Wirtschaftsuniversität Wien
Vienna University of Economics and Business Administration

Working Papers Series:

Growth and Employment in Europe: Sustainability and Competitiveness

Working Paper No. 30

SOME CURRENT ISSIUES IN THE STATISTICAL ANALYSIS OF SPILLOVERS

Daniela Gumprecht, Nicole Gumprecht and Werner G. Müller

September, 2003

This working paper series presents research results of the WU-Research Focus:
Growth and Employment in Europe, Sustainability and Competitiveness
The papers are available online under: http://www.wu-wien.ac.at/inst/vw1/gee/workp.html

SOME CURRENT ISSUES IN THE STATISTICAL ANALYSIS OF SPILLOVERS

by

Daniela Gumprecht
Department of Statistics and Decision Support Systems
University of Vienna
Universitätsstraße 5/3
A-1010 Vienna, Austria

Nicole Gumprecht
Department of Statistics and Decision Support Systems
University of Vienna
Universitätsstraße 5/3
A-1010 Vienna, Austria

Werner G. Müller
Department of Statistics
Vienna University of Economics and Business Administration (WU)
Augasse 2-6
A-1090, Vienna, Austria
email: werner.mueller@wu-wien.ac.at

Abstract

Spillover phenomena are usually statistically estimated on the basis of regional and temporal panel data. In this paper we review and investigate exploratory and confirmatory statistical panel data techniques. We illustrate the methods by calculations in the stetting of the well known Research and Development Spillover study by Coe and Helpman, 1995. It will be demonstrated that alternative estimation techniques that are well compatible with the data can lead to opposite conclusions.

Acknowledgements

We are most grateful to Werner Hölzl for a number of useful comments that led to an improvement in the paper.

Keywords

Panel data, fixed effects, random coefficients, DOLS, R&D spillover.

JEL

C33

Some Current Issues in the Statistical Analysis of Spillovers

Daniela Gumprecht*, Nicole Gumprecht*, Werner G. Müller**

*Department of Statistics and Decision Support Systems, University of Vienna
**Department of Statistics, University of Economics Vienna

Abstract

Spillover phenomena are usually statistically estimated on the basis of regional and temporal panel data. In this paper we review and investigate exploratory and confirmatory statistical panel data techniques. We illustrate the methods by calculations in the stetting of the well known Research and Development Spillover study by Coe and Helpman (1995). It will be demonstrated that alternative estimation techniques that are well compatible with the data can lead to opposite conclusions.

1. Coe & Helpman's R&D Spillover Study

Most theories of growth explain economic growth in terms of the accumulation of capital and the growth of the labor force and exogenous technological progress captured by a time trend. In recent formulations these variables are quality adjusted (human capital, embodied technological progress). In contrast the new growth theory (Romer 1990; Grossman and Helpman, 1991) tries to explain the growth record in terms of endogenous R&D decisions. Productivity depends therefore on the amount of knowledge generated by innovation activities and productivity increases depend on current R&D efforts which translate into increased technical knowledge. By building on these theories Coe and Helpman (1995) claimed that the productivity of an economy depends on its own R&D as well as the R&D spendings of its trade partners. A direct advantage is a more effective use of resources by the application of new technologies, materials, production processes and organisation methods. Indirect benefits come from the import of goods and services from trade partners.

In their meanwhile classical paper Coe and Helpman (1995) used a panel dataset to study the extent to which a country's productivity level depends on domestic and foreign stock of knowledge. They used the cumulative spendings for R&D of a country to measure the domestic stock of knowledge of this country. As a representative for the foreign stock of knowledge, Coe and Helpman used the import-weighted sums of cumulated R&D expenditures of the trade partners of the country. The importance of the R&D capital stock is measured by the elasticity of total factor productivity with respect to the R&D capital stock. A panel dataset with 22 countries (21 OECD countries plus Israel) during the period from 1971 to 1990 was used[1]. The variables total factor productivity (TFP), domestic R&D capital stock (DRD) and foreign R&D capital stock (FRD) are constructed as indices with basis 1985 (1985 = 1).

[1] All data can be found on the homepage of Elhanan Helpman (Helpman, 2003), which is accessible via the internet address http://post.economics.harvard.edu/faculty/helpman/data.html

In their papers Coe and Helpman have used a variety of specifications to model the effects on TFP. To simplify the exposition we will here only regard one of those. Our conclusions, however, are not limited to this particular case but rather apply to all of the suggested models (for a more complete analysis see D.Gumprecht, 2003). Our illustrative model contains three variables: total factor productivity (TFP) as the regressand, domestic R&D capital stock (DRD) and foreign R&D capital stock (FRD) as the regressors. The impact of domestic and foreign R&D expenditures is supposed to be the same for all countries. The equation – with regional index *i* and temporal index *t* – has the following form:

$$log\, F_{it} = \alpha_{it}^{0} + \alpha_{it}^{d}\, log\, S_{it}^{d} + \alpha_{it}^{f} m_{i,t-1}\, log\, S_{it}^{f},$$

where

F_{it} denotes total factor productivity (TFP),

S_{it}^{d} domestic R&D capital stock (DRD), and

S_{it}^{f} foreign R&D capital stock (FRD). FRD is defined as the import-share-weighted average of the domestic R&D capital stocks of trade partners.

α_{it}^{0} stands for the intercepts, which are allowed to vary across countries for two reasons: first, there may exist country specific effects on productivity that are not included in the variables of our model; and second, all variables are transformed into index numbers and TFP is measured in country specific currency whereas DRD and FRD are measured in U.S. dollars.

α_{it}^{d} then denotes the regression coefficient, which corresponds to the elasticity of TFP with respect to DRD, and

α_{it}^{f} determines the elasticity of TFP with respect to FRD, which equals $\alpha_{it}^{f} m_{i,t-1}$. Finally

$m_{i,t-1}$ denotes the fraction of imports in GDP.

According to standard practice Coe and Helpman (1995) used a panel data model with fixed effects, which is described in detail in the next subsection, for their estimations. They were especially focussed on the time dimension of the data and therefore used time series methods and analysis for their panel data model. As they were interested to identify a long-run relationship between TFP and domestic and foreign R&D spendings, and as TFP, DRD and FRD showed a clear temporal trend, they estimated cointegrated equations.

"The basic idea of cointegration is that if there is a long-run relationship between two or more trended variables, a regression contain all the variables – the cointegration equation – will have a stationary error term, even if none of the variables taken alone is stationary. If the error term is not stationary, the estimated relationship may be spurious." (Coe and Helpman, 1995: 867-868 according to Granger and Newbold, 1974). Cointegrated equations have the important econometric property that OLS estimates are 'super consistent' (Stock, 1987). This means if the number of observations increases, the OLS estimator of the cointegrating equation converges to the true parameter value much faster than in the case where the variables are stationary. The idea of cointegration comes from time-series analysis and it seems natural for Coe and Helpman to use this technique for their R&D Spillovers problem. Because of the relative small number of time-series observations for each country, Coe and Helpman estimated their equations from panel data and interpreted the results as pooled cointegration equations (Coe and Helpman, 1995: 868).

Conditions for the existence of cointegration are the following: first the separate variables have to be nonstationary; and second, the error term of a linear combination of the variables has to be stationary. Nonstationarity of each single time-series was tested with the Dickey-Fuller, the augmented Dickey-Fuller, the Levin and Lin (1992)- and the Levin and Lin (1993) unit root tests. The Levin and Lin unit root tests on the pooled data confirm the nonstationarity of the variables. Nonstationarity of the error term was tested with Levin and Lin (1992)-, Levin and Lin (1993) unit root tests and a test from Engle and Granger (1987). These tests provided different results (Coe and Helpman, 1995: Table 3). Because of these mixed results and the fact that the econometrics of pooled cointegration were not fully worked out at that time, Coe and Helpman concentrated more on the theoretical model and on the a priori plausibility of the estimated parameters rather than on the tests for cointegration (Coe and Helpman, 1995: 870).

