SEEDS
Sustainability Environmental Economics and Dynamics Studies

Working Paper Series

Model uncertainty, nonlinearities and out-of-sample comparison: evidence from international technology diffusion

by

Georgios Gioldasis, Antonio Musolesi, Michel Simioni

01/2020

SEEDS Working Paper 01/2020
January 2020
By Georgios Gioldasis, Antonio Musolesi, Michel Simioni

Model uncertainty, nonlinearities and out-of-sample comparison: evidence from international technology diffusion

Georgios Gioldasis
Department of Economics and Management (DEM),
University of Ferrara and SEEDS, Ferrara, Italy

Antonio Musolesi
Department of Economics and Management (DEM),
University of Ferrara and SEEDS, Ferrara, Italy

Michel Simioni
MOISA, INRA, University of Montpellier, Montpellier, France

December 4, 2019

Abstract

This paper reconsiders the international technology diffusion model. Because the high degree of uncertainty surrounding the Data Generating Process and the likely presence of nonlinearities and latent common factors, it considers alternative nonparametric panel specifications which extend the Common Correlated Effects approach and then contrasts the out-of-sample performance of them with those of more common parametric models. To do so, we adopt an approach recently proposed within the literature of nonparametric regression. This approach is based on a pseudo Monte Carlo experiment that takes its roots on cross validation and aims at testing whether two competing approximate models are equivalent in terms of their expected true error. Our results indicate that the adoption of a nonparametric approach provides better performances. This work also refines previous results by showing threshold effects, nonlinearities and interactions, which are obscured in parametric specifications and which have relevant implications for policy.

Keywords: large panels; cross-sectional dependence; factor models; nonparametric regression; spline functions; approximate model; predictive accuracy, international technology diffusion.

JEL classification: C23; C5; F0; O3.

1 Introduction

With the development of endogenous growth theory since the nineties, there has been an increasing interest in estimating the effect research and development (R&D) on growth and productivity. A pioneering empirical work by Coe and Helpman (1995), revisited by Coe et al. (2009) – henceforth CH and CHH, respectively – relates total factor productivity (TFP) to both domestic and foreign R&D and, assuming that technology spills over across countries through the channel of trade flows, constructs foreign R&D capital stock as the import-share-weighted average of the domestic R&D capital stocks of the trading partners. Subsequent studies consider other factors as channels of international spillovers, such as foreign direct investment, bilateral technological proximity, patent citations between countries, language skills or geographic proximity.

In recent years, with the increasing relevance of globalisation, panel data approaches have focused on the issue of cross-country dependence arising from the interactions among economic units or the consideration of global unobservable factors (Pesaran, 2006; Bai, 2009). While alternative approaches have been developed to estimate a model with a multifactor error structure, an extremely appealing one is the common correlated effects (CCE) approach developed by Pesaran (2006), which has been further developed and proved to be valid in a variety of situations (Chudik et al., 2011; Pesaran and Tosetti, 2011; Kapetanios et al., 2011). In particular, Pesaran and Tosetti (2011) prove that the CCE estimator provides consistent estimates of the slope coefficients and their standard errors under the more general case of a multifactor error structure and spatial error correlation. Moreover, Chudik et al. (2011), after introducing the concept of strong and weak factors and clarifying that Pesaran (2006) introduces CCE estimators in a panel model with a fixed number of strong factors and without weak (semi-weak or semi-strong) factors, demonstrate that the CCE method still yields consistent estimates of the mean of the slope coefficients when there are an infinite number of factors, a fixed number of which are strong, with the rest being weak, semi-weak or semi-strong. Additionally, the asymptotic normal theory continues to be applicable even in this extended framework. Finally, Kapetanios et al. (2011) provide both analytical results and a simulation study indicating that the CCE approach is still valid when the unobserved factors are allowed to follow unit root processes.

With the availability of these new methods, recent studies extended the literature on international R&D spillovers (see, *e.g.* Ertur and Musolesi, 2017). To the best of our knowledge, however, all the empirical literature adopted parametric specifications. Although a log-log specification is customary in the literature, more flexible functional forms could be suitable to model the likely complex relation between research activity and economic performances. This relevant issue was recognised in closely related literatures on economics of innovation. In an early work, (Griliches, 1998, p. 1674), relating R&D to productivity suggests that *"Given the nonlinearity and the noisiness in this relation, the finding of "diminishing returns" is quite sensitive to functional form, weighting schemes, and the particular point at which the elasticity is* *evaluated'*. As highlighted by Charlot et al. (2014), it can be expected, for instance, that a critical mass of R&D or human capital is necessary to make such inputs truly effective. Moreover, not only the linearity but also the additivity assumption implicit in the linear model, might be too restrictive and should be relaxed, as suggested by (Hall et al., 2010, p. 1074): *"Because the additive model is not really a very good description of knowledge production, further work on the best way to model the R&D input would be extremely desirable'*.

Because of the possible existence of nonlinearities, threshold effects, non-additive relations, estimating a nonparametric relation could be important to avoid a functional form bias. Moreover, nonparametric approaches have been shown to provide new and useful insights in topics very closely related to the present one (Ma et al., 2015; Maasoumi et al., 2007). These methods are recently developing also in the context of panel data (Rodriguez-Poo and Soberon, 2017; Parmeter and Racine, 2018). Our econometric approach builds on the nonparametric model by Su and Jin (2012), which allows for a multifactor error structure and extends the approach by Pesaran (2006). The approach adopted in this paper combines the flexibility of nonpamaetric models with the ability of factor models to allow for cross-sectional dependence and to account for endogeneity due to unobservables, whereby the explanatory variables are allowed to be correlated with the unobserved factors. Following Su and Jin (2012), the nonparametric component is estimated using splines. Specifically, we adopt a regression splines framework, which provides computationally attractive low rank smoothers. We also employ penalized regression splines, as they combine the features of regression splines and smoothing splines, and have proven to be useful empirically in many aspects (Ruppert et al., 2003) while their asymptotic properties have been studied in recent years (see, *e.g.* Li and Ruppert, 2008, Wood et al., 2016). The choice of the knots is avoided by using knot-free bases for smooths (Wood, 2003).

There exists a high degree of uncertainty surrounding the Data Generating Process (GDP) and it can be expected a bias-efficiency trade-off when comparing parsimonious to complex models (Ma et al., 2015; Racine and Parmeter, 2014; de Almeida et al., 2018). Considering flexible models is appealing but, because of the curse of dimensionality, it may come at the price of unfeasible or extremely inefficient estimates (Ma et al., 2015; Racine and Parmeter, 2014; Baltagi et al., 2002, 2003, 2004). For this reason, we perform model selection by comparing the out-of-sample performances of some alternative models. In particular, we consider a fully nonparametric relation between TFP and the explanatory variable, we then avoid the curse of dimensionality problem by relying on a specification with additive smooth terms and finally consider the more constrained parametric CCE specification. To do so, we adopt the approach recently proposed by Racine and Parmeter (2014), which is based on a pseudo Monte Carlo experiment and takes its roots on cross validation. While in consistent model selection, it is assumed that exists a finite-dimensional 'true model', in Racine and Parmeter (2014) fitted econometric models are view as approximations, as suggested by (Hansen, 2005, p. 60) *"econometric model selection methods should be based on a semiparametric vision, models should be viewed as approximations'*, and the goal is to test whether one approximate model performs better than another on data drawn from the same data generating process. In such a framework it is common adopting a sample-splitting mechanism whereby one splits the full sample into two sub-samples where one sub-sample is used for estimation and the other for out-of-sample evaluation. To avoid that the results reflect a particular division of the data into two sub-samples, the main idea by Racine and Parmeter (2014) is to repeat this process a large number of times becausee it can provide significant power improvements over existing single-split techniques. This is a new line of research, which has been recently pursued with nonparametric panel data (Ma et al., 2015; Delgado et al., 2014).

