Interactive R&D Spillovers: an estimation strategy based on forecasting-driven model selection

Georgios Gioldasis, Antonio Musolesi, Michel Simioni

Working Paper Series

Interactive R&D Spillovers: An estimation strategy based on forecasting-driven model selection

by

Georgios Gioldasis, Antonio Musolesi, Michel Simioni

06/2021

SEEDS is an interuniversity research centre. It develops research and higher education projects in the fields of ecological and environmental economics, with a special focus on the role of policy and innovation. Main fields of action are environmental policy, economics of innovation, energy economics and policy, economic evaluation by stated preference techniques, waste management and policy, climate change and development.

This paper is a substantially revised version of the SEEDS Working Paper 01/2020

Interactive R&D Spillovers: An estimation strategy based on forecasting-driven model selection

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, INRAe, University of Montpellier, Montpellier, France

April 16, 2021

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 extend a recently proposed data-driven model choice approach, which takes its roots on cross validation and aims at testing whether two competing approximate models are equivalent in terms of their expected true error, to the case of cross-sectionally dependent panels, by exploiting moving block bootstrap resampling methods and assessing forecasting performances of competing models. Our results indicate that the adoption of a fully nonparametric specification 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.

1 Introduction

With the development of endogenous growth theory since the nineties, there has been an increasing interest in estimating the effect of research and development (R&D) on growth and productivity (Jones, 1995). At present, however, as also stressed by Keller (2004), while only few rich countries account for most of the world's creation of new technology, for most countries, foreign technology accounts for a large part of domestic productivity. Studying international technology diffusion, which can be defined as the process by which innovation spreads within and across economies (Stoneman, 1985; Karshenas and Stoneman, 1995) and estimating its impact on domestic productivity is consequently of crucial relevance.

A pioneering empirical work by Coe and Helpman (1995), revisited by Coe et al. (2009) – henceforth CH and CHH, respectively – builds on innovation-driven endogenous growth models (Helpman, 1992) and relates total factor productivvity (TFP) not solely on domestic R&D but also on 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 (Keller, 2002; Potterie and Lichtenberg, 2001; Lee, 2006).

As for econometric estimation and testing, the most relevant issues raised by the literature are the one of heterogeneous slope parameters and the non-stationarity of the variables (Coe et al., 2009) and, more recently, the existence of cross-country dependence arising from the interactions among economic units or from the presence of latent common factors (Ertur and Musolesi, 2017).

To the best of our knowledge, however, the existing literature adopted parametric specifications and did not provide any attempt to check the ability of the CH/HH specification to correctly approximate the underlying DGP and to test whether other specifications may be more suitable, while the likely complex relation between research activity and economic performances was often recognised in closely related literatures on economics of innovation (Griliches, 1998; Hall et al., 2010; Charlot et al., 2014).

This paper aims to contribute to the existing literature by proposing an estimation strategy based on forecasting-driven model selection, which exploits recent panel data nonparametric methods and extends existing data-driven model selection procedures to the case of cross-sectionally dependent panels.

First, we adopt a nonparametric approach to avoid a possible functional form bias which may arise when estimating the standard parametric model. 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). Following Su and Jin (2012), the nonparametric component is estimated using splines. Specifically, we adopt 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).

Second, because the high degree of uncertainty surrounding the Data Generating Process (DGP) (Ma et al., 2015; Racine and Parmeter, 2014; de Almeida et al., 2018), we perform model selection by comparing the forecasting performances of some alternative models. Specifically, we compare the forecasting performances of a fully nonparametric specification of the relationship between TFP and domestic R&D, foreign R&D and human capital, with those of two competing specification: an usual log-log parametric one and a nonparametric additive one.

To do so, we build on a method 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), 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 because it can provide significant power improvements over existing single-split techniques. Drawing from a recent literature focusing on cross-sectional dependence in panels and on factor models (Gonçalves, 2011; Palm et al., 2011; Smeekes and Westerlund, 2019), we extend the data-driven model choice method proposed by Racine and Parmeter (2014) by considering a moving block bootstrap resampling scheme and assessing performance using the accuracy of forecasts at various time horizons.

The econometric analysis is conducted using new annual country-level data for 24 OECD countries from 1971 to 2014, which extends the time-coverage of the data used by CHH and Ertur and Musolesi (2017). Results clearly conclude in the superiority of a fully nonparametric specification of the relationship between TFP and domestic R&D capital, foreign R&D and human capital, over parametric and even semiparametric specifications. The nonparametric specification allows for a richer view of the resulting spillover effects, showing nonlinear effects and thresholds not previously detected, ultimately suggesting the presence of *(nonlinear) interactive spillover effects*, i.e. a situation in which the estimated output elasticity of domestic R&D (nonlinearly) increases with the amount of foreign R&D. These effects are illustrated by a detailed comparison of the results obtained with those presented in the literature or obtained by the estimation of usual parametric models.

The paper is organized as follows. Section 2 provides a short literature review focusing on the alternative channels of technology diffusion, and on the main econometric issues, highlighting finally existing research gaps. Section 3 focuses on the econometric methodology, and, more specifically, on the issues of model specification and estimation. The forecasting-driven model selection procedure is presented in section 4. Section 5 describes the data. The out-of-sample comparison of the alternative specifications as well as the estimation results are presented in section 6. Finally, section 7 concludes and additional information is provided in the appendices.

2 Literature review

Beyond its plausibility with respect to endogenous growth theory (see, e.g. Keller, 2004, p. 762), the influence of the CH/CHH model was due to its versatility in allowing for both the consideration of alternative channels of international technology diffusion and the utilisation of modern large panel data econometric methods. In this section, we first briefly review these two areas of the literature and finally highlight the research gaps that this paper tries to address.

2.1 The uncertain effect of international technological spillovers

As noted by Potterie and Lichtenberg (2001, p. 490), *"International technological spillovers have no widely accepted measures"*. According to Keller (2004), the main channels of international technology diffusion are trade, foreign direct investment (FDI), and language skills. For instance, CHH and Lichtenberg and van Pottelsberghe de la Potterie (1998) use alternative definitions of weights based on imports. Potterie and Lichtenberg (2001) focus on FDI, and Musolesi (2007) adopts a weighting scheme that takes language skills into account. Keller (2002) and Ertur and Musolesi (2017) use geographic proximity, while Spolaore and Wacziarg (2009) suggest genetic distance as a barrier to the diffusion of development.

Most previous studies focus on rent spillovers, which are spillovers that originate solely from economic transactions such as trade and FD, while analyses of technology diffusion originating from knowledge spillovers that are not necessarily embodied in specific economic transactions are rarer (see, e.g., Griliches (1979) for a discussion of rent and knowledge spillovers).

As for rent spillovers, the body of literature focusing on trade reports results that are characterized by a high degree of instability of the estimated output elasticity with respect to foreign R&D, (see Table A1 of Appendix A), as the estimated elasticity ranges from being non-significant in some cases (Kao et al., 1999; Lee, 2006) to be high in magnitude with some estimates about 0.2-0.3 (Coe et al., 2009; Engelbrecht, 1997; Barrio-Castro et al., 2002). As Fracasso and Marzetti (2013) note, analyzing spillovers generated by FDI also presents conflicting results. For instance, Potterie and Lichtenberg (2001) report that outward FDI flows are conducive to international knowledge spillovers while inward FDI flows do not have a significant effect. This result has been questioned by subsequent papers obtaining the opposite finding (Lee, 2006; Bitzer and Kerekes, 2008).

As for knowledge spillovers that are not necessarily embodied in specific economic transactions, Lee (2006) finds that disembodied channels such as bilateral technological proximity and patent citations between countries have a significant and quite substantial effect in the range of 0.15-0.18. This result is broadly the same as that obtained by Musolesi (2007), who focuses on knowledge spillovers incorporated into language skills.

The empirical literature has often allowed the impact of foreign R&D to differ between the G7 countries and the others. According to the pioneering CH/CHH studies, the effect of trade-related foreign R&D on TFP is much higher for non-G7 than for G7 countries. The same result can be found in Bitzer and Kerekes (2008), who find that non-G7 countries benefit more from inward FDI than G7 countries. By contrast, Potterie and Lichtenberg (2001) find the opposite result for outward

FDI. Ertur and Musolesi (2017) find that richer countries benefit more from geographic spillovers, that are not necessarily embodied in specific economic transactions, than poorer countries, while smaller countries benefit more from rent spillovers originating from trade.

Overall, the literature does not seem to provide a robust evidence regarding the magnitude of spillovers on TFP. Moreover, a high variability in the estimates seems to be present i) irrespective of whether spillovers are supposed to originate solely from economic transactions such as trade and FDI or they are knowledge spillovers not necessarily embodied in specific economic transactions and ii) also when researchers have allowed the effect of foreign R&D to vary between G7 and non-G7 countries. This ultimately suggests that a functional misspecification bias could be present in standard parametric CH/CHH specifications.

2.2 Econometric issues

As far as econometric estimation and testing are concerned, the most relevant issues raised by the existing literature are i) heterogeneous slope parameters; ii) integration-cointegration; and iii) cross-sectional dependence.

Heterogeneous slope parameters. The homogeneity of the slope parameters implicit in the use of a pooled version of the CH/CHH specification (see Eq.(1) below) has been questioned from both an econometric and an economic perspective. Econometrically speaking, when the DGP is characterized by heterogeneous slopes, fixed effects estimators yield consistent estimates of the mean coefficients only when the number of cross-sectional units approaches infinity. From an economic standpoint and closely related to the present topic, a theoretical justification for heterogeneous slope parameters across countries can be found in the 'new growth' literature, which argues that technology differs across countries (Brock and Durlauf, 2001; Durlauf et al., 2001). Concerning international technology diffusion in particular, some studies have challenged the assumption of a common technology by arguing that technologies are specific to particular combinations of inputs (Basu and Weil, 1998). While, some previous econometric studies have introduced some degree of heterogeneity by allowing the impact of the explanatory variables to differ between the G7 countries and the others (see, e.g. Potterie and Lichtenberg, 2001, p. 762), other studies have instead adopted methods allowing for heterogeneous slopes such as the Mean Group estimator (Pesaran and Smith, 1995). Although approaches allowing for slope-specific parameters are conceptually appealing, they may face the problem of parameter estimate instability caused by the estimation of several parameters with relatively short time series (Baltagi et al., 2002, 2004). In particular, the Mean Group estimator has asymptotic justification only for $T \to \infty$ and generally suffers from the problem of parameter estimate instability, which can in turn produce non-significant estimates. Moreover, pooled and fully heterogeneous estimators have relative advantages and disadvantages, as *"the truth probably lies somewhere in between. The parameters are not exactly the same, but there is some similarity between them"* (Maddala et al., 1997, p. 91) and, for such a reason, some previous studies have applied the shrinkage estimators described in Maddala et al. (1997) and the hierarchical Bayes approach (Hsiao et al., 1999). The former can be viewed as a compromise

between the unrealistic homogeneity assumption and unstable heterogeneous estimates, while the latter makes use of Markov Chain Monte Carlo methods via Gibbs sampling. According to Hsiao et al. (1999), this estimator is asymptotically equivalent to the Mean Group estimator but performs better in small samples.

Integration-cointegration. The large time dimension of macro-panel data makes extremely relevant the adoption of time-series procedures applied to panel data to deal with nonstationarity, spurious regression and cointegration. In this respect, some of the previous works find evidence of nonstationary variables by applying first-generation tests (Coe et al., 2009) and, after having checked for the existence of a valid cointegration relation, they applied estimation methods for cointegrated regression models in panel data, such as the dynamic OLS estimator proposed by Kao and Chiang (2001). Recently, Ertur and Musolesi (2017) provided a more nuanced and thorough picture by adopting second-generation tests decomposing the panel into deterministic, common and idiosyncratic components (Bai and Ng, 2004), suggesting that while the unobserved idiosyncratic component of the variables under study is stationary, the unobserved common factors component is nonstationary.

Cross-sectional dependence. In the last decade, with the raising process of globalization and the growing importance of the economic and social interconnections across economic agents, the issue of cross-sectional dependence (CSD) in panel data models has become of crucial relevance from both a theoretical and an empirical point of view. CSD has been typically introduced as a result of a finite number of unobservable (and/or observed) common factors or by the consideration of spatial model. Ertur and Musolesi (2017) focus on the detection of CSD and then on the estimation under CSD. They first find evidence of strong cross-sectional dependence in the data and then estimate the model by using the Common Correlated Effects (CCE) approach as it remains valid in a variety of situations that are likely to occur, such as the presence of both forms of dependence (Pesaran and Tosetti, 2011) or the existence of nonstationary factors (Kapetanios et al., 2011) and also allows for the estimation of both pooled and heterogeneous specifications.

2.3   Research gaps

In summary, the existing literature has adopted alternative measures of international technological spillovers and has focused on various methodological issues, such as allowing for heterogeneous slopes, addressing non-stationary variables and handling cross-sectional dependence. Although the likely complex relation between research activity and economic performances, which was pointed out in related literatures on economics of innovation, and the high degree of instability of the estimated output elasticity with respect to foreign R&D suggest that the traditional CH/CHH parametric specification may suffer of a functional misspecification bias, previous works did not check the ability of this specification to correctly approximate the underlying DGP or test whether more flexible specifications may be more suitable. This is the main purpose of the present work.

3 Model specifications and their estimation

3.1 The classical parametric specification

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,

$$S_{it}^f = \sum_{j \neq i} \omega_{ij} S_{jt}^d \tag{2}$$

where ω_{ij} represents the weighting scheme.