In what follows we present the recalculation of the OLS estimators for the model of Coe and Helpman (1995) with corrected degrees of freedom in the calculation of the t-values making use of the Helpman (2003) data. Kao et al. (1999) have re-estimated Coe and Helpman's equations (with corrected t-values, see a discussion of their approach later). However they made a mistake when implementing the calculation in GAUSS (a commonly used statistically oriented matrix language package, see www.aptech.com). They used wrong degrees of freedom for the calculation of the t-values, namely

instead of `vb1=inv(x1'*x1)*((u1'*u1)/(N*T-N-k1),`

they used `vb1=inv(x1'*x1)*((u1'*u1)/(N*T-1)).`

The corrected estimation (with no substantial difference in significances) yields (t-values in parentheses)

$$\log F_{it} = \hat{\alpha}_{it}^{0} + \underset{(12,8885)^{**}}{0,10511 \log S_{it}^{d}} + \underset{(5,8011)^{**}}{0,2665 m_{i,t-1} \log S_{it}^{f}},$$

with $i = 1, \ldots, N$ (=21) and $t = 1, \ldots, T$ (=20) and a coefficient of determination $R^2 = 0,5576$ and an Adjusted R^2 of $0,5331$. Note that in panel models the definition of the coefficient of determination is not without ambiguity and we have calculated all R^2 throughout the paper as the squared correlations of $\hat{y}_{it}$ and y_{it}.

Coe and Helpman (1995) took these estimation results, with both a positive and statistically significant regression coefficients as a confirmation of their hypothesis that TFP of a country depends on both domestic R&D capital stock and foreign R&D capital stock.

A corresponding exploratory study seems to confirm these conclusions. The simple time-series scatterplots of TFP and DRD and TFP and FRD are given in Figures 1 and 2, respectively. To simplify the plots we have only included the G7 countries[2], without restriction on generality. The plots show the time paths from lower left (1971) to upper right (1990), which all exhibit an upward slope as an indication of a positive relationship between these variables.

[2] U.S.A., Japan, Germany, France, Italy, U.K., Canada

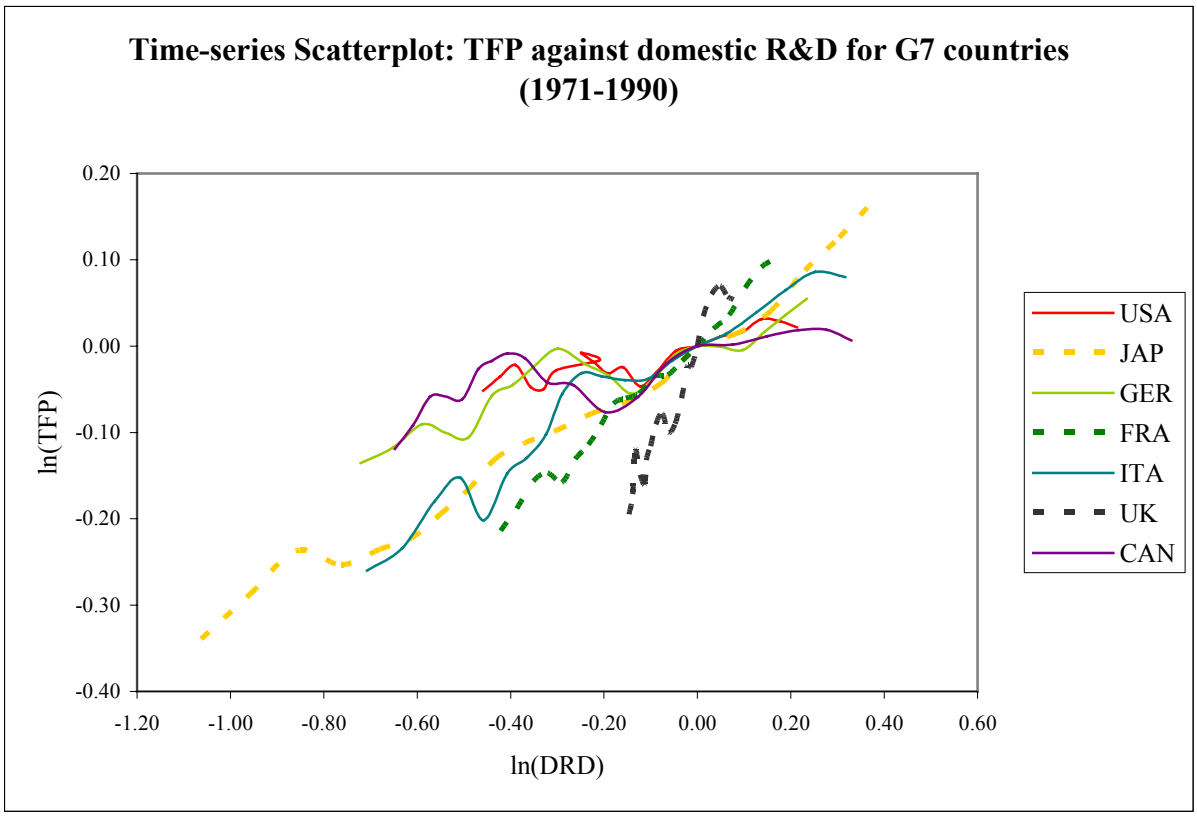

Figure 1: Time-series scatterplot TFP against DRD for G7 countries (1971-90).

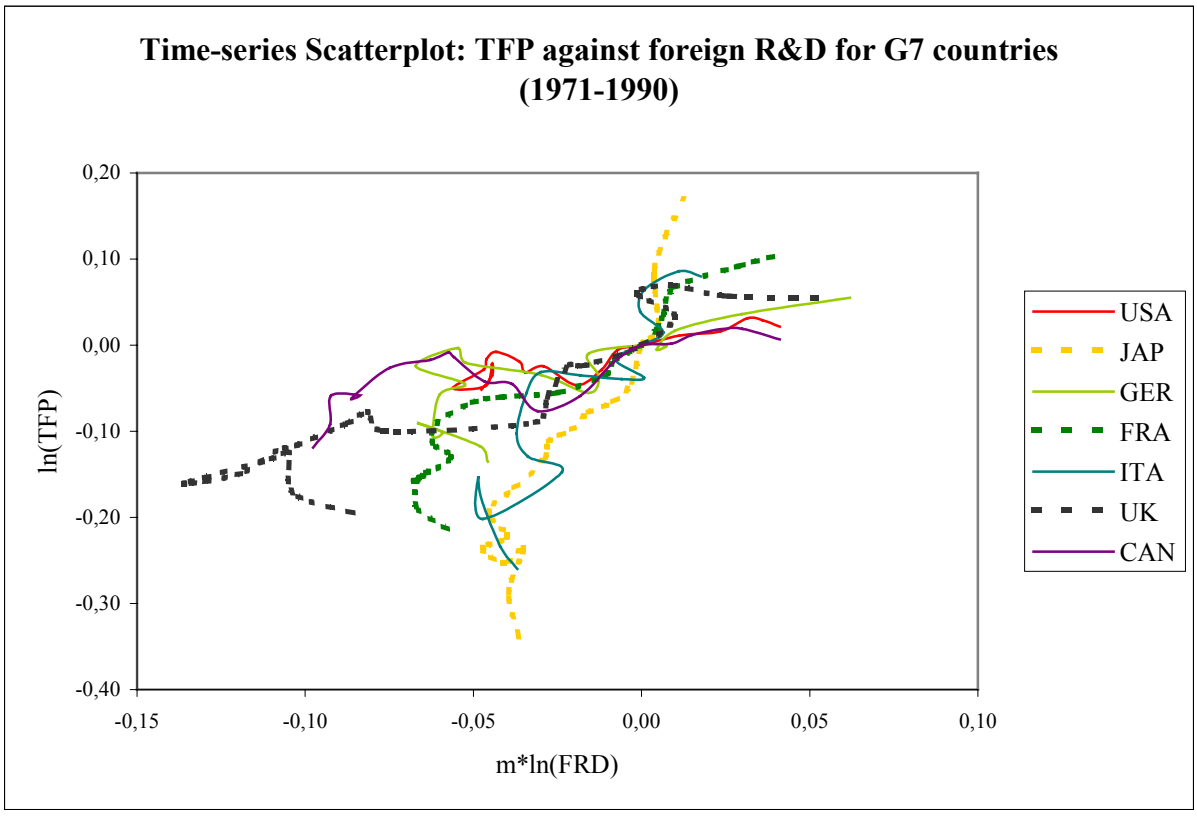

Figure 2: Time-series scatterplot TFP against FRD for G7 countries (1971-90).

