International R&D spillovers, absorptive capacity and relative backwardness: a panel smooth transition regression model

Andrea Fracasso

Department of Economics, University of Trento, via Inama 5, 38122 Trento, Italy

Giuseppe Vittucci Marzetti*

Department of Sociology ad Social Research, University of Milano-Bicocca, via Bicocca degli Arcimboldi 8, 20126 Milan, Italy

Abstract

We investigate how a country's absorptive capacity and relative backwardness affect the impact of international R&D spillovers on domestic Total Factor Productivity (TFP). To account for nonlinearities, we adopt a Panel Smooth Transition Regression approach, where a country's TFP elasticity to the foreign R&D stock is allowed to change smoothly across various identified extreme values, and the change is related to observable transition variables: human capital (capturing the country's absorptive capacity) and relative backwardness. The results suggest that absorptive capacity is positively associated with international R&D spillovers. In contrast with previous results, relative backwardness is instead found to have a negative and significant impact on international knowledge spillovers.

Key words: Absorptive capacity, International R&D spillovers, Nonlinear panel, Smooth Transition Regression, Total Factor Productivity

JEL Classification: C23, C24, F43, O30, O47

*Corresponding author: Department of Sociology ad Social Research, University of Milano-Bicocca, via Bicocca degli Arcimboldi 8, 20126 Milan, Italy Phone: +39 0264487457, Fax: +39 0264487561

Email addresses: andrea.fracasso@unitn.it (Andrea Fracasso), giuseppe.vittucci@unimib.it (Giuseppe Vittucci Marzetti)

1. Introduction

Several empirical and theoretical contributions (e.g. Grossman and Helpman, 1991; Rivera-Batiz and Romer, 1991; Aghion and Howitt, 1992; Eaton and Kortum, 1999; Howitt, 2000; Keller, 2004) have shown that the impact of international R&D spillovers on Total Factor Productivity (TFP) is non-negligible. In particular, most studies have focused on the specific channels through which foreign knowledge is transferred across countries and revealed that international trade is an important vehicle of transmission (e.g. Coe and Helpman, 1995; Coe et al., 1997, 2009; Keller, 1998, 2002; Guellec and Van Pottelsberghe de la Potterie, 2004; Lumenga-Neso et al., 2005; Krammer, 2010; Franco et al., 2011; Fracasso and Vittucci Marzetti, 2012). Fewer studies have instead looked at the local conditions which make foreign knowledge appropriable and domestically implementable.

In fact, since the seminal work by Cohen and Levinthal (1989), there has been a widely-held consensus that international spillovers depend on the ability of the recipient country to identify, assimilate and exploit foreign knowledge. This ability has been called, following Cohen and Levinthal, absorptive capacity.[1] As argued by Keller (1996), absorptive capacity most likely depends on the country's stock of human capital for the labor skills determine the extent to which foreign knowledge is assimilated by the receiving country. Accordingly, one would expect international knowledge spillovers to have a greater impact on local productivity (or its growth) in those countries where human capital is more abundant.

In a similar vein, the literature on catching-up and knowledge transmission has tackled the issue of whether the relative backwardness of a country, i.e. its distance from the technological leader, can affect the impact of foreign knowledge flows on its growth performance. Countries closer to the technological frontier could either benefit little from the accumulation of knowledge in relatively less developed foreign countries, or rather make the best out of foreign improvements. On the one hand, developed countries could have little external knowledge to absorb because they are already at the frontier (Gerschenkron, 1962). On the other hand, relative backwardness could make it more difficult for the laggard countries to borrow the technology developed elsewhere (Matthews, 1969).[2] This issue has not been fully settled at the empirical level, as indicated by the contrasting conclusions of Crespo-Cuaresma et al. (2004), on the one hand, and Kneller (2005) and Falvey et al. (2007), on the other.

In this paper, we empirically investigate how absorptive capacity and relative

---

[1]The intuition dates back to Abramovitz (1986), who argued that "social capability" affects the different strength of the catching-up processes across countries and over time.

[2]The latter case recalls to mind the very concept of absorptive capacity: if laggard countries are hampered in making the best foreign knowledge flows by their technological backwardness, the latter could be interpreted as a measure of absorptive capacity. It should be noted, however, that while Abramovitz (1986) and subsequent authors link absorptive capacity to absolute levels of human capital, backwardness is a purely relative concept.

backwardness impact on the international knowledge diffusion to shed some light on the above mentioned predictions of the theory, to clarify the contrasting findings in previous works, and to provide a better approach to address the complex nonlinear relationship between international R&D and domestic TFP.

We are not the first to tackle these issues. As to the role of absorptive capacity, Kwark and Shyn (2006) analyzed its relationship with R&D spillovers by adding an interaction term between the human capital and the foreign knowledge stocks to a regression à la Coe and Helpman (1995). They estimated the specification in first difference using five-year changes on a sample encompassing both developed and developing countries, as in Coe et al. (1997), and found that human capital is important for the absorption of foreign knowledge, thereby corroborating the hypothesis that absorptive capacity matters in the international transmission of knowledge.[3]

In fact, among the various empirical findings that Kwark and Shyn (2006) reported, some clash against well established empirical results. For instance, when the interaction between human capital and foreign knowledge is added to the specification, the positive and significant elasticity of TFP with respect to the average years of schooling (i.e., a widely used measure of human capital) disappears. This is most likely due to the fact that a linear specification does not allow the TFP elasticity to foreign knowledge to reflect properly all possible nonlinear (country- and time-specific) effects. Nonlinearities, and in particular threshold effects, are very likely to exist. For instance, there might be a minimum level of human capital that is necessary for any country to assimilate foreign knowledge. Similarly, the marginal importance of absorptive capacity may not be constant and may start diminishing once a certain level of development is reached.[4] These examples call for a specification more flexible than a linear form with interaction terms; what is needed is a specification that allows the TFP elasticity to foreign knowledge to reflect more complex nonlinear effects.

In addressing how absorptive capacity affects the way international knowledge spillovers impact on medium-term output growth, Crespo-Cuaresma et al. (2004) and Falvey et al. (2007) acknowledged the possible presence of nonlinear effects. They adopted a Threshold Regression (TR) model, where the knowledge absorption parameters can change across two regimes: one associated with the average years of secondary schooling above a critical value, the other one with the values below it. Albeit informative, these studies do not allow to conclude much about the precise impact of absorptive capacity on how international spillovers

---

[3]In a similar vein, Kneller (2005) investigated the impact of absorptive capacity on international R&D spillovers at the sectoral level in a sample of 12 developed economies over 1972–1992. In his empirical specification, the highest level of productivity in the sample is adopted as a measure of foreign knowledge, following Benhabib and Spiegel (1994), and three proxies of absorptive capacity are linearly interacted with it. This is in contrast with the literature on R&D spillovers à la Coe and Helpman (1995), which we follow, where spillovers depend on foreign R&D stocks. Similar considerations hold for Madsen et al. (2010), where foreign knowledge is not considered at all.

[4]Focusing on different manufacturing sectors, Girma (2005), for instance, found evidence that the impact of FDI on TFP is affected by absorptive capacity in a nonlinear fashion.

affect productivity. First, the dependent variable in these studies is the growth rate of GDP per capita, rather than TFP. Second, TR models are restrictive in so far as they allow the parameters to change only across a limited number of regimes and in a dichotomous fashion.[5]

As said, also the relationship between relative backwardness and growth remains theoretically unclear, particularly because it is very likely to be characterized by nonlinearities as well. This point has been recognized in some previous empirical works. In fact, as shown by the contrasting results in Benhabib and Spiegel (1994), Falvey et al. (2007), and Mancusi (2008), the role played by relative backwardness in the international transmission of knowledge still remains an open empirical issue.

In this work, we investigate the effects of both absorptive capacity and relative backwardness on international knowledge diffusion. In so doing, we achieve several goals: we address the limitations of the empirical specification adopted by Kwark and Shyn (2006); we revisit the analysis of Crespo-Cuaresma et al. (2004) and Falvey et al. (2007) in an estimation framework that follows more closely the seminal work of Coe and Helpman (1995) on international R&D spillovers (rather than those in the GDP convergence literature); finally, we adopt a flexible estimation strategy so as to capture complex, economically-meaningful, nonlinear effects. More precisely, we study how absorptive capacity and relative backwardness affect the impact of international R&D spillovers on TFP by means of a general nonlinear regression method applied to Coe and Helpman's (1995) specification with human capital (as in Engelbrecht (1997) and Coe et al. (2009)). The Panel Smooth Transition Regression (PSTR) approach we adopt allows relaxing the hypotheses of homogeneity and time-stability of the parameters in a convenient and flexible way: the parameter of interest (i.e, the elasticity of TFP to foreign knowledge) is let free to change smoothly across various identified regimes, and the transition across them is related to meaningful observable variables—in this study, variables proxying absorptive capacity and relative backwardness—which vary across countries and over time.