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

$$y_{it} = \alpha_i' \mathbf{d}_t + \beta_i' \mathbf{x}_{it} + e_{it}, \tag{3}$$

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 (3), 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 general 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 (3) with both N and T large has shown that ignoring cross-sectional dependence can seriously impair the properties of usual panel data estimators (Andrews, 2005; Phillips and Sul, 2007; Sarafidis and Wansbeek, 2012). 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:

$$e_{it} = \gamma_i' \mathbf{f}_t + \varepsilon_{it}, \tag{4}$$

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 errors while Pesaran and Tosetti (2011) focus on the more general case of a multifactor error structure and spatial error correlation. Combining (3) and (4), we obtain the following:

$$y_{it} = \alpha_i' \mathbf{d}_t + \beta_i' \mathbf{x}_{it} + \gamma_i' \mathbf{f}_t + \varepsilon_{it}. \tag{5}$$

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. (4) 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 $\mathbf{f}_t$:

$$\mathbf{x}_{it} = \mathbf{A}'_i \mathbf{d}_t + \mathbf{\Gamma}'_i \mathbf{f}_t + \mathbf{v}_{it}, \tag{6}$$

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. (5) and (6), we get the following system of equations

$$z_{it} = \begin{pmatrix} y_{it} \\ x_{it} \end{pmatrix} = \mathbf{B}'_i \mathbf{d}_t + \mathbf{C}'_i \mathbf{f}_t + \xi_{it} \tag{7}$$

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. (7) can be written as

$$\overline{z}_t = \overline{\mathbf{B}}' \mathbf{d}_t + \overline{\mathbf{C}}' \mathbf{f}_t + \overline{\xi}_t \tag{8}$$

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. (8) by $\overline{\mathbf{C}}$ and solve for $\mathbf{f}_t$. We get

$$\mathbf{f}_t = \left(\overline{\mathbf{C}} \, \overline{\mathbf{C}}' \right)^{-1} \overline{\mathbf{C}}' \left(\overline{z}_t - \overline{\mathbf{B}}' \mathbf{d}_t - \overline{\xi}_t \right). \tag{9}$$

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

$$\mathbf{f}_t - \left(\overline{\mathbf{C}} \, \overline{\mathbf{C}}' \right)^{-1} \overline{\mathbf{C}}' \left(\overline{z}_t - \overline{\mathbf{B}}' \mathbf{d}_t \right) \xrightarrow{\text{q.m.}} 0, \text{ as } N \to 0 \tag{10}$$

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 et al., 2011; 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, Kapetanios et al., 2011; 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) (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.

3.2   Alternative nonparametric specifications

Specifications   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,

$$y_{it} = \alpha_i' \mathbf{d}_t + g_i(\mathbf{x}_{it}) + \gamma_i' \mathbf{f}_t + \varepsilon_{it}, \tag{11}$$

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,

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

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

$$\log f_{it} = \alpha_i + \phi(\log S_{it}^d) + \xi(\log S_{it}^f) + \psi(\log H_{it}) + \gamma_i' \mathbf{f}_t + \varepsilon_{it}, \tag{12}$$

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.

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

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

Sieve approximation   Su and Jin (2012) extend CCE approach to the estimation of heterogenous panel data model (11). First, following Pesaran (2006), they proxy the unobservable common factors $\mathbf{f}_t$ in (11) by the cross-sectional averages $\bar{\mathbf{z}}_t = N^{-1} \sum_{j=1}^{N} \mathbf{z}_{jt}$, where $\mathbf{z}_{it} = [y_{it}, \mathbf{x}_{it}']'$. Second, they approximate the nonparametric part of the model, $g_i(.)$, 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 $g_i(.)$, or of N and T when estimating the homogenous model with $g(.)$. 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.

$$g_i(.) \approx \delta_{g_i}' \pi^K(x) \text{ (heterogenous case), or } g(.) \approx \delta_g' \pi^K(x) \text{ (homogenous case)} \tag{14}$$

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

$$y_{it} = \alpha_i' \mathbf{d}_t + \delta_{g_i}' \pi^K(x) + \psi_i' \bar{\mathbf{z}}_t + u_{it}, \tag{15}$$

or, to estimate δ_g, the regression

$$y_{it} = \alpha_i' \mathbf{d}_t + \delta_g' \pi^K(x) + \psi_i' \bar{\mathbf{z}}_t + u_{it}, \tag{16}$$

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.

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). TPRS are a low rank eigen-approximation to thin plate splines. Thin plate splines are somehow ideal smoothers (see Wood, 2017) but are not computationally attractive because they require the estimation of as many parameters as the number of data points. TPRS avoid the problem of knot placement that usually complicates modeling with RS or PRS and more generally have some optimality properties, as they provide optimal low rank approximations to thin-plate splines, while they also are computationally efficient (see Wood, 2003). Since our explanatory variables have different units, in the case of the non-additive specification (13), we avoid isotropy by considering a tensor product basis, which is constructed by assigning TPRS as the basis for the

marginal smooth function 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 function 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). Since TPRS have been developed and mostly adopted in statistical science, a more detailed description of them is provided in Appendix C.

4   Forecasting-driven model selection procedure

To compare the alternative specifications, we perform a pseudo Monte Carlo experiment. 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 since a model is mostly selected to fit data best.

In the case of independent and identically distributed data, Racine and Parmeter (2014) proposed a method 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. In the approach by Racine and Parmeter (2014) the observations are randomly shuffled at a given percentage level into training points and into evaluation points. Each model is then fitted according to the training sample and 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 S, so that a $S \times 1$ vector of prediction errors is created for each model.[1] As highlighted in Racine and Parmeter (2014), the method can provide significant power improvements over existing single-split techniques.

In the context of time series, the performance of a model is usually assessed on its ability to forecast new observations over a given time horizon, and resampling methods are then used to evaluate forecasting performance. Racine and Parmeter (2014) extend their method to time-series data using recent advances in time-series resampling methodology based on block bootstrapping which was introduced by Carlstein (1986), and developed by Künsch (1989).[2] The main idea underlying block resampling is that individual blocks of observations that are separated far enough in time will be approximately uncorrelated and can be treated as exchangeable. Suppose the time-series has length $T = b \times l$. We can generate b nonoverlapping blocks each of length l. Moreover, if

---

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

[2]Applications of the various types of block bootstrap methods for time series and other models of dependent data (including spatial data) are extensively presented in Lahiri (2003).

the blocks are sufficiently long, each block preserves, in the resampled series, the dependence present in the original data sequence. The resampling or bootstrap scheme here, named nonoverlapping block bootstrap, is to resample with replacement from the set of b blocks. Moving block bootstrap (MBB) generalizes this scheme by allowing the blocks to overlap (Künsch, 1989). Resampling now involves more blocks, a total of $T - l + 1$ overlapping blocks.

Although Racine and Parmeter (2014) do not provide any extension of their method to panel data, for which the random resampling originally proposed for cross-sections has been employed by Ma et al. (2015) and Delgado et al. (2014), they provide enough insights as to the generalization of their method to this type of data. Indeed, block bootstrapping approaches, and, more specifically, MBB, have been recently generalized to panel data, especially when cross-sectional dependence is present, eventually arising from a multifactor error structure. For instance, Gonçalves (2011) proposed panel MBB, i.e. standard MBB applied to the vector containing all the individual observations at each point of time (see Appendix D for a complete description). She proved that panel MBB is robust to both serial and cross-sectional dependence of unknown form when applied to the fixed effects estimator. The same year, Palm et al. (2011) presented panel unit root tests that can deal not only with common factors, but also with a wide range of other plausible dynamic dependencies. They use MBB to achieve this goal. Recently, Smeekes and Westerlund (2019) extended the work of Palm et al. (2011) to panel predictability tests, making also use of block bootstrapping.

Drawing from the above mentioned literature, we extend the data-driven model choice method proposed by Racine and Parmeter (2014). As measures of model forecasting performance, we focus attention on both the average square prediction error (ASPE), which was adopted by Racine and Parmeter (2014), and the mean absolute percentage error (MAPE), which is often recommended for forecasting exercises (see e.g. Bowerman et al., 2005, p. 18, and Hyndman and Koehler, 2006) and has the interesting property of being scale independent. These two measures are defined as:

$$ASPE \;\; = \;\; \frac{1}{N} \sum_{i=1}^{N} \sum_{f=1}^{F} (y_{if} - \widehat{y}_{if})^2 \tag{17}$$

$$MAPE \;\; = \;\; \frac{1}{N} \sum_{i=1}^{N} \sum_{f=1}^{F} \frac{100 \times |y_{if} - \widehat{y}_{if}|}{y_{it}} \tag{18}$$

where F is the forecast horizon. Our forecasting-data driven model selection procedure then works as follows:

1. Apply MBB to resample from original panel data $\mathbf{z}_{it} = (y_{it}, \mathbf{x}'_{it})'$, $i = 1, \ldots, n$, and $t = 1, \ldots, T$, and call these $\mathbf{z}^*_{it} = (y^*_{it}, \mathbf{x}^{*'}_{it})'$, $i = 1, \ldots, n$, and $t = 1, \ldots, T$.

2. Let the first T_1 years for the n countries form a training sample, i.e. $\mathbf{z}^*_{it}$, $i = 1, \ldots, n$, and $t = 1, \ldots, T_1$, and let the remaining observations, or $\mathbf{z}^*_{it}$, $i = 1, \ldots, n$, and $t = T_1 + 1, \ldots, T$, form an evaluation sample.

3. Consider two competing models A and B, fit each model on the $n \times T_1$ training observations. Let $\widehat{g}^A_{T_1}(.)$ and $\widehat{g}^B_{T_1}(.)$ denote the estimated models.

4. Generate predictions for evaluation observations, or $\widehat{g}_{T_1}^A(\mathbf{x}_{it}^*)$ and $\widehat{g}_{T_1}^B(\mathbf{x}_{it}^*)$, $i = 1, \ldots, n$, and $t = T_1 + 1, \ldots, T$. Generation of predictions requires knowledge of factor values, i.e. $\mathbf{f}_t$, $t = T_1 + 1, \ldots, T$. These values are easily obtained using their approximation by cross-sectional averages (see Eq.(10)).

5. Compute the ASPE, and MAPE, of each model, i.e.

$$ASPE^L = \frac{1}{n \times T_1} \sum_{i=1}^{n} \sum_{t=T_1+1}^{T} (y_{it}^* - \widehat{g}_{T_1}^L(\mathbf{x}_{it}^*))^2, \ L = A, B, \text{ and}$$

$$MAPE^L = \frac{1}{n \times T_1} \sum_{i=1}^{n} \sum_{t=T_1+1}^{T} \frac{100 \times |y_{it}^* - \widehat{g}_{T_1}^L(\mathbf{x}_{it}^*)|}{y_{it}^*}, \ L = A, B$$

6. Repeat this a large number of times, say S, yielding $(ASPE_s^A, ASPE_s^B)$, $s = 1, \ldots, S$, and $(MAPE_s^A, MAPE_s^B)$, $s = 1, \ldots, S$.

7. Compare the predictive performances of the two models, using their respective ASPE, or MAPE, empirical distribution functions. For instance, comparison can be achieved using paired-t or dominance tests.

Appendix D explains how MBB can be applied to panel data while Appendix E provides further insights about the validity of the proposed forecasting-driven model selection procedure, by conducting a small simulation study.

5 Data

We build a new annual country-level balanced panel data set covering 24 OECD countries from 1971 to 2014, which extends the time-coverage of the data used by CHH and Ertur and Musolesi (2017). As for the dependent variable (f_{it}), we use TFP at constant prices from Penn World Table version 9.0 (Feenstra et al., 2015). As for the explanatory variables, to build the domestic R&D stock (S_{it}^d), we consider Business Expenditure on Research and Development (BERD) flow values, collected from OECD-STATS database, using total data as source of funds. Missing values are filled and then the BERD stock is calculated using the perpetual inventory method assuming the depreciation rate to be 0.05. In doing these steps, we follow CH and CHH. For the human capital variable (H_{it}^d), we build a stock measure using the same method by Ertur and Musolesi (2017) which is grounded on earlier works on the return to schooling and then assumed a log-(piecewise)linear relationship between H and the average number of years of schooling, which were obtained from Barro and Lee (2013). Note that for the last three years of our sample, which are not covered by Barro and Lee (2013), we fill the missing values by linear extrapolation.

Finally, as far a the foreign R&D is concerned, we focus on geographic proximity as a channel for technology diffusion. This choice has been done not only because it is theoretically consistent (Keller, 2002; Eaton and Kortum, 2002) but mainly because traditional channels of international

technology diffusion, such as trade, FDI or patent activity, might create reverse causality problems when included in econometric specifications as they may depend on the country's technological level and, in turn, may be endogenous with respect to TFP (Hong and Sun, 2011). In contrast, geographic distance is generally considered as exogenous (Keller, 2004, p. 772) and may also proxy some endogenous measures of socioeconomic, institutional, cultural or linguistic similarities that might enhance the diffusion of technology. Following Keller (2002) and Ertur and Musolesi (2017), we propose a specification of foreign R&D that incorporates the notion that the impact of foreign R&D is a decreasing function of geographic distance from foreign economies. Therefore, the foreign R&D capital stock for each country i is obtained by weighting the domestic R&D capital stocks of every other country $j \neq i$ in the sample using an exponential distance decay function, $\omega_{ij} = \exp(-d_{ij})$, with d_{ij} being the spherical distance between the capitals of countries. Therefore,

$$S_{it}^f = \sum_{j \neq i} \exp(-d_{ij}) S_{jt}^d. \tag{19}$$

It is also interesting to note that a possible alternative approach, which has never been investigated within this literature, involves in calculating a multidimensional distance by adopting the Mahalanobis method, as proposed by Berry et al. (2010). This method allows summarizing many dimensions and is preferable to other approaches when the alternative measures of cross-national distance are highly correlated, as it is often the case. This may provide further and interesting insights. However, in this methodological paper, which aims at comparing alternative specifications, geographic distance is preferred in light of exogeneity reasons.

