Modeling Spatiotemporal Polarization of Resident Foreign Population

Giampiero Marra, David L. Miller, and Luca Zanin

July 15, 2010

1 Introduction

Migrants play an important socio-economic role in all countries. They perform vital functions in boosting the economy by working in jobs that others cannot (or would not) do, however they also cause controversy due to their increasing use of public services. Increased political interest, as well as academic interest has lead to a large volume of both theoretical and applied work regarding international migration: its causes and implications for both natives and foreigners.

Several migration theories have been proposed in the literature, from both a macro- and micro-economic point of view, based on different schools of thought. Examples include: the neoclassical economic theory (e.g., Borjas 1989; Massey et al. 1993), the assumptions regarding the new economics of migration(WHAT? This doesn't make sense in English, can you clarify? –DLM) (e.g., Stark and Bloom 1985; Massey et al. 1993), dual labor markets (e.g., Piore 1979; Massey et al. 1993) and world systems theory (AGAIN, I'm not sure what this is, expand or elaborate? –DLM) (e.g., Wallerstein 1974; Portes and Walton 1981; Morawska 1990; Hooghe et al. 2008). As pointed out, e.g., by Massey et al. (1993), migration perpetuates migration via migrant networks which connect migrants and non-migrants in both origin and destination regions through interpersonal ties (for example by friendship, shared community origin or employment).

In addition to the theory, a large number of empirical cases have been analyzed. A non-exhaustive list of examples includes: studies investigating the impact of foreigners on the labor market of the host country (e.g., Borjas 1994, 2003, 2005; Borjas, Freeman and Katz 1996; Card 2005; Fullin and Reyneri 2010), underemployment (Slack and Jensen 2007), possible connections between immigration and nations' poverty rates (Raphael and Smolensky 2009), transfer of identity from the first to second generation immigrants (Casey and Dustmann 2010), difference in education, earnings and employment between first and second generation immigrants (Algan et al. 2010), problems of integration of different cultures and languages (e.g., Lazer 1999; Contucci and Ghirlanda 2007), and the tendency for immigrants to live in ethnic "enclaves" (Edin, Fredriksson and Aslund 2003).

Despite the increasing amount of literature on international migration, to the best of our knowledge no studies have attempted to model the spatiotemporal polarization of the incidence of resident foreigners in the total population of a country. The results of such an investigation would complement previous findings, and would better help guide governmental policy. This is particularly relevant given the work of Massey et al. (1993), since migration patterns will tend to be similar year-on-year due to the network effect, the facility for policy-makers to map where migrants are particularly prevalent.

The incidence of resident foreigners can be calculated as follows

$$\%\ incidence\ of\ resident\ foreigners_j = \frac{number\ of\ resident\ foreigners_j}{total\ resident\ population_j} \times 100 \quad (1)$$

where j denotes a specific area such municipality, province, or region (e.g., De Bartolo 2007; Lowell 2007; Coleman 2008; Miguet 2008).

In the special issue of *The Economic Journal* dedicated to 'the integration of immigrants and its consequences', Manning (2010) observes that many European countries are currently experiencing a significant increase in immigrants in their populations. Here we consider the Italian case.

According to the official statistics of the Italian National Statistical Office (ISTAT), over the period $2003 - 2008$, the incidence of resident foreigners in the total population has increased from 3.4% to 6.5%. This large increase has been partly attributed to the regularization of previously illegal immigrants as part of new legislation (law no. 189/2002; see Zincone (2002) for more details). Even taking this into account, the change is remarkable if we consider that, according to the census (carried out every ten years), during the period $1991 - 2001$ the incidence increased from 0.6% to 2.3%.

Recent trends in migration have caused important socio-demographic and economic changes. For example, recent official statistics have highlighted an increase in marriage between Italians and foreigners, as well as an increased number of entrepreneurs with foreign citizenship. The projections up to 2050 produced by ISTAT suggest a continuing, substantial increase in numbers of resident foreigners (about 17% of the total population by 2050 as shown in Figure 1). Such a tendency is not uncommon in European countries (e.g., Alders, Keilman and Cruijsen 2007). It is therefore of paramount importance to formulate adequate social policies which can support, for example, the process of integrating immigrants into their host country. It is crucial to promote recognition and respect for different cultural identities with the aim of ensuring community cohesion between foreign and native people, and thus avoid social unrest (e.g., Cheong et al. 2007; Van Der Veer 2003). Immigrants' rights and integration policies typically relate to specific laws regarding public health and education (e.g., Zincone 2006). Taking a wider view, though, policies may also relate to professional training and the right to housing, as well as guidance and counselling on legal issues and administrative procedures. Key to policy makers decisions must be the knowledge of which areas of the country such policies should be enacted and strengthened due to a higher incidence of foreigners; the spatiotemporal maps produced using the proposed approach provide important insights into this matter for practitioners.

Measure (1) represents a simple well known demographic indicator for comparing different regions of a country in terms of number of foreigners per 100 resident inhabitants. It can be thought of as indicative of the spatial distribution of foreigners and thus identify areas of the country with a greater or lesser presence of resident foreigners in the population (e.g., OECD 2004). If such distribution is unbalanced in space and time, then there are most likely factors that attract more foreigners to that area as opposed to another. In fact, inhomogeneous spatial incidence may be explained by several characteristics of an area. Example of such characteristics include the presence of foreigners with the same origin (recall, e.g., the network theory, above), and an area's relative abundance of affordable (or socialised) housing.

To asses what draws foreigners to a particular area over anther, it is desirable to have some quantitative measure of the attractiveness of each area. We focus on factors of economic and labour market nature, construct a composite indicator that we have named 'Index of Spatial Economic Attractiveness' and produce spatiotemporal maps of the index. Three important variables constitute the index: (i) added value per

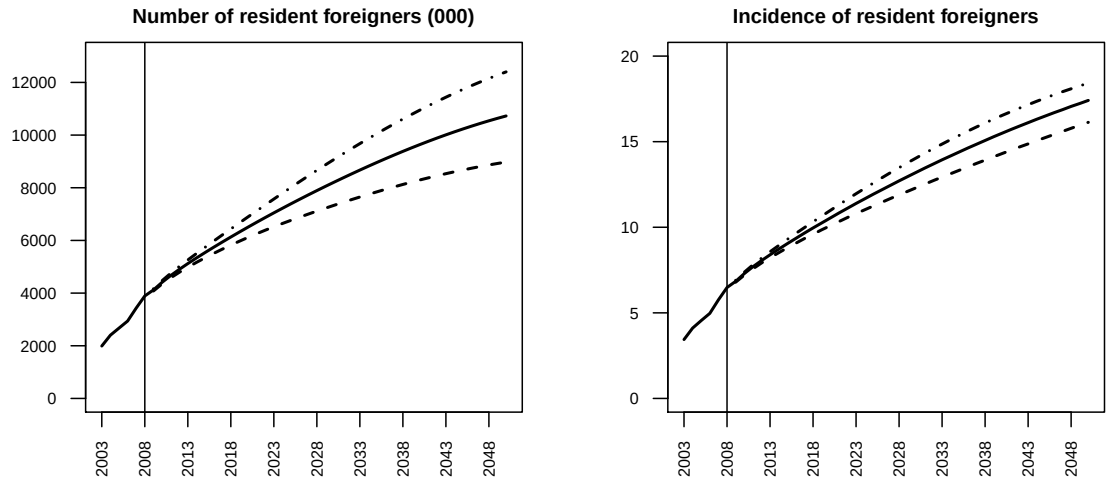

Figure 1: The graphs show the trend of resident foreigners in Italy during $2003-2008$ and in forecast for the period $2009-2050$, based on the official statistics and projections of ISTAT. The information is reported in terms of number (divided by 1000) and incidence of resident foreigners. Both variables are calculated at the end of December of each year, except for 2001 which is obtained from the population census (October 21^{th}). The average, low and high scenarios are by the solid and dotted lines, respectively.

capita, (ii) unemployment rate and (iii) employment rate. Since these are typically readily available, this composite indicator can be easily constructed for many countries. The results can be analyzed in the light of the spatiotemporal maps of the incidence of resident foreigners.

