), Ciccone and Hall (1996) and Fujita and Thisse (2002), among others. The theory commences with a constant returns to scale Cobb–Douglas production function for the output $\widetilde{q}$ of the competitive final goods and services sector

$$\widetilde{q} = \left(m^{\widetilde{\beta}} i^{1-\widetilde{\beta}} \right)^\alpha h^{1-\alpha}$$

in which m denotes sector-specific labour efficiency units and the level of composite services is given by i, and h is area of land. Assume $h = 1$ then

$$\widetilde{q} = \left(m^{\widetilde{\beta}} i^{1-\widetilde{\beta}} \right)^\alpha \tag{25}$$

Assume that the equilibrium output of each service firm is i_e and there are g firms, depending on the total services effective labour. We obtain g by dividing the total services effective labour by the effective labour per firm, thus

$$g = \frac{(1 - \widetilde{\beta})\widetilde{e}}{ai_e + s} \tag{26}$$

In (26), $(1 - \widetilde{\beta})$ equals the services labour share of total effective labour $\widetilde{e}$ under competitive equilibrium in the labour market. For each services firm, we have internal returns to scale, with s denoting the fixed labour requirement and a the marginal labour requirement, so that $ai_e + s$ is the effective labour per firm at equilibrium. From the CES production function, we obtain

$$i = g^{\widetilde{\mu}} i_e \tag{27}$$

where $\widetilde{\mu}$ is a measure of monopoly power in the monopolistically competitive services sector, and $\frac{\widetilde{\mu}}{\widetilde{\mu}-1}$ is the constant elasticity of substitution. Substituting i into equation (25) gives

$$\widetilde{q} = \left(m^{\widetilde{\beta}} g^{\widetilde{\mu} - \widetilde{\mu}\widetilde{\beta}} i_e^{1-\widetilde{\beta}} \right)^{\alpha} \tag{28}$$

Substituting for g in Eq. (28) gives

$$\widetilde{q} = \widetilde{e}^{\alpha\left(\widetilde{\beta} + \widetilde{\mu} - \widetilde{\mu}\widetilde{\beta}\right)} \widetilde{\beta}^{\widetilde{\alpha}\widetilde{\beta}} \left(ai_e + s \right)^{\alpha\widetilde{\mu}\left(\widetilde{\beta} - 1\right)} i_e^{\alpha\left(1-\widetilde{\beta}\right)} \left(1 - \widetilde{\beta} \right)^{-\widetilde{\alpha}\widetilde{\mu}\left(\widetilde{\beta}-1\right)} \tag{29}$$

which simplifies to

$$\widetilde{q} = \phi\widetilde{e}^{\alpha\left(1+\left(1-\widetilde{\beta}\right)\left(\widetilde{\mu}-1\right)\right)} = \phi\widetilde{e}^{\widetilde{\gamma}} \tag{30}$$

This shows that with $\alpha = 1$ there are increasing returns $(\widetilde{\gamma} > 1)$ if service firms are relevant to output in the competitive sector $(\widetilde{\beta} < 1)$ and also possess monopoly power $(\widetilde{\mu} > 1)$. However, $\alpha < 1$ indicates that land is also a relevant factor, and depending on the value of α a tendency to increasing returns could be offset by the congestion effect caused by the restriction of production to a unit of land, leading to $\widetilde{\gamma} < 1$ hence diminishing returns.

Other estimates

The Chow–Lin approach

This commences with aggregate trade values between $\widetilde{N} = 21$ EU counties (denoted by the square $(\widetilde{N} \times \widetilde{N})$ matrix $\mathbf{T}_{\widetilde{N}}$, in which the subscript $\widetilde{N}$ means national). There are $\widetilde{N} = 21$ unobserved intra-country trade flows, thus giving $p = 420$ observations for the year 2000. The $\widetilde{N}^2 = 441$ 'observations' are the dependent variable in a Weighted Least Squares regression model. All observed trade flows are given the