The econometric analysis is conducted using annual country-level data for 24 OECD countries from 1971 to 2004. This dataset is also used, among others, in Coe et al. (2009) and in Ertur and Musolesi (2017) and this allows for a comparability with previous studies. The paper is organized as follows.

In section 2 we describe the model specifications that we employ as well as the adopted estimation approach. The out-of-sample comparison of the alternative specifications is presented in section 3. Because the nonparametric specifications outperform the parametric one, we also discuss the main results obtained from these smooth regression models, which have relevant implication for public policies. Finally, section 4 concludes.

2 Model specifications and estimation methods

2.1 The classical parametric approach

The classical parametric specification. The standard parametric specification *à la* CH/CHH can be expressed as:

$$\log f_{it} = \alpha_i + \theta \log S_{it}^d + \gamma \log S_{it}^f + \delta \log H_{it} + e_{it}, \tag{1}$$

where f_{it} is the TFP of country $i = 1, ..., N$ at time $t = 1, ..., T$, α_i are individual fixed effects, S_{it}^d and S_{it}^f are domestic and foreign R&D capital stocks, respectively, H_{it} is a measure of human capital, and e_{it} is the error term. Foreign capital stock S_{it}^f is defined as the weighted arithmetic mean of S_{jt}^d for $j \neq i$, that is $S_{it}^f = \sum_{j \neq i} \omega_{ij} S_{jt}^d$, where ω_{ij} represents the weighting scheme. We adopt the definition of weights proposed by Lichtenberg and van Pottelsberghe de la Potterie (1998), which has been previously adopted in many other papers (Coe et al., 2009; Lee, 2006; Ertur and Musolesi, 2017), incorporating information on bilateral imports. Specifically, we use bilateral weight, i.e. the ratio of country i's imports of goods and services from country j and nominal GDP of country$j$.

Model in Eq. (1) can be written as a special case of the heterogenous panel data model,

$$y_{it} = \alpha_i' \mathbf{d}_t + \beta_i' \mathbf{x}_{it} + e_{it}, \tag{2}$$

with $y_{it} = \log f_{it}$, α_i is a constant term as $\mathbf{d}_t = d_t = 1$, $\mathbf{x}_{it} = [\log S_{it}^d, \log S_{it}^f, \log H_{it}]'$ and $\beta_i = \beta = [\theta, \gamma, \delta]'$.

In the general specification (2), it is usually assumed that $\beta_i = \beta + \mu_i$ where the deviations, μ_i, are independently and identically distributed with mean 0. Moreover these deviations are distributed independently of e_{jt}, $\mathbf{d}_t$, and $\mathbf{x}_{jt}$, for all i, j and t. In this specification, $\mathbf{d}_t$ denotes a $l \times 1$ vector of observed common effects (including deterministics such as intercepts and seasonal dummies), and α_i is the associated vector of parameters.

CCE estimators. Panel data literature dealing with models like (2) with both N and T large has shown that ignoring cross-sectional dependence of individual errors can seriously impair the properties of usual panel data estimators (Phillips and Sul, 2003; Andrews, 2005; Phillips and Sul, 2007; Sarafadis and Robertson, 2009). Cross-sectional dependence can be due to unobserved common factors such as economy-wide shocks (for instance, oil price rise), that affect all countries albeit with different intensities. The errors e_{it} are then assumed to have the following common factor structure:

$$e_{it} = \gamma_i^{'} \mathbf{f}_t + \varepsilon_{it}, \tag{3}$$

in which $\mathbf{f}_t$ is an $m \times 1$ vector of unobserved common factors with associated country-specific factor loadings γ_i. The number of factors, m, is assumed to be fixed relative to the number of countries N, and in particular $m << N$. These factors $\mathbf{f}_t$ are supposed to have a widespread effect, as they heterogeneously affect every country in the sample. ε_{it} is an idiosyncratic error term. Pesaran (2006) considers the case of i.i.d erros while Pesaran and Tosetti (2011) focus on the more general case of a multifactor error structure and spatial error correlation. Combining (2) and (3), we obtain the following:

$$y_{it} = \alpha_i^{'} \mathbf{d}_t + \beta_i^{'} \mathbf{x}_{it} + \gamma_i^{'} \mathbf{f}_t + \varepsilon_{it}. \tag{4}$$

This model cannot be estimated using traditional panel data estimators due to unobservability of common factors $\mathbf{f}_t$. Pesaran (2006) suggests the Common Correlated Effects (CCE) estimation procedure to deal with that issue. CCE consists of approximating the linear combination of the unobserved factors by cross-sectional averages of the dependent and explanatory variables, and then running standard panel regressions augmented with these cross-sectional averages.

CCE estimator can be motivated as follows. The idiosyncratic errors ε_{it} in Eq. (3) are assumed to be independently distributed over $(\mathbf{d}_t, \mathbf{x}_{it})$, whereas the unobserved factors $\mathbf{f}_t$ can be correlated with the observed variables $(\mathbf{d}_t, \mathbf{x}_{it})$. This correlation is allowed by modeling the explanatory variables as linear functions of the observed common factors $\mathbf{d}_t$ and the unobserved common factors $\mathbf{f}_t$:

$$\mathbf{x}_{it} = \mathbf{A}_i' \mathbf{d}_t + \mathbf{\Gamma}_i' \mathbf{f}_t + \mathbf{v}_{it}, \tag{5}$$

where $\mathbf{A}_i$ and $\mathbf{\Gamma}_i$ are $l \times 3$ and $m \times 3$ factor loading matrices, and $\mathbf{v}_{it} = (v_{i1t}, v_{i2t}, v_{i3t})'$. $\mathbf{v}_{it}$ is assumed to be distributed independently of ε_{it} and is allowed to be serially correlated, and cross-sectionally weakly correlated.

Combining Eqs. (4) and (5), we get the following system of equations

$$z_{it} = \begin{pmatrix} y_{it} \\ x_{it} \end{pmatrix} = \mathbf{B}_i' \mathbf{d}_t + \mathbf{C}_i' \mathbf{f}_t + \xi_{it} \tag{6}$$

where

$$\mathbf{B}_i = (\alpha_i \ \mathbf{A}_i) \begin{pmatrix} 1 & 0 \\ \beta_i & \mathbf{I}_3 \end{pmatrix}, \mathbf{C}_i = (\gamma_i \ \mathbf{\Gamma}_i) \begin{pmatrix} 1 & 0 \\ \beta_i & \mathbf{I}_3 \end{pmatrix}, \text{ and } \xi_{it} = \begin{pmatrix} \varepsilon_{it} + \beta_i' \mathbf{v}_{it} \\ \mathbf{v}_{it} \end{pmatrix}$$

Using sample cross-sectional averages, Eq. (6) can be written as

$$\overline{z}_t = \overline{\mathbf{B}}' \mathbf{d}_t + \overline{\mathbf{C}}' \mathbf{f}_t + \overline{\xi}_t \tag{7}$$

where

$$\overline{\mathbf{z}}_t = \frac{1}{N} \sum_{i=1}^{N} \mathbf{z}_{it}, \ \overline{\mathbf{B}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{B}_i, \ \overline{\mathbf{C}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{C}_i, \text{ and } \overline{\xi}_t = \frac{1}{N} \sum_{i=1}^{N} \xi_{it}$$

Following Pesaran (2006), we can premultiply both sides of Eq. (7) by $\overline{\mathbf{C}}$ and solve for $\mathbf{f}_t$. We get

$$\mathbf{f}_t = \left(\overline{\mathbf{C}} \, \overline{\mathbf{C}}' \right)^{-1} \overline{\mathbf{C}}' \left(\overline{z}_t - \overline{\mathbf{B}}' \mathbf{d}_t - \overline{\xi}_t \right). \tag{8}$$