Some descriptive statistics are provided in Table 1. A relevant information provided by this Table is that both domestic and foreign R&D present a between variation that is almost twice the within variation, whereas for both TFP and human capital, between and within variations are of the same order of magnitude.

===== Insert Table 1 =====

In addition, Figure 1 depicts univariate and bivariate plots for few representative countries. We picked six countries: USA, Germany, South Korea, Italy, Spain and Sweden. USA and Germany are the worldwide and European leader, respectively, with respect to technology. South Korea has become over the years one of the most innovative countries in the world with a massive increase in R&D stock; Italy and Spain are two South-European countries suffering of under investments in R&D and Sweden is a representative Nordic country. Figure 1 shows relevant nonlinear and heterogeneous dynamics across countries, which may be considered as a first descriptive indication that flexible nonparametric models could be suitable in such a framework.

===== Insert Figure 1 =====

14

Appendix B provides some additional insights about the time series properties of the main variables under investigation through application of the PANICCA test by Reese and Westerlund (2016), which is a PANIC approach (Bai and Ng, 2004) based on cross-sectional averages augmentation rather than on principal component estimation. Application of the test confirms the results of Ertur and Musolesi (2017) by using an extended dataset and indicates that the variables are nonstationary and that such a nonstationarity arises as a result of the combination of a stationary idiosyncratic component with nonstationary common factors.

6   Results

6.1   Model selection

We adopt the forecasting data-driven model selection approach previously described. Three specifications are considered:[3]

$$
\log f_{it} = \alpha_i + \theta \log S_{it}^d + \gamma \log S_{it}^f + \delta \log H_{it} + \gamma_i' \mathbf{f}_t + \varepsilon_{it}, \tag{20}
$$

$$
\log f_{it} = \alpha_i + \phi(\log S_{it}^d) + \xi(\log S_{it}^f) + \psi(\log H_{it}) + \gamma_i' \mathbf{f}_t + \varepsilon_{it}, \text{ and} \tag{21}
$$

$$
\log f_{it} = \alpha_i + g(\log S_{it}^d, \log S_{it}^f, \log H_{it}) + \gamma_i' \mathbf{f}_t + \varepsilon_{it}. \tag{22}
$$

Additive and non-additive specifications (20) and (21) are estimated using either regression splines or penalized regression splines.

As far as the blocks are concerned, we fix the length of the blocks to eleven years. This allows a simple split of our sample of 44 years, and also provides blocks that should be long enough to preserve the dependence structure of the original data. Overall, we draw with replacement from $44 - 11 + 1 = 34$ blocks. The number of draws S is fixed to 1000. We consider different out-of-sample performances horizons and similar to Baltagi et al. (2004) we consider both a one-year and a three-year horizon.

A first relevant result is that the non-additive penalized model has the smallest median for both ASPEs and MAPEs. In particular, for the 1-year horizon, the median ASPEs (MAPEs) 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 0.552, 0.634, 0.636, and 0.617 (0.830, 0.913, 0.888, and 0.791), respectively. For the 3-year horizon, these ratios do not change significantly and take the following values: 0.521, 0.656, 0.656, and 0.631 (0.870, 0.978, 0.930, and 0.844). It can be expected that the out-of-sample performances will deteriorate when the horizon increases. Interestingly, it can be noted that when moving from 1- to 3-year horizon the median of the ASPEs and MAPEs is almost unaffected, showing a low increase, while the variability increases substantially as the standard deviation almost doubles in all models.

A second interesting result is that when looking at the median, the parametric model is almost always outperformed by the alternative nonparametric specifications, the only exception being

---

[3]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. The model selection procedure has been coded by the authors using R programming language.

the non-additive unpenalized model, which provides the worst results in terms of MAPE. Third, while when imposing an additive structure, the penalized regression modeling and its unpenalized counterpart present very similar median ASPEs and MAPEs, when estimating the non-additive specification, which typically suffers more from the curse of dimensionality problem, an extremely pronounced gain in terms of predictive ability from using PRS over RS is observed.

In order to test whether the above differences across alternative specifications in terms of their median are statistically significant, we also performed paired Wilcoxon test, which is a nonparametric alternative to paired t-test when comparing paired data. The results of these tests are presented in Table 2. In all cases, when contrasting the non-additive penalized model with the others, the null hypothesis that the difference in medians of the ASPEs/MAPEs is zero is strongly rejected, thus indicating that this difference is statistically significant. The same happens when contrasting the additive models, both penalized and unpenalized, with the standard parametric specification.

===== Insert Table 2 =====

Next, Figure 2 shows the empirical cumulative distribution functions of the ASPEs (MAPEs) for each model. Clearly, the ASPE (MAPE) 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. Even if visual differences seem to appear between the distributions for the different competing models in the figures presented above, it is interesting to confirm their existence by a statistical test. The most popular approach for distribution homogeneity testing is based on the Kolmogorov-Smirnov (KS) statistic, which is obtained as the largest discrepancy of the empirical distribution functions by these models (see, e.g. Lehmann and Romano, 2006). Below we use KS test to choose the best model in terms of predictive performance.

===== Insert Figure 2 =====

The first set of results presented in Table 3 are those of the KS tests of equality of the ASPE distributions for the estimated models. These tests clearly conclude by rejecting the null hypothesis of equality of distributions for each pair of models, whatever the forecasting time horizon, with the noticeable exception of generalized additive unpenalized and penalized specifications. The second part of Table 3 reports results of KS test of dominance for each forecasting time-horizon. We focus on the distribution of the penalized tensor model because this distribution sets oneself apart from those of the other models estimated in the figures presented above. We first test if the distribution of the penalized tensor model dominates the distribution of each of the competing models. Results clearly show that this hypothesis is rejected, whatever the considered competing model and time horizon. We then test the reverse hypothesis, i.e. that the distribution of the penalized tensor model is dominated by the distribution of each of the competing models. Test values do not allow to conclude in the rejection of this hypothesis. The penalized tensor model therefore appears to

be the model which generates the forecasts closest to the observed values, whatever the considered forecasting time horizon, when the quality of the forecasts is measured by the ASPE criterion.

===== Insert Tables 3 and 4 =====

A similar analysis can also be performed using MAPE criterion. Corresponding results are reported in Table 4. The first set of results given in this table shows that there are only significant differences between the distribution of the penalized tensor model and those of its competitors, whatever the forecasting time-horizon. Moreover, dominance tests, here too, clearly conclude that the MAPE distribution of the non-additive penalized model is stochastically dominated by the MAPE distribution of any of the remaining models. The penalized tensor model therefore appears to be the model which generates the forecasts closest to the observed values, whatever the considered forecasting time horizon, when the quality of the forecasts is measured by the MAPE criterion.

To sum up, these results clearly indicate that i) the parametric specification underperforms with respect to the nonparametric penalized models and ii) there is very robust evidence that the penalized non-additive model outperforms all the others. This is a similar result than Ma et al. (2015) who use a similar macro panel data set. As far as nonparametric specifications are concerned, PRS 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 contrasting 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.

6.2 Estimation results

We now present the estimation results of the non-additive penalized specification, which, according to our findings, clearly provides the best predictive performance and thus is the most suitable to approximate the underlying DGP.

As a preliminary step, in order to provide a straightforward comparison with previous works such as Ertur and Musolesi (2017), CHH and others, we also estimate the model exploiting our extended dataset and adopting some parametric estimators, which have been already employed in the literature. The results in Table 5 are structured as follows. Column (i) and (ii) report the results obtained by using the pooled CCE by Pesaran (2006), which was employed by Ertur and Musolesi (2017). While in column (i), we consider the standard parametric specification, which was previously discussed, in column (ii), we allow the impact of the explanatory variables to differ

between the G7 countries and the others, as sometimes done in the literature. In columns (iii) to (vi), following CHH, Kao et al. (1999) and Lee (2006) we instead focus on panel cointegration estimators. We consider both the pooled Fully Modified OLS (FMOLS) by Phillips and Moon (1999) and the pooled Dynamic OLS (DOLS) by Kao and Chiang (2001), without and with a linear time trend in the deterministic component.

===== Insert Table 5 =====

The results are as follows. When employing the pooled CCE in (i), all the estimated parameters are close to zero and are not significant at standard significance levels. A similar result was pointed out by Eberhardt et al. (2013) when estimating the effect of R&D within a Griliches-type knowledge capital production function, and can be partially explained as a result of an *"omitted time-related factors bias"* that may arise in common panel data models that do not account for common factors. Allowing the impact of the explanatory variables to differ between the G7 countries and the others (model (ii)) provides additional insights and it particularly may suggest that model (i) suffers of a misspecification bias as it is now found that the coefficient associated with domestic R&D for G7 countries is positive (0.04) and significant at 10%. It is also found a positive (0.07) and significant effect of foreign R&D for non-G7 countries. Finally, an implausible negative effect of human capital for non-G7 countries is also obtained, with an estimated elasticity equals to -0.31. These results partially contradict the findings by Ertur and Musolesi (2017), who exploited a shorten dataset (1971-2004) and found totally different results for foreign R&D and human capital. Human capital was found to have a positive and significant effect only for G7 countries while the opposite happened for foreign R&D. Overall, such a parameters' estimates instability may suggest that the traditional CH/CHH parametric model likely suffers of a *"functional misspecification bias"*. The results obtained adopting panel cointegrated estimators provide similar evidence. Indeed, while the model without a trend provides results that are consistent with CHH and Kao et al. (1999), who adopted the same panel cointegration estimators, the inclusion of a linear trend in the cointegration relation makes all the estimated parameters close to 0 and statistically non-significant. Including a linear trend seems to be consistent with the time series properties of the data and also makes the estimated specification closer to the factor model. The results do not significantly change when moving from pooled to mean-group estimators.[4]

In order to discuss the results for preferred non-additive model, we first provide a visual inspection of the estimated smooth function and then focus attention on the estimated elasticities. Also note that the estimated (multivariate) smooth function appears to be highly significant by adopting the Wald-type test suggested by Wood (2012).[5]

As for the visual inspection of the estimated multivariate function, we depict the joint effects of two variables at a time, fixing the level of the third one to the first, fifth (the median) and ninth

---

[4]FMOLS and DOLS estimates were obtained by exploiting Eviews 11. The coefficient covariances are computed using the default settings, and the long-run covariances used a Bartlett kernel and a Newey-West fixed bandwidth. As for the DOLS, leads and lags were selected using BIC. Finally note that it was computationally unfeasible estimating the model with G7 interactions.

[5]Test results are available upon request from authors.

decile. At first glance, looking at Figures 3, 4 and 5, it easy to see that complex interactive effects are a relevant feature of the data, which are obscured in parametric specifications.

===== Insert Figures 3, 4 and 5 =====

When focusing on the joint effect of the R&D variables on TFP (see Figure 3), it can be noted that such interactive effect increases with the level of human capital. As depicted in panel (a), for low level of human capital and for most part of the domain of the R&D variables, the estimated function is flat. Only for high values of foreign R&D, domestic R&D has a positive effect on TFP. Panels (b) and (c) show the estimated bivariate function when human capital is fixed to the median and to the ninth decile, respectively. The results in both panels suggest a positive and nonlinear effect of domestic/foreign R&D on TFP, which seems to increase as human capital moves from the median the ninth decile.

A similar pattern appears for the domestic R&D – human capital bivariate function (see Figure 4), as it clearly increases by increasing the level of foreign R&D. Indeed, when foreign R&D is fixed to the first decile (panel (a)), the estimated function is very flat; it then becomes an increasing function when foreign R&D is fixed to the median and to the ninth decile. However, while panel (b) (foreign R&D fixed at the median) indicates a rather additive effect, panel (c) (foreign R&D fixed at the ninth decile), suggests a more complex nonlinear non-additive interactive effect.

Finally, as far as the joint human capital – foreign R&D effect is concerned, a different pattern is observed as the estimated relation (see Figure 5) is always very flat for any level of domestic R&D. Relevant exceptions happen for high levels of foreign R&D, where in some specific regions, human capital has a pronounced positive effect on TFP.