weight 1, and those unobserved weighted zero. In terms of parameter estimates, an entirely equivalent procedure is to estimate the 420 observed trade flows by OLS. In the regression, the explanatory regressors are great circle distances between country pairs ($\mathbf{G}_{\widetilde{N}}$), and the product of each pair of country's national GVA level ($\widetilde{\mathbf{q}}_{\widetilde{N}}$) in the year 2000, so that given the ($\widetilde{N} \times 1$) vector $\widetilde{\mathbf{q}}_{\widetilde{N}}$, $\mathbf{Q}_{\widetilde{N}} = \widetilde{\mathbf{q}}_{\widetilde{N}} \widetilde{\mathbf{q}}_{\widetilde{N}}^\top$ is an ($\widetilde{N} \times \widetilde{N}$) matrix. Subsequently, $\mathbf{Q}_{\widetilde{N}}, \mathbf{G}_{\widetilde{N}}$ and $\mathbf{T}_{\widetilde{N}}$ are reshaped as ($\widetilde{N}^2 \times 1$) vectors $\mathbf{q}_{\widetilde{N}}, \mathbf{g}_{\widetilde{N}}$ and $\mathbf{t}_{\widetilde{N}}$, and together with $\mathbf{c}_{\widetilde{N}}$ which is an ($\widetilde{N}^2 \times 1$) vector of ones, these variables provide the input for the regression denoted by Eq. (31), thus giving the estimates $\widehat{\boldsymbol{\beta}}_{\widetilde{N}}$. Interregional trade estimates are based on the (national-level) regression parameter estimates $\widehat{\boldsymbol{\beta}}_{\widetilde{N}}$ and on the estimated regression residuals $\widehat{\mathbf{e}}_{\widetilde{N}}$. Thus, trade flows between regions, denoted by $\mathbf{t}_{\widetilde{R}}$, are obtained by applying the national-level estimates $\widehat{\boldsymbol{\beta}}_{\widetilde{N}}$ to the regressors measured at the regional level, denoted by $\mathbf{g}_{\widetilde{R}}$ and $\mathbf{q}_{\widetilde{R}}$. Also, an equal share of the national-level residuals $\widehat{\mathbf{e}}_{\widetilde{N}}$ is added to regions corresponding to country pairs. This process is summarized by the two equations:

$$\ln \mathbf{t}_{\widetilde{N}} = \beta_{\widetilde{N},1} \mathbf{c}_{\widetilde{N}} + \beta_{\widetilde{N},2} \ln \mathbf{g}_{\widetilde{N}} + \beta_{\widetilde{N},3} \ln \mathbf{q}_{\widetilde{N}} + \mathbf{e}_{\widetilde{N}} \tag{31}$$

with log bilateral interregional trade flows $\ln \mathbf{t}_{\widetilde{R}}$ then being obtained using

$$\ln \mathbf{t}_{\widetilde{R}} = \beta_{\widetilde{N},1} \mathbf{c}_{\widetilde{R}} + \beta_{\widetilde{N},2} \ln \mathbf{g}_{\widetilde{R}} + \beta_{\widetilde{N},3} \ln \mathbf{q}_{\widetilde{R}} + \widehat{\mathbf{e}}_{\widetilde{R}}. \tag{32}$$

To obtain the ($\widetilde{R}^2 \times 1$) vector $\widehat{\mathbf{e}}_{\widetilde{R}}$, we calculate the ($\widetilde{R} \times \widetilde{R}$) matrix $\widehat{\mathbf{V}}_{\widetilde{R}} = \mathbf{h} \widehat{\mathbf{R}}_{\widetilde{N}} \mathbf{h}^\top$ where $\mathbf{h} = \mathbf{d}^\top (\mathbf{d} \mathbf{d}^\top)^{-1}$ and $\mathbf{d}$ is an ($\widetilde{N} \times \widetilde{R}$) indicator matrix with ones in each country's row indicating which regions are within that country, and zeros indicating those which are not. The ($\widetilde{N} \times \widetilde{N}$) matrix of residuals $\widehat{\mathbf{R}}_{\widetilde{N}}$ is formed by reshaping the ($\widetilde{N}^2 \times 1$) vector $\widehat{\mathbf{e}}_{\widetilde{N}}$ so that $\widehat{\mathbf{R}}_{\widetilde{N}}(1 : \widetilde{N}, 1) = \widehat{\mathbf{e}}_{\widetilde{N}}(1 : \widetilde{N}), \widehat{\mathbf{R}}_{\widetilde{N}}(1 : \widetilde{N}, 2) = \widehat{\mathbf{e}}_{\widetilde{N}}(\widetilde{N} + 1 : 2\widetilde{N}), \dots,$ $\widehat{\mathbf{R}}_{\widetilde{N}}(1 : \widetilde{N}, \widetilde{N}) = \widehat{\mathbf{e}}_{\widetilde{N}}(\widetilde{N}^2 - \widetilde{N} + 1 : \widetilde{N}^2)$.