2. Fixed effects panel regression

In the following section we will review the estimation techniques employed in most of the spillover studies. We will hereby largely follow the exposition of standard econometrics textbooks such as e.g. Greene (2000). More detailed material on the various specifications used in panel data regressions can e.g. be found in the monographs by Hsiao, 1986 and Baltagi, 2001.

2.1. Simple OLS estimators

In fixed effect panel models differences between cross-section units (individuals, regions, etc.) are shown by differences in the constant terms. Each α_i is an unknown parameter and must be estimated. This approach is suitable for models where the differences between individuals can be interpreted as parametrical shifts of the regression function.

There are three different ways to specify the regression model.

1. Original form:

$$y_{it} = \alpha + \boldsymbol{\beta}' \boldsymbol{x}_{it} + \varepsilon_{it}$$

The total sums of squares and total cross products are given by

$$\boldsymbol{S}_{xx}^{t} = \sum_{i=1}^{N}\sum_{t=1}^{T_i}(\boldsymbol{x}_{it} - \bar{\bar{\boldsymbol{x}}})(\boldsymbol{x}_{it} - \bar{\bar{\boldsymbol{x}}})' \qquad \boldsymbol{S}_{xy}^{t} = \sum_{i=1}^{N}\sum_{t=1}^{T_i}(\boldsymbol{x}_{it} - \bar{\bar{\boldsymbol{x}}})(y_{it} - \bar{\bar{y}})$$

where:

$$\bar{\bar{\boldsymbol{x}}} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i}\boldsymbol{x}_{it}}{\sum_{i=1}^{N}T_i} = \frac{\sum_{i=1}^{N}T_i\bar{\boldsymbol{x}}_{i\cdot}}{\sum_{i=1}^{N}T_i} = \sum_{i=1}^{N}w_i\bar{\boldsymbol{x}}_{i\cdot} \text{ and } \bar{\bar{y}} = \sum_{i=1}^{N}w_i\bar{y}_{i\cdot} \text{ with } w_i = \frac{T_i}{\sum_{i=1}^{N}T_i}$$

and the LS-total estimator follows:

$$\boldsymbol{b}^{t} = (\boldsymbol{S}_{XX}^{t})^{-1}\boldsymbol{S}_{XY}^{t}$$

2. Departure from group-mean form:

$$y_{it} - \bar{y}_{i\cdot} = \boldsymbol{\beta}'(\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_{i\cdot}) + \varepsilon_{it} - \bar{\varepsilon}_{i\cdot}$$

Here the so called sums of squares within and cross products within are given by:

$$\boldsymbol{S}_{xx}^{w} = \sum_{i=1}^{N}\sum_{t=1}^{T_i}(\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_{i\cdot})(\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_{i\cdot})' \qquad \boldsymbol{S}_{xy}^{w} = \sum_{i=1}^{N}\sum_{t=1}^{T_i}(\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_{i\cdot})(y_{it} - \bar{y}_{i\cdot})$$

with the corresponding LS-within estimator:

$$\boldsymbol{b}^{w} = (\boldsymbol{S}_{XX}^{w})^{-1}\boldsymbol{S}_{XY}^{w}$$

3. Group mean form:

$$\bar{y}_{i\cdot} = \alpha + \boldsymbol{\beta}'\bar{\boldsymbol{x}}_{i\cdot} + \bar{\varepsilon}_{i\cdot}$$

with only *N* observations because there are only *N* groups. The corresponding so called sums of squares between and cross products between are given by

$$\boldsymbol{S}^{b}_{xx} = \sum_{i=1}^{N}\sum_{t=1}^{T_i} T_i(\bar{\boldsymbol{x}}_{i\cdot} - \bar{\bar{\boldsymbol{x}}})(\bar{\boldsymbol{x}}_{i\cdot} - \bar{\bar{\boldsymbol{x}}})' \quad \boldsymbol{S}^{b}_{xy} = \sum_{i=1}^{N}\sum_{t=1}^{T_i} T_i(\bar{\boldsymbol{x}}_{i\cdot} - \bar{\bar{\boldsymbol{x}}})(\bar{y}_{i\cdot} - \bar{\bar{y}})$$

And the LS-between estimator follows

$$\boldsymbol{b}^{b} = (\boldsymbol{S}^{b}_{XX})^{-1}\boldsymbol{S}^{b}_{XY}\,.$$

The group means are calculated in the following way:

$$\bar{y}_{i\cdot} = \frac{1}{T_i}\sum_{t=1}^{T_i} y_{it}\,,\ \bar{\boldsymbol{x}}_{i\cdot} = \frac{1}{T_i}\sum_{t=1}^{T_i} \boldsymbol{x}_{it}\,,\ \bar{\varepsilon}_{i\cdot} = \frac{1}{T_i}\sum_{t=1}^{T_i} \varepsilon_{it}\,.$$

The terminology "within" and "between" stems from the fact that the estimators are determined by the variation within and between the specific groups as opposed to the "total" variation.

In a panel data model with fixed effects the within-estimator ($\boldsymbol{b}^{w}$) is the BLUE (best linear unbiased estimator). The proof follows directly from the Least Square Dummy Variable (LSDV) form of the fixed effects model (see e.g. Greene 2000: chapter 14.3.).

Suggestions for improvement of Coe and Helpman's estimation came - amongst others - from Kao *et al.* (1999). They criticized two things: First, Coe and Helpman presented their results without any t-values because the asymptotic distribution of the t-statistic for estimates in cointegrated panel data was not known at that time. Therefore no exact statements about the significance of the OLS estimators could be made. As Coe and Helpman's resulting estimates were both relatively small one can't safely conclude that even one of the true coefficients was bigger than zero. Second, due to the unit-root in the time dimension and in spite of the super consistence of the time-series estimator, the upward bias of the estimate can be quite substantial for small samples and there is no reason to assume that this bias becomes negligible by the inclusion of a cross section dimension in panel data. Kao *et al.* (1999) argue that it is quite possible that the estimators even change their sign when introducing a bias correction in the calculation.

For those reasons Kao *et al.* (1999) used different estimation methods for Coe and Helpman's International R&D Spillovers regression and compared the empirical consequences from the different estimation methods. They claim that the DOLS (dynamic OLS) estimation is the best solution for this problem because in the given setting the DOLS estimator exhibits no bias and is asymptotically normal.

2.2. Corrected OLS estimators

Kao, Chiang and Chen also used a panel data model with fixed effects for their estimations. The regression function has again the following specification:

$$y_{it} = \alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}$$

where now

y_{it}	again denotes the dependent variable,
$\boldsymbol{\beta}$	$M \times 1$ the vector of slope parameters,
α_i	the region specific intercepts,
ε_{it}	stands for a stationary error term, but now
$\boldsymbol{x}_{it}$	is regarded as an $M \times 1$ first order integrated process, with $\boldsymbol{x}_{it} = \boldsymbol{x}_{i,t-1} + \boldsymbol{\xi}_{it}$.