We then provide a thorough analysis of the estimated elasticities and compare them with both previous results and our parametric estimates in Table 5. As far as domestic R&D is concerned, the literature provides an estimated elasticity that typically is bounded in the range 0.05-0.2 and also suggests that the effect of domestic R&D on TFP is much higher for G7 than for non-G7 (Ertur and Musolesi, 2017; Coe et al., 2009; Kao et al., 1999). While the parametric CCE in Table 5 confirms this findings, our nonparametric results are somehow consistent with it but are also much richer as shown in Table 6 and in Figure 6, where we report estimated elasticity of one domestic R&D as a function of the potential values of this factor, the two other ones being fixed to some quantile values.[6] Indeed, the output elasticity of domestic R&D capital stock from the preferred non-additive model ranges between about 0 and 0.21. There are some combinations of factors for which the estimated elasticity is very close to zero and is not significant. This happens for low values of human capital associated with high values of domestic R&D. This pattern can be viewed as consistent with the existence of a minimum level of absorptive capacity, arising from the human

---

[6]Standard errors are reported in Table 6, in addition to estimated values of elasticities. Their computation takes advantage of the underlying parametric representation of sieve approximations (see Eq. (14)). It is based on the covariance matrix of estimated parameters $\widehat{\delta}_g$ and derivatives of the elements belonging to the chosen basis $\pi^K(x) = (\pi_1(x), \pi_2(x), \ldots, \pi_K(x))'$. 90% Confidence intervals in Figure 6 are computed in the same way. (See Wood, 2017, p. 341, for more details.)

capital endowment, that allows a country to benefit from technology. Similarly, the estimated output elasticity of domestic R&D is close to zero and non-significant for high levels of foreign R&D associated with low levels of domestic R&D. At the opposite, high values of domestic and foreign R&D in combination with average levels of human capital are associated with an estimated elasticity about 0.2. We label this situation in which the estimated output elasticity of domestic R&D (nonlinearly) increases with the amount of foreign R&D, as *(nonlinear) interactive spillover effects* in opposition to standard spillover effects, which generally refer to the direct ceteris paribus effect of foreign R&D on TFP. We also note that for median/high values of human capital, threshold effects appear as a relevant feature of the data, as the estimated elasticity is positive (in the range 0.05-0-.10) and significant for low values of domestic R&D, then it decreases around zero and finally increases up again for high values of domestic R&D.

===== Insert Table 6 =====

As for the output elasticity of foreign R&D, as also stressed in Section 2, a high degree of variability is observed when looking at the existing literature. Looking also at the results in Table 5, such a variability could be explained as a consequence of a functional misspecification bias, and our results from the preferred nonparametric model are consistent with this view. Three main findings arise from Figure 7. First, there are many combination of inputs for which the estimated elasticity is about 0 and is not statistically significant. This happens for instance for median levels of both domestic R&D and human capital (irrespective of the level of foreign R&D). Second, for high levels of both domestic R&D and human capital, the estimated elasticity is always positive and significant, ranging from about 0.1 to 0.2, depending on the level of foreign R&D. Third, there are some combinations of domestic R&D-human capital for which a certain amount of foreign R&D is necessary to observe a significantly positive elasticity. On one side, these results are consistent with the existence of a certain level of absorptive capacity arising from domestic technology that is necessary for a country to benefit from foreign technology (see, e.g., Xu, 2000) and with theories suggesting that technology that is invented in frontier countries is less appropriate for poorer countries (see, e.g., Basu and Weil, 1998). On the other, however, they suggest that this arises with complex nonlinearities that clearly make the parametric model misspecified and that may explain both the high variability in previous estimates arising from parametric specifications and our parametric estimates in Table 5.

===== Insert Figures 6, 7, and 8 =====

Finally, as for the output elasticity of human capital, typical approaches do not accounting for common latent factors generally conclude in a high and significant estimated elasticity, which is often in the range 0.5-0.7 (see, e.g., CHH). However, Ertur and Musolesi (2017) cast considerable doubt on the idea that the stock of human capital – constructed using the average years of schooling – significantly affects TFP. Their result complements some previous studies on growth emphasizing the key role of the quality of education (see, e.g., Hanushek and Kimko, 2000) and finally suggests

that the quantity of education no longer has a significant effect when omitted time-related variable bias is addressed. While our parametric results in Table 5 totally confirm such a finding, the nonparametric non-additive model provides further insights as depicted in Figure 8. Indeed, for low-median levels of both R&D variables, the estimated elasticity is generally non-significantly different from zero, thus confirming the parametric results. However, for high values of domestic and/or foreign R&D, some locally significant positive effect is documented. In particular, when both R&D variables are high, the estimated elasticity presents an interesting nonlinear pattern as, for low-median values of human capital, the estimated function is flat and non-significantly different from zero, it then increases and become positive, high in magnitude, and significant for the last two deciles of human capital. This pattern seems to be empirically consistent with the Korean technological development that occurred over the last decades.

Overall, this evidence appears to be consistent with the existence of threshold effects, nonlinear interactive spillover effects, complex absorptive capacity effects, finally being consistent with the idea of localized technical change, as suggested in the seminal paper by Atkinson and Stiglitz (1969).

7 Concluding remarks

This paper revisits the analysis of international technology diffusion. It aims to contribute to the existing literature by proposing an estimation strategy based on forecasting-driven model selection, which exploits recent panel data nonparametric methods and extends existing data-driven model selection procedures to the case of cross-sectionally dependent panels. Results clearly conclude in the superiority of a fully nonparametric (non-additive) specification of the relationship between TFP and domestic R&D capital, foreign R&D and human capital, over parametric and even semi-parametric (additive) specifications. The nonparametric specification allows for a richer view of the resulting spillover effects, showing nonlinear effects and thresholds not previously detected, ultimately suggesting the presence of *(nonlinear) interactive spillover effects*, i.e. a situation in which the estimated output elasticity of domestic R&D (nonlinearly) increases with the amount of foreign R&D. 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 predictive ability.

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. (2005). Cross-section regression with common shocks. *Econometrica*, 73(5):1551–1585.

Atkinson, A. and Stiglitz, J. (1969). A new view of technological change. *The Economic Journal*, 79(315):573–578.

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

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bai, J. and Ng, S. (2004). A panic attack on unit roots and cointegration. *Econometrica*, 72(4):1127–1177.

Bai, J. and Ng, S. (2010). Panel unit root tests with cross-section dependence: a further investigation. *Econometric Theory*, pages 1088–1114.

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. J. and Lee, J. W. (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of development economics*, 104:184–198.

Basu, S. and Weil, D. N. (1998). Appropriate technology and growth. *The Quarterly Journal of Economics*, 113(4):1025–1054.

Berry, H., Guillén, M. F., and Zhou, N. (2010). An institutional approach to cross-national distance. *Journal of international business studies*, 41(9):1460–1480.

Bitzer, J. and Kerekes, M. (2008). Does foreign direct investment transfer technology across borders? new evidence. *Economics Letters*, 100(3):355–358.

Bowerman, B. L., O'Connell, R. T., and Koehler, A. B. (2005). *Forecasting, time series, and regression: an applied approach*, volume 4. South-Western Pub.

Brock, W. A. and Durlauf, S. N. (2001). What have we learned from a decade of empirical research on growth? growth empirics and reality. *the world bank economic review*, 15(2):229–272.

Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Annals of Statistics*, 14:1171–1194.

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.

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.

Durlauf, S. N., Kourtellos, A., and Minkin, A. (2001). The local solow growth model. *European Economic Review*, 45(4-6):928–940.

Eaton, J. and Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, 70(5):1741–1779.

Eberhardt, M., Helmers, C., and Strauss, H. (2013). Do spillovers matter when estimating private returns to r&d? *Review of Economics and Statistics*, 95(2):436–448.

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.

Feenstra, R. C., Inklaar, R., and Timmer, M. P. (2015). The next generation of the penn world table. *American economic review*, 105(10):3150–82.

Fracasso, A. and Marzetti, G. V. (2013). An empirical note on international r&d spillovers. *Empirical Economics*, 45(1):179–191.

Gonçalves, S. (2011). The moving blocks bootstrap for panel linear regression models with individual fixed effects. *Econometric Theory*, pages 1048–1082.

Griliches, Z. (1979). Issues in assessing the contribution of research and development to productivity growth. *The bell journal of economics*, pages 92–116.

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.

Hanushek, E. A. and Kimko, D. D. (2000). Schooling, labor-force quality, and the growth of nations. *American economic review*, pages 1184–1208.

Helpman, E. (1992). Endogenous macroeconomic growth theory. *European Economic Review*, 36(2-3):237–267.

Hong, E. and Sun, L. (2011). Foreign direct investment and total factor productivity in china: A spatial dynamic panel analysis. *Oxford Bulletin of Economics and Statistics*, 73(6):771–791.

Hsiao, C., Pesaran, M. H., and Tahmiscioglu, A. K. (1999). *Bayes estimation of short-run coefficients in dynamic panel data models*, pages 268–296. Cambridge University Press.

Hyndman, R. J. and Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4):679–688.

Jones, C. I. (1995). R & d-based models of economic growth. *Journal of political Economy*, 103(4):759–784.

Kao, C. and Chiang, M.-H. (2001). On the estimation and inference of a cointegrated regression in panel data. *Advances in econometrics*, 15:179–222.

Kao, C., Chiang, M.-H., and Chen, B. (1999). International r&d spillovers: an application of estimation and inference in panel cointegration. *Oxford Bulletin of Economics and statistics*, 61(S1):691–709.

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

Karshenas, M. and Stoneman, P. (1995). Technological diffusion. *Handbook of the economics of innovation and technological change*, pages 265–297.

Keller, W. (2002). Geographic localization of international technology diffusion. *The American Economic Review*, 92(1):120–142.

Keller, W. (2004). International technology diffusion. *Journal of economic literature*, 42(3):752–782.

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

Künsch, H. R. (1989). The jacknife and the bootsrap for general stationary observations. *Annals of Statistics*, 17:1217–1241.

Lahiri, S. N. (2003). *Resampling methods for dependent data*. Springer-Verlag, Springer-Verlag , New York.

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

Lehmann, E. L. and Romano, J. P. (2006). *Testing statistical hypotheses*. Springer Science & Business Media.

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.

Maddala, G. S., Trost, R. P., Li, H., and Joutz, F. (1997). Estimation of short-run and long-run elasticities of energy demand from panel data using shrinkage estimators. *Journal of Business & Economic Statistics*, 15(1):90–100.

Musolesi, A. (2007). Basic stocks of knowledge and productivity: Further evidence from the hierarchical bayes estimator. *Economics Letters*, 95(1):54–59.

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

Palm, F. C., Smeekes, S., and Urbain, J. P. (2011). Cross-sectional dependence robust block bootstrap panel unit root tests. *Journal of Econometrics*, 163:85–104.

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 multifactor error structure. *Econometrica*, 74(4):967–1012.

Pesaran, M. H. and Smith, R. (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of econometrics*, 68(1):79–113.

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. and Moon, H. R. (1999). Linear regression limit theory for nonstationary panel data. *Econometrica*, 67(5):1057–1111.

Phillips, P. C. and Sul, D. (2007). Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence. *Journal of Econometrics*, 137(1):162–188.

Potterie, B. v. P. d. l. and Lichtenberg, F. (2001). Does foreign direct investment transfer technology across borders? *Review of Economics and statistics*, 83(3):490–497.

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.

Reese, S. and Westerlund, J. (2016). Panicca: Panic on cross-section averages. *Journal of Applied Econometrics*, 31(6):961–981.

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.

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

Sarafidis, V. and Wansbeek, T. (2012). Cross-sectional dependence in panel data analysis. *Econometric Reviews*, 31(5):483–531.

Smeekes, S. and Westerlund, J. (2019). Robust block bootstrap panel predictability tests. *Econometric Reviews*, 38:1089–1107.

Spolaore, E. and Wacziarg, R. (2009). The diffusion of development. *The Quarterly journal of economics*, 124(2):469–529.

Stoneman, P. (1985). Technological diffusion: the viewpoint of economic theory. Technical report.

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.-P. (2015). Cross-sectional averages versus principal components. *Journal of Econometrics*, 185(2):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.

Xu, B. (2000). Multinational enterprises, technology diffusion, and host country productivity growth. *Journal of development economics*, 62(2):477–493.

Table 1: Descriptive statistics

Variable		Mean	SD	Min	Max
$\log TFP$	Overall	-0.090	0.146	-0.645	0.256
	Between		0.097	-0.308	0.113
	Within		0.110	-0.427	0.218
$\log\ domestic\ R\&D$	Overall	10.190	2.223	-1.545	14.925
	Between		2.099	3.793	3.793
	Within		0.846	4.851	12.771
$\log\ foreign\ R\&D$	Overall	11.210	1.109	6.634	14.793
	Between		0.962	8.300	14.137
	Within		0.585	7.420	13.738
$\log human\ capital$	Overall	1.038	0.141	0.532	1.295
	Between		0.108	0.798	1.236
	Within		0.093	0.741	1.300

Table 2: Paired Wilcoxon-tests of factor models

	Horizon	Additive unpenalized	Additive penalized	Non-additive unpenalized	Non-additive penalized
ASPE					
Parametric	1-year	394916($< 2.2e − 16$)	412761($< 2.2e − 16$)	370720($< 2.2e − 16$)	464626($< 2.2e − 16$)
	3-year	411659($< 2.2e − 16$)	432372($< 2.2e − 16$)	367896($< 2.2e − 16$)	451674($< 2.2e − 16$)
Additive unpenalized	1-year		294332($1.398e − 06$)	164042($< 2.2e − 16$)	459080($< 2.2e − 16$)
	3-year		309488($8.916e − 11$)	140828($< 2.2e − 16$)	438475($< 2.2e − 16$)
Additive penalized	1-year			152358($< 2.2e − 16$)	452034($< 2.2e − 16$)
	3-year			127792($< 2.2e − 16$)	427726($< 2.2e − 16$)
Non-additive unpenalized	1-year				473174($< 2.2e − 16$)
	3-year				461908($< 2.2e − 16$)
MAPE					
Parametric	1-year	208088(0.020)	220794($5.548e − 05$)	196197(0.4649)	220797($5.539e − 05$)
	3-year	179872(0.057)	185370(0.007)	169856(0.680)	194917($3.691e − 05$)
Additive unpenalized	1-year		189658(0.8835)	179751(0.140)	219520(0.000)
	3-year		171769(0.4868)	157148(0.141)	183474(0.015)
Additive penalized	1-year			177662(0.079)	213788(0.002)
	3-year			152699(0.033)	176860(0.147)
Non-additive unpenalized	1-year				223099($1.421e − 05$)
	3-year				196897($9.896e − 06$)

Null hypothesis: The true difference in medians of the ASPEs/MAPEs of the compared models is zero.