This nonlinear technique, albeit only recently extended by González et al. (2005) to panel data models, has been already employed in other economic fields with interesting results (see, for instance, Fok et al., 2005; Fouquau et al., 2008; Béreau et al., 2010; Delatte and Fouquau, 2011; Alcidi et al., 2011, for time series applications). However, in order to account better for the econometric issues arising from the identification problem in testing the hypothesis of linearity, we adopt state-of-the-art econometric techniques to perform reliable inference (e.g. González and Teräsvirta, 2006; Hurn and Becker, 2009). Furthermore, to make inference more robust and to account for unspecified forms of heteroskedasticity and serial and simultaneous correlation in the data,

---

[5]In addition, as they looked at five-year non-overlapping average values, these studies focused on long-term effects, while in this paper we exploit yearly data to unveil short-term effects. Another difference of this work with Crespo-Cuaresma et al. (2004) is that the latter does not include human capital among the regressors. This is at odds with what usually done after Engelbrecht (1997).

we combine Bravo and Godfrey's (2011) double bootstrap method with the panel moving blocks bootstrap recently proposed by Gonçalves (2011). To the best of our knowledge, this approach has never been used before in applied works and represents a further contribution of the paper.

To anticipate our results, we find that absorptive capacity and relative backwardness are, respectively, positively and negatively related with the impact of the trade-weight measure of foreign knowledge on domestic TFP. Given our sample of OECD countries, this entails that trade-related international R&D spillovers are influenced by the conditions of the recipient countries even if the latter are developed economies. We also find that the elasticity of TFP to foreign knowledge varies considerably across countries and over time. Finally, our findings suggest that failing to account for nonlinearities, as done in much of the previous literature, leads to the overestimation of the direct impact of domestic R&D and human capital on TFP.

The paper proceeds as follows. In Section 2, we illustrate the PSTR model and discuss the methodological issues regarding the tests of (no remaining non) linearity (2.2.1 and 2.2.3), the model estimation (2.2.2), and the bootstrap methods used to robustify inference (2.3). We present and discuss our findings in Section 3. In Section 4 we conclude.

2. Empirical methodology

2.1. Specification

In their seminal paper, Coe and Helpman (1995) adopted a specification that is now the workhorse of the empirical studies in the literature on international R&D spillovers:

$$\ln F_{it} = \alpha_i + \beta^d \ln S_{it}^d + \beta^f \ln S_{it}^f + \epsilon_{it} \tag{1}$$

where i is the country index, t is the time index, $\ln F_{it}$ is the log of domestic TFP, S_{it}^d the domestically produced R&D stock, and S_{it}^f an import-weighted sum of the R&D stock produced abroad (i.e., $S_{it}^f = \sum_{j \neq i} \frac{M_{ijt}}{\sum_{j \neq i} M_{ijt}} S_{jt}^d$, where M_{ijt} is the import of country i from country j at time t). Engelbrecht (1997) showed that human capital affects domestic productivity and should accordingly be included in the model. Equation (1) is thus modified as follows:[6]

$$\ln F_{it} = \alpha_i + \beta^d \ln S_{it}^d + \beta^h \ln H_{it} + \beta^f \ln S_{it}^f + \epsilon_{it} \tag{2}$$

where H_{it} is human capital, proxied by the average years of schooling in country i at time t.

Such specification clearly imposes that the elasticities, that is the β coefficients, are constant across countries and over time. As said in the previous section, this

---

[6]Lichtenberg and van Pottelsberghe de la Potterie (1998) observed that, to weight foreign R&D stocks, one should not use import, but export, shares (i.e. the ratio of bilateral imports over the GDP of the exporting country). As shown by Coe et al. (2009), this modification does neither invalidate nor does it weaken what found with the specification of equation (1).

might be unwarranted as the recipient country's conditions may play a role in the knowledge diffusion process and nonlinear effects may be important. The simple inclusion of interacting terms between absorptive capacity and foreign R&D stock (as done for instance in Kwark and Shyn, 2006) might not address properly the possible nonlinearities in the relationship. Indeed, the inclusion of quadratic and interaction terms might be well suited to model U-shaped relations, but fail to account properly for more complex nonlinear effects. The TR model proposed by Hansen (1999b) represents a better solution, as it allows the parameters to change according to specific economic conditions: the regression coefficients take a discrete number of different values (so called "regimes"); the switch of a cross-sectional unit from one regime to another is not probabilistic, but driven by the fact that a chosen "transition variable" is above or below a given identified critical value (the "threshold"). Though more flexible than the linear estimation, this approach has limitations too: in particular, it imposes abrupt transitions across regimes (i.e. sudden jumps in the coefficients when the transition variable crosses the identified thresholds), so that the groups of observations can be clearly distinguished as belonging to either one regime or another. This excludes any gradual variation of the parameters as well as any smooth transition of the cross-sectional units from one regime to another.

To address these limitations, we adopt the PSTR model developed by González et al. (2005) and Fok et al. (2005) following the work of Granger and Teräsvirta (1993) on Smooth Transition Autoregressive (STAR) models. This approach allows the parameters of interest to change smoothly between the values associated with two (or more) extreme regimes. As in the TR model, the change of the coefficients is driven by a transition variable, but in the PSTR model the transition can be smooth. Both the threshold level (here called "location parameter") and the smoothness of the transition can be estimated from the data. Since the transition variable is time-varying and unit specific, the regression coefficients for each of the cross-sectional units in the panel change gradually over time. Thus, the coefficients are not forced to assume at each point in time and for each unit the values associated with either of the extreme regimes, but they are let free to vary within them. The adoption of a bounded and continuous (typically, a logistic) function of the transition variable to guide the fluctuations of the coefficients across the extreme regimes guarantees their gradual variation (in contrast with the dichotomous switches superimposed in the TR model).[7]

In this paper, we focus on how absorptive capacity and relative backwardness can affect the elasticity of TFP to foreign R&D spillovers and we aim at accounting for those possible nonlinear effects not fully explored in the existing literature on international R&D spillovers. More precisely, we estimate a PSTR model to deal with the potential heterogeneity and time-instability of the coefficients and the nonlinear impact of absorptive capacity and relative backwardness on

---

[7]Differently from random coefficient models (Hsiao and Pesaran, 2004), PSTR makes the transition across regimes depend on an observable variable.

them. Accordingly, absorptive capacity (measured as the lagged value of the country's human capital) and relative backwardness (measured as the gap in GDP per capita in PPP with respect to the leading country) are the variables that drive—in line with the theoretical arguments summarized in Section 1—the transition across regimes.

We modify equation (2) so as to have a PSTR model and thus adopt the following specification:

$$\ln F_{it} = \alpha_i + \beta^d \ln S_{it}^d + \beta^h \ln H_{it} + \beta_{it}^f \ln S_{it}^f + \epsilon_{it} \tag{3}$$

where

$$\beta_{it}^f = \beta_0^f + \sum_{j=1}^{r} \beta_j^f \, g(q_{it}^{(j)}; \gamma_j, c_j) \tag{4}$$

and

$$g(q_{it}^{(j)}; \gamma_j, c_j) = \frac{1}{1 + \mathrm{e}^{-\gamma_j(q_{it}^{(j)} - c_j)}}. \tag{5}$$

Equation (4) shows that the time- and country-varying elasticity of TFP to foreign R&D stock is a weighted average of the coefficients associated with the $r + 1$ regimes, with weights given by the equations (5). These weights are logistic functions of a transition variable $q_{it}^{(j)}$—that is, the variable that guides the transition across regimes (e.g. the proxies of absorptive capacity and relative backwardness)—, where c_j is a location parameter (i.e., the critical level separating two contiguous regimes) and the parameter $\gamma_j \, (> 0)$ determines the smoothness of the transition across the regimes.

When $\gamma_j \to \infty$, for all j, the r transition functions $g(.)$ become indicator functions and the model reduces to a simple panel TR model. On the contrary, when $\gamma_j \to 0$, the r functions $g(.)$ become constant and the model collapses to a simple panel linear regression model with fixed effects. The procedure provides for the estimation of all the parameters of interest in the model, including any γ_j and c_j, so that no *a priori* identification of the number and values of the (extreme) regimes, or the shape of the transitions between them, is required.

2.2. Estimation procedure and methodological issues

2.2.1. Testing for linearity

The first step in the procedure—thoroughly discussed in González et al. (2005)—is to test the linearity of equation (2) against a PSTR model with two regimes (equation (3) with $r = 1$) and a given candidate transition variable $(q_{it}^{(1)})$, that is:

$$\ln F_{it} = \alpha_i + \beta^d \ln S_{it}^d + \beta^h \ln H_{it} + \beta_0^f \ln S_{it}^f + \beta_1^f \, g(q_{it}^{(1)}; \gamma_1, c_1) \ln S_{it}^f + \epsilon_{it} \tag{6}$$

Notably, if more candidate transition variables $q_{it}^{(1)}$ exist, the procedure is repeated for each of them: the variable that leads to the strongest rejection of the null is selected as the transition variable for the estimation of equation (6).