Under these assumptions the panel regression describes a system of cointegrated regressions, this means y_{it} is cointegrated with $\boldsymbol{x}_{it}$. Furthermore y_{it} and $\boldsymbol{x}_{it}$ are independent between different cross section units and $\boldsymbol{w}_{it} = (\varepsilon_{it} \quad \boldsymbol{\xi}'_{it})'$ is a linear process that fulfils the assumptions of Kao and Chiang (1997). The asymptotic covariance matrix $\boldsymbol{\Omega}$ of $\boldsymbol{w}_{it}$ can be written in the following form:

$$\boldsymbol{\Omega} = \sum_{j=-\infty}^{\infty} E(\boldsymbol{w}_{ij}\boldsymbol{w}'_{i0})$$

$$= \boldsymbol{\Sigma} + \boldsymbol{\Gamma} + \boldsymbol{\Gamma}'$$

$$= \begin{bmatrix} \boldsymbol{\Omega}_{\varepsilon} & \boldsymbol{\Omega}_{\xi\varepsilon} \\ \boldsymbol{\Omega}_{\xi\varepsilon} & \boldsymbol{\Omega}_{\xi} \end{bmatrix},$$

where

$$\boldsymbol{\Gamma} = \sum_{j=1}^{\infty} E(\boldsymbol{w}_{ij}\boldsymbol{w}'_{i0}) = \begin{bmatrix} \boldsymbol{\Gamma}_{\varepsilon} & \boldsymbol{\Gamma}_{\varepsilon\xi} \\ \boldsymbol{\Gamma}_{\xi\varepsilon} & \boldsymbol{\Gamma}_{\xi} \end{bmatrix}$$

and

$$\boldsymbol{\Sigma} = E(\boldsymbol{w}_{i0}\boldsymbol{w}'_{i0}) = \begin{bmatrix} \boldsymbol{\Sigma}_{\varepsilon} & \boldsymbol{\Sigma}_{\varepsilon\xi} \\ \boldsymbol{\Sigma}_{\varepsilon\xi} & \boldsymbol{\Sigma}_{\xi} \end{bmatrix}$$

are partitioned according to $\boldsymbol{w}_{it}$.

The one-sided asymptotic covariance is defined as:

$$\boldsymbol{\Delta} = \boldsymbol{\Sigma} + \boldsymbol{\Gamma}$$

$$= \sum_{j=0}^{\infty} E(\boldsymbol{w}_{ij}\boldsymbol{w}'_{i0})$$

with

$$\boldsymbol{\Delta} = \begin{bmatrix} \boldsymbol{\Delta}_{\varepsilon} & \boldsymbol{\Delta}_{\varepsilon\xi} \\ \boldsymbol{\Delta}_{\varepsilon\xi} & \boldsymbol{\Delta}_{\xi} \end{bmatrix}.$$

With this "long run correction" the correct t-values can be calculated.

Kao and Chiang (1997) defined the limiting distribution of the OLS and a so called DOLS (dynamic ordinary least squares) estimator of a cointegrated regression. They also showed that these limiting distributions are asymptotically normal and analysed the characteristics of these estimators in finite samples. They found out that the OLS estimator has a non-neglible bias and that the DOLS estimator is therefore preferable for estimating cointegrated panel regressions. The OLS estimator is given by

$$\hat{\boldsymbol{\beta}}_{OLS} = \left[\sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_{i.})(x_{it} - \bar{x}_{i.})' \right]^{-1} \left[\sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_{i.})(y_{it} - \bar{y}_{i.}) \right],$$

where $\bar{x}_{i.}$ and $\bar{y}_{i.}$ are the respective group means (see Kao et al., 1999: 697). The asymptotic distribution of this estimator is, according to Kao and Chiang (1997),

$$\sqrt{NT}(\hat{\boldsymbol{\beta}}_{OLS} - \boldsymbol{\beta}) - \sqrt{N}\delta_{NT} \rightarrow N(\mathbf{0}, 6\boldsymbol{\Omega}_{\xi}^{-1}\boldsymbol{\Omega}_{\varepsilon\xi}); \qquad \text{convergence in distribution,}$$

where

$$\boldsymbol{\Omega}_{\varepsilon\xi} = \boldsymbol{\Omega}_{\varepsilon} - \boldsymbol{\Omega}_{\varepsilon\xi}\boldsymbol{\Omega}_{\xi}^{-1}\boldsymbol{\Omega}_{\xi\varepsilon}$$

and

$$\boldsymbol{\delta}_{NT} = \left[\frac{1}{N} \sum_{i=1}^{T} \frac{1}{T^2} \sum_{t=1}^{T} (x_{it} - \bar{x}_{i.})(x_{it} - \bar{x}_{i.})' \right]^{-1} \times \left[\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\Omega}_{\xi}^{1/2} \left(\int_0^1 \widetilde{\boldsymbol{W}}_i(\boldsymbol{r}) d\boldsymbol{W}_i'(\boldsymbol{r}) \right) \boldsymbol{\Omega}_{\xi}^{-1/2} \boldsymbol{\Omega}_{\xi\varepsilon} + \boldsymbol{\Delta}_{\xi\varepsilon} \right],$$

$\boldsymbol{W}_i(r)$ being a standard Brownian motion, and

$$\widetilde{\boldsymbol{W}}_i(\boldsymbol{r}) = \boldsymbol{W}_i(\boldsymbol{r}) - \int_0^1 \boldsymbol{W}_i(\boldsymbol{r}) d\boldsymbol{r}.$$

2.3. The DOLS estimator

This estimator, which was employed in Kao et al. (1999), can be obtained by running the regression:

$$y_{it} = \alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \sum_{j=-q_1}^{q_2} \Delta \boldsymbol{x}_{i,t+j}' \boldsymbol{c}_{ij} + v_{it}, \qquad q_1, q_2 \in \{0, 1, 2, \ldots\}$$

The DOLS estimation as used by Kao et al. (1999) is also based on a fixed effect regression model:

$$y_{it} = \alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it} \quad i = 1, \ldots, N, \; t = 1, \ldots, T.$$

We assume that $\{\boldsymbol{x}_{it}\}$ are $k \times 1$ integrated processes of order one for all i, where

$$\boldsymbol{x}_{it} = \boldsymbol{x}_{i,t-1} + \boldsymbol{\xi}_{it}$$

and

$$\Delta \boldsymbol{x}_{it} = \boldsymbol{x}_{it} - \boldsymbol{x}_{i,t-1},$$

where $\Delta \boldsymbol{x}_{it}$ denotes the difference of $\boldsymbol{x}_{it}$ to $\boldsymbol{x}_{i,t-1}$.

If we assume that the process $\{\varepsilon_{it}\}$ can be projected on to $\{\boldsymbol{\xi}_{it}\}$, we get

$$\varepsilon_{it} = \sum_{j=-\infty}^{\infty} \boldsymbol{\xi}'_{i,t+j} \boldsymbol{c}_{ij} + v_{it}$$

where

$$\sum_{j=-\infty}^{\infty} \left\| \boldsymbol{c}_{ij} \right\| < \infty,$$

$\{v_{it}\}$ is stationary with mean zero, and $\{v_{it}\}$ and $\{\boldsymbol{\xi}_{it}\}$ are uncorrelated, both contemporaneously and in all lags and leads (see Saikkonen, 1991: 11).