The novel empirical analysis proposed in this article addresses the points discussed above. Our study is carried out for Italy by using data from ISTAT which, since 2002, has provided a public database with annual frequency of the number of resident foreigners at a municipal level. This new data represents the current maximum spatial and temporal detail available. Spatiotemporal modelling can be achieved via the use of splines which can perform smoothing over some bounded domain $\Omega \subseteq \mathbb{R}^2$, over time (e.g., Hastie and Tibshirani 1990; Ruppert, Wand, and Carroll 2003; Wood 2006). Here, the response (incidence of resident foreigners) is modelled as a function of time and spatial coordinates. The model can then be used to create smooth maps of the geographical area of interest over time. Two points are noteworthy here. First, when a geographical region has complex boundaries, as in many applied contexts, features from one part of the domain can unduly influence other parts giving rise to a phenomenon known as *leakage* (e.g., Ramsay 2002). Leakage typically occurs when a smoother inappropriately links two parts of a domain causing the fitted surface to be mis-estimated, hence leading to incorrect inference (e.g., biased incidence estimates), which is clearly not desirable. This issue can be overcome if the smoother used can account for the structure of the domain that is under investigation; several solutions have been proposed (Ramsay 2002; Wood, Bravington and Hedley 2008, and references therein). Second, given that it is unlikely that the spatial trend of incidence of resident foreigners is additive in time, the employed approach should account for a space-time interaction. These two goals can be jointly achieved by using a scale-invariant three-dimensional tensor product smoother combining a cubic regression spline basis function for time trend and a soap film spline basis for the spatial dimensions (Wood 2006; Wood, Bravington and Hedley 2008).

The remainder of the article is organized as follows.

2 Data

2.1 Overview

The entry of foreigners from some countries of the European Union (EU) into Italy is regulated by the Schengen agreements. These allow for the elimination of border controls, hence creating a common area of free movement among the state members (e.g., Broeders 2007). Foreigners from a non-EU country must possess a visa or the necessary documents (for further details see the official website of the Ministry of the Interior http://www.interno.it). Most foreigners enter Italy for one or more of the following reasons: study, tourism, work and family (NEED to CITE this –DLM).

At this point, it is important to distinguish between resident foreign citizens and immigrants born in foreign countries. The term resident foreign citizens refers specifically to people with non-native (non-Italian) citizenship with residence in the country, entered into the official population registry of the municipality in which they live. Their nationality corresponds to the country of origin, unless otherwise stated on the identity document. (CAN we have a definition of immigrants born in foreign countries HERE? –DLM)

In many countries, official statistics report the number of resident foreign citizens, but not the number of immigrants born in foreign countries (Coleman (2008)). Currently, the dynamics of immigration to Italy still allows the acceptance of a definition which does not distinguish between these two groups. This simplification is reasonable in the first phase of the immigration process in a country, and when the majority of immigrants keep their own citizenship for a long period. Indeed, this is the case in Italy. The population census gives the best distinction between immigrants and foreigners. In this case the data can be analysed by birthplace, citizenship at birth, and citizenship at the census time.

Italian official statistics on resident foreign population are mainly based on administrative sources which are linked to specific laws. Since 2002, ISTAT has integrated the official statistics with a new public database (http://demo.istat.it). The database gives the total number of resident foreigners calculated at the end of December of each year, for each municipality (of which there are around 8100). For each municipality, measure (1) was calculated for each year in the period $2003 - 2008$. Compared to provincial or regional data, municipal-level data offers far more detail on the spatial distribution of the foreigners in the total population.

As well known in demographic science, populations are dynamic over time as well as space. To account for this we also include temporal information in our model. The change in number of foreign people, at the end of a year, is determined by

$$\mathrm{RFP}_{jy}^{31^{\mathrm{th}}} = \mathrm{RFP}_{jy}^{1^{\mathrm{st}}} + (I_1 + I_2 + I_3 + I_4) - (D_1 + D_2 + D_3 + D_4 + AC).$$

$\mathrm{RFP}_{jy}^{31^{\mathrm{th}}}$ indicates the total number of resident foreigners at the end of December in the j^{th} Italian municipality for the y^{th} year, and $\mathrm{RFP}_{jy}^{1^{\mathrm{st}}}$ the number of resident foreigners (in the same municipality, j and year, y) on the 1^{st} of January. I_1 denotes the number of people whose parents are foreigners (at least one of them being resident in the municipality), I_2 the number of foreign citizens who asked to transfer their residence from another Italian municipality to the current one, I_3 those who asked to transfer their residence from abroad, and I_4 refers to recording operations due to several reasons (ERM! Do you mean it's a miscellaneous term for mopping up things that aren't in that category? –DLM). D_1 represents the number of resident foreigners who died during the year, D_2 those who moved to a different municipality, D_3 those who moved abroad, D_4 refers to cancellations (MIGHT need to elaborate on that –DLM), and AC denotes

those resident foreigners who obtained Italian citizenship during the year.

Acquisition of Italian citizenship by foreigners is regulated by law 91 of the 5th of February 1992 and its subsequent amendments and additions. Foreigners can acquire the Italian citizenship by marriage to an Italian or by 'naturalisation'. This last case refers to the situation in which the foreigner has lived in the country from a certain period of time. Such a period varies dependent on the status of the person in question. For example the period is at least 10 years for non-EU citizens, at least 4 years for EU citizens or at least 5 years for political refugees or stateless persons and adult individuals who been adopted by Italian citizens. An exhaustive list of ways to obtain Italian citizenship is available on the Ministry of the Interior's website.

Once foreigners have acquired Italian citizenship, they are included in the Italian population's statistics. Currently, no information is available on the number of resident Italian citizens separated by origin. In the opinion of the authors, were this information available a very interesting study could be conducted. For instance, this information could be useful for monitoring the ethnic and cultural diversity in the population. However, the multi-ethnic or multi-cultural aspect is not addressed here.

The phenomenon of illegal immigrants (e.g., Carter 1999; Hillman and Weiss 1999; De Bartolo 2007; Bchir 2008; Cangiano 2008) and its respective quantification (e.g., Strozza 2004) is also a very important issue for policy-makers and scholars. In this case, those entering the country are not in possession of a visa, the necessary documents or a valid permission to be in that country. Several approaches have been proposed to quantify this phenomenon, but all are subject to criticism (especially with regard to the magnitude of errors in the estimates). For this reason, we only make use of the official data on the resident foreign population, hence avoiding the possible introduction of factors which are not directly controllable and quantifiable which could bias our results.