Testing the null hypothesis of linearity is a non-standard problem because under the null there are unidentified nuisance parameters.[8] The identification problem can be solved in two ways. The first approach, proposed by Luukkonen et al. (1988), tests the null $\gamma_1 = 0$ with a m-order Taylor expansion of the nonlinear model around this point. An auxiliary regression is run:

$$\ln F_{it} = \alpha_i + \beta^d \ln S_{it}^d + \beta^h \ln H_{it} + \delta_0 \ln S_{it}^f + \sum_{p=1}^{m} \delta_p \, q_{it}^p \ln S_{it}^f + \nu_{it} \qquad (7)$$

where ν_{it} is the sum of the residuals of (6) and the remainder of the series expansion. The null hypothesis ($\delta_1 = \ldots = \delta_m = 0$) can be tested by using a (heteroskedasticity-robust) LM-test statistic. Under the null, the test statistic follows asymptotically a χ^2 distribution with m degrees of freedom. In small samples, the authors suggest to use the F-version of the LM test (LM_F) by dividing the latter by the number of restrictions.[9] Usually, a third-order Taylor approximation—$m = 3$ in equation (7)—is chosen.

The second approach, applied by Hansen (1999a,b, 2000) in the context of TR models, tests the null $\beta_1^f = 0$ and circumvents the identification problem by computing the supremum LR test statistic (SupLM). Andrews and Ploberger (1994), instead, suggested to use alternative statistics, i.e. AveLM, ExpLM or wLM, that are weighted averages of (heteroskedasticity-robust) LM-test statistics computed for several combinations of γ_1 and c_1 spanning the parameter space.[10] Given that these statistics have (asymptotically) pivotal but non-standard distributions, which depend also on the moments of the distribution of the nonlinear parameters and whose critical values cannot therefore be tabulated, to obtain the critical values one has to bootstrap the tests.

Hansen's testing approach was recently extended to STR models by González and Teräsvirta (2006), Hurn and Becker (2009) and Becker and Osborn (2010).

---

[8]Linearity follows imposing either $\beta_1^f = 0$ or $\gamma_1 = 0$. When the null is $\beta_1^f = 0$, c_1 and γ_1 are unidentified nuisance parameters. When the null is $\gamma_1 = 0$, the unidentified nuisances are c_1 and β_1^f.

[9]The F-version in our case is approximately F-distributed with m and $(TN - N - m - 3)$ degrees of freedom, where T is the time length of the panel and N the number of cross-sectional units.

[10]AveLM, ExpLM and wLM are, respectively, the average test statistic, the exponential average test statistic and the weighted average test statistic. They are defined as follows:

$$\mathrm{AveLM} = \int_{\gamma_1} \int_{c_1} \mathrm{LM}(\gamma_1, c_1) \, dW(\gamma_1, c_1)$$

$$\mathrm{ExpLM} = \ln \left(\int_{\gamma_1} \int_{c_1} \exp \left(\frac{1}{2} \mathrm{LM}(\gamma_1, c_1) \right) dW(\gamma_1, c_1) \right)$$

$$\mathrm{wLM} = \int_{\gamma_1} \int_{c_1} \omega(\gamma_1, c_1) \, \mathrm{LM}(\gamma_1, c_1) \, dW(\gamma_1, c_1)$$

where $W(\gamma_1, c_1)$ is the weight function that allocates weights on the pairs (γ_1, c_1), while $\omega(\gamma_1, c_1)$ is the weight function on LM, with weights proportional to the magnitude of the values of the LM statistic, for the test not to be too heavily influenced by redundant values of γ_1 and c_1, that may have a negative effect on its power.

In particular, González and Teräsvirta (2006) studied the finite sample properties of Andrews (1993) and Andrews and Ploberger's (1994) test statistics (SupLM, AveLM, ExpLM or wLM) and compared them with the Taylor expansion-based linearity test of Luukkonen et al. (1988) for STR models. They showed that AveLM, ExpLM or wLM are always more powerful than the SupLM and the Taylor expansion-based tests.

Hurn and Becker (2009) and Becker and Osborn (2010), instead, dealt with the problem of heteroskedasticity and the distortions the latter causes to the size of the test in small samples. Indeed, dealing with heteroskedasticity in nonlinearity tests can be problematic. On the one hand, neglecting heteroskedasticity may lead to reject the null of linearity when it is not the case; on the other hand, robustification can remove most of the test power as shown by Lundbergh and Teräsvirta (1998). To cope with this, Hurn and Becker (2009) computed heteroskedasticity-robust test statistics and calculate the critical values of the tests using fixed-design wild bootstrap (Gonçalves and Kilian, 2004). They showed via simulation that this leads to a significant reduction in the distortions of the test. Accordingly, we follow their approach and use heteroskedasticity-robust test statistics.[11]

2.2.2. Estimation of the PSTR specification

If the null of linearity is rejected, a two-regime PSTR model (equation (3) with $r = 1$) is estimated. The estimation is carried out minimizing a concentrated Sum of Squared Residuals (SSR) via Nonlinear Least Squares (NLS). The SSR is concentrated with respect to the fixed effects α_i and the linear coefficients β's applying a standard fixed effects estimator for panel data conditional on a given combination of the nonlinear parameters (c_1 and γ_1). The panel fixed effects estimates are recomputed at each iteration in the nonlinear optimization.[12]

Notably, to select the starting values of the nonlinear coefficients, we do not use a grid search over a limited number of values, as usually done in the literature, because this approach may easily lead to local minima in the estimation. Instead, following the suggestions of González et al. (2005) and González and Teräsvirta (2006), we implement and apply the Simulated Annealing (SA) algorithm proposed by Corana et al. (1987) (see also Goffe et al., 1994, for an application to M-estimation problems).[13]

---

[11]The alternative heteroskedasticity-robust bootstrap procedure, discussed in Hansen (1999a) for TR models, is able to preserve the observed heteroskedasticity, but it does not exactly reproduce the heteroskedastic pattern of the observed data.

[12]As noted by González et al. (2005), with normally distributed errors this estimation procedure is equivalent to the maximization of a concentrated log-likelihood.

[13]Simulated Annealing—so named as it resembles the process undergone by the atoms in a heated metal when it cools slowly—denotes a large class of probabilistic algorithms used to locate global minima/maxima of functions in large search spaces, when the problem is unmanageable using combinatorial or analytical methods. SA improves more standard iterative optimization algorithms by introducing the "Metropolis criterion": some steps are taken in the "wrong direction" with a certain probability, as they serve to better explore the possible solution space. The probability for this to happen decreases when one makes many iterations

2.2.3. Testing for no remaining nonlinearity

After the estimation of a two-regime PSTR model, it is necessary to test the hypothesis that it adequately captures the nonlinearities in the panel. We follow González et al. (2005) and perform a test of (no remaining) nonlinearity on the following specification:

$$
\begin{aligned}
\ln F_{it} = {} & \alpha_i + \beta^d \ln S_{it}^d + \beta^h \ln H_{it} + \beta_0^f \ln S_{it}^f + \beta_1^f \, g(q_{it}^{(1)}; \hat{\gamma}_1, \hat{c}_1) \ln S_{it}^f \\
& + \beta_2^f \, g(q_{it}^{(2)}; \gamma_2, c_2) \ln S_{it}^f + \epsilon_{it}
\end{aligned}
\tag{8}
$$

where $q_{it}^{(1)}$ is the transition variable driving the fluctuation of the parameters between the two previously estimated extreme regimes, and $\hat{\gamma}_1$ and $\hat{c}_1$ are the previously estimated values of the nonlinear parameters. $q_{it}^{(2)}$ is instead an additional (candidate) transition variable, potentially associated with nonlinearities that are not captured by the two-regime PSTR model with $q_{it}^{(1)}$ as transition variable.[14] The rejection of the null at this stage implies that the variation of the parameters of the model is not fully captured by a two-regime PSTR model. This suggests that the parameters should be let change across time and individuals in a more complex way. More precisely, in the case of a three-regime model with two transition variables ($q_{it}^{(1)}$ and $q_{it}^{(2)}$), the individual and time-varying parameters become weighted averages of the three β^f parameters characterizing the three extreme regimes.

Proceeding as before, this test of (no remaining) nonlinearity is performed testing either $\gamma_2 = 0$ with the LM_F test statistic on the Taylor-based expansion around this point, or testing $\beta_2^f = 0$ by computing AveLM, ExpLM or wLM test statistics. Also in this case, we pursue both testing methods. When the null hypothesis of no remaining nonlinearity is rejected for more than one of the alternative candidate transition variables (the $q_{it}^{(2)}$'s), these tests are used to select among them the most appropriate transition variable for the additional nonlinear regime. As before, the candidate transition variable associated with the lowest p-value/highest value of the test statistic is chosen for the estimation of the three-regime model.

---

observing no significant improvements in the solution (this is regulated by the decrease in the "temperature", again by analogy with the annealing of a metal). Corana et al.'s (1987) algorithm is just one of the many proposed in the literature. See, for instance, Otten and van Ginneken (1989).

[14]The transition $q_{it}^{(2)}$ may be also equal to $q_{it}^{(1)}$. If this is the case, there are two critical values (c_1 and c_2) of $q_{it}^{(1)}$ associated with the three (extreme) regimes. To gain some intuition, consider an example in which the level of development of a country is allegedly associated with some nonlinear effects of debt on growth. Let us assume that: i) there are three regimes ($r = 2$); ii) the only transition variable is the country's per capita income ($q_{it}^{(1)} = q_{it}^{(2)}$); iii) transitions across regimes are rather sharp (i.e. γ_1 and γ_2 are quite high), so that the PSTR can be conveniently approximated by a TR model. Then the three-regime model could imply that the effect of debt on growth changes across time and individual countries depending on whether the country has a low, middle or high income per capita. The location parameters c_1 and c_2 would then be the thresholds distinguishing the three (income-related) regimes.