It is possible to show that $\overline{\xi}_t$ converges to 0 in quadratic mean as $N \to \infty$ (Pesaran and Tosetti, 2011). Accordingly, it can be shown that

$$\mathbf{f}_t - \left(\overline{\mathbf{C}} \, \overline{\mathbf{C}}' \right)^{-1} \overline{\mathbf{C}}' \left(\overline{z}_t - \overline{\mathbf{B}}' \mathbf{d}_t \right) \stackrel{\text{q.m.}}{\longrightarrow} 0, \text{ as } N \to 0 \tag{9}$$

or, put differently, the unobservable common factors, $\mathbf{f}_t$, can be well approximated by a linear combination of observed common factors $\mathbf{d}_t$, the cross-sectional averages of the dependent variable, $\overline{y}_t$, and those of the country-specific regressors, $\overline{x}_t$. Two alternative estimators have been proposed in the literature: the CCE Mean Group (CCEMG) estimator and the CCE Pooled (CCEP) estimator. It has been shown that CCE estimators yield consistent estimates under a large variety of situations (Kapetanios and Pesaran, 2009; Pesaran and Tosetti, 2011; Chudik et al., 2011). Moreover, small sample properties of CCE estimators have also been investigated in various papers (see, among others, Coakley et al., 2002; Kapetanios and Pesaran, 2009; Chudik et al., 2011; Westerlund and Urbain, 2015). More specifically, these papers compare the small sample properties of CCE estimators to their competitors, i.e. estimators based on principal components (PC) (Coakley et al., 2002; Bai, 2009), and show that, although the PC estimates of factors are more efficient than the cross-sectional averages, the CEE estimators of slope coefficients generally perform the best. To conclude, a significant advantage of CCE estimators is that they do not require a priori knowledge of the number of unobserved common factors.

2.2 Alternative nonparametric approach

Sieve estimation Recently, Su and Jin (2012) consider a panel data model that extends the multifactor linear specification proposed by Pesaran (2006). Specifically, Su and Jin (2012) consider the following panel data model, which allows for a nonparametric relation between the dependent variable and the regressors, while the common factors enter the model in a parametric way,

$$y_{it} = \alpha_i^{'}\mathbf{d}_t + g_i\left(\mathbf{x}_{it}\right) + \gamma_i^{'}\mathbf{f}_t + \varepsilon_{it}, \tag{10}$$

where $g_i(.)$ are unknown smooth continuous functions (heterogenous case). In the homogenous case, $g_i(.) = g(.)$, for all $i = 1, 2, \ldots, N$. For identification purposes, the following condition is necessary,

$$E(g_i\left(\mathbf{x}_{it}\right)) = 0.$$

Su and Jin (2012) extend CCE approach to the estimation of heterogenous panel data model (10). First, following Pesaran (2006), they proxy the unobservable common factors $\mathbf{f}_t$ in (10) by the cross-sectional averages $\overline{\mathbf{z}}_t = N^{-1}\sum_{j=1}^{N}\mathbf{z}_{jt}$, where $\mathbf{z}_{it} = [y_{it}, \mathbf{x}_{it}']'$. Second, they approximate the nonparametric part of the model, $g_i\left(.\right)$, using sieve approximation.

Sieve approximation proceeds as follows. First, we must choose an infinite sequence of known basis functions, we denote by $\{\pi_l(x), l = 1, 2, \ldots\}$, that can approximate any square-integrable function of x very well. Different choices are possible, including spline approximation (see below). Second, the order of approximation must be defined. Let K denote this order that is a function of T when estimating the heterogenous model with $g_i\left(.\right)$, or of N and T when estimating the homogenous model with $g\left(.\right)$. This integer number will tend to infinity as $N \to \infty$ (heterogenous case), or $(N, T) \to \infty$ (homogenous case). Third, under fairly weak conditions, we can approximate the unknown function very well by a linear combination of the K first elements of the chosen basis, or $\pi^K(x) = (\pi_1(x), \pi_2(x), \ldots, \pi_K(x))^{'}$, i.e.

$$g_i\left(.\right) \approx \ \delta_{g_i}{'}\pi^K(x) \text{ (heterogenous case), or } g\left(.\right) \approx \delta_g{'}\pi^K(x) \text{ (homogenous case)}$$

Finally, to estimate δ_{g_i}, we run the regression

$$y_{it} = \alpha_i^{'}\mathbf{d}_t + \delta_{g_i}{'}\pi^K(x) + \psi_i^{'}\overline{\mathbf{z}}_t + u_{it}, \tag{11}$$

or, to estimate δ_g, the regression

$$y_{it} = \alpha_i^{'}\mathbf{d}_t + \delta_g{'}\pi^K(x) + \psi_i^{'}\overline{\mathbf{z}}_t + u_{it}, \tag{12}$$

Su and Jin (2012) show that the extended CCE estimators of both the heterogenous and homogenous regression functions are consistent as N and T tend to infinity, and establish asymptotic normality of these estimators.

Nonparametric specifications In our empirical framework, we consider two alternative specifications where $\mathbf{x}_{it}$ enter the model nonparametrically. Because of the relatively small time timension, we restrict our analysis to the the homogenous case, $g_i(.) = g(.)$ (see Su and Jin, 2012, p. 41) and propose two alternative specifications. The first specification assumes an additive structure of $g(.)$, as follows:

$$\log f_{it} = \alpha_i + \phi(\log S_{it}^d) + \xi(\log S_{it}^f) + \psi(\log H_{it}) + \gamma_i^{'}\mathbf{f}_t + \varepsilon_{it}, \tag{13}$$

where $\phi(.)$, $\xi(.)$ and $\psi(.)$ are unknown univariate smooth continuous functions of interest.

The second specification assumes instead a non-additive structure of $g(.)$, i.e.

$$\log f_{it} = \alpha_i + g(\log S_{it}^d, \log S_{it}^f, \log H_{it}) + \gamma_i^{'}\mathbf{f}_t + \varepsilon_{it}. \tag{14}$$

Relaxing additivity may suffer of the curse of dimensionality but, at the same time, may allow to detect relevant intercation effects, which are not allowed in the additive specification.

Thin plate regression splines Su and Jin (2012) estimate the nonparametric component of the model using sieves, and particularly splines, as they typically provide better approximations (see, *e.g.*, Hansen, 2014). Following Su and Jin (2012), we adopt a regression splines (RS) framework. We also employ penalized regression splines (PRS), as they combine the features of both regression splines, which use less knots than data points but do not penalize roughness, and smoothing splines, which control the smoothness of the fit through a penalty term but use all data points as knots. PRS have proven to be useful empirically in many aspects (see, e.g. Ruppert et al., 2003) and, in recent years, their asymptotic properties have been studied and then connected to those of regression splines, to those of smoothing splines and to the Nadaraya - Watson kernel estimators (Claeskens et al., 2009; Li and Ruppert, 2008). In this work, for both RS and PRS we use thin plate regression splines (TPRS), which are introduced by Wood (2003). Since TPRS have been developed and mostly adopted in the statistical science, a short introduction of them is in order.