2.2 Demographic contribution of foreign population

Bijak et al. (2007) shows that in the absence of migration, many of the 27 European countries could see a decline in the resident population up to 2052. Typically, the main causes of population decline are an ageing population and a low fertility rate. It has been estimated that the population of Italy could decrease from 57 million in 2002 to about 43 million in 2052. The ratio of old (over 64) to young (up to 14) in the population was 143.4 at the end of 2008 (156.2 if we only consider individuals with Italian citizenship). In this case the elderly population is about 43% larger than the young population. The resident foreign population contributes 11.2 to the ratio (WE are talking about the ratio here not the percentage, right? –DLM) (the average age of resident foreigners is about 30). The latest data from ISTAT gives the average number of children for Italian women of childbearing age (from 15 to 49 years) as 1.33; lower than the average of 2.05 children for foreign women living in Italy. Combining the fertility rate for both Italian and foreign women, we obtain a total of 1.41. Coleman (2008) states that "typically, in low-mortality populations a total fertility of about 2.04 is regarded as the 'replacement level', that is, a level sufficient to replace the population in the long-run, ignoring migration". Given the impact of foreigners on the population structure, the measure of total fertility (and hence the level of replacement) must take into account the contribution of immigrants (Coleman (2008). In recent years, the presence of foreigners in Italy has contributed not only to population growth, but also to a reduction in average (IS this what I mean? –DLM) age and an increase in the total fertility rate.

The profile of the nationalities of resident foreigners in Italy is changing. In 2003, the top three countries

of origin were Albania (13.6%), Morocco (12.7%) and Romania (8.9%). In 2008, Romania was top (20.5%), followed by Albania (11.3%) and Morocco (10.4%). The significant increase of the Romanian population in Italy coincides with the country's entry to the European Union (1^{th} January 2007). Currently, it is the predominant nationality in 14 of the 20 Italian regions (YOU mean after Italian, right? –DLM). Italy hosts the largest Romanian community outside Romania and almost half of the entire Romanian migrant stock (Ban (2009)). The choice of Italy as primary destination for Romanians can be explained by many factors including the presence of small Italian companies, the economic links with in Romania, and the accessibility of the Italian language for Romanians. By continent in 2008 the resident foreign population in Italy was made up of 53.6% individuals from the Europe (EXCLUDING Italy? –DLM), 22.4% from Africa, 15.8% from Asia, 8.1% from America, and 0.1% from Oceania.(CITE this too, is it Ban 2009? –DLM)

3 Methodology

3.1 Model specification

The model we employ belongs to the class of generalized additive models (GAMs, Hastie and Tibshirani 1990). They allow for complex relationships between covariates and response variable, which are crucial to uncover interesting features in the data. Notice that the incidence of resident foreigners exhibits a positively skewed distribution. For this reason, a gamma distribution provides a realistic description of the response variable. The proposed model is as follows

$$\log\{\mathbb{E}(\texttt{irf}_{it})\} = f(\texttt{year}_t, \texttt{n}_i, \texttt{e}_i), \quad \texttt{irf}_{it} \sim \text{Gamma}, \tag{2}$$

for $i = 1, \ldots, 8094$ and $t = 1, \ldots, 6$. The log link function ensures positive fitted values. $\texttt{irf}_{it}$, $\texttt{n}_i$, and $\texttt{e}_i$ represent the variables percentage incidence of resident foreigners, northing, and easting, respectively. The function f is a multidimensional smooth of `year`, `n`, and `e` which models the joint effect of these variables on `irf`. Notice that we want both the space and time dimension to have an optimal degree of smoothness in terms of the bias variance trade-off. This means that the chosen smoother has to be invariant to the relative scaling of space (km) and time (years). This can be achieved by using a multidimensional tensor product smooth combining a cubic regression spline basis for `year` and an *isotropic* soap film spline basis for the two spatial dimensions `n` and `e`, since the smoother should not depend on the coordinate system used and account for the structure of the geographical domain under investigation (details are given in the next two sections). The smooth component in (2) is subject to identifiability constraints; see Wood (2006) for more details.

3.2 A three-dimensional tensor product smoother for time and space

The construction of a three-dimensional scale invariant tensor product smoother of time and space is based on a marginal one-dimensional spline basis for time and a two-dimensional marginal smooth for space, with associated quadratic penalties measuring their roughness. We omit the subscripts i and t for simplicity. Let us assume that we have two low-rank regression spline bases of any type to represent the smooth functions

f_{year} and f_{space}, we can write (e.g., Ruppert, Wand and Carroll 2003; Wood 2006)

$$f_{\text{year}}(\texttt{year}) = \sum_{l=1}^{L} \alpha_l a_l(\texttt{year}) = \mathbf{X}_{\text{year}}\boldsymbol{\alpha} \quad \text{and} \quad f_{\text{space}}(\texttt{n},\texttt{e}) = \sum_{r=1}^{R} \gamma_r d_r(\texttt{n},\texttt{e}) = \mathbf{X}_{\text{space}}\boldsymbol{\gamma},$$

where the $a_l(\texttt{year})$ and $d_r(\texttt{n},\texttt{e})$ are known cubic regression spline and soap film basis functions, with corresponding parameters α_l and γ_r, L and R are the spline dimensions of the two smooth components, and $\mathbf{X}_{\text{year}}$ and $\mathbf{X}_{\text{space}}$ are marginal model matrices evaluating the basis functions with parameter vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$. Notice that the expansion for f_{space} and its roughness penalty (described in the next paragraph) do not correspond exactly to the expressions used to set up the model. In this section, we are concerned with illustrating the construction of a three-dimensional scale invariant tensor product smoother, hence it is convenient to keep the description of the two-dimensional marginal smooth for space and its quadratic penalty as simple as possible. A rigorous description will be given in the next section. In order to set up a three-dimensional tensor product smoother for time and space we need $f_{\text{year}}(\texttt{year})$ to vary smoothly within the space dimensions. This can be achieved by allowing the parameters α_l to vary smoothly with `n` and `e`. Using the spline set-up for $f_{\text{space}}(\texttt{n},\texttt{e})$ we can write (e.g., Wood 2006, p. 163)

$$\alpha_l(\texttt{n},\texttt{e}) = \sum_{r=1}^{R} \gamma_{lr} d_r(\texttt{n},\texttt{e})$$

which results in

$$f(\texttt{year},\texttt{n},\texttt{e}) = \sum_{l=1}^{L}\sum_{r=1}^{R} \gamma_{lr} d_r(\texttt{n},\texttt{e}) a_l(\texttt{year}).$$

For any particular set of observations of `year`, `n`, and `e`, there exists a simple relationship between the matrix $\mathbf{X}$ evaluating the tensor product smooth at these observations, and the model matrices $\mathbf{X}_{\text{year}}$ and $\mathbf{X}_{\text{space}}$ evaluating the marginal smooths at the same observations. Ordering appropriately the parameters γ_{lr} into a vector $\boldsymbol{\theta}$, the i^{th} row of $\mathbf{X}$ is given by $\mathbf{X}_i = \mathbf{X}_{\text{year,i}} \otimes \mathbf{X}_{\text{space,i}}$, where $\otimes$ is the Kronecker product.