Following a sequential procedure, as in González et al. (2005), we generalize the test to a generic number of regimes to determine the number of transitions in the model. After the estimation of a model with $r + 1$ regimes, we perform a nonlinearity test on:

$$
\ln F_{it} = \alpha_i + \beta^d \ln S_{it}^d + \beta^h \ln H_{it} + \beta_0^f \ln S_{it}^f + \sum_{j=1}^{r} \beta_j^f \, g(q_{it}^{(j)}; \hat{\gamma}_j, \hat{c}_j) \ln S_{it}^f \\
+ \beta_{r+1}^f \, g(q_{it}^{(r+1)}; \gamma_{r+1}, c_{r+1}) \ln S_{it}^f + \epsilon_{it}
\tag{9}
$$

where the null is $\gamma_{r+1} = 0$ or $\beta_{r+1}^f = 0$. If it is rejected, we estimate a $(r + 2)$-regime PSTR model with the transition variable for the $(r+2)$-th regime leading to the strongest rejection of the null. We continue adding regimes until the first acceptance of the null of no remaining nonlinearity.[15]

In estimating models with more than two regimes, we first search via the SA the starting values of the nonlinear parameters for the additional regime keeping constant those of the previously estimated regimes. Then, we let the gradient-based algorithm freely search the entire parameter space for the combination of the nonlinear coefficients that minimizes the concentrated SSR. By doing so, we can also check whether the addition of a further regime affects the estimates of the nonlinear parameters of the other (previously estimated) regimes. Moreover, since it is not desirable for a regime to be estimated with only few observations, we check that the estimated location parameters are within the 5–95th percentiles of the sample values of the transition variables and that they are not too close each other, so that each regime can be estimated using at least 5% of all the observations.[16]

2.3. Robust inference

In order to achieve asymptotic refinements and account for the presence of unspecified forms of heteroskedasticity, and serial and simultaneous correlation in the data, we build on the recent contributions of Gonçalves (2011) and Bravo and Godfrey (2011) to estimate the statistical significance of the linear coefficients β's conditional on the nonlinear parameters γ_j and c_j $(j = 1, \ldots, r)$ in a way that, to our knowledge, has never been used. More precisely, to perform the t

---

[15]To avoid excessively large models, González et al. (2005) suggested to adjust the initial significance level α of the test, multiplying it by a factor τ $(0 < \tau < 1)$ after every regime addition.

[16]As pointed out by an anonymous referee, this restriction on the values that the location(s) can take has the side-effect of preventing, especially in the case of very smooth transitions, the detection of location(s) in the tails of the distribution of the transition variables. Given the need of a sufficiently large number of observations in each regime for the estimation of its parameters, in the literature the 5–95th percentiles of the transition variable are typically used to determine the boundaries of the location parameters. The size of our panel would allow for percentiles slightly lower (higher) than the 5th (95th) as each 5%-tail contains 40 observations. That being said, we follow the standard approach for the starting values and we shall discuss the implications of relaxing these limiting values in the estimation.

tests, we combine Bravo and Godfrey's (2011) double bootstrap method—that uses (first- and second-level) Moving Blocks Bootstrap (MBB) (Liu and Singh, 1992) and quasi-estimators (Hu and Zidek, 1995)—with the Panel MBB (PMBB) recently put forward by Gonçalves (2011) in the context of large n, large T balanced panels with weak time series dependence (of the mixing type) and (weak or strong) cross-sectional dependence.

In particular, to calculate the p-value of the t statistic for any coefficient β, we first compute the panel fixed-effects estimator $\hat{\beta}$. Using the actual data, we then generate B first-level bootstrap samples by applying Gonçalves's (2011) PMBB, that is applying the standard MBB to all the individual observations at each point in time. On each bootstrap sample, we compute the bootstrap fixed-effects estimator $\hat{\beta}_b^*$, the quasi-estimator $\tilde{\beta}_b^*$ and the discrepancy $(\hat{\beta}_b^* - \hat{\beta})$.[17] For each first-level bootstrap sample, we generate D second-level bootstrap samples by applying PMBB to the first-level sample and compute D quasi fixed-effects estimates $\tilde{\beta}_{bd}^{**}$[18] and their sample variance $\tilde{C}_b^{**} = \sum_{d=1}^{D} (\tilde{\beta}_{bd}^{**} - \bar{\tilde{\beta}}_b^{**})^2/D$, where $\bar{\tilde{\beta}}_b^{**} = \sum_{d=1}^{D} \tilde{\beta}_{bd}^{**}/D$. Then, we compute the B bootstrap t statistics $t^* = (\hat{\beta}_b^* - \hat{\beta})/\sqrt{\tilde{C}_b^{**}}$. To calculate the bootstrap p-value, we compute the fraction of these statistics greater than $t = \hat{\beta}/\sqrt{C^*}$, with C^* being the variance of the first-level bootstrap quasi-estimator $\tilde{\beta}_b^*$.[19]

For the nonlinear parameters, we instead use the robust quasi-ML sandwich estimator.[20] Contrary to TR models, in STR models the estimates of the location parameters c_j are asymptotically normally distributed and conventional hypothesis testing is possible.[21]

---

[17] The quasi-estimator is computed as follows:

$$\tilde{\beta}_b^* = \hat{\beta} + (\dot{\boldsymbol{X}}' \dot{\boldsymbol{X}})^{-1} \dot{\boldsymbol{X}}_b^{*'} (\dot{\boldsymbol{y}}_b^* - \dot{\boldsymbol{X}}_b^* \hat{\beta})$$

where $\boldsymbol{X}$ is the $NT \times (r+3)$ matrix of regressors ($\ln S_{it}^d$, $\ln H_{it}$, $\ln S_{it}^f$, $g(q_{it}^{(1)}; \hat{\gamma}_1, \hat{c}_1) \ln S_{it}^f$, ..., $g(q_{it}^{(r)}; \hat{\gamma}_r, \hat{c}_r) \ln S_{it}^f$), $\boldsymbol{y}$ is the $NT \times 1$ vector $\ln F_{it}$, $\boldsymbol{X}_b^*$ and $\boldsymbol{y}_b^*$ are the series resulting from the first-level bootstrap, and the dot denotes the standard within-transformation needed to remove individual means.

[18]The second-level bootstrap quasi-estimator $\tilde{\beta}_{bd}^{**}$ is computed as follows:

$$\tilde{\beta}_{bd}^{**} = \hat{\beta}_b^* + (\dot{\boldsymbol{X}}_b^{*'} \dot{\boldsymbol{X}}_b^*)^{-1} \dot{\boldsymbol{X}}_{bd}^{**'} (\dot{\boldsymbol{y}}_{bd}^{**} - \dot{\boldsymbol{X}}_{bd}^{**} \hat{\beta}_b^*)$$

where $\boldsymbol{X}_{bd}^{**}$ and $\boldsymbol{y}_{bd}^{**}$ are the series as they result in each second-level MBB replication made on each first-level bootstrap sample and the dot denotes the within transformation.

[19]This procedure is different from the one discussed by Gonçalves (2011), who used the sandwich form of the covariance matrix and a kernel variance estimator to studentize the test statistic, and a multivariate analogue of the estimator of the MBB variance proposed by Götze and Künsch (1996) to studentize the bootstrap test statistics. It is also different from the naive bootstrap discussed by Gonçalves and Vogelsang (2011), which uses the same sandwich covariance matrix estimator to studentize both the original statistic and the bootstrap ones.

[20]An alternative procedure could have been to compute bootstrap t statistics also in this case as Gonçalves and White (2004) proved the first-order asymptotic validity of the bootstrap distribution of MBB analogs of Wald and LM statistics for hypotheses testing in quasi-ML estimates. We nevertheless decided not to adopt it here, as it is not yet entirely clear if and when this method leads to asymptotic refinements.

[21]We must warn the reader that, for the parameters inducing non-linearity the standard

In Section 3, we apply these techniques to estimate the possible nonlinear effects of absorptive capacity and relative backwardness on the elasticity of countries' TFP to foreign knowledge.[22]

3. Data and results

To preserve the comparability of this investigation with the seminal work of Coe and Helpman (1995) and with most of the subsequent literature on international R&D spillovers, we focus on the sample of 24 OECD countries over the period 1971–2004 analyzed by Coe et al. (2009).[23] Accordingly, domestic and (trade-weighted) foreign R&D stocks, human capital (average years of schooling), and TFP indexes come from Coe et al. (2009). In accordance with the discussion in Section 1, we take absorptive capacity and relative backwardness as possible factors affecting in a nonlinear fashion the impact of foreign knowledge on the countries' TFP: therefore, their proxies are our candidate transition variables. We use each country's lagged human capital ($H_{i,t-1}$) as a measure of its absorptive capacity, while to proxy its relative backwardness we use the lagged percentage difference of each country's and the highest GDP per capita in constant Purchasing Power Parity (PPP) in each period ($\text{gap}_{i,t-1}$) .[24]

Following the procedure outlined in Section 2.2.1, we start by testing the null of linearity against the PSTR model of equation (3) with two regimes ($r = 1$) and each of the two candidate transition variables H_{t-1} and gap_{t-1}. The test statistics, along with the associated p-values, are reported in Table 1. In particular, we report the asymptotic p-value of the heteroskedasticity-robust Taylor expansion-based LM_F test statistic (LST-LM_F) and the bootstrap p-values of the heteroskedasticity-robust ExpLM and wLM (with critical values calculated via fixed-design wild bootstrap).[25]

---

approach to make inference may encounter problems because of the identification issues discussed before, as testing the null $\beta_j = 0$ or $\gamma_j = 0$ is equivalent to a linearity test. Indeed, when $\beta_j^f = 0$ ($\gamma_j = 0$), c_j and γ_j (β_j^f) are not identified, hypothesis testing for β_j^f and γ_j could be therefore not straightforward. The general problem of inference with weak identification has been very recently addressed by Andrews and Cheng (2011). An example of the problems one might encounter will be briefly discussed in the next section.