The outcome $\widehat{\mathbf{V}}_{\widetilde{R}}$ is an ($\widetilde{R} \times \widetilde{R}$) matrix containing the inter-country residuals $\widehat{\mathbf{e}}_{\widetilde{N}}$ allocated equally to all regions corresponding to country pairs. Also, there are block diagonal zeros as a result of the unobserved intra-country trade flows, and zeros across four rows representing four regions equal to the countries Estonia, Lithuania, Latvia and Luxembourg which were excluded from the initial regression. Finally, $\widehat{\mathbf{V}}_{\widetilde{R}}$ is reshaped as the ($\widetilde{R}^2 \times 1$) vector $\widehat{\mathbf{e}}_{\widetilde{R}}$ of Eq. (32) to give $\ln \mathbf{t}_{\widetilde{R}}$. The resulting ($\widetilde{R} \times \widetilde{R}$) matrix of interregional trade flows is $\mathbf{W}_N^* = \mathbf{T}_{\widetilde{R}}$.

References

Abdel-Rahman H, Fujita M (1990) Product variety, Marshallian externalities, and city sizes. J Reg Sci 30(2):165–183

Abeysinghe T, Lee C (1998) Best linear unbiased disaggregation of annual GDP to quarterly figures: the case of Malaysia. J Forecast 17:527–537

Anselin L (1988) Spatial econometrics: methods and models. Kluwer Academic Publishers, Dordrecht

Arellano M, Bond S (1991) Some tests of specification for panel data: Monte Carlo evidence and an application to employment. Rev Econ Stud 58:277–297

Bailey N, Holly S, Pesaran MH (2016) A two-stage approach to spatio-temporal analysis with strong and weak cross-sectional dependence. J Appl Econom 31:249–280

Baltagi BH (2013) Econometric analysis of panel data, 5th edn. Wiley, New York

Baltagi BH, Fingleton B, Pirotte A (2014) Estimating and forecasting with a dynamic spatial panel model. Oxf Bull Econ Stat 76:112–138

Baltagi B, Fingleton B, Pirotte A (2018) A time-space dynamic panel data model with spatial moving average errors. Reg Sci Urban Econ. https://doi.org/10.1016/j.regsciurbeco.2018.04.013

Bond S (2002) Dynamic panel data models: a guide to micro data methods and practice. Port Econ J 1:141–162

Bowsher CG (2002) On testing overidentifying restrictions in dynamic panel data models. Econ Lett 77:211–220

Box GEP, Draper NR (1986) Empirical model-building and response surfaces. Wiley, Chichester

Bouayad-Agha S, Védrine L (2010) Estimation strategies for a spatial dynamic panel using GMM. A new approach to the convergence issue of European regions. Spat Econ Anal 5:205–227

Chamberlain G (1984) Panel data. In: Griliches Z, Intriligator M (eds) The handbook of econometrics, chapter 22. North-Holland, Amsterdam, pp 1247–1318

Chow G, Lin A-I (1971) Best linear unbiased interpolation, distribution and extrapolation of time series by related series. Rev Econ Stat 53:372–375

Ciccone A, Hall RE (1996) Productivity and the density of economic activity. Am Econ Rev 86(1):54

Debarsy N, Ertur C, LeSage JP (2012) Interpreting dynamic space-time panel data models. Stat Methodol 9:158–171

Derbyshire J, Gardiner B, Waights S (2010) Estimating the capital stock for the NUTS 2 regions of the EU-27. European Union working papers no. 01/2011

Dhingra S, Huang H, Ottaviano G, Paulo Pessoa J, Sampson T, Van Reenen J (2017a) The costs and benefits of leaving the EU: trade effects. Econ Policy 32(92):651–705

Dhingra S, Machin S, Overman H (2017b) Local economic effects of Brexit. Natl Inst Econ Rev 242:R24–R36

Doran J, Fingleton B (2014) Economic shocks and growth: spatio-temporal perspectives on Europe's economies in a time of crisis. Pap Reg Sci 93(S1):S137–S165

Elhorst JP (2001) Dynamic models in space and time. Geogr Anal 33:119–140

Elhorst JP (2014) Spatial econometrics: from cross-sectional data to spatial panels. Springer, Berlin

Ertur C, Musolesi A (2017) Weak and strong cross-sectional dependence: a panel data analysis of international technology diffusion. J Appl Econom 32:477–503

Feenstra RC, Lipsey RE, Deng H, Ma AC, Mo H (2005) World trade flows: 1962–2000. NBER Working paper series, working paper 11040. http://www.nber.org/papers/w11040

Fingleton B (2008) A generalized method of moments estimator for a spatial panel model with an endogenous spatial lag and spatial moving average errors. Spat Econ Anal 3:28–44