Within the GAM context, it is necessary to quantify the roughness of the smooth functions in the model so that over-fitting can be accounted for and hence avoided during the parameter estimation process (e.g., Marra and Radice 2010). As for the penalty associated with this tensor product basis, it is possible to start from roughness measures associated with the marginal smooths $f_{\text{year}}(\texttt{year})$ and $f_{\text{space}}(\texttt{n},\texttt{e})$. Suppose that functionals Js measuring the roughness of the smooth terms are available, and that these can be written as quadratic forms in the marginal parameters, we have that

$$J_{\text{year}}(f_{\text{year}}) = \boldsymbol{\alpha}^{\mathsf{T}}\mathbf{S}_{\text{year}}\boldsymbol{\alpha} \quad \text{and} \quad J_{\text{space}}(f_{\text{space}}) = \boldsymbol{\gamma}^{\mathsf{T}}\mathbf{S}_{\text{space}}\boldsymbol{\gamma},$$

where the $\mathbf{S}$ matrices contain known coefficients whose values depend on the chosen bases for time and space. For instance, the second-order cubic spline penalty for $f_{\text{year}}(\texttt{year})$ evaluates $J_{\text{year}}(f_{\text{year}}) = \int \left(\partial^2 f_{\text{year}}/\partial \texttt{year}^2\right)^2 d\texttt{year}$, but it may be more complex (Wood 2006, 2008). Following, e.g., Zanin and Marra (2010), an overall penalty for the tensor product smooth can be obtained by applying the penalties of $f_{\text{space}}(\texttt{n},\texttt{e})$ to the varying coefficients of the marginal smooth $f_{\text{year}}(\texttt{year})$, $\alpha_l(\texttt{n},\texttt{e})$,

$$\sum_{l=1}^{L} J_{\text{space}}\left\{\alpha_l(\texttt{n},\texttt{e})\right\},$$

and the penalties of $f_{\text{year}}(\texttt{year})$ to the varying coefficients of the marginal smooth $f_{\text{space}}(\texttt{n}, \texttt{e})$, $\gamma_r(\texttt{year})$,

$$\sum_{r=1}^{R} J_{\text{year}} \left\{ \gamma_r(\texttt{year}) \right\}.$$

It follows that the roughness penalty of $f(\texttt{year}, \texttt{n}, \texttt{e})$ can be measured as

$$J(f) = \lambda_{\text{space}} \sum_{l=1}^{L} J_{\text{space}} \left\{ \alpha_l(\texttt{n}, \texttt{e}) \right\} + \lambda_{\text{year}} \sum_{r=1}^{R} J_{\text{year}} \left\{ \gamma_r(\texttt{year}) \right\},$$

which can also be written as

$$\lambda_{\text{space}} \boldsymbol{\theta}^{\mathsf{T}} \mathbf{I}_L \otimes \mathbf{S}_{\text{space}} \boldsymbol{\theta} + \lambda_{\text{year}} \boldsymbol{\theta}^{\mathsf{T}} \mathbf{S}_{\text{year}} \otimes \mathbf{I}_R \boldsymbol{\theta},$$

where, once again, the vector $\boldsymbol{\theta}$ contains the tensor product smooth parameters. The λ are smoothing parameters controlling the trade-off between model fit and model smoothness. The next section shows how f_{space} and $\mathbf{S}_{\text{space}}$ can be constructed.

3.2.1 Soap film

As alluded to above, the smoother used for the spatial part of the model must take into account of the fact that the borders of Italy represent both physical and administrative geographic features. Clearly a rather naïve model would smooth over the bounding box encompassing all of the region we wish to draw inference on, this is not useful; first, because there are no foreigners in the sea (at best there are merely potential foreigners) and second, because smoothing over this whole area would cause leakage, as mentioned in the introduction.

Leakage occurs when a smoother inappropriately links two pats of a domain, this can happen when a two peninsulae jut out into the sea with different population densities on either side. Say one has a very high density, where as the other is significantly lower. Most smoothers will not respect that the two areas are different and should be treated so. Rather, the model will "smooth across" this gap, causing the high functional values to "leak" into the low valued peninsula and vice versa. There are, of course, situations in which leakage is appropriate. For example, in a study of the propagation of a chemical through a river system, there are several mechanisms to transport the chemical (e.g., surface water flow, animals) other than the river itself. Therefore there must be a motivation for why we wish to use a model that specifically prevents leakage. This is the case here, as there is no particular reason we should believe that the foreign population should be continuous across physical boundaries such as the Mediterranean Sea.

The soap film smoother (Wood, Bravington and Hedley 2008) uses a rather simple physical model to prevent leakage from occurring. First, consider the domain boundary to be made of wire, then dip this wire into a bucket of soapy water, you will then have a soap film in the same shape as the boundary (provided it doesn't pop(!)). Now consider the wire to lie in the `n-e` plane and the height of the soap film at a given point to be the functional value of the model. This film is then distorted smoothly by moving it vertically toward each datum locally, while minimising the surface tension in the film as a whole. The domain (Ω) is bounded by some polygon (B) which, in this case is the coastlines. The boundary conditions on B are estimated using a cyclic spline (Wood 2006, p. 151).

In order to perform soap film smoothing, we must first construct two sets of basis functions to form a smoother that conforms to the necessary boundary conditions. The first basis is used for the smoothing within the region of interest (Ω); the second is for finding the values on the boundary (i.e., smoothing on B itself). These bases are then summed to form

$$f_{\text{space}}(\mathtt{n}, \mathtt{e}) = \sum_{j=1}^{J} \alpha_j a_j(\mathtt{n}, \mathtt{e}), + \sum_{k=1}^{n} \gamma_k g_k(\mathtt{n}, \mathtt{e}), \tag{3}$$

where the γ_k and α_j are the parameters to be estimated. One can think of the $a_j(\mathtt{n}, \mathtt{e})$ as an offset dictated by the estimated boundary conditions on B and the sum of the $g_k(\mathtt{n}, \mathtt{e})$ as the smooth to the data inside Ω. We now show how these bases are constructed.

For the internal part of the smooth we first find a set of functions $\rho_k(\mathtt{n}, \mathtt{e})$, which are each solutions to the Laplace equation in two dimensions

$$\frac{\partial^2 \rho}{\partial \mathtt{n}^2} + \frac{\partial^2 \rho}{\partial \mathtt{e}^2} = 0 \tag{4}$$

except at one of the knots $(\mathtt{n}_k^*, \mathtt{e}_k^*)$. Then, we solve Poisson's equation in 2-dimensions

$$\frac{\partial^2 g_k}{\partial \mathtt{n}^2} + \frac{\partial^2 g_k}{\partial \mathtt{e}^2} = \rho_k(\mathtt{n}, \mathtt{e}) \tag{5}$$

for k indexing the n knots. When the boundary condition $\rho_k(\mathtt{n}, \mathtt{e}) = 0$ is applied, the set of basis functions for the soap film smoother, $g_k(\mathtt{n}, \mathtt{e})$ is found. The PDEs are solved numerically using successive overrelaxation (Wood, Bravington and Hedley 2008, for further details).