Consider the generic problem of finding the smooth function g of $y = g(\mathbf{x}) + \epsilon$ from n observations, where $\mathbf{x}$ is a vector of d variables. Thin plate splines (TPS) can be employed to estimate g by finding the function $\hat{g}$ that minimizes the quantity

$$||\mathbf{y} - \boldsymbol{\chi}|| + \lambda J_{md}(\chi), \tag{15}$$

where $\mathbf{y}$ and $\boldsymbol{\chi}$ are n-dimentional vectors of the y_i and $\chi(\mathbf{x_i})$, $i = 1, 2, ..., n$, respectively. $|| \cdot ||$ is the Euclidean norm. $J_{md}(\chi)$ is a penalty functional that is related to the order m of differentiation in J_{md} and to the dimension d, as described below.

$$J_{md}(\chi) = \int_{\mathbb{R}^d} \sum_{\nu_1+\dots+\nu_d=m} \frac{m!}{\nu_1! \dots \nu_d!} \left(\frac{\partial^m \chi}{\partial x_1^{\nu_1} \dots \partial x_d^{\nu_d}} \right)^2 dx_1 \dots dx_d.$$

It is proven that the function that minimizes the expression above is of the form

$$\hat{g}(\mathbf{x}) = \sum_{i=1}^{n} \delta_i \eta_{md}(||\mathbf{x} - \mathbf{x}_i||) + \sum_{j=1}^{M} \alpha_j \pi_j(\mathbf{x}), \tag{16}$$

under the constraint that $\boldsymbol{T'\delta} = \mathbf{0}$, $T_{ij} = \pi_j(\mathbf{x}_i)$. $\boldsymbol{\delta}$ and $\boldsymbol{\alpha}$ are vectors of unknown parameters and π_j, $j = 1, 2, ..., M$, are $M = \binom{m+d-1}{d}$ are linearly independent polynomials of degree less than m that span the $\mathbb{R}^d$ space. η_{md} is a specific function associated with m and d (see Wood, 2003). Then, (15) translates to minimizing with respect to $\boldsymbol{\delta}$ and $\boldsymbol{\alpha}$

$$||\boldsymbol{y} - \mathbf{E}\boldsymbol{\delta} - \boldsymbol{T\alpha}||^2 + \lambda \boldsymbol{\delta}' \mathbf{E} \boldsymbol{\delta}, \text{ subject to } \boldsymbol{T}'\boldsymbol{\delta} = \mathbf{0}, \tag{17}$$

where $\mathbf{E}$ is the matrix with elements $E_{ij} = \eta_{md}(||\mathbf{x}_i - \mathbf{x}_j||)$, $i, j = 1, 2, ..., n$.

In contrast to typical RS and PRS, the estimation of g using TPS does not require the choice of knots or the selection of basis functions. Moreover, TPS do not impose any restriction in the number of predictor variables and allow some flexibility to the selection of m. Nevertheless, TPS are not computationally attractive because, as implied by (16) and (17), except for the case when $d = 1$, they require the estimation of as many parameters as the number of data points n.

To overcome this computational difficulty, Wood (2003, 2017) starts from the smoothing problem (17) and truncates the space of the components with parameters $\boldsymbol{\delta}$, which are the ones associated with the wiggliness of the spline. Following Wood (2017), let $\mathbf{E} = \mathbf{UDU}'$ be the eigen-decomposition of $\mathbf{E}$, where $\mathbf{D}$ is a diagonal matrix of eigenvalues of $\mathbf{E}$ such that $|D_{i,i}| \geq |D_{i-1,i-1}|$ and $\mathbf{U}$ the corresponding eigenvectors. Denote by $\mathbf{U}_k$ the matrix of the first k columns of $\mathbf{U}$ and by $\mathbf{D}_k$ the upper left $k \times k$ submatrix of $\mathbf{D}$. Restrict $\boldsymbol{\delta}$ to the column space of $\mathbf{U}_k$, so that $\boldsymbol{\delta} = \mathbf{U}_k\boldsymbol{\delta}_k$, where $\boldsymbol{\delta}_k$ is a k-dimensional vector with $k > M$. Then, within the space spanned by $\mathbf{U}_k$, problem (17) is replaced by minimizing

$$||\boldsymbol{y} - \mathbf{U}_k\mathbf{D}_k\boldsymbol{\delta}_k - \boldsymbol{T\alpha}||^2 + \lambda\boldsymbol{\delta}_k'\mathbf{D}_k\boldsymbol{\delta}_k, \text{ subject to } \boldsymbol{T}'\mathbf{U}_k\boldsymbol{\delta}_k = \mathbf{0}. \tag{18}$$

Having fitted (18), the spline is evaluated from (16) after estimating $\boldsymbol{\delta}$ from $\boldsymbol{\delta}_k$.

It is worth to note that while TPS are optimal with respect to minimizing (15), the low rank smoothers resulting from the truncation described above do not inherit such an optimality property. Moreover, this low rank approximation would be ideal only if, for any given $\boldsymbol{\delta}$, it would result in minimum change in the goodness of fit and, simultaneously, in the penalty term. Nevertheless, no single basis can achieve the above for all $\boldsymbol{\delta}$. This fact raises the need to define a way by which (17) is approximated by (18). Wood (2003, 2017) proposes an approach that is associated to minimizing the largest possible change of the goodness of fit, that is $\hat{e}_k = \max_{\boldsymbol{\delta}=\mathbf{0}} \frac{||(\mathbf{E}-\hat{\mathbf{E}}_k)\boldsymbol{\delta}||}{||\boldsymbol{\delta}||}$, as well as minimizing the largest change in wiggliness, that is $\tilde{e}_k = \max_{\boldsymbol{\delta}=\mathbf{0}} \frac{\boldsymbol{\delta}'(\mathbf{E}-\tilde{\mathbf{E}}_k)\boldsymbol{\delta}}{||\boldsymbol{\delta}||^2}$. In these quantities, $\hat{\mathbf{E}}_k = \mathbf{E}\mathbf{U}_k\mathbf{U}_k'$ and $\tilde{\mathbf{E}}_k = \mathbf{U}_k'\mathbf{U}_k\mathbf{E}\mathbf{U}_k\mathbf{U}_k'$. Further, Wood (2003) shows that the choice of $\mathbf{U}_k$ as the truncated basis for $\boldsymbol{\delta}$ minimizes simultaneously both $\hat{e}_k$ and $\tilde{e}_k$. This approximation that also considers the minimization criteria of $\hat{e}_k$ and $\tilde{e}_k$ results in the definition of the TPRS.

Since our explanatory variables have different units, in the case of the non-additive specification (14), we avoid isotropy by considering a tensor product basis, which is constructed by assigning TPRS as the basis for the marginal smooth of each covariate and then creating their Kronecker product. The tensor product smooths are invariant to the linear rescaling of covariates, and for this reason, they are appropriate when the arguments of a smooth have different units (Wood, 2006). Finally note that in the PRS framework, the smoothing parameter is selected by the restricted maximum likelihood (REML) estimation, which, relative to other approaches, is less likely to develop multiple minima or to undersmooth at finite sample sizes (see, e.g. Reiss and Todd Ogden, 2009).

3 Results

In this section, we compare the results from the estimation of the three aforementioned specifications:[1]

$$\begin{aligned}
\log f_{it} &= \alpha_i + \theta \log S^d_{it} + \gamma \log S^f_{it} + \delta \log H_{it} + \gamma_i' \mathbf{f}_t + \varepsilon_{it}, \\
\log f_{it} &= \alpha_i + \phi(\log S^d_{it}) + \xi(\log S^f_{it}) + \psi(\log H_{it}) + \gamma_i' \mathbf{f}_t + \varepsilon_{it}, \text{ and} \\
\log f_{it} &= \alpha_i + g(\log S^d_{it}, \log S^f_{it}, \log H_{it}) + \gamma_i' \mathbf{f}_t + \varepsilon_{it}.
\end{aligned}$$

We use data on a balanced panel of 24 OECD countries observed over the period 1971-2004 from Coe et al. (2009) and revisited in Ertur and Musolesi (2017). Total factor productivity, f_{it}, is measured as the log of output minus a weighted average of labor and capital inputs using factor shares as weights. S^d_{it} is total domestic R&D capital stock computed using perpetual inventory procedure. S^f_{it} is foreign R&D capital stock defined as the weighted average of S^d_{jt}, $j \neq i$, using bilateral imports as weights Lichtenberg and van Pottelsberghe de la Potterie (1998), i.e.

$$S^f_{it} = \sum_{j \neq i} (M_{ijt}/Y_{jt}) S^d_{jt}$$

with M_{ijt} = country i's imports of good and services from country j and Y_{jt} = nominal GDP of country j. H_{it} is the stock of human capital used by Ertur and Musolesi (2017), which is obtained starting from the average years of schooling provided by Barro and Lee (2000).