[22] All the computations were made using gretl 1.9.5. Code available at request.

[23] The panel is balanced. The countries are: Australia, Austria, Belgium-Luxembourg, Canada, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Israel, Italy, Japan, South Korea, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, UK, US.

[24] Data on countries' GDP per capita in constant PPP come from the OECD.

[25] To calculate ExpLM and wLM, we first compute a heteroskedasticity-robust LM test statistic for each of 1000 pairs (γ_1, c_1): $LM(\gamma_1^{(j)}, c_1^{(j)})$. Each pair is built as follows: γ_1 is drawn from a uniform distribution 0-100; c_1 is drawn uniformly at random from the set of observed values of the transition variables within the 5-95th percentile in the sample. Then we apply the following formulas:

$$\text{ExpLM} = \ln \left(\sum_{j=1}^{1000} \frac{\exp \left(0.5 LM(\gamma_1^{(j)}, c_1^{(j)}) \right)}{1000} \right)$$

Table 1: Tests of linearity

Transition variable			p-value
H_{t-1}	LST-LM$_F$	3.77022	0.0105
	ExpLM	22.7020	0.0000
	wLM	0.03310	0.0000
gap$_{t-1}$	LST-LM$_F$	25.6022	0.0000
	ExpLM	16.0325	0.0000
	wLM	0.02139	0.0000

All the test statistics strongly reject the null, thus corroborating the working hypothesis of a nonlinear impact of human capital (i.e., absorptive capacity) and technological gap (i.e. relative backwardness) on the TFP elasticity to foreign knowledge.

As to the choice of the transition variable to be included in the PSTR estimation, the various test statistics lead to different conclusions: the strongest rejection of the null is obtained for H_{t-1} (gap$_{t-1}$) using wLM and ExpLM (LST-LM$_F$). As wLM and ExpLM are usually more reliable than LST-LM$_F$, we estimate equation (3) setting $r = 1$ and $q_{it}^{(1)} = H_{i,t-1}$.

For the estimation of the nonlinear parameters associated with the two regimes (see Section 2.2.2), we find the starting values by means of the SA.[26] The results of the NLS are reported in the second column of Table 2.[27] This is not the conclusion of the procedure which, as illustrated in the previous Section, ends when the null of no remaining nonlinearity cannot be rejected. Accordingly, we proceed now with the tests of no remaining nonlinearity and with the estimation of any necessary further regimes (thereby concluding the whole estimation procedure and identifying our preferred specification) before discussing point estimates and inference.[28]

$$\text{wLM} = \frac{1}{1000} \sum_{j=1}^{1000} \omega_j LM(\gamma_1^{(j)}, c_1^{(j)})$$

where $\omega_j = LM(\gamma_1^{(j)}, c_1^{(j)}) / \sum_{j=1}^{1000} LM(\gamma_1^{(j)}, c_1^{(j)})$. We decided not to use AveLM since its power could be negatively affected by the presence of possible redundant values of γ_1 and c_1. To calculate bootstrap p-values via fixed-design wild bootstrap, we compute ExpLM and wLM for 999 bootstrap replications, where, in each replication, we randomize the sign of the residuals of the estimated linear model. The bootstrap p-value is equal to the fraction of bootstrap test statistics greater than the original one.

[26] In implementing Corana et al.'s (1987) algorithm, we set the initial temperature at 100, far above the average difference in SSR. The temperature reduction factor is 0.85. The algorithm adjusts the step-size vector every 20 parameter changes and this loop is repeated 50 times before each temperature reduction. In all the cases, the procedure converges after on average 6 million function evaluations.

[27] As a robustness check, the third column of the Table also reports the results of the alternative estimate of a two-regime PSTR model with gap$_{t-1}$ as transition variable.

[28] For the sake of brevity, we will not discuss the individual results of all the intermediate steps, such as the two-regime models in Table 2 or each of the alternative tests of remaining no linearity in Table 3. These are to be considered auxiliary parts of the procedure leading to our specification. One only exception will be made for the four-regime specification which we

Table 2: Estimates of nonlinear parameters (24 countries 1972-2004: 792 obs.)

Regimes $(r+1)$	1	2	2	3	4
Transition variables		H_{t-1}	gap_{t-1}	$H_{t-1}, \mathrm{gap}_{t-1}$	$H_{t-1}, \mathrm{gap}_{t-1}, \mathrm{gap}_{t-1}$
γ_1		3.3552 (0.7493)		4.4459 (3.8759)	4.1366 (4.8629)
c_1		8.3815 (0.0976)		8.3307 (0.1512)	8.3270 (0.1755)
γ_2			0.1028 (0.0100)	0.1164 (0.0140)	0.1218 (0.0122)
c_2			-59.440 (1.1701)	-58.809 (1.2736)	-58.777 (1.2327)
γ_3					27.941 (3.7788)
c_3					-1.5029 (0.0188)
SSR	5.43068	4.94314	3.14351	2.97204	2.78815
log-L	849.269	886.518	1065.77	1088.14	1113.28

Robust standard errors in brackets (QML estimator).

Table 3: Tests of no remaining nonlinearity

Hypothesis	Transition variables				p-value
	Estimated regimes	Additional			
H_0: $r=1$; H_1: $r=2$	H_{t-1}	H_{t-1}	LST-LM$_F$	4.3843	0.0045
			ExpLM	7.4068	0.0000
			wLM	0.0112	0.0000
		gap_{t-1}	LST-LM$_F$	24.739	0.0000
			ExpLM	21.315	0.0000
			wLM	0.0190	0.0000
H_0: $r=2$; H_1: $r=3$	$H_{t-1}, \mathrm{gap}_{t-1}$	H_{t-1}	LST-LM$_F$	0.2815	0.8388
			ExpLM	1.7287	0.0601
			wLM	0.0048	0.0350
		gap_{t-1}	LST-LM$_F$	4.1860	0.0059
			ExpLM	4.5622	0.0030
			wLM	0.0061	0.0170
H_0: $r=3$; H_1: $r=4$	$H_{t-1}, \mathrm{gap}_{t-1}, \mathrm{gap}_{t-1}$	H_{t-1}	LST-LM$_F$	0.6573	0.5785
			ExpLM	1.1691	0.2092
			wLM	0.0034	0.1732
		gap_{t-1}	LST-LM$_F$	0.7128	0.5445
			ExpLM	2.1588	0.0301
			wLM	0.0037	0.1201

On the basis of the tests of no remaining nonlinearity (Table 3), we end up estimating a PSTR model with four regimes ($r = 3$): the linear one, one regime associated with human capital (the first transition variable we discussed above) and two regimes associated with the technological gap. In all the estimates of the nonlinear parameters, reported in the various columns of Table 2, the location parameter associated with human capital ($c_1 = 8.33$) is rather close to the sample mean (8.54).[29] The estimated smoothness parameter γ_1 is roughly equal to 4, indicating that the transition is moderately smooth.[30]

On the contrary, the location parameters of the regimes associated with the technological gap are much closer to the minimum admissible values for the locations. In particular, the location parameter of the third regime ($c_2 = -58.8$) is near (though higher than) the lower bound of the 5th percentile, that is -59.2.

Although the above mentioned tests suggest that a four-regime ($r = 3$) specification fits the data well, it should be noted that the location parameter of the last regime ($c_3 = -1.5$) is above the 95th percentile, i.e. -2.[31] Moreover, this last regime has an associated smoothness parameter which is quite high ($\gamma_3 = 27.9$), suggesting that the transition is rather sharp. This entails that only few observations—namely, those of just three countries (Norway, Switzerland and US) in few years—end up in this additional regime. On these grounds we conclude that the estimates of the fourth regime should not be considered as fully reliable and, therefore, be neglected.

In fact, the contrasting indications from the nonlinearity tests (calling for a fourth regime) and the evidence above (supporting a three regimes model) suggest to look at the impact of excluding (or including) the fourth additional nonlinear component in the regression in greater detail. In this respect, the omission of the fourth regime does not affect much the estimates on the nonlinear parameters of the other regimes: as can be seen in Table 2, the estimated values of c_1, γ_1, c_2, and γ_2 are almost unchanged both in the three- and four-regime specification. Nor are seriously affected the estimates of the implied linear coefficients β's, reported in Table 4. This table shows the conditional fixed effects estimates in the PSTR model of equation (3) with, respectively, three and four regimes—second and third column—, along with those associated with the linear model of equation (2)—first column—, which is used as a traditional benchmark to gauge the other results.