In practice, the lags and leads are restricted to a range from q_1 to q_2. Retaining the former assumption approximately, it follows that

$$\varepsilon_{it} = \sum_{j=-q_1}^{q_2} \boldsymbol{\xi}'_{i,t+j} \boldsymbol{c}_{ij} + v_{it}\ .$$

This follows the assumption that $\{c_{it}\}$ are absolutely summable, which means

$$\sum_{j=-\infty}^{\infty} \left\| \boldsymbol{c}_{ij} \right\| < \infty .$$

After substitution

$$\varepsilon_{it} = \sum_{j=-\infty}^{\infty} \boldsymbol{\xi}'_{i,t+j} \boldsymbol{c}_{ij} + v_{it}$$

and

$$\boldsymbol{\xi}_{i,t+j} = \Delta \boldsymbol{x}_{i,t+j} = \boldsymbol{x}_{i,t+j} - \boldsymbol{x}_{i,t+j-1}$$

into the initial model

$$y_{it} = \alpha_i + \boldsymbol{x}'_{it} \boldsymbol{\beta} + \varepsilon_{it} \quad i = 1, ..., N,\ t = 1, ..., T$$

we yield the specification

$$y_{it} = \alpha_i + \boldsymbol{x}'_{it} \boldsymbol{\beta} + \sum_{j=-q_1}^{q_2} \Delta \boldsymbol{x}'_{i,t+j} \boldsymbol{c}_{ij} + v_{it}\ .$$

This is the regression model for the DOLS estimation (Kao and Chiang 1997: 9). The asymptotic distribution of a corresponding (now unbiased) estimator $\hat{\boldsymbol{\beta}}_D$ is given by

$\sqrt{NT}(\hat{\boldsymbol{\beta}}_D - \boldsymbol{\beta}) \rightarrow N(0, 6\boldsymbol{\Omega}_{\xi}^{-1}\boldsymbol{\Omega}_{\varepsilon\xi})$; convergence in distribution as N$\rightarrow\infty$ and T $\rightarrow\infty$.

For definition of $\boldsymbol{\Omega}_{\varepsilon\xi}$ see the OLS estimator of $\boldsymbol{\beta}$ (section 2.2.) .

Kao *et al.* (1999) reported the following results for their DOLS estimation of Coe and Helpman's R&D Spillovers model:

$$\log F_{it} = \hat{\alpha}_{it}^{0} + 0{,}1237 \log S_{it}^{d} + 0{,}0682 m_{i,t-1} \log S_{it}^{f} + Rest$$
$$\qquad\qquad (5{,}9572)^{**} \quad (0{,}6333)$$

with an R^2 of *0,5016*.

They concluded from this estimation that domestic R&D expenditures affect TFP of a country but foreign R&D expenditures do not have a significant effect on TFP of the country. Thus, they argue, Coe and Helpman's (1995) conclusions should be rejected. However, Kao *et al.* (1999) wanted to estimate a fixed effect regression model (a model with county-specific intercepts) but erroneously they implemented a common coefficient model (a model with a common intercept). Furthermore, they left out the lag zero, which is not backed up by the corresponding theory by Saikkonen (1991). Additionally, their R^2 is calculated by ESS (Explained Sums of Squares) divided by TSS (Total Sums of Squares) but the wrong numbers of degrees of freedom were used.

The result of the correct implementation of the fixed effect model with dynamic regressors is now the following:

$$\log F_{it} = \hat{\alpha}_{it}^{0} + 0{,}1284 \log S_{it}^{d} + 0{,}1321 m_{i,t-1} \log S_{it}^{f} + Rest$$
$$\qquad\qquad (18{,}3164)^{**} \quad (3{,}8375)^{**}$$

with an R^2 (according to our definition) of *0,8755* and an Adjusted R^2 of *0,8689*.

The correct estimated coefficient for foreign R&D expenditures is again – as in the original paper – significant. Domestic- and foreign R&D expenditures still seem to affect TFP of a country, which supports Coe and Helpman's conclusions. The R^2, calculated as the square of the correlation between $\hat{y}_{it}$ and y_{it}, is much better than the R^2 of Kao *et al.* (1999), calculated as ESS divided by TSS.

Nevertheless, a considerable innovation of Chiang and Kao's (1999, 2002) implementation is the use of the so called "long run correction" (see section 2.2.) for the correct calculation of the t-values of the coefficients, a suggestion, which will be taken up in our final model.

3. An Alternative View

There are many debates in the panel data estimation literature, whether to regard the region specific or other effects as random outcomes poses a valuable alternative to the fixed coefficient model. In the present context Müller and Nettekoven (1999) have suggested a so called random coefficient model to analyse the R&D Spillovers model of Coe and Helpman (1995) and conclude that although the alternative specification is well compatible with the data, one astonishingly has to draw contradictory conlusions.

3.1. The Random Coefficient Model

Here, the parameters $\boldsymbol{\beta}_i$ are assumed to vary randomly around a common mean $\boldsymbol{\beta}$. This model can be described in the form:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i,$$

where

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{v}_i$$

with

$$E[\boldsymbol{v}_i] = \boldsymbol{0},$$

$$E[\boldsymbol{v}_i\boldsymbol{v}_i'] = \boldsymbol{\Gamma}.$$

Under the assumption that there is no autocorrelation and no correlation between the cross section units, $\boldsymbol{\beta}_i$ (that applies for a particular cross section unit) can be considered the result of a random process with mean $\boldsymbol{\beta}$ and covariance matrix $\boldsymbol{\Gamma}$.

If $\boldsymbol{\beta}_i$ is express by the relation $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{v}_i$ the following model results:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + (\boldsymbol{\varepsilon}_i + \boldsymbol{X}_i\boldsymbol{v}_i) = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{w}_i,$$

where

$$E[\boldsymbol{w}_i] = \boldsymbol{0}$$

and

$$E[\boldsymbol{w}_i\boldsymbol{w}_i'] = \sigma_i^2\boldsymbol{I} + \boldsymbol{X}_i\boldsymbol{\Gamma}\boldsymbol{X}_i' = \boldsymbol{\Pi}_i$$

The covariance matrix for all observations ($\boldsymbol{V}$) has the following form:

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{\Pi}_1 & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Pi}_2 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ & & \vdots & & \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{\Pi}_n \end{bmatrix}.$$

Now, the (best linear unbiased) GLS estimator can be expressed by a matrix weighted average of the OLS estimators:

$$\hat{\boldsymbol{\beta}} = \sum_{i=1}^{N} \boldsymbol{W}_i\boldsymbol{b}_i$$

where $\boldsymbol{b}_i$ is the i-th OLS coefficient estimator and

$$\boldsymbol{W}_i = \left[\sum_{j=1}^{N} (\boldsymbol{\Gamma} + \boldsymbol{V}_j)^{-1}\right]^{-1} (\boldsymbol{\Gamma} + \boldsymbol{V}_i)^{-1}$$

where

$$\boldsymbol{V}_i = \sigma_i^2 (\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}.$$

The estimator of $\boldsymbol{\beta}$ can also be expressed in the usual form of the GLS estimator:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}.$$

As $\boldsymbol{V}$ is a block diagonal matrix it follows that:

$$\hat{\boldsymbol{\beta}} = \left[\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{\Pi}_i^{-1}\boldsymbol{X}_i\right]^{-1}\left[\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{\Pi}_i^{-1}\boldsymbol{y}_i\right]$$

where

$$\boldsymbol{\Pi}_i = \sigma_i^2\boldsymbol{I} + \boldsymbol{X}_i\boldsymbol{\Gamma}\boldsymbol{X}_i'$$

This representation of $\hat{\boldsymbol{\beta}}$ follows the fact that $\hat{\boldsymbol{\beta}}$ is a weighted average of the OLS estimators, (for a detailed proof, see Greene, 2000: 610).

To estimate the unknown parameters in $\boldsymbol{\Gamma}$ and $\boldsymbol{V}_i$ Swamy (1971) suggested the following procedure. Let $\boldsymbol{b}_i$ be the group specific OLS coefficient vector and let $\hat{\boldsymbol{V}}_i$ be the sample covariance matrix,

$$s_i^2(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1},$$

where

$$s_i^2 = \frac{\boldsymbol{e}_i'\boldsymbol{e}_i}{T_i - K};$$

now

$$\bar{\boldsymbol{b}} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{b}_i,$$

then

$$\hat{\boldsymbol{\Gamma}} = \frac{1}{N-1}\left(\sum_{i=1}^{N}\boldsymbol{b}_i\boldsymbol{b}_i' - N\bar{\boldsymbol{b}}\bar{\boldsymbol{b}}'\right) - \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{V}}_i$$

If the second matrix in $\hat{\boldsymbol{\Gamma}}$ is quite big it is possible that $\hat{\boldsymbol{\Gamma}}$ is not positive definite anymore. In big samples the second matrix is negligibly small but in small samples $\hat{\boldsymbol{\Gamma}}$ might become not positive definite. A simple and asymptotical valid solution for this problem is, just to drop the second matrix. For the calculations in this paper this asymptotical valid form of $\hat{\boldsymbol{\Gamma}}$ was used, i.e. the matrix $\hat{\boldsymbol{V}}_i$ was not included in the estimation of $\hat{\boldsymbol{\Gamma}}$.