Number of resampling iterations B: 1000.

p-values between brackets.

Table 3: Kolmogorov-Smirnov test results: $ASPE$

	Additive unpenalized	Additive penalized	Non-additive unpenalized	Non-additive penalized
—— Forecasting horizon = 1 year ——				
Two-sided tests				
Parametric	0.002	< 0.001	0.006	< 0.001
Additive unpenalized	—	0.999	0.061	< 0.001
Additive penalized	—	—	0.020	< 0.001
Non-additive unpenalized	—	—	—	< 0.001

One-sided tests				
H$_A$: The distribution of $ASPE$ for non-additive penalized is lower than the distribution of:				
	Parametric	Additive unpenalized	Additive penalized	Non-additive unpenalized
Non-additive penalized	< 0.001	< 0.001	< 0.001	< 0.001
H$_A$: The distribution of $ASPE$ for non-additive penalized is greater than the distribution of:				
	Parametric	Additive unpenalized	Additive penalized	Non-additive unpenalized
Non-additive penalized	≈ 1	0.999	0.996	≈ 1

	Additive unpenalized	Additive penalized	Non-additive unpenalized	Non-additive penalized
—— Forecasting horizon = 3 years ——				
Two-sided tests				
Parametric	< 0.001	< 0.001	0.003	< 0.001
Additive unpenalized	—	0.954	0.055	< 0.001
Additive penalized	—	—	0.020	< 0.001
Non-additive unpenalized	—	—	—	< 0.001

One-sided tests				
H$_A$: The distribution of $ASPE$ for non-additive penalized is lower than the distribution of:				
	Parametric	Additive unpenalized	Additive penalized	Non-additive unpenalized
Non-additive penalized	< 0.001	< 0.001	< 0.001	< 0.001
H$_A$: The distribution of $ASPE$ for non-additive penalized is greater than the distribution of:				
	Parametric	Additive unpenalized	Additive penalized	Non-additive unpenalized
Non-additive penalized	≈ 1	≈ 1	≈ 1	≈ 1

Table 4: Kolmogorov-Smirnov test results: $MAPE$

	Additive unpenalized	Additive penalized	Non-additive unpenalized	Non-additive penalized
—— Forecasting horizon = 1 year ——				
Two-sided tests				
Parametric	0.573	0.370	0.859	0.008
Additive unpenalized	—	0.954	0.610	0.078
Additive penalized	—	—	0.200	0.069
Non-additive unpenalized	—	—	—	0.069
One-sided tests				
H$_A$: The distribution of $MAPE$ for non-additive penalized is lower than the distribution of:				
	Parametric	Additive unpenalized	Additive penalized	Non-additive unpenalized
Non-additive penalized	0.004	0.039	0.034	0.034
H$_A$: The distribution of $MAPE$ for non-additive penalized is greater than the distribution of:				
	Parametric	Additive unpenalized	Additive penalized	Non-additive unpenalized
Non-additive penalized	0.952	0.845	0.866	0.940
—— Forecasting horizon = 3 years ——				
Two-sided tests				
	Additive unpenalized	Additive penalized	Non-additive unpenalized	Non-additive penalized
Parametric	0.573	0.288	0.648	0.005
Additive unpenalized	—	0.999	0.401	0.029
Additive penalized	—	—	0.219	0.020
Non-additive unpenalized	—	—	—	0.015
One-sided tests				
H$_A$: The distribution of $MAPE$ for non-additive penalized is lower than the distribution of:				
	Parametric	Additive unpenalized	Additive penalized	Non-additive unpenalized
Non-additive penalized	0.023	0.015	0.009	0.007
H$_A$: The distribution of $MAPE$ for non-additive penalized is greater than the distribution of:				
	Parametric	Additive unpenalized	Additive penalized	Non-additive unpenalized
Non-additive penalized	0.975	0.845	0.845	0.952

Appendices

A Summary of previous works

The Table A1 below summarizes some previous works and highlights some relevant methodological aspects.

Table 5: Parametric estimation results

	CCE	CCE_G7	FMOLS	DOLS	FMOLS_T	DOLS_T
	(i)	(ii)	(iii)	(iv)	(v)	(vi)
$\log S^d$	-0.0004		0.071***	0.062***	0.005	0.006
	(0.0056)		(0.0086)	(0.0097)	(0.0102)	(0.0111)
$\log S^f$	0.009		0.013	0.005	-0.030	-0.033
	(0.0122)		(0.0115)	(0.0149)	(0.0186)	(0.0195)
$\log H$	-0.095		0.357***	0.361***	0.225	0.178
	(0.0779)		(0.0680)	(0.0719)	(0.1366)	(0.1431)
$\mathbf{1}_{G7} \log S^d$		0.041*				
		(0.0248)				
$\mathbf{1}_{NOG7} \log S^d$		0.008				
		(0.0060)				
$\mathbf{1}_{G7} \log S^f$		0.039				
		(0.0281)				
$\mathbf{1}_{NOG7} \log S^f$		0.065***				
		(0.0184)				
$\mathbf{1}_{G7} \log H$		0.26				
		(0.2357)				
$\mathbf{1}_{NOG7} \log H$		-0.306**				
		(0.121)				

Notes:

CCE: Pooled CCE ; CCE_G7: Pooled CCE with G7 interactions;

$1_{G7} = \begin{cases} 1 & \text{if country} \in \text{G7 group} \\ 0 & \text{if country} \notin \text{G7 group} \end{cases}$, and: $1_{NOG7} = 1 - 1_{G7}$

FMOLS: Fully Modified OLS; FMOLS_T: Fully Modified OLS with linear trend;

DOLS: Dynamic OLS; leads and lags selection based on BIC;

DOLS_T: Dynamic OLS with linear trend;

***, **, *: significant at 1%, 5%, 10%, respectively; Standard errors in brackets.

TABLE A1

Some previous studies on R&D international spillovers

Author	Sample	Technology transfer	Method	COINT	HET	CSD	NL	Foreign R&D
Coe and Helpman (1995)	22 countries, 1971-90	trade	LSDV					.06-.092
Coe et al. (2009)	22 countries, 1971-90	trade	DOLS					.165-.186
	24 countries, 1971-2004	trade	LSDV	x				.185-.206
	24 countries, 1971-2004	trade	DOLS					.206-.213
Kao et al. (1999)	22 countries, 1971-90	trade	BC-OLS					.09-.125
	22 countries, 1971-90	trade	FM-OLS	x				.075-.103
	22 countries, 1971-90	trade	DOLS					.044NS-.056NS
Lichtenberg and Van Pottelsberghe (1997)	22 countries, 1971-90	trade	LSDV					.058-.276
Lichtenberg and Van Pottelsberghe (2001)	23 countries, 1971-90	trade	LSDV					.154
	23 countries, 1971-90	FDI	LSDV					-.06NS-.072
	23 countries, 1971-90	trade	FD					.067
	23 countries, 1971-90	FDI	FD					.006NS-.039
Musolesi (2007)	13 countries, 1981-98	trade	HB					.09
	13 countries, 1981-98	FDI	HB		x			-0.01NS-.004NS
	13 countries, 1981-98	language	HB					.23
Lee (2006)	16 countries, 1981-2000	trade	DOLS					-.02NS-.17
	16 countries, 1981-2000	FDI	DOLS	x				-.017NS-.034
	16 countries, 1981-2000	patents	DOLS					.157-.183
Keller (2002)	14 countries, 1970-95	geography	NLS				x	.843
	14 countries, 1970-95	language	NLS					.565
Engelbrecht (2002)	21 countries, 1971-85	trade	LSDV					.220-.305
Bitzer and Kerekes (2008)	17 countries, 1973-2000	trade	FGLS					.009-.02
	17 countries, 1973-2000	FDI	FGLS					-.016-.012
Barrio-Castro et al. (2002)	21 countries, 1971-85	trade	LSDV					.094-.225
	21 countries, 1966-95	trade	LSDV					0.016-0.106
	21 countries, 1966-95	trade	DOLS					.092-.141
Fracasso and Vittucci (2013)	24 countries, 1971-2004	trade	DOLS					.107-.126
	24 countries, 1971-2004	geography	NLS		x		x	0.016-0.106
Ertur and Musolesi (2017)	24 countries, 1971-2004	geography	CCEP					0.094NS-0.329
	24 countries, 1971-2004	geography	CCEMG	x	x	x		0.240NS-0.510NS
	24 countries, 1971-2004	geography	SEM					0.011-NS-0.22

Notes:

COINT: Integration-Coitegration analysis; HET: Heterogeneous slopes; CSD: Cross-Sectional Dependence; NL: Nonlinear models.

LSDV: Least Squares Dummy Variable ; DOLS: Dynamic Ordinary Least Squares; BC-OLS: Bias Corrected OLS; FM-OLS: Fully Modified OLS;

FD: First Difference; HB: Hierarchical Bayes; NLS: Non Linear Least Squares; FGLS: Feasible Generalized Least Squares;

CCEP:Pommon Correlated Effects Pooled; CCEMG: Pommon Correlated Effects Mean Group; SEM: Spatial Error Model; NS: not significant.

B   Panel unit roots tests

Some of the previous works find evidence of nonstationary variables by applying first-generation tests (Coe et al., 2009). Recently, Ertur and Musolesi (2017), by adopting the PANIC approach (Bai and Ng, 2004), which decomposes the panel into deterministic, common and idiosyncratic components, not only confirmed that the series are nonstationary but also suggested that that such a nonstationarity arises as a result of a stationary idiosyncratic component combined with nonstationary common factors. Despite providing a thorough investigation of the time series properties of the variables is somehow outside the scope of the paper, it is of some interest to check whether this result holds with this extended dataset, also in light of the results provided by Kapetanios et al. (2011) showing that the CCE approch remains valid when the nonstationarity arises because of nonstationary factors.

We adopt the PANICCA test by Reese and Westerlund (2016), which is a PANIC approach based on cross-sectional averages augmentation rather than on principal component estimation. PANICCA maintains the generality of PANIC as it allows testing for unit roots in both the common and idiosyncratic components of the data, but it has better small samples properties and avoids the determination of the total number of factors, which is usually done by adopting information criteria (Bai and Ng, 2002). Nevertheless, practical implementation of such criteria is difficult as they may tend to overestimate the number of factors and the results are known to be sensitive to the maximum number of factors which should be arbitrarily fixed (Ertur and Musolesi, 2017).

The results in Table B1 are clear-cut. The statistics proposed by Bai and Ng (2010), denoted P_a and P_b and $PMSB$ provide evidence for rejection of the null hypothesis of nonstationarity of the idiosyncratic components for all variables. The rejection of the nonstationarity of the idiosyncratic component does not imply that the series are stationary, as some of the common factors may be nonstationary. To determine how many of these factors are nonstationary, we follow Reese and Westerlund (2016) and we consider the MQ_f and MQ_c statistics. The limiting distributions of these statistics are nonstandard, and critical values are reported in Bai and Ng (2002) for up to six factors. The results provide a very clear picture. For all variables, regardless of the test used, the number of nonstationary common factors is almost always equal to the total number of common factors, which given the cross-sectional averages augmentation is equal to four. The application of the PANICCA approach thus suggests that the variables are nonstationary and that this property is the result of multiple nonstationary common factors combined with stationary idiosyncratic components.

TABLE B1

PANICCA test

	Idiosyncratic component			Nonstationary factors	
	P_a	P_b	$PMSB$	MQ_f	MQ_c
	p-value				
log TFP	0	0	0.008	4	4
log $domestic\ R\&D$	0	0.004	0.07	4	4
log $foreign\ R\&D$	0	0	0.059	4	4
log $human\ capital$	0	0.003	0.086	4	4

The estimated number of independent stochastic trends is reported (5% level).

C   Thin plate regression splines

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

$$||\mathbf{y} - \boldsymbol{\chi}|| + \lambda J_{md}(\chi), \tag{23}$$

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.

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

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

$$\hat{g}(\mathbf{x}) = \sum_{i=1}^{n} \delta_i \eta_{md}(||\mathbf{x} - \mathbf{x}_i||) + \sum_{j=1}^{M} \alpha_j \pi_j(\mathbf{x}), \tag{24}$$

under the constraint that $\boldsymbol{T}'\boldsymbol{\delta} = \mathbf{0}$, $T_{ij} = \pi_j(\mathbf{x}_i)$. $\boldsymbol{\delta}$ 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, (23) translates to minimizing with respect to $\boldsymbol{\delta}$ and $\boldsymbol{\alpha}$

$$||\boldsymbol{y} - \mathbf{E}\boldsymbol{\delta} - \boldsymbol{T}\boldsymbol{\alpha}||^2 + \lambda \boldsymbol{\delta}' \mathbf{E} \boldsymbol{\delta}, \text{ subject to } \boldsymbol{T}'\boldsymbol{\delta} = \mathbf{0}, \tag{25}$$

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 (24) and (25), 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 (25) 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 $\mathbf{E} = \mathbf{U}\mathbf{D}\mathbf{U}'$ 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 (25) is replaced by minimizing

$$||\boldsymbol{y} - \mathbf{U}_k\mathbf{D}_k\boldsymbol{\delta}_k - \boldsymbol{T}\boldsymbol{\alpha}||^2 + \lambda \boldsymbol{\delta}_k'\mathbf{D}_k\boldsymbol{\delta}_k, \text{ subject to } \boldsymbol{T}'\mathbf{U}_k\boldsymbol{\delta}_k = \mathbf{0}. \tag{26}$$

Having fitted (26), the spline is evaluated from (24) after estimating $\boldsymbol{\delta}$ from $\boldsymbol{\delta}_k$.