---

will report (and in part discuss) before rejecting it on the basis of a series of considerations spelled out in the text.

[29]Human capital ranges from 2.51 (Portugal in 1971) to 12.29 (USA in 2003) and its 5-95th percentiles are, respectively, 4.98 and 11.53.

[30]Although the standard errors of γ_1 might look quite high, they are actually not so when compared with the results usually obtained in STR models. As well known, it is very difficult to obtain a precise estimate of γ in these models (see, for instance, the discussion in Franses and van Dijk, 2003, Section 3.2.2). Moreover, one should not judge the significance of this coefficient by looking at the associated t-statistic, because of the identification problem discussed in Section 2.2.1.

[31]The variable *gap* ranges from -89.9 (South Korea in 1971) to 0 with a mean about -32.6 and a standard deviation equal to 16.0.

Table 4: Estimates of linear parameters (24 countries 1972-2004: 792 obs.)

Regimes $(r+1)$	1	3	4
Transition variables		H_{t-1}, gap_{t-1}	H_{t-1}, gap_{t-1}, gap_{t-1}
β^d	0.0450	0.0278	0.0294
	(0.0259)	(0.0140)	(0.0129)
	[0.0051]	[0.0040]	[0.0010]
β^h	0.6555	0.3408	0.2922
	(0.1249)	(0.1034)	(0.0992)
	[0.0000]	[0.0560]	[0.0901]
β^f_0	0.1491	0.0939	0.1052
	(0.0558)	(0.0391)	(0.0391)
	[0.0000]	[0.0000]	[0.0000]
β^f_1		0.0070	0.0063
		(0.0038)	(0.0032)
		[0.0000]	[0.0010]
β^f_2		0.0692	0.0705
		(0.0109)	(0.0103)
		[0.0000]	[0.0000]
β^f_3			0.0090
			(0.0011)
			[0.1041]

Unreported country dummies. Asymptotic HACC standard errors in round brackets. Bootstrap p-values of t statistics calculated via PMBB double-bootstrap in square brackets: 999 first-level replications; 1000 second-level replications; block size 3.

An additional piece of evidence in favor of discarding the fourth regimes is the fact that, once robust inference is done, the coefficient β^f_3 is not statistically significant. Table 4 reports in square brackets the bootstrap p-value of the t statistics computed via the double bootstrap discussed in Section 2.3 (with 1000 second-level bootstrap replications for each of 999 first-level bootstrap replications using PMBB resampling with block size equal to 3) and the p-value of the test for $\beta^f_3 = 0$ is 0.1041. This result, in fact, is not very strong because of the identification problems (discussed in Section 2.2.1) in testing the null $\beta^f_3 = 0$ in the presence of nuisance parameters. The bootstrap procedure used to compute the p-values of the test for $\beta^f_3 = 0$ is indeed robust to serial and simultaneous correlation in the data, but it is not free from nuisance parameter issues. The tests of no remaining nonlinearity in Table 3, on the contrary, address the nuisance parameter problem, but are not robustified to serial and simultaneous correlation. As we cannot get rid of both nuisance and autocorrelation problems at the same time and given the contrasting indications coming from the two tests, we cannot rely exclusively on them for the choice of our preferred specification. Hence, despite the failure to reject some residual nonlinearity associated with gap_{t-1} for the three-regime specification (with H_{t-1} and gap_{t-1} as transition variables), the long series of considerations put forward in the paragraph above led us to choose the model with three regimes as the preferred specification.

It is worth stressing that this choice is in line with what done in the (P)STR literature. As shown in Franses and van Dijk (2003), outliers may affect the tests of no remaining nonlinearity without leading to the discovery of any sensible additional regime: the tests are thus to be interpreted with a grain of salt and with reference to the overall fit of the nonlinear specifications they would lead to estimate. Moreover, it is inappropriate to use the PSTR model to single out very few observations in a regime (as it would be the case were we including a fourth regime in our estimation). Accordingly, in the discussion of the main results below, we will focus on the main differences between the benchmark linear model and the three-regime PSTR model.

3.1. Discussion of the results for the preferred specification

The estimates of the three-regime PSTR model—fourth column in Table 2 for the nonlinear part and third column in Table 4 for the linear part—support the idea that the human capital of a country, by determining its absorptive capacity, positively affects the exploitation of foreign knowledge. The TFP elasticity to the trade-weighted foreign R&D stock is positively related with human capital: β_1^f is positive and significant, although the overall impact of absorptive capacity on the TFP elasticity is not huge (the maximum relative difference accounted by human capital in the parameter β^f is approximately 7.5%). These results are in line with Crespo-Cuaresma et al. (2004), who found some evidence of a positive impact of human capital on the effects of foreign knowledge on growth. Their estimated effects were however not statistically significant and, thus, weaker than ours. Also Kwark and Shyn (2006) found a positive and significant impact of the interaction between human capital and foreign R&D stock on TFP; however, as mentioned in Section 1, in their specification the positive and significant impact of the human capital and of the trade-weighted foreign R&D stock on TFP disappears. Notably, we do not encounter any of these problems here.[32]

The effects of the regime switching due to H_{t-1} are depicted in Figure 1 which plots the values of $\hat{\beta}_0^f + \hat{\beta}_1^f g(\bar{H}_i; \hat{\gamma}_1, \hat{c}_1) + \hat{\beta}_2^f g(\text{g}\bar{\text{a}}\text{p}; \hat{\gamma}_2, \hat{c}_2)$ for each country, where g$\bar{\text{a}}$p is the sample mean of gap and $\bar{H}_i$ is the average human capital by country in each of two sub-periods: 1971-1987 and 1988-2003.

A much greater difference in the elasticity of TFP to foreign knowledge across countries and periods is related with the countries' relative backwardness. According to our estimates, the TFP elasticity to foreign knowledge in laggard countries (e.g. South Korea, Portugal and Greece) is significantly lower than the same elasticity in leading countries (e.g. Norway, Switzerland and US). For instance, the TFP elasticity to foreign knowledge for South Korea in the 1970s would have been on average 70% higher if South Korea had filled the technological gap with the leading country before.[33] The effects for the different

---

[32]Our findings are also qualitatively in line with those in Kneller (2005), which however measured foreign knowledge in a different way and adopted just a linear interacting term.

[33]These results apply to, but are not due to, the inclusion of the country in the panel. Indeed, they do hold even if South Korea, which exhibits a low human capital and the largest technological gap, is excluded from the sample.
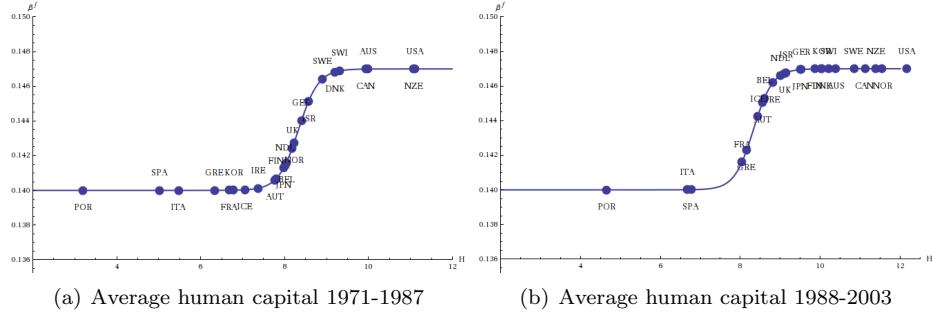
(a) Average human capital 1971-1987

(b) Average human capital 1988-2003

Figure 1: Effect of the regime switching associated with human capital on TFP elasticity to import-weighted foreign R&D stock

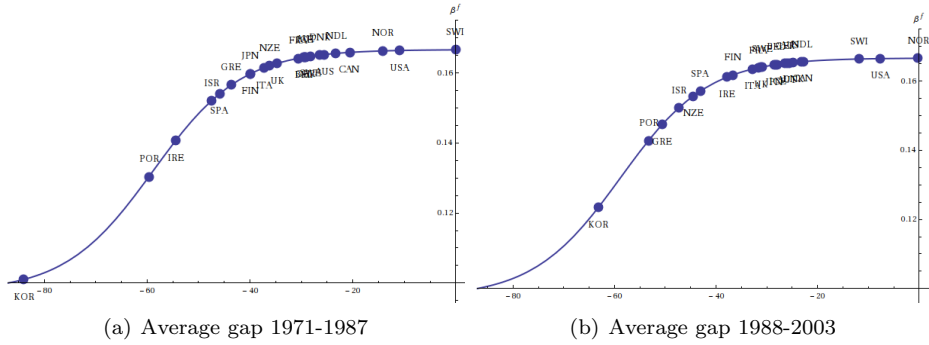(a) Average gap 1971-1987

(b) Average gap 1988-2003

Figure 2: Effect of the regime switching associated with relative backwardness on TFP elasticity to import-weighted foreign R&D stock

countries of the regime switches associated with *gap* are summarized in Figure 2, which plots the values of $\hat{\beta}_0^f + \hat{\beta}_1^f g(\bar{H}; \hat{\gamma}_1, \hat{c}_1) + \hat{\beta}_2^f g(\text{g\={a}p}_i; \hat{\gamma}_2, \hat{c}_2)$ for each country, with $\bar{H}$ being the sample mean of human capital and $\text{g\={a}p}_i$ the mean gap by country during 1971-1987 and 1988-2003.