Now predictors for the individual parameter vectors can be calculated. The best linear predictor for $\boldsymbol{\beta}_i$ is:

$$\hat{\boldsymbol{\beta}}_i = [\boldsymbol{\Gamma}^{-1} + \boldsymbol{V}^{-1}]^{-1}[\boldsymbol{\Gamma}^{-1}\hat{\boldsymbol{\beta}} + \boldsymbol{V}_i^{-1}\boldsymbol{b}_i] = \boldsymbol{A}_i\hat{\boldsymbol{\beta}} + [\boldsymbol{I} - \boldsymbol{A}_i]\boldsymbol{b}_i$$

where

$$\boldsymbol{A}_i = (\boldsymbol{\Gamma}^{-1} + \boldsymbol{V}_i^{-1})^{-1}\boldsymbol{\Gamma}^{-1}$$

and

$$\boldsymbol{V}_i = \sigma_i^2(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}.$$

This predictor is again a matrix weighted average. The weights are the inverse of the covariance matrix of $\hat{\boldsymbol{\beta}}_i$ and $\boldsymbol{b}_i$. In practice the estimators $\hat{\boldsymbol{\Gamma}}$ and $\hat{\boldsymbol{V}}_i$ are used for $\boldsymbol{\Gamma}$ and $\boldsymbol{V}_i$.

The variance of the predictor $\hat{\boldsymbol{\beta}}_i$ is given by

$$Var[\hat{\boldsymbol{\beta}}_i] = \begin{bmatrix} \boldsymbol{A}_i \\ \boldsymbol{I} - \boldsymbol{A}_i \end{bmatrix}' \begin{bmatrix} \sum_{i=1}^{n} \boldsymbol{W}_i(\boldsymbol{\Gamma} + \boldsymbol{V}_i)\boldsymbol{W}_i' & \boldsymbol{W}_i(\boldsymbol{\Gamma} + \boldsymbol{V}_i) \\ (\boldsymbol{\Gamma} + \boldsymbol{V}_i)\boldsymbol{W}_i' & (\boldsymbol{\Gamma} + \boldsymbol{V}_i) \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_i \\ \boldsymbol{I} - \boldsymbol{A}_i \end{bmatrix}.$$

Parameters estimated according to this specification partly differ considerably from the OLS estimators of Coe and Helpman (1995) (as well as the DOLS estimators of Kao et al., 1999). Even to such an extent that the sign of single parameters may depend on the choice of the model – a model with fixed or a model with random coefficients (cf. Müller and Nettekoven, 1999). Other than the fixed effect model the random effect (coefficient) one assumes the existing array of countries a random draw of a (fictitious) population of similar economies.

A correct random coefficient estimation yields

$$\begin{array}{l} \log F_{it} = \hat{\alpha}_{it}^{0} + 0{,}2475 \log S_{it}^{d} - 0{,}0841 m_{i,t-1} \log S_{it}^{f} \\ \qquad\qquad (7{,}7578)^{**} \quad (-0{,}5087) \end{array}$$

with an R^2 of *0,9122* and an Adjusted R^2 of *0,9074*.

The estimates for the random coefficient model differ decisively from the fixed coefficient model and especially the estimator of the foreign R&D expenditures changed sign, although this is not statistically significant. Values for R^2 and Adjusted R^2 raised, both are now around *0,91*, so the explanatory power of the model is quite good. Contrary to the other estimations so far (and Coe and Helpman's conclusions) this model indicates that the foreign spillover effect is not significant!

Note that although Müller and Nettekoven (1999) have already identified this effect, they report other estimates for the random coefficient model. This is due to erroneously relating foreign R&D expenditures of some countries to domestic R&D expenditures and TFP of other countries.

4. Time Series Added-Variable-Plots

In this section we will demonstrate, that it could have been possible by a proper exploratory analysis to detect the reported inconsistencies of the data with the posited models and the achieved results.

Although at first glance Figures 1 and 2 seem to confirm Coe and Helpman's conclusions, we propose in a second step to construct so called Added-Variable-Plots of the data. Such plots are used to analyse the importance of additional variables for the explanation of the variation of a dependent variable and they are capable of discovering masking effects.

The relevant question that should be answered is, whether foreign R&D expenditures can provide any explanation of the variation of TFP additional to the explanation provided by domestic R&D expenditures. Is it useful to add FRD into the model when it already includes DRD? To answer this question additional to the time-series scatterplot of TFP against domestic R&D expenditures for G7 countries (see Figure 1), time-series scatterplot of TFP against foreign R&D expenditures for G7 countries (see Figure 2) a third time-series scatterplots is of relevance, the one of foreign R&D expenditures against domestic R&D expenditures (see Figure 3).

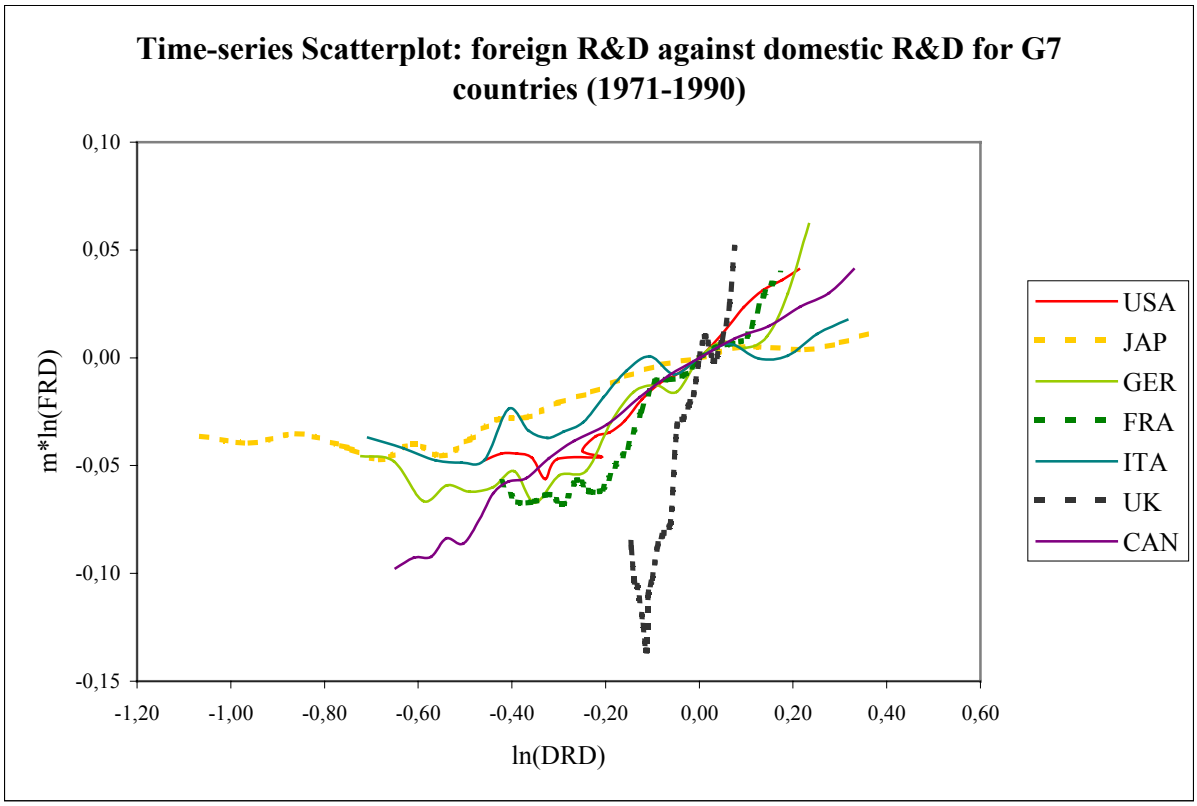

Figure 3: Time-series scatterplot FRD against DRD for G7 countries (1971-90).