3.1 Out-of-sample comparison

To compare the aforementioned specifications, we perform a pseudo Monte Carlo experiment. In particular, along the lines depicted by Racine and Parmeter (2014), Ma et al. (2015) and Delgado et al. (2014), using similar macro panel data variables related to economic growth, the observations are randomly shuffled at 90% into training points and at 10% into evaluation points. Each model is fitted according to the training sample. Then, the average out-of-sample squared prediction error (ASPE) is computed using the evaluation sample. The above steps are repeated a large number of times $B = 1000$, so that a $B \times 1$ vector of prediction errors is created for each model.[2]

According to the statistical literature dealing with apparent versus true error (see e.g. Efron, 1982), the true error is associated with out-of-sample measures of fit, contrasted to the apparent error, which is associated with within sample measures. Typically, the latter is smaller than the former and frequently overly optimistic. The method proposed by Racine and Parmeter

[1]The parametric specification is estimated using CCEP estimator already coded in R within the package plm while the nonparametric specifications are estimated by exploiting the R package mgcv.

[2]See also Baltagi et al. (2003) who contrast the out-of-sample forecast performance of alternative parametric panel data estimators.

(2014) is linked to cross validation (CV), in the original formulation of which a regression model fitted on a randomly selected first half of the data was used to predict the second half. The division into equal halves is not necessary. For instance, a common variant is the leave-one-out CV, which fits the model to the data excluding one observation each time and then predicts the remaining point. The average of the prediction errors is the CV measure of the true error. As highlighted in Racine and Parmeter (2014), the method can provide significant power improvements over existing single-split techniques.

FIGURE 1

Figure 1 presents the box-and-whisker plots of the ASPE distributions for the different specifications. A first relevant result is that the median that corresponds to the parametric model is the largest among the different specifications, while the non-additive penalized model has the smallest median. In particular, the median ASPEs of the non-additive penalized model relative to the other models – the parametric, the additive unpenalized, the additive penalized and the non-additive unpenalized – is 0.6023, 0.9284, 0.9409 and 0.8278, respectively. A second interesting result is that the penalized regression modeling has a smaller median ASPE than its unpenalized counterpart for both additive and non-additive specifications. However, although when imposing an additive structure, the two approaches provide quite similar performances, the gain in terms of predictive ability from using PRS over RS is extremely pronounced when estimating the non-additive specification, which typically suffers more from the curse of dimensionality problem. Also, it is worth noting that within the RS framework, the additive specification provides a better performance than the non-additive one.

FIGURE 2

Next, figure 2 shows the empirical cumulative distribution functions of the ASPEs for each model. Clearly, the ASPE of the non-additive penalized model is stochastically dominated by the ASPE of any of the remaining models. This indicates that the non-additive penalized model outperforms all others in terms of predictive ability. It is also evident that the parametric model underperforms with respect to the nonparametric ones.

Finally, we compare the different specifications using the test of revealed performance (TRP) proposed by Racine and Parmeter (2014). The TRP involves estimating the distribution of the true errors for the different models and testing whether their expectations are statistically different. The results of these paired t-tests are presented in Table 1. In all cases, the null hypothesis that the difference in means of the ASPEs is zero is rejected. Thus, the tests complement the above presented results, indicating that this difference is statistically significant in all cases.

TABLE 1

In summary, these results clearly indicate that the parametric specification underperforms with respect to the nonparametric ones. This is a similar results than Ma et al. (2015) who use a similar macro panel data set. As far as nonparametric specifications are concerned, PRS always perform better than unpenalized RS. The improvement achieved when using PRS is much more pronounced when focusing on the nonadditive specification where RS suffer more of the curse of dimensionality problem while PRS apper to be extremely efficient. While there exists a number of studies comparing alternative spline methods by using Monte Carlo simulations (see e.g. Nie and Racine, 2012; Wood, 2003, 2006), to the best of our knowledge, this is the first paper constrasting PRS and RS in terms of their predictive ability and may provide some guidance for future works. The nonadditive specification is indeed the best one when using PRS while with unpenalized RS, the best model is the one with additive smooth terms. These results thus suggest adopting a nonparametric nonadditive specification and that PRS are more efficient than their unpenalized counterparts, especially for nonadditive specifications when the curse of dimensionality is a concern.

3.2 Estimation results

In this subsection, we present the main estimation results and specifically focus attention on the nonparametric specifications. These results have relevant implication for public policies. We only consider PRS, since they outperform their unpenalized counterparts. We first provide the results obtained using the additive specification (13) because, due to the additive structure, the results are directly comparable to those ones of the parametric specifications adopted in previous studies. Then, we present the results of the non-additive specification (14), which, according to our findings, provides the best predictive performance and thus is more suitable to approximate the underlying DGP.

The results concerning the nonparametric part of the additive specification are presented in figure 3. The three graphs depict the estimated univariate smooth functions. Following Marra and Wood (2012), the estimated smooths are shown with confidence intervals that include the uncertainty about the overall mean, which provides better coverage performance. We also computed the p-values for smooth terms using a Wald test statistic that is motivated by an extension of Nychka (1988) analysis of the frequentist properties of Bayesian confidence intervals for smooths as suggested by Wood (2012). These are p-values associated with Wald test that the whole function equals zero. Low p-values indicate low likelihood that the splines of the function are jointly zero. Also note that smooths are subject to sum-to-zero identifiability constraints as detailed in Cardot and Musolesi (2019).

FIGURE 3

All the estimated smooths appear to be highly significant, with extremely low p-values associated with the Wald test. Moreover, using an approximate ANOVA test procedure (see Wood, 2017), the parametric model is strongly rejected in favour of model (13). It is worth mentioning that because the response as well as the explanatory variables are in logs, the slope of the estimated smooth functions represents the estimated elasticity, which are depicted in the bottom panel of 3. The first plot shows the effect of domestic R&D on TFP. It appears that for low values of R&D, where data are sparse and large confidence interval bands are present, the relation is flat. Then, for intermediate values of domestic R&D, the function is monotonic increasing, with a steep rise in approximately the last two deciles. The policy implications resulting from the above are clear: an increase in domestic R&D has an effect on productivity only above a threshold, thus suggesting that a critical mass of investments in R&D is crucial for R&D to become effective. After this threshold, the estimated output elasticity becomes positive and increases even more for very high levels of domestic R&D. This can be seen as a refinement of the results of the existing empirical literature on R&D spillovers, which is based on parametric models and generally distinguishes between G7 and non-G7 countries. Indeed, Ertur and Musolesi (2017), employing the CCE approach, show that the estimated output elasticity of domestic R&D is positive and significant for G7 countries, while it is non-significant for non-G7 countries. Similar results are also found by Coe et al. (2009), who adopt the dynamic OLS for cointegrated panels, and by Barrio-Castro et al. (2002), who use a standard fixed effects approach.

The second graph shows the effect of foreign R&D on TFP. Again, for low levels of the variable, data are sparce, making it difficult to identify a clear pattern. Then, the relation is positive and roughly concave for intermediate values, while it becomes flat for high levels of foreign R&D. The results show that an increase in foreign R&D affects TFP positively, but only up to a certain level. They complement previous empirical literature such as Coe et al. (2009), who indicate that trade-related foreign R&D is a significant determinant of TFP. More specifically, our findings improve the results of Ertur and Musolesi (2017), among others, who find a small, positive and significant effect of R&D on TFP in non-G7 countries, but no significant effect in the case of the G7. Nevertheless, in all previous studies, the linearity assumption obscures the fact that the output elasticity of foreign R&D is not constant but varies with respect to the different levels of foreign R&D. Indeed, looking at the bottom panel of figure 3, it can be seen that the estimated elasticity constantly decreases over the range of foreign R&D up to a level where it becomes not significantly different from zero.