These findings indicate that the elasticity of TFP to foreign knowledge is non marginally affected by the internal conditions of the recipient country and that, even in a sample of developed economies, leading countries have an advantage over the others. This is supportive of the intuition of Matthews's (1969). Despite the non-negligible differences in the relationships estimated in this and in other works, our results seem qualitatively in line with those in Falvey et al. (2007) whereas they contrast with Crespo-Cuaresma et al. (2004).[34]

---

[34]In fact, although we find as Falvey et al. (2007) that the marginal increase in the TFP elasticity to foreign knowledge deriving from further reductions in the technological gap is decreasing, we cannot find evidence of the hump-shaped relationship implied by the three-regime specification estimated by Falvey et al. (2007). It is also worth noticing, in passing,

Having illustrated and commented on the estimates associated with each individual transition variable, we move now to a more general discussion of the results. To gauge the overall statistical fit of the PSTR models, in line with the literature we focus on the Sums of Square Residuals (SSR) and the log-likelihood values reported in Table 2. The former passes from 5.43 in the linear estimation to 2.97 in our preferred specification and the log-likelihood moves from 848.27 to 1065.77.[35] These findings, together with the strong statistical significance of the individual coefficients even after a proper robustification of the tests, support the adoption of a PSTR model where absorptive capacity and relative backwardness drive two regime changes.

Although not the main object of the investigation, we now briefly comment on how the parameters of domestic R&D stock and human capital (respectively, β^d and β^h) are affected by moving from the linear to the nonlinear specification. As to the domestic R&D stock, the observed decline in β^d is in line with Kneller (2005) and compatible with the standard errors of the estimated coefficients. This implies that failing to reckon the nonlinearities investigated in this work leads to an overestimation of the impact of the domestic R&D stock. The discussion of the coefficient for human capital is slightly more complicated because the variable enters in the specification both as regressor and (once lagged) as transition variable. It follows that its overall impact on TFP is not directly comparable in the baseline and in the nonlinear estimations. In the linear specification the estimate of the (constant) TFP elasticity to human capital is $\hat{\beta}^h = 0.6555$ and in the nonlinear specification $\hat{\beta}^h = 0.3408$. The latter is smaller than the former, but $\hat{\beta}^h$ in the nonlinear specification should be interpreted as the lower bound of the total impact of human capital: in other words, $\hat{\beta}^h = 0.3408$ is the elasticity for the countries having H greater than 10 or smaller than 7 (i.e., values that are far away from the location parameter $c_1 = 8.33$). For any observation of H taking values between 7 and 10, the elasticity is significantly higher than $\hat{\beta}^h$. When evaluated at the sample means, the point estimate of the long-run TFP elasticity to human capital implied by our specification happens to be very high (higher than 1).[36]

Finally, we summarize in Figure 3 the total effect on β_{it}^f of the inclusion of three regimes for the different countries over time. This figure allows to grasp the extent of variation of the elasticity of TFP to foreign knowledge both over

---

that our specification builds on those adopted to study international R&D spillovers and differs from medium-term growth convergence models. Accordingly, one should be careful not to interpret our results as implying lack of convergence, as they refer to short-term dynamics.

[35]In line with the more moderate impact of relative backwardness found above, the fit of the model improves more because of the inclusion of gap_{t-1} than that of H_{t-1}.

[36]The long-run TFP elasticity to human capital of country i is given by:

$$\frac{\partial \ln F_i}{\partial \ln H_i} = \beta^h + \beta_1^f \gamma_1 \frac{\mathrm{e}^{-\gamma_1(H_i - c_1)}}{\left(1 + \mathrm{e}^{-\gamma_1(H_i - c_1)}\right)^2} H_i \ln S_i^f$$

This formula has been evaluated using our point estimates of the parameters and at the sample mean of H and $\ln S^f$.
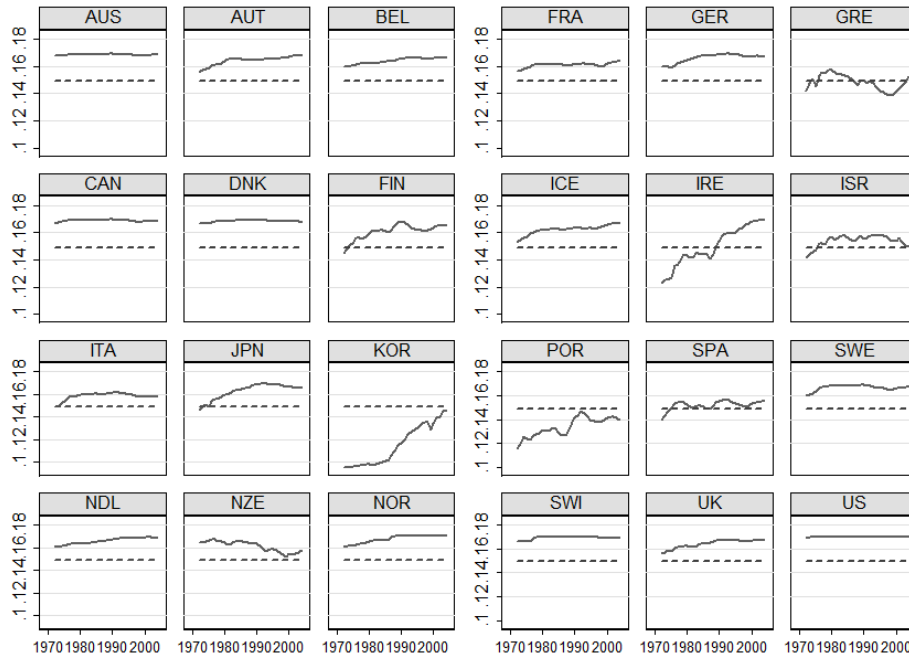
Figure 3: TFP elasticity to import-weighted foreign R&D stock by country

time and across countries. As to the cross-section dimension, it is worth noticing that: i) the elasticity of Portugal, South Korea, and, in part, Ireland is lower than the value of β^f estimated with a linear panel; ii) the linear estimates seem to fit Greece, Israel, Italy, Spain and, in the last periods, New Zealand; iii) in all the other cases, the linear estimates fall short of the value obtained taking nonlinear effects into account.

4. Conclusions

In this paper, we investigate whether domestic factors affect the impact that foreign knowledge has on the TFP of the recipient countries. Although economic intuition and theoretical modeling suggest that both absorptive capacity and relative backwardness of the countries exposed to foreign knowledge can affect TFP and growth, the literature has not yet reached conclusive empirical results.

Applying the PSTR model to the workhorse empirical specification in this strand of the literature (Coe and Helpman, 1995; Coe et al., 2009), we test on a sample of OECD countries over the period 1971-2004 whether absorptive capacity and relative backwardness have nonlinear effects on the TFP elasticity to foreign R&D stock. This nonlinear approach relaxes the assumptions of parameter homogeneity across countries and parameter constancy over time, and allows to relate the nonlinear effects to observable proxies of both absorptive capacity and relative backwardness (measured, respectively, as the level of human capital and

the gap with respect to the leading country in terms of GDP per capita in PPP terms).

Despite its flexibility, PSTR modeling requires the adoption of non trivial econometric techniques to test properly the hypothesis of linearity in the presence of nuisance parameters. Accordingly, we run a battery of tests used in the literature (González and Teräsvirta, 2006) and take heteroskedasticity into account by means of a fixed-design wild bootstrap procedure (Hurn and Becker, 2009).

The results of our nonlinear estimations suggest that absorptive capacity is significantly associated with higher R&D spillovers. The indirect impact of human capital on TFP via the increase in the absorption of foreign knowledge can be quite significant. In addition, and in contrast with previous results in the literature, we find that relative backwardness has a negative and significant impact on the transmission of foreign knowledge. Laggard countries seem to derive lower benefits from foreign R&D stocks than more advanced countries do.

The adoption of a technique encompassing nonlinear effects allows differentiating the TFP elasticity to foreign knowledge across countries and over time. A good number of countries in the sample exhibit an elasticity higher than that estimated by means of a linear (homogeneous parameter) panel model, whereas South Korea, Portugal and (in part) Ireland score worse than all the other countries. Interestingly, but not surprisingly, Greece, Italy and Spain occupy an intermediate position.

All these results are strongly statistically significant even though we adopt a series of up-to-date econometric measures to make inference robust to unspecified forms of heteroskedasticity, and serial and simultaneous correlation in the data. In particular, we combine the Bravo and Godfrey's (2011) double bootstrap method with the panel moving block bootstrap of Gonçalves (2011). This is, to our knowledge, the first time such method is used in an applied empirical work.