To find the boundary basis we first take a a 'boundary function', $f_{\text{bnd}}(r)$, using cyclic splines. The function evaluates the height of the function at each point around the boundary. $f_{\text{bnd}}(r)$ will have the expansion

$$f_{\text{bnd}}(r) = \sum_{j=1}^{J} \alpha_j \delta_j(r), \tag{6}$$

where r is the distance along the boundary, the α_j are parameters and δ_j are known cubic spline basis functions. To ensure that the spline is cyclic the usual constraint is enforced: that the value of the function at the first knot is the same as that at the last knot up to their second derivatives. Note that we don't find the α_j at this stage, for now we are only interested in the expansion. The basis functions $a_j(\mathtt{n}, \mathtt{e})$ themselves can be found by solving (5) for $\rho_k(\mathtt{n}, \mathtt{e}) \equiv 0$ with the boundary condition resulting from setting $\alpha_j = 1$ (and all other α_i to zero) in (6), using the same methods as for the $g_k(\mathtt{n}, \mathtt{e})$, above. The $a_j(\mathtt{n}, \mathtt{e})$ can be thought of as the set of functions with a peak at each of the J points around the boundary in turn which are smooth across the whole of Ω.

We have now found the set of basis functions for the inside of the domain and also the boundary-induced-smooth which acts as a base for the soap film smoother. Although this seems like a rather esoteric setup, all the procedure above is effectively doing is setting up a basis in order that we can use standard penalized regression techniques. Just as some spline bases are isotropic or non-isotropic, the soap film basis has the property that it obeys the boundary of the region we are smoothing over.

As with the basis, the penalty is split into two parts: one for the cyclic smooth around the boundary and one for the internal smooth. Writing $\boldsymbol{\beta}$ as the vector of all smooth coefficients for the soap film, $\boldsymbol{\gamma}$ for

the boundary smooth parameters and $\boldsymbol{\alpha}$ for the interior, we have

$$\boldsymbol{\beta}^{\mathrm{T}}\mathbf{S}_{\text{space}}\boldsymbol{\beta} = \lambda_{\text{bnd}}\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{S}_{\text{bnd}}\boldsymbol{\gamma} + \lambda_{\text{int}}\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{S}_{\text{int}}\boldsymbol{\alpha},$$

where

$$\mathbf{S}_{\text{space}} = \lambda_{\text{bnd}}\mathbf{S}_{\text{bnd}} + \lambda_{\text{int}}\mathbf{S}_{\text{int}}.$$

$\mathbf{S}_{\text{space}}$ is the total penalty for the spatial part of the smooth, $\mathbf{S}_{\text{int}}$ is the interior penalty and $\mathbf{S}_{\text{bnd}}$ is the cyclic spline boundary penalty. The λ are the smoothing parameters for the boundary and interior smooths respectively. We now look at the two parts of the penalty individually. Letting,

$$f_{\text{int}} = \sum_{k=1}^{n} \gamma_k g_k(\mathtt{n}, \mathtt{e}),$$

the isotropic interior penalty term is calculated as

$$\mathbf{S}_{\text{int}} = \int_{\Omega} \Big(\frac{\partial^2 f_{\text{int}}}{\partial \mathtt{n}^2} + \frac{\partial^2 f_{\text{int}}}{\partial \mathtt{e}^2} \Big)^2 \mathrm{d}\mathtt{n}\mathrm{d}\mathtt{e},$$

Differing from the standard thin plate regression spline penalty (**??? we suddenly introduce TPRS, we need to explain at least that TPRS is another approach and that we do better by using soap**) since: (*i*) the integration occurs only over Ω, (*ii*) there is no mixed derivative term, and (*iii*) the whole integrand is squared rather than each term individually. This allows the $\mathtt{n}$ and $\mathtt{e}$ term's derivatives to be traded off against each other so the nullspace of the penalty is infinite dimensional. This allows those functions in the nullspace to be sufficiently wiggly to meet any boundary conditions. The penalty for the cyclic spline running about the boundary, used to calculate the α_j is calculated as

$$\mathbf{S}_{\text{bnd}} = \int_{B} \Big(\frac{\partial^2 f_{\text{bnd}}}{\partial r^2} \Big)^2 \mathrm{d}r, \tag{7}$$

as normal.

The solution of the PDEs above, yielding the basis and penalty, is the most computationally expensive part of the procedure. Knots to use for $\mathtt{n}_k^*$ and $\mathtt{e}_k^*$ must be specified, here we use a grid (further detail on our exact setup is given below). In summary, we now have both the basis functions $\mathbf{A}_{ij} = a_j(\mathtt{n}_i, \mathtt{e}_i)$ and $\mathbf{G}_{ij} = g_j(\mathtt{n}_i, \mathtt{e}_i)$, which make up the i^{th} row of the model matrix $\mathbf{X}_{\text{space}}$, and the penalties $\mathbf{S}_{\text{bnd}}$ and $\mathbf{S}_{\text{int}}$, forming $\mathbf{S}_{\text{space}}$, for the soap film smoother.

3.3 Parameter estimation

In model (2), replacing f with its tensor product expression yields a generalized linear model (McCullagh and Nelder, 1989) whose design matrix contains the spline bases representing the smooth component in the model. This means that in principle such a model can simply be estimated by maximum likelihood (ML). However, in a smoothing spline context, unpenalized parameter estimation is likely to result in smooth component estimates that are too 'wiggly', hence undermining the utility of such a model. This can be overcome by penalized ML, where the use of penalties allows for the suppression of that part of smooth term complexity which has no support from the data (e.g., Marra and Radice 2010). Specifically, the model

can be fitted by minimization of

$$D(\boldsymbol{\theta}) + \boldsymbol{\theta}^{\mathsf{T}}\mathbf{S}\boldsymbol{\theta} \quad \text{w.r.t.} \quad \boldsymbol{\theta}, \tag{8}$$

where $\boldsymbol{\theta}$ contains all tensor product smooth parameters associated with the one-dimensional spline basis for time and two-dimensional smooth for space, and $\mathbf{S} = \sum_i \lambda_i \mathbf{S}_i$, where the $\mathbf{S}_i$ are matrices of known coefficients properly defined according to the results of the previous sections with associated smoothing parameters λ_i. The model deviance, D, is defined as $2\phi(l_{\text{sat}} - l)$, where ϕ is a dispersion parameter, l is the log-likelihood of the model and l_{sat} the maximum value for the log-likelihood of the model with one parameter per datum.

Given values for the λ_i, minimization of (8) is straightforward. However, smoothing parameter estimation has to be addressed. This can be achieved by minimization of a prediction error estimate, such as the generalized cross-validation (GCV) (Craven and Wahba 1979), or by approximate restricted ML (REML) estimation (Wood 2010). Smoothing parameter selection via the GCV consists of minimizing

$$V(\boldsymbol{\lambda}) = \frac{nD(\hat{\boldsymbol{\theta}})}{\{n - \text{tr}(\mathbf{A})\}^2}, \tag{9}$$

where $\mathbf{A}$ is the usual hat matrix, and $\text{tr}(\mathbf{A})$ represents the effective degrees of freedom (edf) or number of estimated parameters of the penalized model. Wood (2006) described a computational procedure to estimate smoothing parameters on the basis of criterion (9). As an alternative, approximate REML can be employed. Within this framework, the penalized likelihood estimates, $\hat{\boldsymbol{\theta}}$, can be seen as the posterior modes of the distribution of $\boldsymbol{\theta}|\mathbf{y}$ if $\boldsymbol{\theta} \sim N(\mathbf{0}, \mathbf{S}^{-1})$, where $\mathbf{S}^{-1}$ is an appropriate generalized inverse. Viewing the spline parameters as random effects allows for the possibility to estimate the λ_i via REML (Wahba 1985). The recent work by Reiss and Ogden (2009) shows that at finite sample sizes GCV is prone to undersmoothing and is more likely to develop multiple minima than REML. Therefore, we employ REML given its practical advantages for smoothing parameter selection. We use the `gam()` function of the `R` package `mgcv` since approximate REML optimization for GAMs is reliably and efficiently implemented (Wood 2010). During the model fitting process, all gamma distributions for all years are assumed to have a common shape parameter since preliminary descriptive analysis suggested that this is the case.