The first two time-series scatterplots show positive correlation and this might lead to the (somewhat premature) conclusion that domestic as well as foreign R&D expenditures might be able to explain TFP. However, the time-series scatterplot of FRD against DRD also shows a positive correlation between those variables, which strongly indicates that one of the regressors might carry mainly redundant information.

Added-Variable-Plots rather than plotting original variables like in usual scatterplots employ the partial effects of the considered regressors manifested by the residuals of corresponding OLS regressions (for a detailed description see Cook and Weisberg, 1994). Thus in our context two time-series Added-Variable-Plots will be useful: one shows the residuals from a simple OLS regression of the natural logarithm of TFP on the natural logarithms of the domestic R&D expenditures (y-axis) against the residuals of a simple OLS regression of the natural logarithms of the foreign R&D expenditures multiplied with the import-shares on the natural logarithms of the domestic R&D expenditures (x-axis), i.e. it will be displaying the

partial effect of DRD on TFP (Figure 4). The other will be constructed vice versa for the partial effect of FRD on TFP (Figure 5). Both OLS regressions were calculated for each country separately.

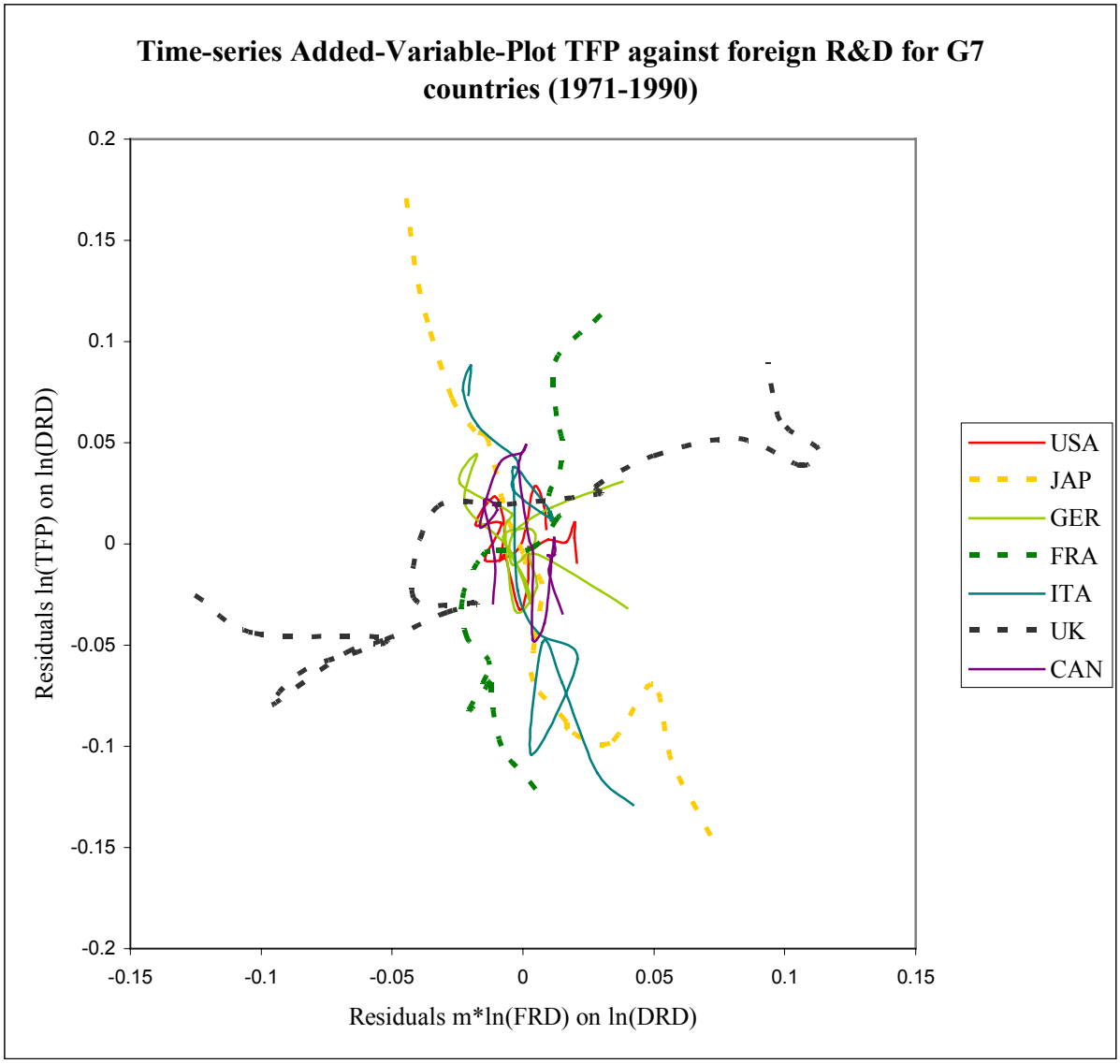

Figure 4: Time-series Added-Variable-Plot TFP against FRD for G7 countries (1971-90).

For the interpretation of these time-series added-variable plots it is useful to have a look at the three extreme cases of Added-Variable Plots. If all points lie exactly on a straight line with nonzero slope, this means all residuals from the A-V-P regression are zero, it is useful to add the second variable (e.g. in Figure 4: $m \times ln(FRD)$) to the first variable (in Figure 4: $ln(DRD)$) because this will give a perfect fit of the model. If all points lie on a horizontal line all variation of the dependent variable is explained by the first variable, there is no need to insert the second variable into the model. If all points lie on a vertical line the second variable is a linear function of the first one and it is not useful to include the second variable into the model because it can not explain the variation any further.

The A-V-P for FRD of the R&D Spillovers data for the G7 countries (Figure 4) shows that the lines for Japan, France and Italy are nearly vertical, for Japan and Italy they even show a negative trend, this means that FRD are nearly a linear function of DRD and therefore they are redundant for the model. The lines for U.S. and Canada do not show any trend, they vary randomly and therefore FRD of U.S. and Canada are also not able to provide additional explanation to the model with DRD only. The line for U.K. shows a slight positive trend but it is still nearly horizontal and therefore can provide nearly no further explanation of the variation of TFP.