It is worth to note that while TPS are optimal with respect to minimizing (23), 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 (25) is approximated by (26). 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.

D   Moving block bootstrap applied to panel data

Moving block bootstrap (MBB) applied to panel data runs as follows:

1. Consider you have one response y_{it} and a $p \times 1$ vector of explanatory variables, or $\mathbf{x}_{it}$, with $i = 1, \ldots, n$, and $t = 1, \ldots, T$.

2. Let $\mathbf{z}_{it} = (y_{it}, \mathbf{x}_{it}')'$, vector whose dimension is $(p+1) \times 1$.

3. Let $Z_{t,n} = (\mathbf{z}_{1t}', \mathbf{z}_{2t}', \ldots, \mathbf{z}_{nt}')'$ whose dimension is $n(p+1) \times 1$. In fact, we stack all observations for the n individuals at period t.

4. Let l the chosen length of a block. A block will thus be defined as the matrix $B_{t,l} = \{Z_{t,n}, Z_{t+1,n}, \ldots, Z_{t+l-1,n}\}$. $T - l + 1$ overlapping blocks are generated.

5. MBB resamples $k = T/l$ blocks with replacement from the set of $T - l + 1$ overlapping blocks.

E   MBB forecasting-driven model selection procedure with generated data

This appendix aims to provide further insights about the validity of the proposed MBB-forecasting-driven model selection procedure by applying such a procedure to generated data. We generate three alternative DGPs: linear, additive and non-additive, respectively, all being characterized by both cross-sectional dependence arising from (stationary and nonstationary) common factors and serial error dependence:

$$y_{it} = \alpha_i + 0.3x_{1it} + 0.7x_{2it} + \gamma_{1i}f_{1t} + \gamma_{2i}f_{2t} + \varepsilon_{it},$$

$$y_{it} = \alpha_i + 0.01x_{1it}^2 - 0.00007x_{1it}^3 + 0.8\sin(x_{1it}) + 0.1x_{2it} + 0.1x_{2it}^2 - 0.005x_{2it}^3 + \cos(x_{2it}) + \gamma_{1i}f_{1t} + \gamma_{2i}f_{2t} + \varepsilon_{it}, \text{ and}$$

$$y_{it} = \alpha_i + x_{1it}^{1.3} * x_{2it}^{0.7} + 0.5\sin(x_{1it}) + 1.5\cos(x_{2it}) + \gamma_{1i}f_{1t} + \gamma_{2i}f_{2t} + \varepsilon_{it}.$$

The error term ε_{it} has been geneurated as a first-order autoregressive process:

$$\varepsilon_{it} = 0.8\varepsilon_{it-1} + v_{it}, \text{ with } v_{it} \sim N(0, 0.7).$$

The explanatory variables are generated as linear functions of the individual effects and of common factors:

$$x_{1it} = c_{1i} + a_{11i}f_{1t} + a_{12i}f_{2t} + e_{1it} \text{ and } x_{2it} = c_{2i} + a_{21i}f_{1t} + a_{22i}f_{2t} + e_{2it},$$

where the error terms are supposed to be white noise processes, i.e. $e_{1it} \sim N(0, 0.4)$ and $e_{2it} \sim N(0, 0.6)$.

The first common factor f_{1t} is stationary and is generated as:

$$f_{1t} = t + v_{1t} \text{ with } v_{1t} \sim N(0, 0.3),$$

while the second one, f_{2t}, is supposed to be nonstationary:

$$f_{2t} = \log(t) + v_{2t} \text{ with } v_{2t} = v_{2t-1} + u_t, \text{ and } u_t \sim N(0, 0.1)$$

The individual effects and the factor loadings are drawn from Gaussian distributions as follows:

$$
\begin{aligned}
\gamma_{1i} &\sim N(0.8, 1), \\
\gamma_{2i} &\sim N(0.4, 0.4), \\
a_{11i} &\sim N(0.6, 0.8), \\
a_{12i} &\sim N(0.9, 0.5), \\
a_{21i} &\sim N(1, 0.8), \\
a_{22i} &\sim N(0.6, 0.6), \\
c_{1i} &\sim N(80, 1), \\
c_{2i} &\sim N(10, 0.5), \\
c_{3i} &\sim N(0, 0.3), \text{ and} \\
\alpha_i &= c_{1i} + c_{2i} + c_{3i}
\end{aligned}
$$

We fix $T = 44$ and $N = 24$ to get the same sample dimension as in the empirical analysis. As in the main analysis, we fix the length of the blocks to eleven years and we draw with replacement from $44 - 11 + 1 = 34$ blocks. The number of draws S is fixed to 1000. Figure A1 shows the ECDFs of the MAPEs (1-year forecast horizon) for each DGP, where, for estimation purposes, we employed the parametric CCE, the semiparametric (additive) CCE approach and the fully nonparametric (non-additive) CCE approach. To significantly reduce the computation time, we only consider PRS.

The results are clear-cut and provide evidence regarding how well the model choice procedure performs when selecting the right DGP underlying observed data in presence of nonlinearities. Indeed, when the underlying DGP is fully nonparametric (resp. semiparametric) the MAPE of the non-additive (resp. additive) penalized estimator is stochastically dominated by the MAPEs of others models, with the parametric estimator that clearly provides the worst forecasting performances. Moreover, results indicates that the gain in terms of forecasting of employing tensor splines with respect to univariate (additive) splines, when the underlying functional relation is non-additive, is much bigger than the gain that is obtained by using univariate splines rather than tensor splines, when the underlying DGP is additive.

When the underlying DGP is characterized by a linear functional relationship, all three estimators generate very similar forecasting performances. The parametric specification being a special case of the additive semiparametric and fully nonparametric ones, this result is not surprising. Indeed, if the underlying DGP is linear and parametric, the estimator of the function $s(x_1, x_2)$ in the nonlinear and non additive specification can be expected to be an additive function with linear components at x_1 and x_2. Therefore, the choice of one of the two nonlinear specifications must be completed by a detailed analysis of the additivity and linearity of the resulting estimated functions. Poor performances of the parametric specification when the underlying DGP is either non linear and additive or nonlinear and non-additive could be similarly motivated by the fact that nor the additive semiparametric or the fully nonparametric specifications are special cases of the linear parametric one. Nevertheless, this last issue would deserve a more in-depth analysis.

===== Insert Figure A1 =====

Table 6: R&D elasticities

Value of domestic R&D fixed at 10% quantile				
		Values of Foreign R&D fixed at		
		Q10	Median	Q90
Values	Q10	0.0342**	0.04601***	0.0616
of		(0.0160)	0.0092)	(0.0276)
Human	Median	0.0393***	0.0320***	-0.0193
Capital		(0.0137)	(0.0086)	(0.0170)
fixed	Q90	0.0876***	0.0921***	0.0241
at :		(0.0229)	(0.0215)	(0.0230)
Value of domestic R&D fixed at median				
		Values of Foreign R&D:		
		Q10	Median	Q90
Values	Q10	0.0083	0.0630***	0.2001***
of		(0.0276)	(0.0185)	(0.0634)
Human	Median	0.0154	-0.0188	-0.0251*
Capital		(0.0210)	(0.0130)	(0.0129)
fixed	Q90	-0.0076	0.0354**	0.0697***
at:		(0.0249)	(0.0194)	(0.0201)
Value of domestic R&D fixed at 90% quantile				
		Values of Foreign R&D:		
		Q10	Median	Q90
Values	Q10	-0.0176	-0.0564	-0.0114
of		(0.0479)	(0.0499)	(0.1017)
Human	Median	0.1619***	0.1977 ***	0.2128***
Capital		(0.0351)	(0.0316)	(0.0469)
fixed	Q90	0.1459***	0.0572*	0.0125
at:		(0.0377)	(0.0302)	(0.0281)
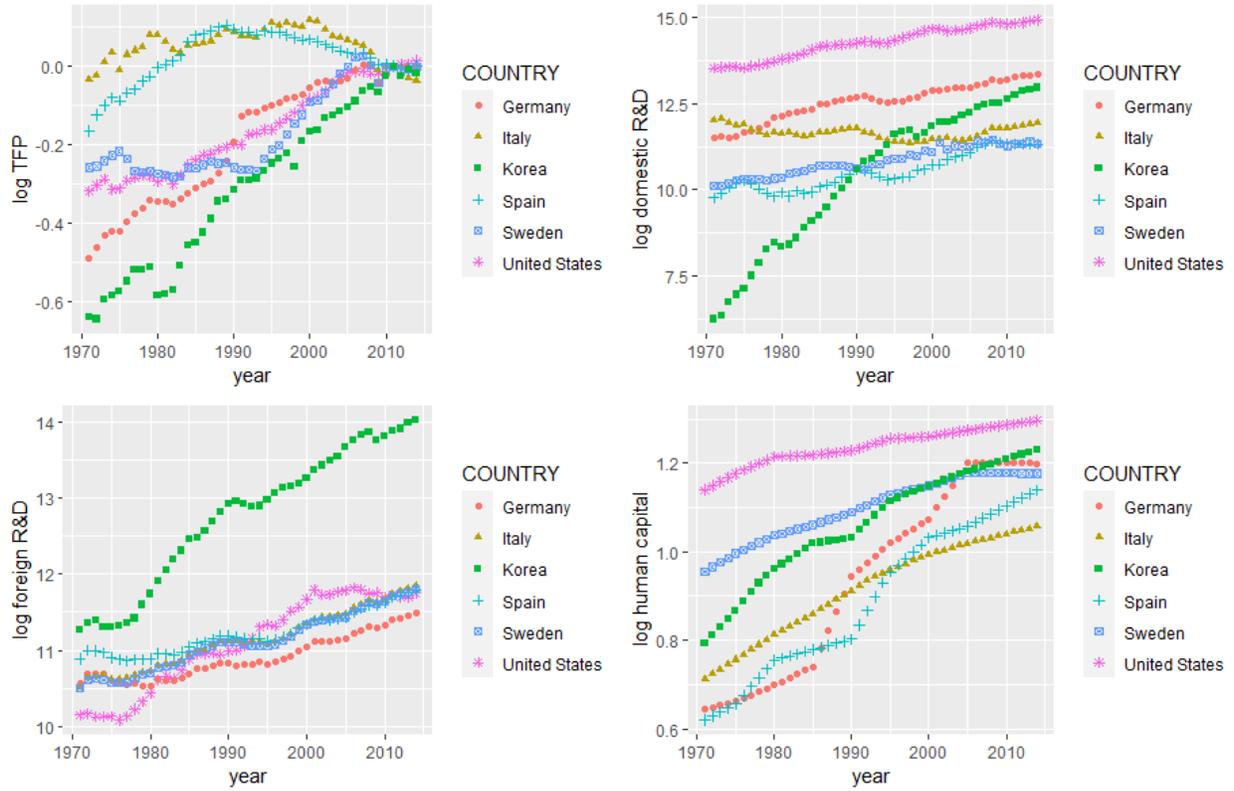
(a)
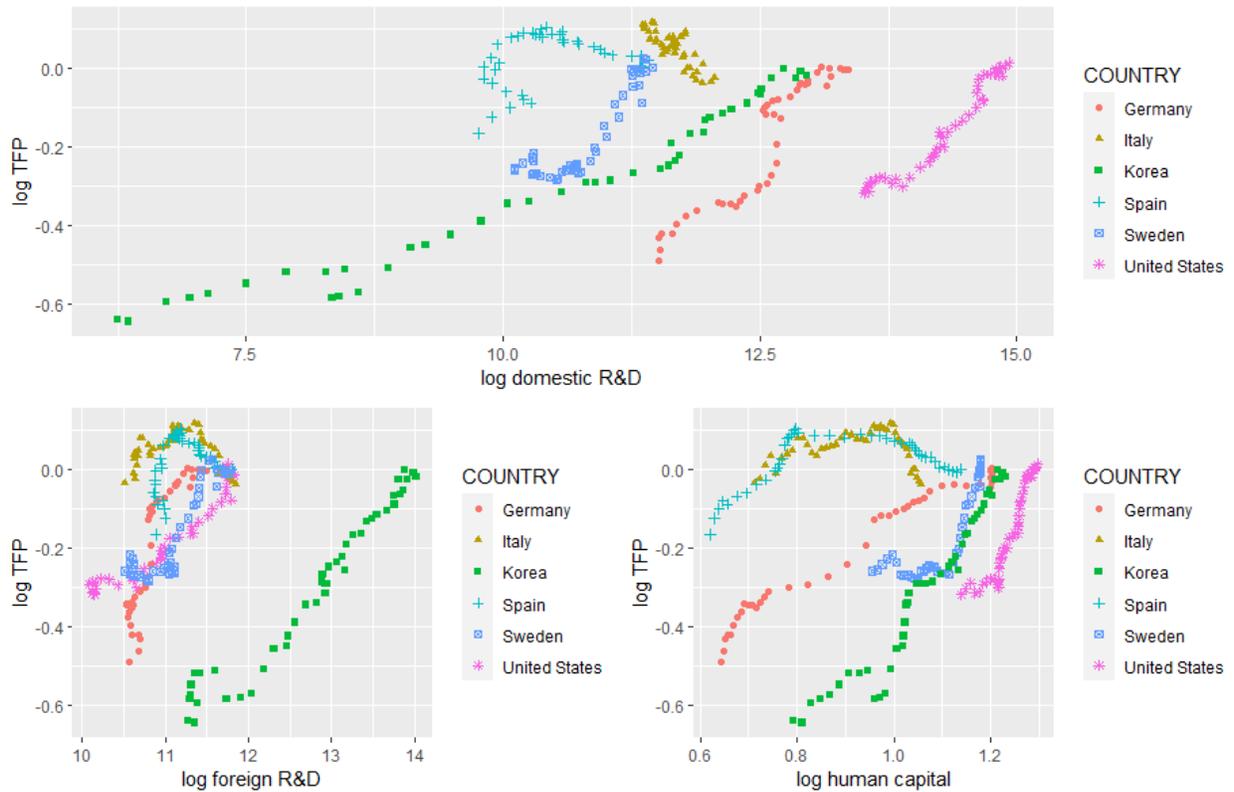
(b)