Given the dimension of the dataset used here, we experienced some memory issues which could not be sorted out. This was eventually dealt with by trying to find a compromise between tensor product basis dimension and dataset reduction. The final model was fitted on data obtained by averaging over a 150×150 grid for all years. The basis sizes (for the cubic regression splines and cyclic splines) and interior knots (for the soap film) are given in Table 1.

Region	Interior knots	Cyclic spline basis size	Cubic spline basis size
Main land	109 (20x25)	50	20
Sardinia	50 (10x10)	20	4
Sicily	40 (10x10)	20	4

Table 1: Basis sizes per region for the smooth functions to be fit to the Italian data. For the interior (soap film) knots, the numbers in brackets show the initial grid, the other number gives the number of knots actually used (i.e., with those outside the boundary or causing numerical problems with the fitting procedure removed).

The spatiotemporal structure of model (2) and the lack of availability of economic variables at municipal level would suggest to fit the model by also taking into account the possible presence of some unexplained

spatial and autocorrelation structure in the data. In a smoothing spline context, this can be consistently achieved by using a mixed modelling approach (Breslow and Clayton 1993). This possibility was explored by using the `gamm()` function of the `mgcv` package, which iteratively calls the `lme()` function of the `nlme` package (Pinheiro et al. 2009) for maximization; no models could be fitted due to convergence failures (see, e.g., Ruppert, Wand, and Carroll (2003) and Wood (2006) for problems and limitations with this approach). To explore the sensitivity of our estimates, we tried out a fixed effect approach (Wooldridge 2002) but the results did not change significantly.

3.4 Variance estimation

Confidence intervals can be effectively constructed using the well known Bayesian credible intervals originally proposed by Wahba (1983) or Silverman (1985) in the univariate spline model context, and then generalized to the component-wise case when dealing with Gaussian and non-Gaussian data (e.g., Gu 2002; Gu and Wahba 1993; Ruppert, Wand, and Carroll 2003). The Bayesian model representation provides a self-consistent framework for constructing intervals since the posterior distribution of the model parameters is known. Also, because these intervals include both a bias and variance component (Nychka 1988), they have good observed *frequentist* coverage probabilities across the function as shown, for example, in the simulation experiments of Wang and Wahba (1995). The posterior distribution is given as

$$\boldsymbol{\theta}|\mathbf{y} \backsim N(\hat{\boldsymbol{\theta}}, \mathbf{V}_{\boldsymbol{\theta}}), \tag{10}$$

where $\hat{\boldsymbol{\theta}}$ is the maximum penalized likelihood estimate of $\boldsymbol{\theta}$ obtained as explained in the previous section, which is of the form $(\mathbf{X}^{\mathsf{T}}\mathbf{W}\mathbf{X} + \mathbf{S})^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{W}\mathbf{z}$, $\mathbf{V}_{\boldsymbol{\theta}} = (\mathbf{X}^{\mathsf{T}}\mathbf{W}\mathbf{X} + \mathbf{S})^{-1}\phi$, $\mathbf{X}$ contains the columns associated with the regression spline bases used to set up the model, and $\mathbf{W}$ and $\mathbf{z}$ are the diagonal weight matrix and the pseudodata vector at convergence of the algorithm used to fit the penalized model (e.g., Wood 2006, 2010).

Given the result above, it is easy to find confidence intervals for linear functions of the parameters such as smooth components. Intervals for non-linear functions of the model coefficients can be efficiently obtained by simulation from the posterior distribution of $\boldsymbol{\theta}$. Result (10) can also be used to produce trend estimates as discussed in the next section.

3.5 Trend estimation

To aid interpretation of the results and also possibly uncover interesting features in the data, trend estimates for some given areas of interest can be produced using the predictive distribution of $\texttt{irf}_{it}$, constructed employing the result of the previous section. This approach has been previously adopted in the literature (e.g., Augustin et al. 2009), and can be implemented as follows:

1. Repeat the following steps for $b = 1, \cdots, N_b$, where N_b represents the number of random draws.
   (a) Simulate a random $N(\hat{\boldsymbol{\theta}}, \hat{\mathbf{V}}_{\boldsymbol{\theta}})$ and call the resulting coefficient vector $\boldsymbol{\theta}_b$.
   (b) Calculate $\widehat{\mathbb{E}(\texttt{irf}_{it})} = \widehat{\texttt{irf}}_{itb} = \exp(\mathbf{X}^*_{it}\boldsymbol{\theta}_b)$, where $\mathbf{X}^*_{it}$ is evaluated at the observed values.

(c) For a given area of interest a and year t, work out the quantity

$$\widehat{\overline{\texttt{irf}}}^{a}_{tb} = \frac{1}{n_a}\sum_{i=1}^{n_a}\widehat{\texttt{irf}}_{itb}.$$

2. Produce the required summary statistics, in this case median, lower and upper 95% quantiles, for the temporal trend $\widehat{\overline{\texttt{irf}}}^{a}_{tb}$.

Small values for N_b are typically tolerable. In practice, N_b can be set to 100. Increasing this value does not change the results which are presented in Section 4.