The third graph in figure 3 depicts the effect of human capital on TFP. It again shows scarce data and large confidence bands for low levels. Then, the relation between human capital and TFP is approximately flat for intermediate values, while for high values, it seems to be monotonic increasing, with a steep rise in approximately the last two deciles. In terms of policy perspectives, the results suggest a threshold that occurs at very high levels of human capital, above which the estimated elasticity becomes positive. Investing in human capital becomes effective only after a certain level is reached. These findings add new insights to Ertur and Musolesi (2017), who find no significant effect of human capital on TFP for both G7 and non-G7 countries and explain their result on the grounds that the quantity of education no longer has a significant effect when omitted variable bias is addressed. We find confirmation of such results for most of the domain of human capital, but we also show that allowing for nonlinearity in the relation between human capital and TFP is crucial in order to highlight a positive effect for the highest levels of human capital.

FIGURE 4

Next, we turn to the estimates of the non-additive specification. Also, in this case, the estimated (multivariate) smooth function appears to be highly significant. Again, an approximate ANOVA test procedure is used and the the model with additive smooth functions (13) is rejected in favour of model with a trivariate smooth function (14). Interestingly, the results of this test confirm the out-of-sample comparison.

In particular, we focus attention on the joint effect of domestic and foreign R&D stocks on TFP. This because the effect of these two variables is the main focus in this literature. Indeed, the inclusion of human capital has been motivated not only because it affects productivity and the ability of firms to absorb information but also because it is likely to be correlated with R&D and estimating the model without human capital should bias the coefficient associated with R&D upward. In some previous studies (see e.g. Barrio-Castro et al., 2002; Frantzen, 2000; Engelbrecht, 1997), within the parametric framework under the linearity and additiity assumptions, this bias has been estimated to be approximately 20% to 30%. The results are presented in figure 4, which shows the joint effect of the R&D variables on TFP for a level of human capital fixed to the first, fifth (the median) and ninth decile. As depicted in the first graph, for low levels of human capital and irrespective of the level of domestic R&D, foreign R&D has almost no effect on TFP. In terms of policy implications, these findings suggest that foreign R&D spillovers cannot be effective if the level of human capital in a country remains low. Moreover, the effect of domestic R&D on TFP seems not to be linked to the level of foreign R&D, which implies an additive pattern when the level of human capital is low. Similar to the additive model presented above, there is a threshold above which domestic R&D becomes effective.

The second and third graphs in figure 4 show the effect on TFP when human capital is fixed to the median and to the ninth decile, respectively. The results in both graphs suggest a complementarity between domestic R&D and foreign R&D. Indeed, while for low levels of domestic R&D, the effect of foreign R&D on TFP is low, and vice versa, domestic and foreign R&D become more and more effective when the levels of both domestic and foreign R&D are increasing. These findings have interesting policy implications; in countries with intermediate or high levels of human capital, investments in R&D are not very effective if the level of foreign R&D is low. Further, the benefits of foreign R&D spillovers cannot be exploited unless both human capital and domestic R&D are above a critical mass. The above results contrast with results from some previous studies such as in Coe et al. (2009), who report that their estimations considering interactions between human capital and domestic and foreign R&D do not yield correctly signed and significant results.

In summary, these results clearly suggest that parametric estimates conceal an important part of the story and that the underlying DGP is not correctly approximated by adopting the classical parametric framework. The relationship under study is is not linear and presents relevant thresholds and complex interactions, which are better modeled by adopting a nonparametric approach.

4 Concluding remarks

In this paper, we revisit the analysis of international technology diffusion. Because the high degree of uncertainty surrounding the DGP and the likely presence of nonlinearities and latent common factors, we consider alternative nonparametric panel specifications which extend the Common Correlated Effects approach and then contrast the out-of-sample performance of them with those of more common parametric models. To do so, we adopt an approach recently proposed within the literature of nonparametric regression. This approach is based on a pseudo Monte Carlo experiment that takes its roots on cross validation. Fitted econometric models are view as approximations and the goal is to test whether one approximate model performs better than another on data drawn from the same DGP. While it is common adopting a sample-splitting mechanism whereby one splits the full sample into two sub-samples where one sub-sample is used for estimation and the other for out-of-sample evaluation, to avoid that the results reflect a particular division of the data into two sub-samples, the main idea by Racine and Parmeter (2014) is to repeat this process a large number of times becausee this can provide significant power improvements over existing single-split techniques.

We first show that a shift from a parametric to a nonparametric framework provides a significant improvement in terms of predictive ability. Moreover, we also documented that penalized regression splines perform significantly better than their unpenalized counterparts, especially in the case of a non-additive model, when the curse of dimensionality is a concern. To the best of our knowledge this is the first paper contrasting penalized and unpenalized regression splines in terms of their predctive ability.

Turning to the estimation results, our findings suggest the presence of threshold effects and nonlinearities, while the adoption of a non-additive specification provides further insights into the interactions among explanatory variables without imposing any parametric restrictions and definitively indicating that a critical mass of human capital is necessary to benefit from R&D spillovers and to observe an interactive effect between domestic and foreign R&D. In general, our findings strongly highlight that the presence of nonlinearities and complex interactions is an important feature of the data; these are obviously hidden within a parametric framework and have relevant implications for policy.

Finally, it is worth mentioning that a further extension of the present study may account for heterogeneous relations across countries. Given the relatively small time dimension, such an extension is outside the realm of the nonparametric estimators presented in this paper, where heterogeneity is addressed by adopting a Mean Group approach, and could be accomplished, for instance, by resorting to Bayesian modeling (Kiefer and Racine, 2017; Parmeter and Racine, 2018) to address the additional curse of dimensionality problem raised by heterogeneity.

References

Andrews, D. W. K. (2005). Cross section regression with common shocks. *Econometrica*, 73:1551–1585.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279.

Baltagi, B. H., Bresson, G., Griffin, J. M., and Pirotte, A. (2003). Homogeneous, heterogeneous or shrinkage estimators? some empirical evidence from french regional gasoline consumption. *Empirical Economics*, 28(4):795–811.

Baltagi, B. H., Bresson, G., and Pirotte, A. (2002). Comparison of forecast performance for homogeneous, heterogeneous and shrinkage estimators: Some empirical evidence from us electricity and natural-gas consumption. *Economics Letters*, 76(3):375–382.

Baltagi, B. H., Bresson, G., and Pirotte, A. (2004). Tobin q: forecast performance for hierarchical bayes, shrinkage, heterogeneous and homogeneous panel data estimators. *Empirical Economics*, 29(1):107–113.

Barrio-Castro, T., López-Bazo, E., and Serrano-Domingo, G. (2002). New evidence on international r&d spillovers, human capital and productivity in the oecd. *Economics Letters*, 77(1):41–45.

Barro, R. and Lee, J. W. (2000). International data on educational attainment: Updates and implications. Working paper 42, Center for International Development, Harvard University.

Cardot, H. and Musolesi, A. (2019). Modeling temporal treatment effects with zero inflated semi-parametric regression models: the case of local development policies in France. *Econometric Reviews (forthcoming).*

Charlot, S., Crescenzi, R., and Musolesi, A. (2014). Econometric modelling of the regional knowledge production function in europe. *Journal of Economic Geography*, 15(6):1227–1259.

Chudik, A., Pesaran, M. H., and Tosetti, E. (2011). Weak and strong cross-section dependence and estimation of large panels.

Claeskens, G., Krivobokova, T., and Opsomer, J. D. (2009). Asymptotic properties of penalized spline estimators. *Biometrika*, 96(3):529–544.

Coakley, J., Fuertes, A.-M., and Smith, R. (2002). A principal components approach to cross-section dependence in panels.

Coe, D. T. and Helpman, E. (1995). International r&d spillovers. *European economic review*, 39(5):859–887.

Coe, D. T., Helpman, E., and Hoffmaister, A. W. (2009). International r&d spillovers and institutions. *European Economic Review*, 53(7):723–741.

de Almeida, D., Hotta, L. K., and Ruiz, E. (2018). Mgarch models: Trade-off between feasibility and flexibility. *International Journal of Forecasting*, 34(1):45–63.