Fingleton B, Szumilo N (2019) Simulating the impact of transport infrastructure investment on wages: a dynamic spatial panel model approach. Reg Sci Urban Econ. https://doi.org/10.1016/j.regsciurbeco.2018.12.004

Fingleton B, Garretsen H, Martin RL (2015) Shocking aspects of monetary union: the vulnerability of regions in Euroland. J Econ Geogr 15:907–934

Fingleton B, Le Gallo J, Pirotte A (2017) A multi-dimensional spatial lag panel data model with spatial moving average nested random effects errors. Empir Econ. https://doi.org/10.1007/s00181-017-1410-7

Fingleton B, Fuerst F, Szumilo N (2018) Housing affordability: is new local supply the key?. Environ Plan A Econ Space. https://doi.org/10.1177/0308518X18798372

Fujita M, Thisse JF (2002) Economics of agglomeration: cities, industrial location, and regional growth. Cambridge University Press, Cambridge

Gianelle C, Goenaga X, González I, Thissen M (2014) Smart specialisation in the tangled web of European inter-regional trade. S3 working paper series no. 05/2014, JRC-IPTS, Sevilla

Kelejian HH, Prucha IR (1999) A generalized moments estimator for the autoregressive parameter in a spatial model. Int Econ Rev 40:509–533

Krueger AO (1999) Trade creation and trade diversion under NAFTA (no. w7429). National Bureau of Economic Research

LeSage JP, Pace RK (2009) Introduction to spatial econometrics. Chapman & Hall, Boca Raton

Los B, McCann P, Springford J, Thissen M (2017) The mismatch between local voting and the local economic consequences of Brexit. Reg Stud 51:786–799

Magee CS (2008) New measures of trade creation and trade diversion. J Int Econ 75(2):349–362

McCann P (2018) The trade, geography and regional implications of Brexit. Pap Reg Sci 97(1):3–8

Ortiz Valverde G, Latorre MC (2018) A computable general equilibrium analysis of Brexit: barriers to trade and immigration restrictions. https://doi.org/10.2139/ssrn.3282231

Pace RK, LeSage JP, Zhu S (2012) Spatial dependence in regressors. In: Terrell D, Millimet D (eds) Advances in econometrics, Thomas B. Fomby, R. Carter Hill, Ivan Jeliazkov, Juan Carlos Escanciano and Eric Hillebrand, vol 30. Emerald Group Publishing Limited, Bingley, pp 257–295

Parent O, LeSage JP (2011) A space-time filter for panel data models containing random effects. Comput Stat Data Anal 55:475–490

Parent O, LeSage JP (2012) Spatial dynamic panel data models with random effects. Reg Sci Urban Econ 42(4):727–738

Pesaran MH (1990) Econometrics. In: Eatwell J, Milgate M, Newman P (eds) The new Palgrave: econometrics. W.W. Norton and Company, New York, pp 25–26

Pesaran MH (2015) Time series and panel data econometrics. Oxford University Press, Oxford

Polasek W, Verduras C, Sellner R (2010) Bayesian methods for completing data in spatial models. Rev Econ Anal 2:194–214

Sevestre P, Trognon A (1996) Dynamic linear models. In: M àtyàs L, Sevestre P (eds) The econometrics of panel data: a handbook of theory with applications, chapter 7. Kluwer Academic Publishers, Dordrecht, pp 121–144

Simini F, González MC, Maritan A, Barabási A (2012) A universal model for mobility and migration patterns. Nature 484:96–100

Thissen M, van Oort F, Diodato D, Ruijs A (2013) Regional competitiveness and smart specialization in Europe: place-based development in international economic networks. Edward Elgar Publishing, Cheltenham

Thissen M, Diodato D, van Oort F (2013a) Integration and convergence in regional Europe: European regional trade flows from 2000 to 2010, PBL publication number: 1036. PBL Netherlands Environmental Assessment Agency, The Hague

Thissen M, Diodato D, van Oort F (2013b) Integrated regional Europe: European regional trade flows in 2000, PBL publication number: 1035. PBL Netherlands Environmental Assessment Agency, The Hague

Vega SH, Elhorst JP (2016) A regional unemployment model simultaneously accounting for serial dynamics, spatial dependence and common factors. Reg Sci Urban Econ 60:85–95

Vidoli F, Mazziotta C (2010) Spatial composite and disaggregate indicators: Chow–Lin methods and applications. In: Proceedings of the 45th scientific meeting of the Italian Statistical Society, Padua