4 Empirical results

XX

4.1 Spatiotemporal trends

XX

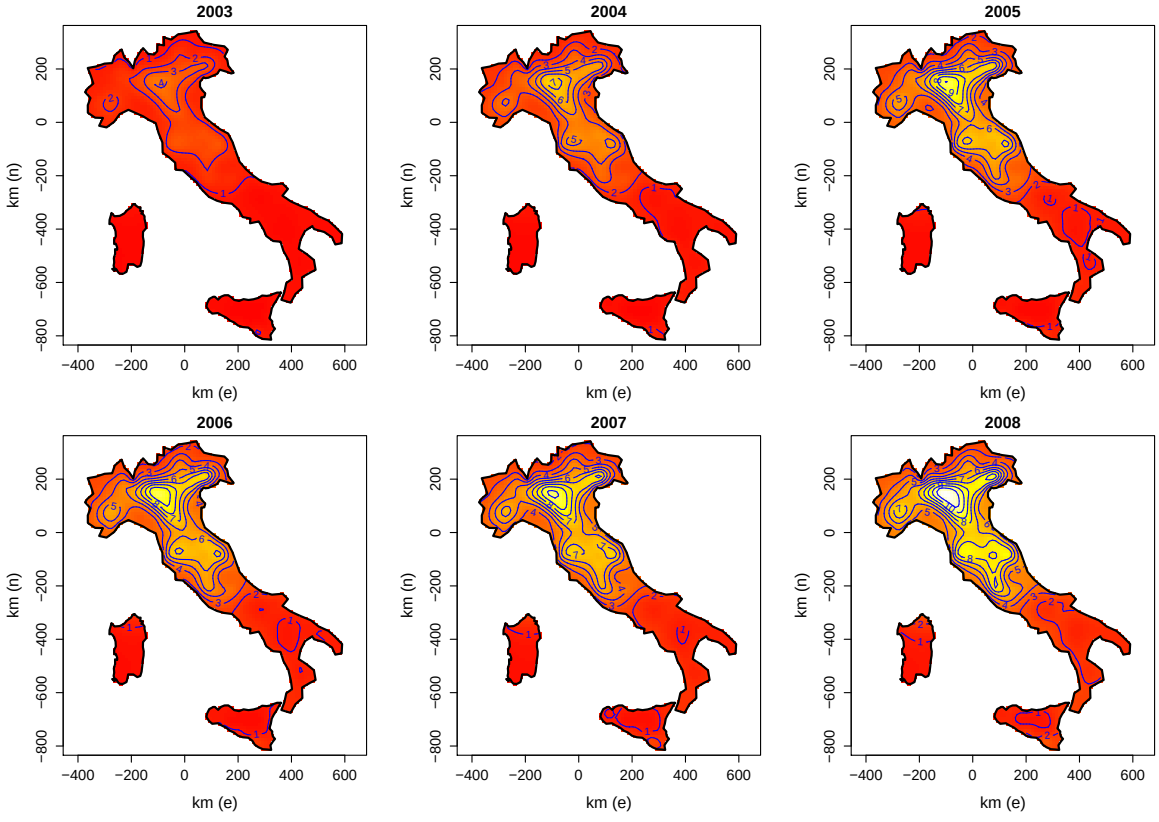

Figure 2: to fill in

4.2 Foreigner presence and spatial attractiveness: is there a nexus?

XX

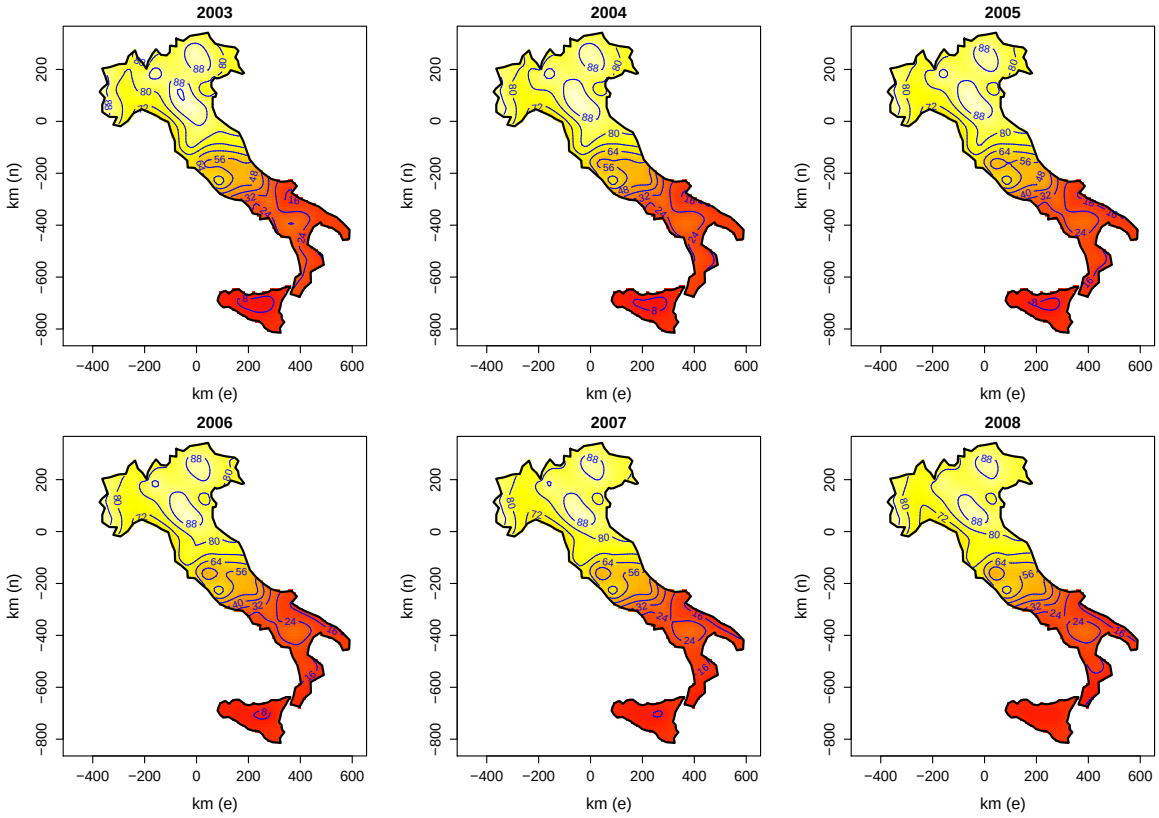

Figure 3: to fill in

4.3 Policy-maker's corner: a point of discussion

XX

5 Conclusions

XX

References

[1] Alders, M., Keilman, N., and Cruijsen, H. (2007) , "Assumptions for long-term stochastic population forecasts in 18 European countries," *European Journal of Population*, 23, 33–69.

[2] Algan, Y., Dustmann, C., Glitz, A., and Manning, A. (2010), "The economic situation of first- and second-generation immigrants in France, Germany, and the UK," *The Economic Journal*, 120, F4–F30.

[3] Augustin, N. H., Musio, M., Wilpert, K., Kublin, E., Wood, S. N., and Schumacher, M. (2009), "Modeling spatiotemporal forest health monitoring data," *Journal of the American Statistical Association*, 104, 899–911.

[4] Ban, C. (2009), "Economic transnationalism and its ambiguities: the case of Romanian Migration to Italy," *International Migration*, DOI: 10.1111/j.1468-2435.2009.00556.x.

[5] Bchir, H. M. (2008), "The effect of mode 4 liberalization on illegal immigration," *Economic modelling*, 25, 1051–1063.

[6] Bijak, J., Kupiszewska, D., Kupiszewski, M., Saczuk, K., and Kicinger, A. (2007), "Population and labour force projections for 27 European countries, 2002–2052: impact of international migration on population ageing," *European Journal of Population*, 23, 1–31.

[7] Borjas, J. G. (1989), "Economic theory and international migration," *International Migration Review*, 23, 457–85.

[8] Borjas, J. G. (1994), "The economics of immigration," *Journal of Economic Literature*, 32, 1667–1717.

[9] Borjas, J. G., Freeman, R. B., and Katz, L. F. (1996), "Searching for the effect of immigration on the labor market," *American Economic Review*, 86, 246–51.

[10] Borjas, J. G. (2003), "The labor demand curve IS downward sloping: reexamining the impact of immigration on the labor market," *The Quaterly Journal of Economics*, 118, 1335–74.

[11] Borjas, J. G. (2005), "The labor market impact of high-skill immigration," *American Economic Review*, 92, 56–60.

[12] Breslow, N. E., and Clayton, D. G. (1993), "Approximate inference in generalized linear mixed models," *Journal of the American Statistical Association*, 88, 9–25.

[13] Broeders, D. (2007), "The new digital borders of Europe: EU database and the surveillance of irregular migrants," *International Sociology*, 22, 71–92.

[14] Cangiano, A. (2008), "Foreign migrants in Southern European countries: evaluation of recent data," in *International Migration in Europe. Data, Models, and Estimates*, Editors: Raymer, J. and Willekens, F., 89–112, Wiley.

[15] Casey, T., and Dustmann, C. (2010), "Immigrants' identity, economic outcomes and the trasmission of identity across generations," *The Economic Journal*, 120, F31–F51.