Figure 1: Univariate and bivariate plots
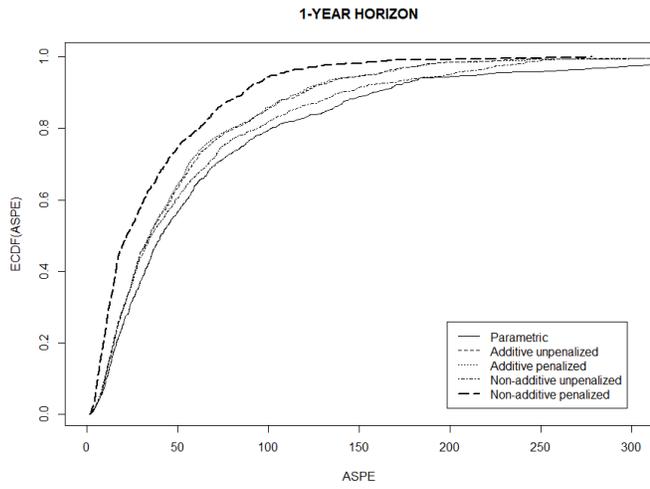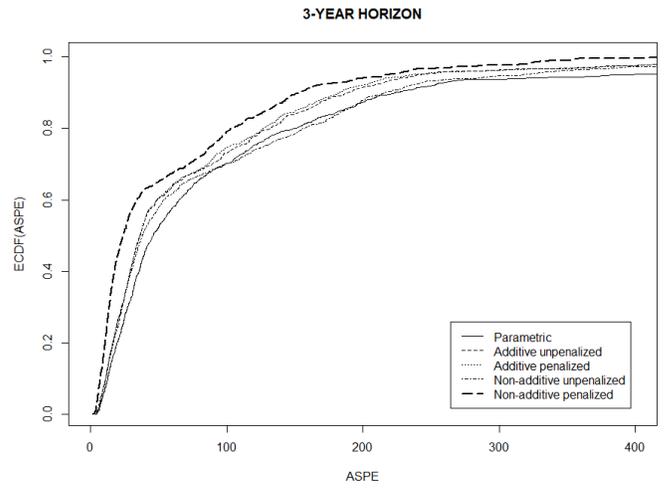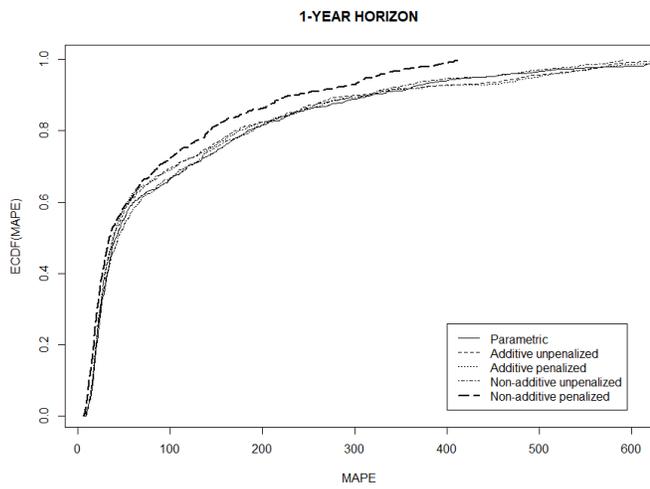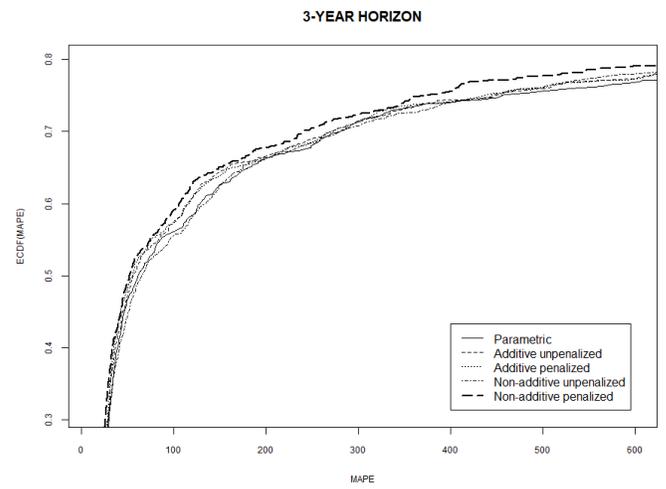
Figure 2: Empirical Cumulative Distribution Functions (ECDFs) of the ASPE for different factor models: the linear, the additive and the non-additive models for the OECD data. (a) and (b) refer to 1-year and 4-year horizon, respectively.

Figure 3: TFP as a function of Domestic and Foreign R&D

Human capital at 10% quantile
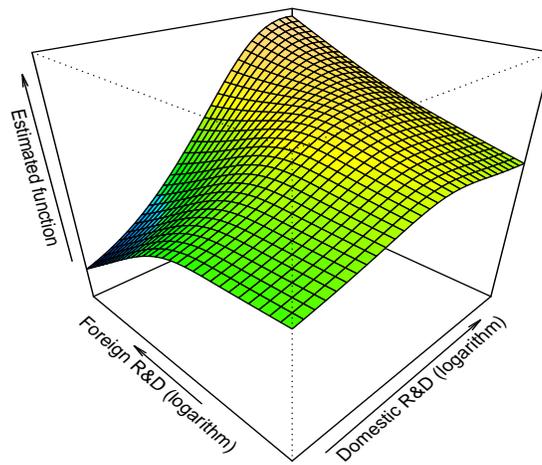
(a)

Human capital at median
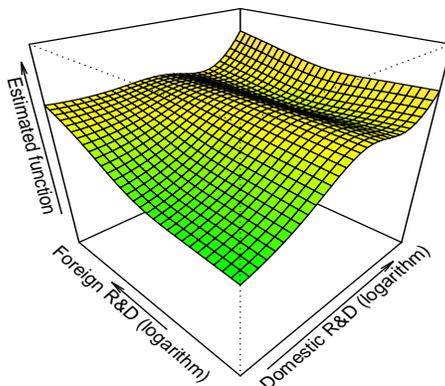
(b)

Human capital at 90% quantile
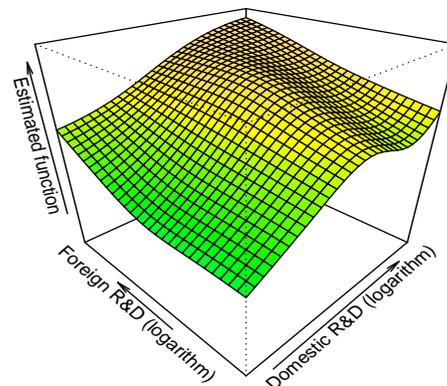
(c)

Figure 4: TFP as a function of Domestic R&D and Human Capital

Foreign R&D at 10% quantile
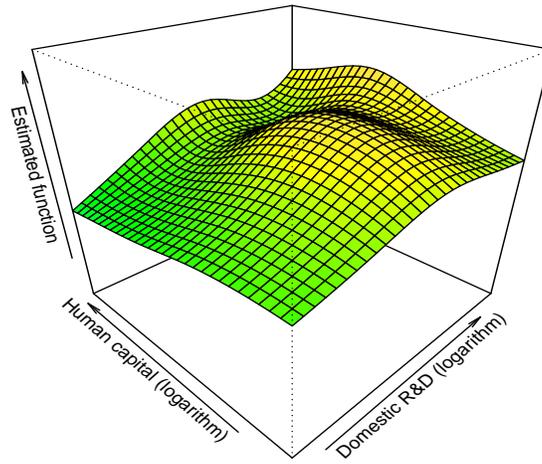
(a)

Foreign R&D at median
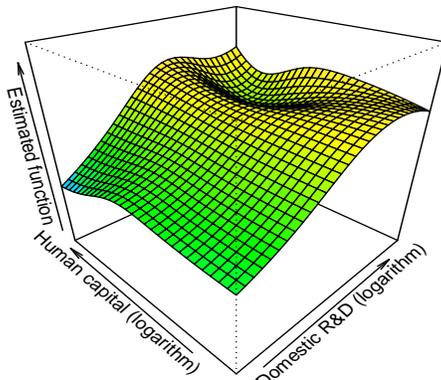
(b)

Foreign R&D at 90% quantile
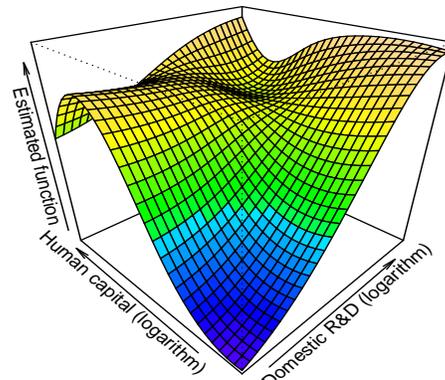
(c)

Figure 5: TFP as a function of Foreign R&D and Human Capital

Domestic R&D at 10% quantile
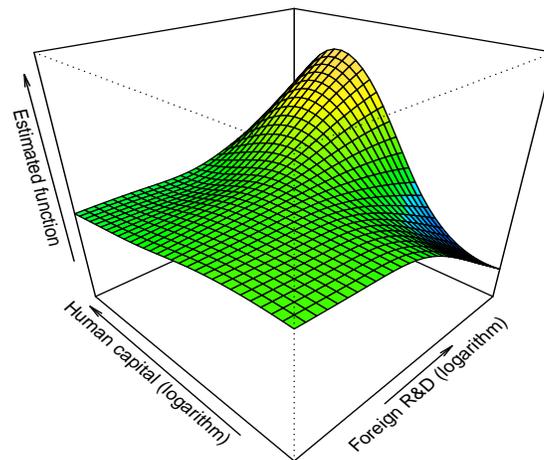
(a)

Domestic R&D at median
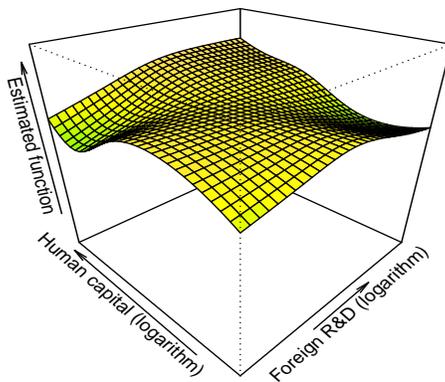
(b)

Domestic R&D at 90% quantile

(c)