Delgado, M. S., McCloud, N., and Kumbhakar, S. C. (2014). A generalized empirical model of corruption, foreign direct investment, and growth. *Journal of Macroeconomics*, 42:298–316.

Efron, B. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans*. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania 19103.

Engelbrecht, H.-J. (1997). International r&d spillovers, human capital and productivity in oecd economies: An empirical investigation. *European Economic Review*, 41(8):1479–1488.

Ertur, C. and Musolesi, A. (2017). Weak and strong cross-sectional dependence: A panel data analysis of international technology diffusion. *Journal of Applied Econometrics*, 32(3):477–503.

Frantzen, D. (2000). R&d, human capital and international technology spillovers: a cross-country analysis. *Scandinavian Journal of Economics*, 102(1):57–75.

Griliches, Z. (1998). Patent Statistics as Economic Indicators: A Survey. In Griliches, Z., editor, *R&D and Productivity: The Econometric Evidence*, pages 287–343. University of Chicago Press, Chicago.

Hall, B. H., Mairesse, J., and Mohnen, P. (2010). Measuring the returns to r&d. In *Handbook of the Economics of Innovation*, volume 2, pages 1033–1082. Elsevier.

Hansen, B. E. (2005). Challenges for econometric model selection. *Econometric Theory*, 21(1):60–68.

Hansen, B. E. (2014). Nonparametric sieve regression: Least squares, averaging least squares, and cross-validation. In Racine, J. S., Su, L., and Ullah, A., editors, *Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, pages 215–248. Oxford University Press.

Kapetanios, G. and Pesaran, M. H. (2009). Alternative approaches to estimation and inference in large multifactor panels: Small sample results with an application to modelling of asset returns. In Phillips, G. and Tzavalis, E., editors, *The refinenment of econometric estimation and test procedures: Finite sampel and asymptotic analysis*, pages 239–281. Cambridge University Press, Cambridge.

Kapetanios, G., Pesaran, M. H., and Yamagata, T. (2011). Panels with non-stationary multi-factor error structures. *Journal of Econometrics*, 160(2):326–348.

Kiefer, N. M. and Racine, J. S. (2017). The smooth colonel and the reverend find common ground. *Econometric Reviews*, 36(1-3):241–256.

Lee, G. (2006). The effectiveness of international knowledge spillover channels. *European Economic Review*, 50(8):2075–2088.

Li, Y. and Ruppert, D. (2008). On the asymptotics of penalized splines. *Biometrika*, 95(2):415–436.

Lichtenberg, F. R. and van Pottelsberghe de la Potterie, B. (1998). International r&d spillovers: a comment. *European Economic Review*, 42(8):1483–1491.

Ma, S., Racine, J. S., and Yang, L. (2015). Spline regression in the presence of categorical predictors. *Journal of Applied Econometrics*, 30:703–717.

Maasoumi, E., Racine, J. S., and Stengos, T. (2007). Growth and convergence: A profile of distribution dynamics and mobility. *Journal of Econometrics*, 136:483–508.

Marra, G. and Wood, S. N. (2012). Coverage properties of confidence intervals for generalized additive model components. *Scandinavian Journal of Statistics*, 39(1):53–74.

Nie, Z. and Racine, J. S. (2012). The crs package: Nonparametric regression splines for continuous and categorical predictors. *R Journal*, 4(2).

Nychka, D. (1988). Bayesian confidence intervals for smoothing splines. *Journal of the American Statistical Association*, 83(404):1134–1143.

Parmeter, C. and Racine, J. (2018). Nonparametric estimation and inference for panel data models.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multi-factor error structure. *Econometrica*, 74(4):967–1012.

Pesaran, M. H. and Tosetti, E. (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics*, 161(2):182–202.

Phillips, P. C. B. and Sul, D. (2003). Dynamic panel estimation and homogeneity testing under cross section dpendence. *The Econometrics Journal*, 6:217–259.

Phillips, P. C. B. and Sul, D. (2007). Bias in dynamic panelm estimation with fixed effects, incidental trends and cross section dependence. *Journal of Econometrics*, 137:162–188.

Racine, J. and Parmeter, C. (2014). Data-driven model evaluation: a test for revealed performance. In Racine, J. S., Su, L., and Ullah, A., editors, *Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, pages 308–345. Oxford University Press.

Reiss, P. T. and Todd Ogden, R. (2009). Smoothing parameter selection for a class of semiparametric linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):505–523.

Rodriguez-Poo, J. M. and Soberon, A. (2017). Nonparametric and semiparametric panel data models: Recent developments. *Journal of Economic Surveys*, 31(4):923–960.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric regression*. Number 12. Cambridge university press.

Sarafadis, V. and Robertson, D. (2009). On the impact of error cross-sectional dependence in short dynamic panel estimation. *The Econometrics Journal*, 12:62–81.

Su, L. and Jin, S. (2012). Sieve estimation of panel data models with cross section dependence. *Journal of Econometrics*, 169(1):34–47.

Westerlund, J. and Urbain, J. (2015). Cross-sectional averages or principal compoments. *Journal of Econometrics*, 185:372–377.

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(1):95–114.

Wood, S. N. (2006). Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics*, 62(4):1025–1036.

Wood, S. N. (2012). On p-values for smooth components of an extended generalized additive model. *Biometrika*, 100(1):221–228.

Wood, S. N. (2017). *Generalized additive models: an introduction with R*. CRC press.

Wood, S. N., Pya, N., and Säfken, B. (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association*, 111(516):1548–1563.

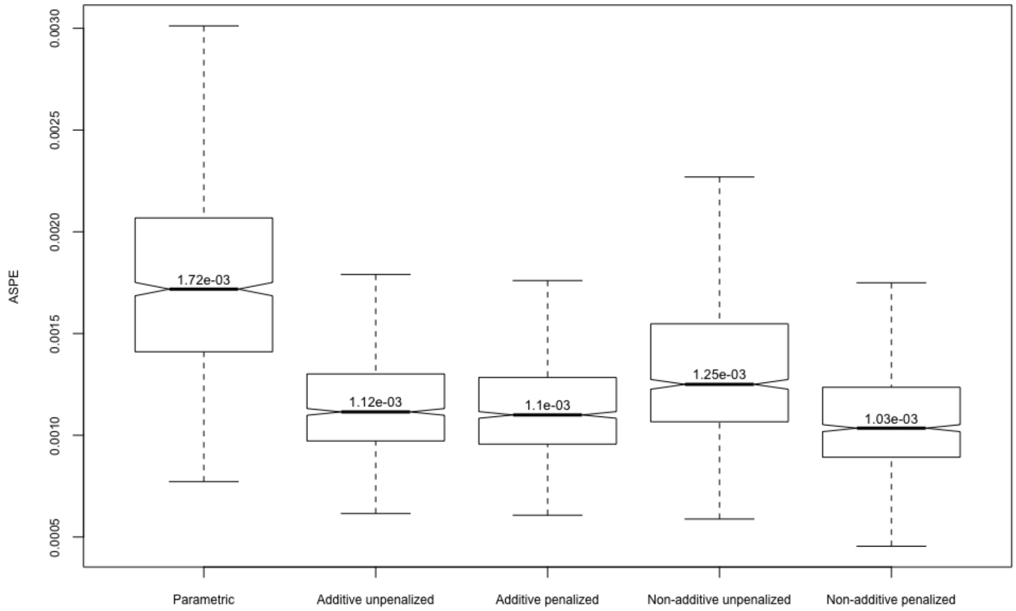

Figure 1: Out-of-sample average square prediction error (ASPE) box plots for different factor models: the parametric, the additive and the non-additive.