[16] Card, D. (2005), "Is the new immigration really so bad?," *The Economic Journal*, 115, 300–323.

[17] Carter, T. J. (1999), "Illegal immigration in a efficiency wage model," *Journal of International Economics*, 49, 385–401.

[18] Cheong, P. H., Edwards, R., Goulbourne, H., and Solomos, J. (2007), "Immigration, social cohesion and social capital: A critical review," *Critical Social Policy*, 27, 24–49.

[19] Coleman, D. (2008), "The demographic effects of international migration Europe," *Oxford Review of Economic Policy*, 24, 452–76.

[20] Contucci, P., and Ghirlanda, S. (2007), "Modelling society with statistical mechanics: an application to cultural contact and immigration," *Quality and Quantity*, 41, 569–78.

[21] Craven, P., and Wahba, G. (1979), "Smoothing noisy data with spline functions," *Numerische Mathematik*, 31, 377–403.

[22] De Bartolo, G. (2007), "Immigration in Italy: the great emergency," Population Association of America - Annual Meeting, March 29-31, New York, Office of Population Research, Princeton University.

[23] Edin, Per-A., Fredriksson, P., and Aslund, O. (2003), "Ethnic enclaves and the economic success of immigrants - evidence from a natural experiment," *The Quarterly Journal of Economics*, 118, 329–57.

[24] Fullin, G., and Reyneri, E. (2010), "Low unemployment and bad jobs for new immigrants in Italy," *International Migration*, In press, DOI: 10.1111/j.1468-2435.2009.00594.x.

[25] Gu, C. (2002), *Smoothing Spline ANOVA Models*, London: Springer-Verlag.

[26] Gu, C., and Wahba, G. (1993), "Smoothing Spline ANOVA with Component-Wise Bayesian Confidence Intervals," *Journal of Computational and Graphical Statistics*, 2, 97–117.

[27] Hastie, T., and Tibshirani, R. (1990), *Generalized Additive Models*, London: Chapman & Hall.

[28] Hillman, A. L., and Weiss, A. (1999), "A theory of permissible illegal immigration," *European Journal of Political Economy*, 15, 585–604.

[29] Hooghe, M., Trappers, A., Meuleman, B., and Reeskens, T. (2008), "Migration to European countries: a structural explanation of patterns 1980-2004," *International Migration Review*, 42, 476–504.

[30] Lazear, E. P. (1999), "Culture and language," *Journal of Political Economy*, 107, S95–S126.

[31] Lowell, L. B. (2007), "Trends in international migration flows and stocks, 1975-2005," OECD Social, Employment and Migration Working Paper 58, OECD, Paris.

[32] Manning, A. (2010), "Feature: the integration of immigrants and their children in Europe: introduction," *The Economic Journal*, 120, F1–F3.

[33] Marra, G, and Radice, R. (2010), "Penalised regression splines: theory and application to medical research," *Statistical Methods in Medical Research*, 19, 107–125.

[34] Massey, D. S., Arago, J., Hugo, G., Kouaouci, A., Pellegrino, A., and Taylor, J. E. (1993), "Theory of international migration: a review and appraisal," *Population and Development Review*, 19, 431–66.

[35] McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models*, London: Chapman & Hall.

[36] Miguet, F. (2008), "Voting about immigration policy: What does the Swiss experience tell us?," *European Journal of Political Economy*, 24, 628–41.

[37] Morawska, E. (1990), "The Sociology and historiography of Immigration," in *Immigration Reconsidered: History, Sociology, and Politics*, ed. Virginia Y. McLaughlin, New York: Oxford University Press, 187–240.

[38] Nychka, D. (1988), "Bayesian Confidence Intervals for Smoothing Splines," *Journal of the American Statistical Association*, 83, 1134–1143.

[39] OECD (2004), "Trends in International Migration 2003," *OECD Pubblishing*, Paris.

[40] Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and the R Core Team (2009), *nlme: Linear and Nonlinear Mixed Effects Models*, R package version 3.1- 96.

[41] Piore, M. J. (1979), *Birds of Passage: Migrant Labor in Industrial Societies*. Cambridge: Cambridge University Press.

[42] Portes, A., and Walton, J. (1981), *Labor, Class, and the International System*. New York: Academic Press.

[43] Ramsay, T. (2002), "Spline smoothing over difficult regions," *Journal of the Royal Statistical Society Series B*, 64, 307–19.

[44] Raphael, S., and Smolensky, E. (2009), "Immigration and poverty in the United States," *American Economic Review*, 99, 41–4.

[45] Reiss, P. T., and Ogden, R. T. (2009), "Smoothing parameter selection for a class of semiparametric linear models," *Journal of the Royal Statistical Society Series B*, 71, 505–524.

[46] Ruppert, D., Wand, M. P., and Carroll, R. J. (2003), *Semiparametric Regression*, London: Cambridge University Press.

[47] Silverman, B. W. (1985), "Some Aspects of the Spline Smoothing Approach to Non-Parametric Regression Curve Fitting," *Journal of the Royal Statistical Society Series B*, 47, 1–52.

[48] Slack, T., and Jensen, L. (2007), "Underemployment across immigrant generations," *Social Science Research*, 36, 1415–30.

[49] Stark, O., and Bloom, E.D. (1985), "The new economics of labor migration," *American Economic Review*, 75, 173–78.

[50] Strozza, S. (2004), "Estimates of the illegal foreigners in Italy: a review of the literature," *International Migration Review*, 38, 309–331.

[51] Van Der Veer, K. (2003), "The future of western societies: multicultural identity or extreme nationalism?," *Futures*, 35, 169–187.

[52] Wahba, G. (1983), "Bayesian 'Confidence Intervals' for the Cross-Validated Smoothing Spline," *Journal of the Royal Statistical Society Series B*, 45, 133–150.

[53] Wahba, G. (1985), "A Comparison of GCV and GML for Choosing the Smoothing Parameter in the Generalized Spline Smoothing Problem," *The Annals of Statistics*, 13, 1378–1402.

[54] Wallerstein, I. (1974), *The Modern World System, Capitalist Agricolture and the Origins of the European World Economy in the Sixteenth Century*. New York: Academic Press.

[55] Wang, Y., and Wahba, G. (1995), "Bootstrap Confidence Intervals for Smoothing Splines and Their Comparison to Bayesian Confidence Intervals," *Journal of Statistical Computation and Simulation*, 51, 263–279.

[56] Wood, S. N. (2006), *Generalized Additive Models: An Introduction with R*, London: Chapman & Hall.

[57] Wood, S. N., Bravington, M. V., and Hedley, S. L. (2008), "Soap film smoothing," *Journal of the Royal Statistical Society Series B*, 70, 931–55.

[58] Wood SN (2010), "Fast stable REML estimation of semiparametric GLMs," *Journal of the Royal Statistical Society Series B*, in press.

[59] Wooldridge, J. M. (2002), *Econometric Analysis of Cross Section and Panel Data*, Cambridge: MIT Press.

[60] Zanin, L., and Marra, G. (2010), "Modelling the probability that a household participates in tourism: the Italian case," *Physica A: Statistical Mechanics and its Applications*, ?.

[61] Zincone, G. (2006), "The making of policies: immigration and immigrants in Italy," *Journal of Ethnic and Migration Studies*, 32, 347–75.