Acknowledgements

References

Abramovitz, M., 1986. Catching-up, forging ahead, and falling behind. Journal of Economic History XLVI (2), 385–406.

Aghion, P., Howitt, P., 1992. A model of growth through creative destruction. Econometrica 60 (2), 323–51.

Alcidi, C., Flamini, A., Fracasso, A., 2011. Policy regime changes, judgment and Taylor rules in the Greenspan era. Economica 78 (309), 89–107.

Andrews, D. W. K., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61 (4), 821–56.

Andrews, D. W. K., Cheng, X., 2011. Estimation and inference with weak, semi-strong, and strong identification. Discussion Paper 1773R, Cowles Foundation.

Andrews, D. W. K., Ploberger, W., November 1994. Optimal tests when a nuisance parameter is present only under the alternative. Econometrica 62 (6), 1383–1414.

Becker, R., Osborn, D. R., 2010. Weighted smooth transition regressions. Journal of Applied Econometrics.

Benhabib, J., Spiegel, M., 1994. The role of human capital in economic development: Evidence from aggregate cross-country data. Journal of Monetary economics 34 (2), 143–173.

Béreau, S., Villavicencio, A. L., Mignon, V., 2010. Nonlinear adjustment of the real exchange rate towards its equilibrium value: A panel smooth transition error correction modelling. Economic Modelling 27 (1), 404–416.

Bravo, F., Godfrey, L. G., 2011. Bootstrap HAC tests for ordinary least squares regression. Oxford Bulletin of Economics and Statistics forthcoming.

Coe, D., Helpman, E., 1995. International R&D spillovers. European Economic Review 39, 859–887.

Coe, D. T., Helpman, E., Hoffmaister, A. W., 1997. North-south R&D spillovers. The Economic Journal 107 (440), 134–149.

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

Cohen, W. M., Levinthal, D. A., 1989. Innovation and learning: The two faces of R&D. Economic Journal 99 (397), 569–96.

Corana, A., Marchesi, M., Martini, C., Ridella, S., 1987. Minimizing multimodal functions of continuous variables with the 'simulated annealing' algorithm. ACM Trans. Math. Softw. 13, 262–280.

Crespo-Cuaresma, J., Foster, N., Scharler, J., 2004. On the determinants of absorptive capacity: Evidence from OECD countries. In: Proceedings of OeNB Workshops. Vol. 2.

Delatte, A.-L., Fouquau, J., 2011. The determinants of international reserves in the emerging countries: a nonlinear approach. Applied Economics 43 (28), 4179–4192.

Eaton, J., Kortum, S., 1999. International technology diffusion: Theory and measurement. International Economic Review 40 (3), 537–70.

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.

Falvey, R., Foster, N., Greenaway, D., 2007. Relative backwardness, absorptive capacity and knowledge spillovers. Economics Letters 97 (3), 230–234.

Fok, D., van Dijk, D., Franses, P. H., 2005. A multi-level panel STAR model for US manufacturing sectors. Journal of Applied Econometrics 20 (6), 811–827.

Fouquau, J., Hurlin, C., Rabaud, I., 2008. The Feldstein-Horioka puzzle: A panel smooth transition regression approach. Economic Modelling 25 (2), 284–299.

Fracasso, A., Vittucci Marzetti, G., 2012. An empirical note on international R&D spillovers. Empirical Economics forthcoming.

Franco, C., Montresor, S., Marzetti, G. V., 2011. On indirect trade-related r&d spillovers: The average propagation length of foreign r&d. Structural Change and Economic Dynamics 22 (3), 227–237.

Franses, P. H., van Dijk, D., 2003. Nonlinear Time Series Models in Empirical Finance. Cambridge University Press, New York.

Gerschenkron, A., 1962. Economic backwardness in historical perspective. Belknap Press, Cambridge MA.

Girma, S., 2005. Absorptive capacity and productivity spillovers from FDI: A threshold regression analysis. Oxford Bulletin of Economics and Statistics 67 (3), 281–306.

Goffe, W., Ferrier, G., Rogers, J., 1994. Global optimization of statistical functions with simulated annealing. Journal of Econometrics 60 (1–2), 65–99.

Gonçalves, S., 2011. The moving blocks bootstrap for panel linear regression models with individual fixed effects. Econometric Theory forthcoming.

Gonçalves, S., Kilian, L., 2004. Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. Journal of Econometrics 123, 89–120.

Gonçalves, S., Vogelsang, T. J., 2011. Block bootstrap HAC robust tests: The sophistication of the naive bootstrap. Econometric Theory 27, 745–791.

Gonçalves, S., White, H., 2004. Maximum likelihood and the bootstrap for nonlinear dynamic models. Journal of Econometrics 119 (1), 199–219.

González, A., Teräsvirta, T., 2006. Simulation-based finite sample linearity test against smooth transition models. Oxford Bulletin of Economics and Statistics 68, 797–812.

González, A., Teräsvirta, T., van Dijk, D., 2005. Panel smooth transition regression models. Research Paper 165, Quantitative Finance Research Centre.

Götze, F., Künsch, H., 1996. Second-order correctness of the blockwise bootstrap for stationary observations. Annals of Statistics 24, 1914–1933.

Granger, C., Teräsvirta, T., 1993. Modelling nonlinear economic relationships. Oxford University Press, New York.

Grossman, G. M., Helpman, E., 1991. Trade, knowledge spillovers, and growth. European Economic Review 35 (2), 517–526.

Guellec, D., Van Pottelsberghe de la Potterie, B., 2004. From R&D to productivity growth: Do the institutional settings and the source of funds of R&D matter? Oxford Bulletin of Economics and Statistics 66 (3), 353–378.

Hansen, B. E., 1999a. Testing for linearity. Journal of Economic Surveys 13 (5), 551–76.

Hansen, B. E., 1999b. Threshold effects in non-dynamic panels: Estimation, testing, and inference. Journal of Econometrics 93 (2), 345–368.

Hansen, B. E., 2000. Sample splitting and threshold estimation. Econometrica 68 (3), 575–603.

Howitt, P., 2000. Endogenous growth and cross-country income differences. American Economic Review 90 (4), 829–846.

Hsiao, C., Pesaran, M., 2004. Random coefficient panel data models. Cambridge Working Papers in Economics 0434, Faculty of Economics, University of Cambridge.

Hu, F., Zidek, J. V., 1995. A bootstrap based on the estimating equations of the linear model. Biometrika 82, 263–275.

Hurn, S., Becker, R., 2009. Testing for nonlinearity in mean in the presence of heteroskedasticity. Economic Analysis and Policy 39 (2), 311–326.

Keller, W., 1996. Absorptive capacity: On the creation and acquisition of technology in development. Journal of Development Economics 49 (1), 199–227.

Keller, W., 1998. Are international R&D spillovers trade-related? Analyzing spillovers among randomly matched trade partners. European Economic Review 42 (8), 1469–1481.

Keller, W., 2002. Trade and the transmission of technology. Journal of Economic Growth 7 (1), 5–24.

Keller, W., 2004. International technology diffusion. Journal of Economic Literature 42 (3), 752–782.

Kneller, R., 2005. Frontier technology, absorptive capacity and distance. Oxford Bulletin of Economics and Statistics 67 (1), 1–23.

Krammer, S. M., 2010. International R&D spillovers in emerging markets: The impact of trade and Foreign Direct Investment. The Journal of International Trade & Economic Development 19 (4), 591–623.

Kwark, N.-S., Shyn, Y.-S., 2006. International R&D spillovers revisited: Human capital as an absorptive capacity for foreign technology. International Economic Journal 20 (2), 179–196.

Lichtenberg, F. R., van Pottelsberghe de la Potterie, B., 1998. International R&D spillovers: A comment. European Economic Review 42 (8), 1483–1491.

Liu, R., Singh, K., 1992. Moving blocks jackknife and bootstrap capture weak dependence. In: LePage, R., Billiard, L. (Eds.), Exploring the limits of the bootstrap. Wiley, New York.

Lumenga-Neso, O., Olarreaga, M., Schiff, M., 2005. On indirect trade-related R&D spillovers. European Economic Review 49, 1785–1798.

Lundbergh, S., Teräsvirta, T., 1998. Modelling high-frequency time series with STAR-GARCH models. SSE/EFI working paper 291, Stockholm School of Economics.

Luukkonen, R., Saikkonen, P., Terasvirta, T., 1988. Testing linearity against smooth transition autoregressive models. Biometrika 75, 491–499.

Madsen, J. B., Islam, M. R., Ang, J. B., 2010. Catching up to the technology frontier: the dichotomy between innovation and imitation. Canadian Journal of Economics 43 (4), 1389–1411.

Mancusi, M. L., 2008. International spillovers and absorptive capacity: A cross-country cross-sector analysis based on patents and citations. Journal of International Economics 76 (2), 155–165.

Matthews, R. C. O., 1969. Why growth rates differ. Economic Journal 79 (314), 261–68.

Otten, R. H. J. M., van Ginneken, L. P. P. P., 1989. The Annealing Algorithm. Kluwer, Boston, MA.

Rivera-Batiz, L. A., Romer, P. M., 1991. Economic integration and endogenous growth. The Quarterly Journal of Economics 106 (2), 531–555.