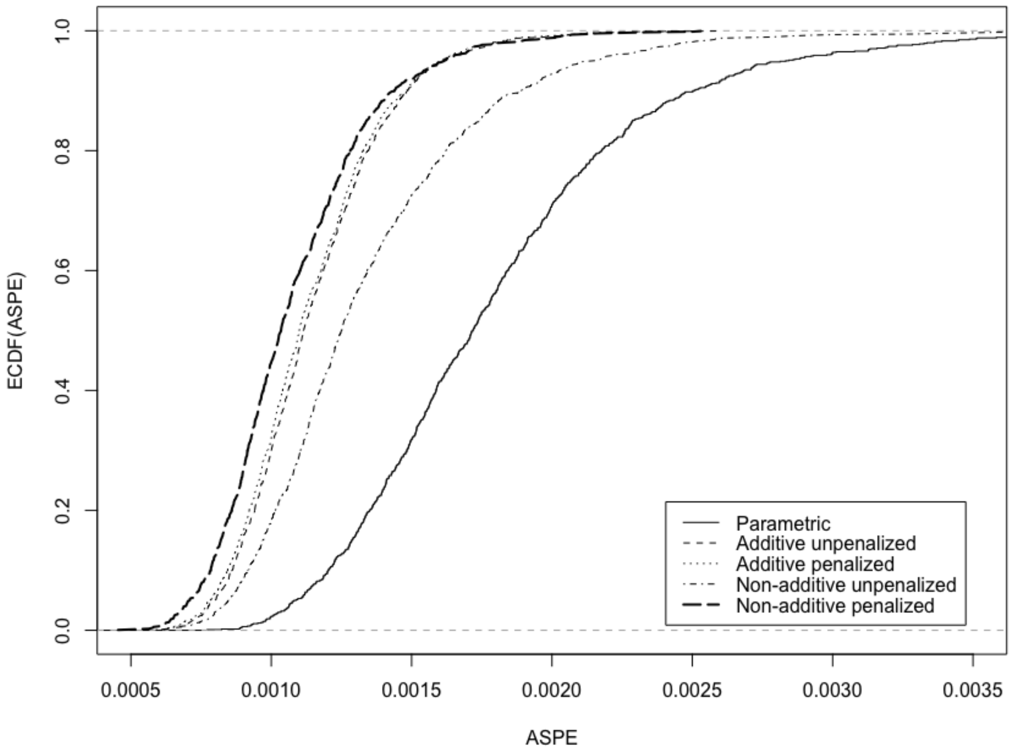

Figure 2: Empirical Cumulative Distribution Functions (ECDFs) of the ASPE for different factor models: the linear, the additive and the non-additive models for the OECD data.

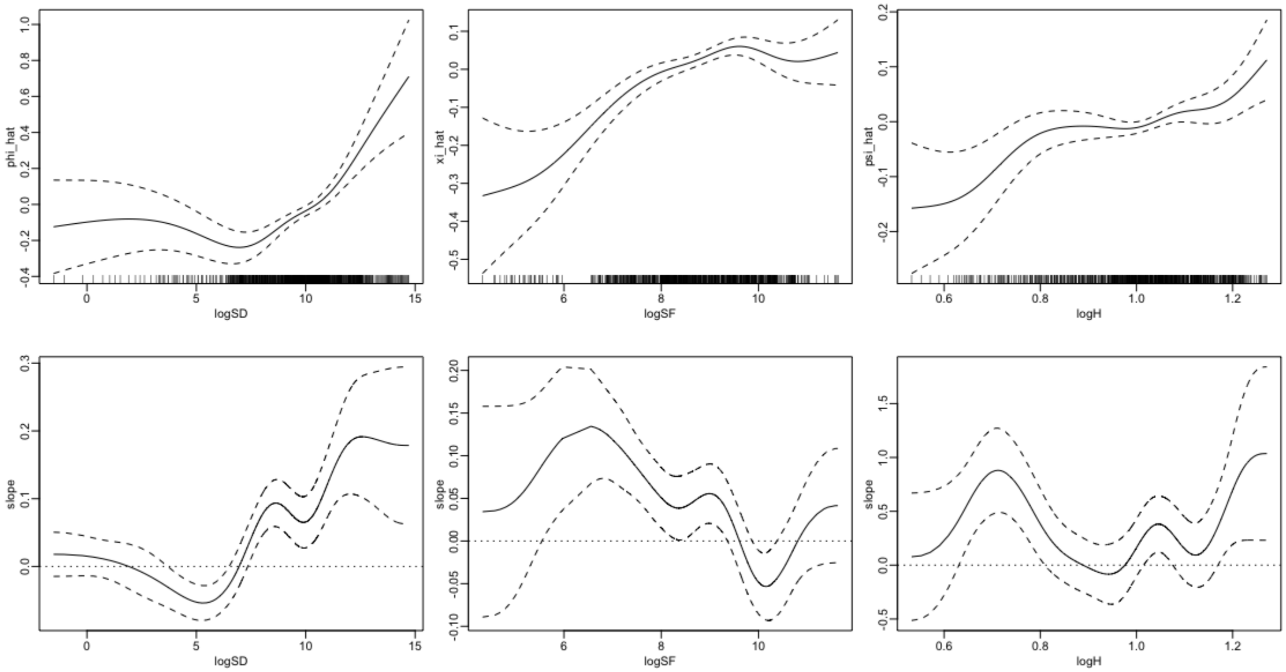

Figure 3: Additive Model. Estimated smooths (top panel) and corresponding derivatives (bottom panel) for the additive penalized regression model. Component smooths are shown with confidence intervals obtained by computing a Bayesian posterior covariance matrix.

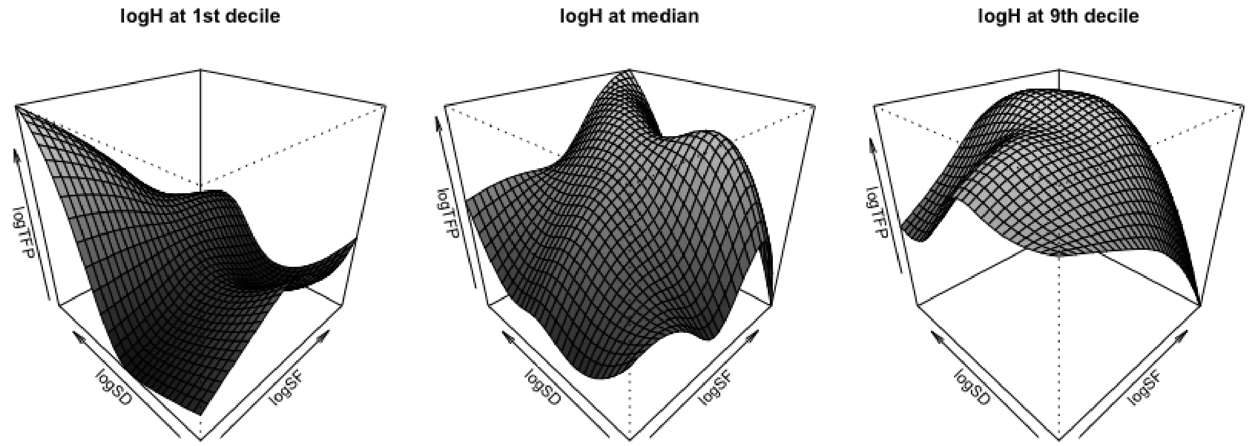

Figure 4: Non-additive model. The effect of domestic and foreign R&D on TFP for different levels of human capital. The log of human capital is fixed to the first, fifth and ninth decile, respectively.

TABLE 1 - Paired t-tests of factor models

models	Additive unpenalized	Additive penalized	Non-additive unpenalized	Non-additive penalized
Parametric	43.683***	45.461***	27.042***	47.992***
Additive unpenalized		9.849***	-18.493***	13.138***
Additive penalized			-20.492***	10.697***
Non-additive unpenalized				32.642***

Null hypothesis: The true difference in means of the ASPEs of the compared models is zero.
The training sample is 90% of the data-sample; number of resampling iterations B: 1.000
Classification of p-value: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$