43
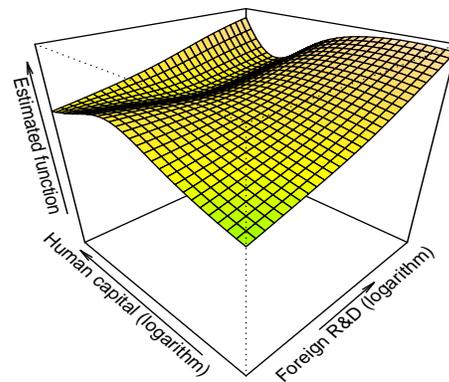
Figure 6: Estimated R&D Elasticities
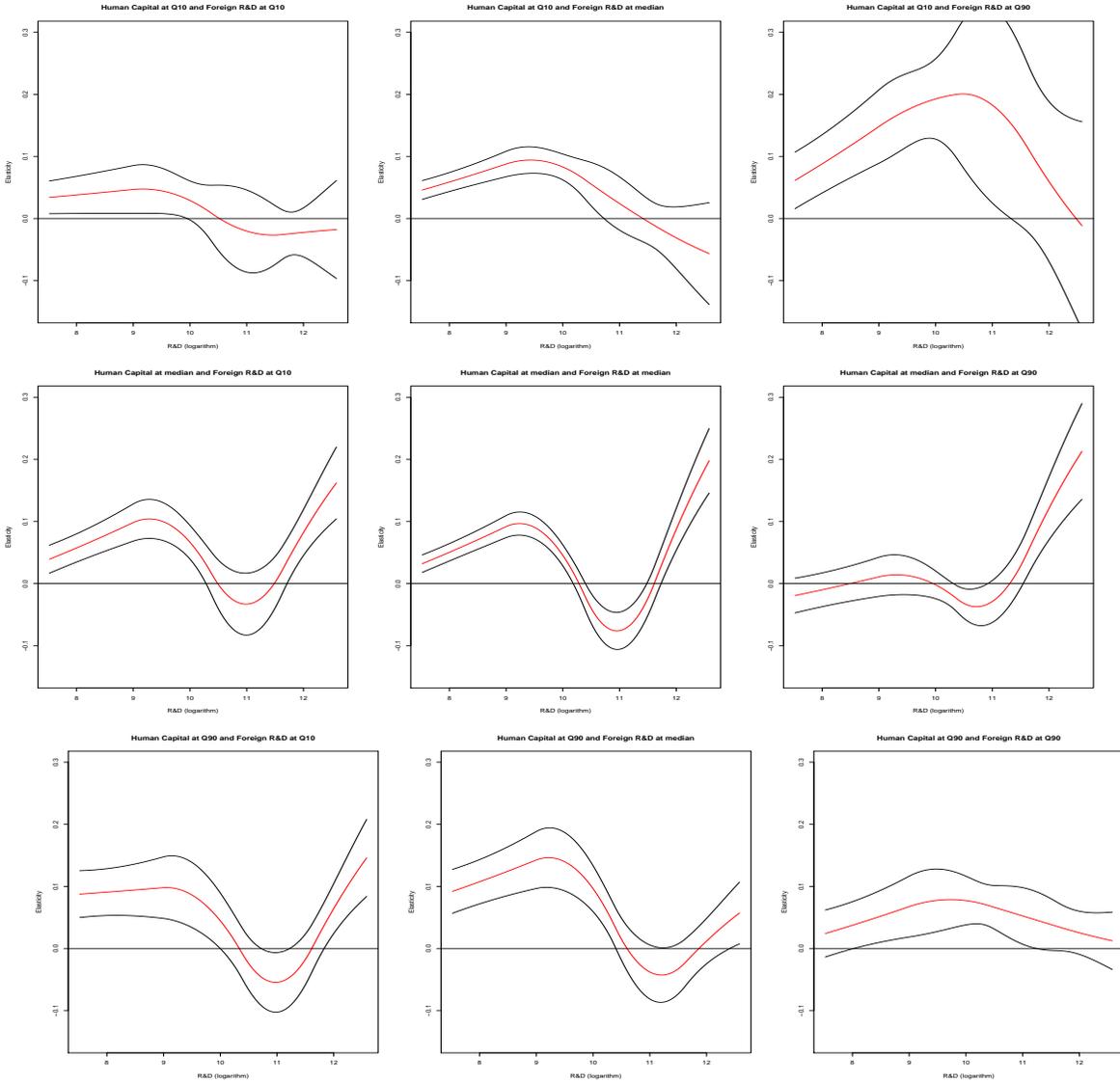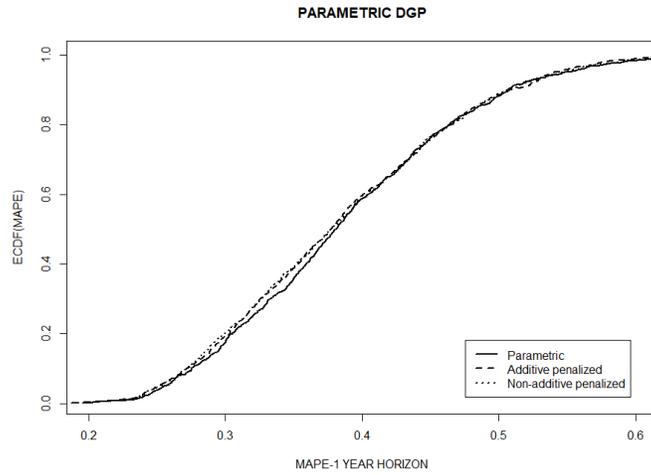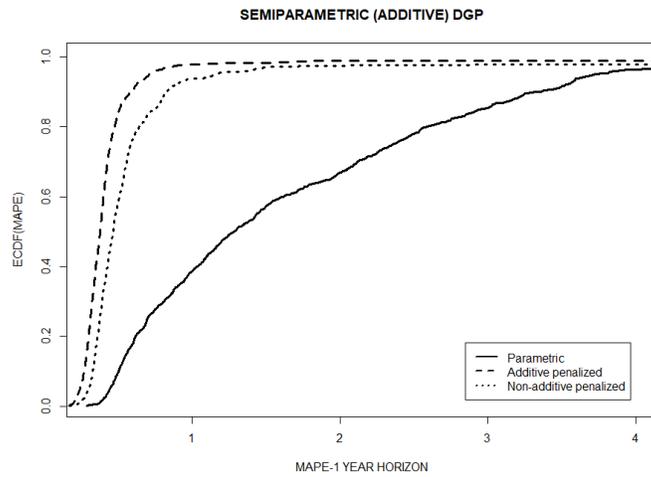
Figure 7: Estimated Foreign R&D Elasticities
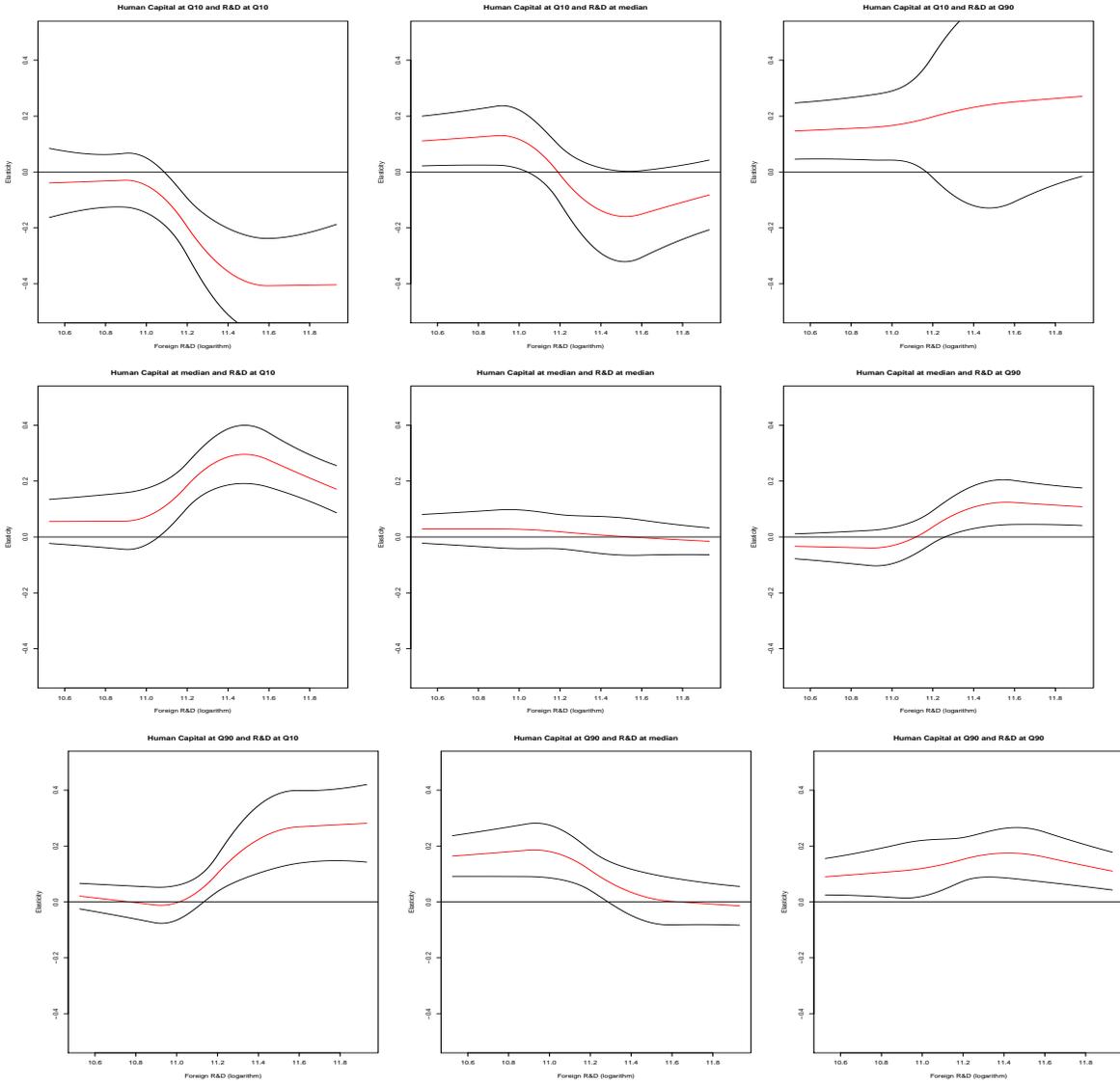
Figure 8: Estimated Human Capital Elasticities
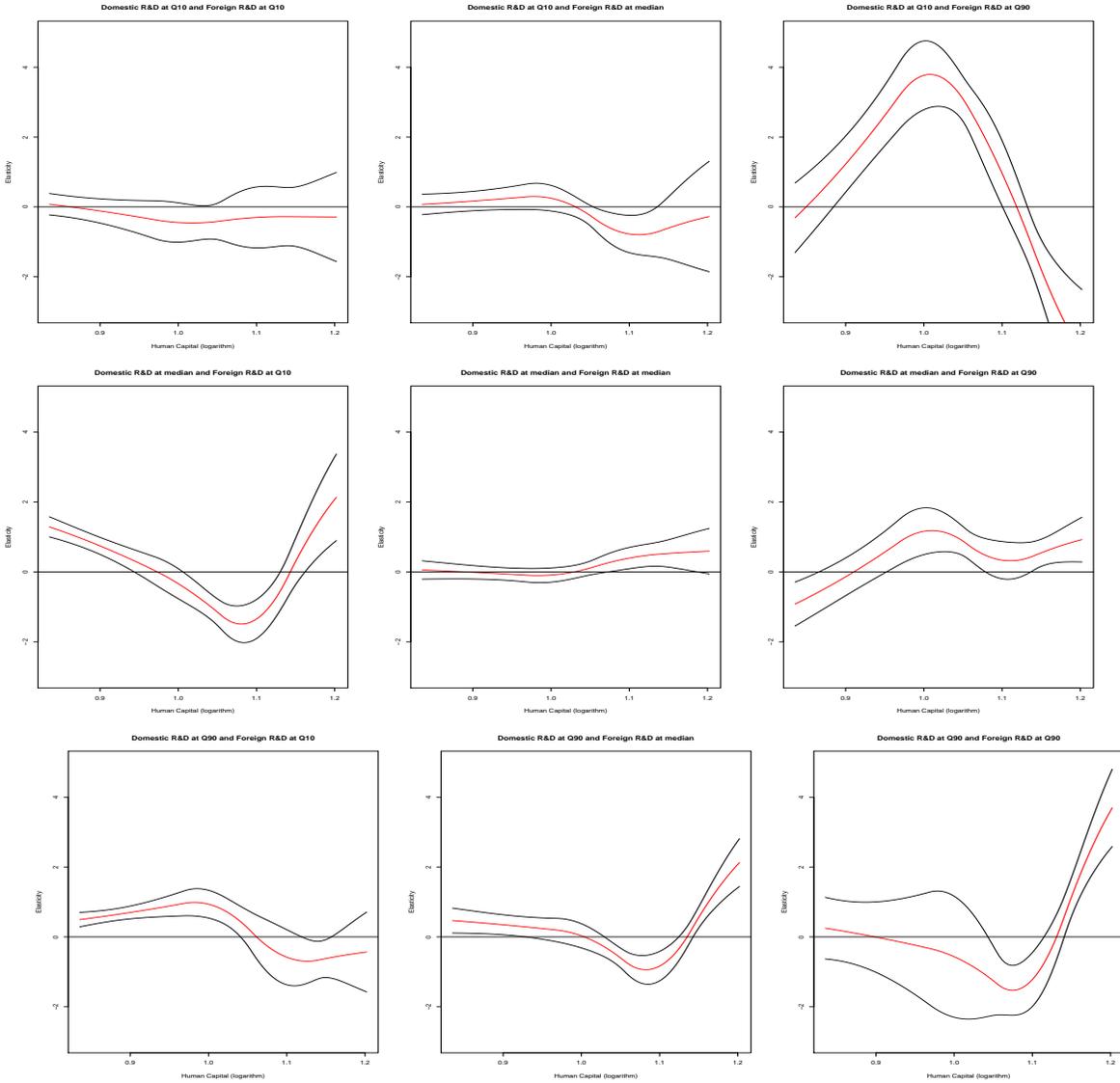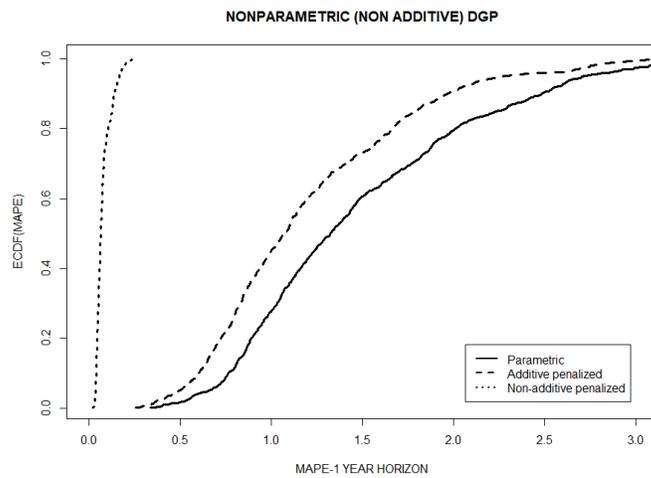
Figure A1: Empirical Cumulative Distribution Functions (ECDFs) of the MAPE for different DGPs: the linear, the additive and the non-additive generated models.