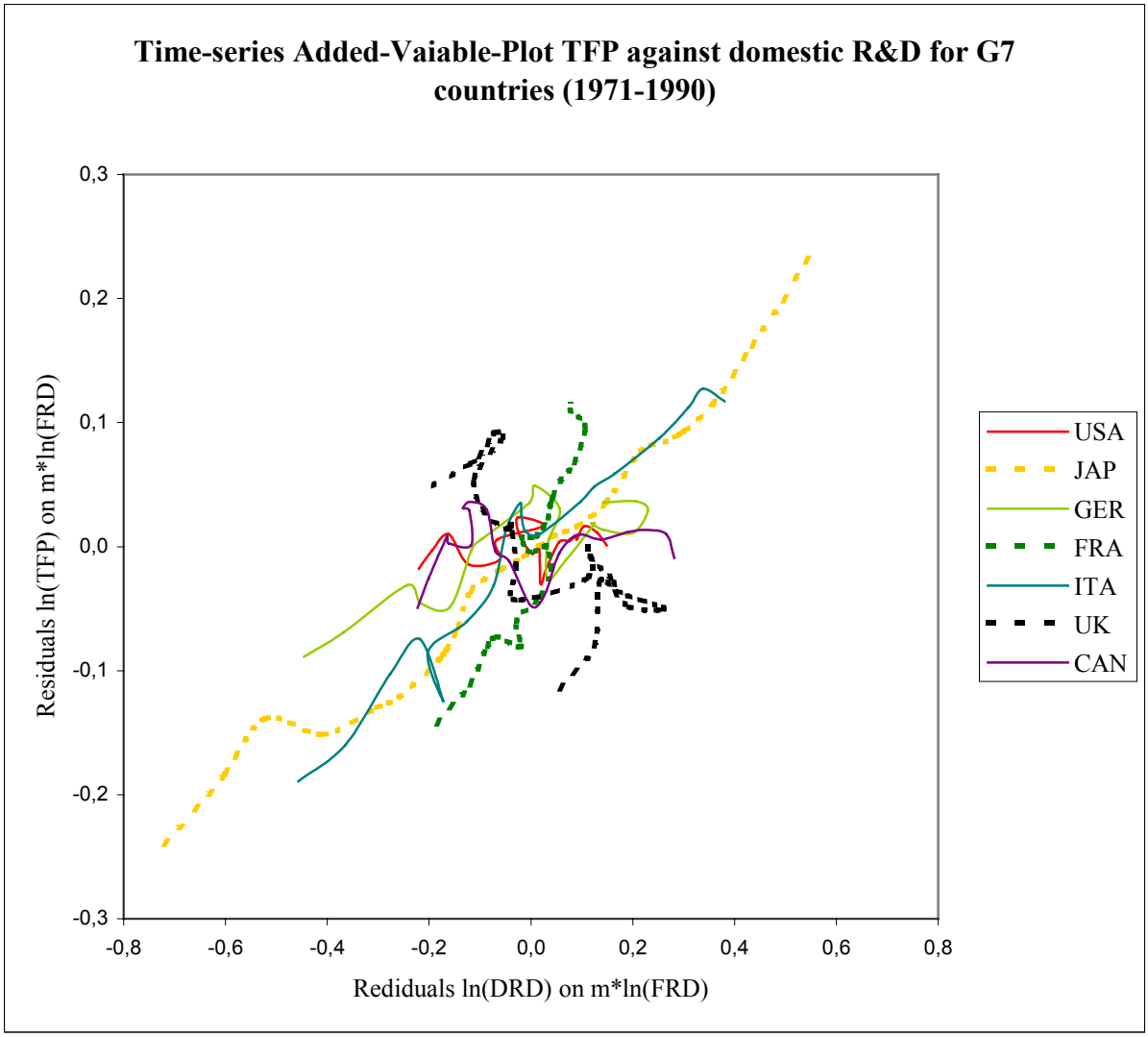

Figure 5: Time-series Added-Variable-Plot TFP against DRD for G7 countries (1971-90).

In the A-V-P for DRD (Figure 5) on the other hand with the exception of the U.K. all countries exhibit an upward slope, which confirms the relevance of this factor.

All in all, it is evident from this exploratory analysis that the use of added-variable-plots in an early phase of their study would have prevented Coe and Helpman (1995) from premature conclusions.

5. A New Model

After a detailed examination of the model of Coe and Helpman (1995) and the various critics of it, the following changes and modifications for this model are suggested:

- use of a random coefficient model,
- use of DOLS regression.

The advantage of the random coefficient model over the fixed effect model is that there is no need for the assumption that there is no variation between the cross section units (countries). This parameter heterogeneity is regarded as a random variation. The fit of the model improves by allowing for the random variation of the single parameters $\boldsymbol{\beta}_i$ around $\boldsymbol{\beta}$.

The use of dynamic regressors is based on the paper of Kao *et al.* (1999) where the DOLS estimator and its advantage over the simple OLS estimator is explained. Thus q_1 lags and q_2 leads of the first differences of the domestic and foreign R&D expenditures should be included as additional dynamic regressors in the model with domestic and foreign R&D expenditures. Kao *et al.* (1999) tested the assumption of cointegration of the estimated equations. They used the panel cointegration test of Kao (1999) and the test of Pedroni (1995). All test-statistics were significant and therefore the null-hypothesis of no cointegration was rejected (see Kao *et al.*, 1999: Table 2). Edmond (2000) used cointegration tests of Pedroni (1997, 1998) and the augmented Dickey-Fuller test to test the assumption of cointegration and came to the same result as Kao, Chiang and Chen, 1999 (Edmond, 2000: Table 2). Because of these results the R&D Spillovers data can be regarded as cointegrated.

The suggestion for the analysis of the international R&D Spillovers is a random coefficient panel cointegration model with dynamic regressors. In this case Coe and Helpman's (1995) model has the following specification:

$$\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i ,$$

with

$\boldsymbol{y}_i$ being the regressand $T-(q_1+q_2+1)$ vector; here the natural logarithm of TFP of country i: $ln\, F_i$,

$\boldsymbol{X}_i$ denotes the $T-(q_1+q_2+1)\times 1+k+(q_1+q_2+1)k$ regressor-matrix; here $\boldsymbol{X}_i = \left[\boldsymbol{1} : ln\, \boldsymbol{S}_i^d : m\, ln\, \boldsymbol{S}_i^f : \Delta \boldsymbol{X}_i\right]$, where $ln\, \boldsymbol{S}_i^d$ and $ln\, \boldsymbol{S}_i^f$ are the natural logarithms of the domestic- and foreign R&D expenditures and m are the import-shares. $\Delta \boldsymbol{X}_i$ is the group-specific matrix of the differences. For all elements of the original matrix $\boldsymbol{X}_i$ ($ln\, \boldsymbol{S}_i^d$ and $ln\, \boldsymbol{S}_i^f$) the values for the respective years get subtracted for all lags and leads (inclusive $lag = 0$). The dimension of matrix $\Delta \boldsymbol{X}_i$ is thus: $T-(q_1+q_2+1)\times(q_1+q_2+1)k$.

$\boldsymbol{\beta}_i$ $1+k+(q_1+q_2+1)k$ is then the vector of parameters. The first entry is the common intercept, the second and third entry are the parameters of the domestic and foreign R&D expenditures and the other entries are the parameters of the differences.

$\boldsymbol{\varepsilon}_i$ $T-(q_1+q_2+1)$ finally denotes the corresponding vector of errors.

In contrast to Coe and Helpman's model, where all years T are included, q_1+q_2+1 years get lost in this model by forming the difference-matrix. Eventually, the DOLS random coefficient estimation yields

$$\log F_{it} = \hat{\alpha}_{it}^{0} + 0{,}3062 \log S_{it}^{d} - 0{,}0756 m_{i,t-1} \log S_{it}^{f} + Rest$$
$$(8{,}0541)^{**} \qquad (0{,}4051)$$

with an R^2 of *0,9766* and an Adjusted R^2 of *0,9750*. The computations were performed with a special GAUSS package, which is described in detail in N.Gumprecht (2003).

The coefficient estimate corresponding to $m \cdot \log S_f$ is not significant. R^2 and Adjusted R^2 are even higher than in the random coefficient model without dynamic regressors.

The results of the panel cointegration model with random coefficient and dynamic regressors do not support Coe and Helpman's hypothesis, that the TFP of a country depends on domestic and foreign R&D knowledge (measured by the R&D expenditures). The effect of the knowledge of the trade-partners of a country is marginal because it is not significant. It seems as foreign R&D do rather not affect the TFP of a country.

6. Conclusions

Coe and Helpman's hypothesis, that the TFP of a country depends on the domestic and foreign R&D knowledge can only be supported partly. Imported knowledge seems to have no effect on TFP of a country.

A summary of different articles about the relationship of imported knowledge and the TFP of a country is provided by Navaretti and Tarr (2000). They concluded that there is a strong evidence for the positive effect of imported technology on TFP of a county. The reason for this completely different conclusion might be the level of aggregation of the analysed data. Navaretti and Tarr (2000) used articles that cared especially about microeconomic relationships between trade and knowledge diffusion whereas the here discussed articles care about a macroeconomic relationship between R&D and TFP. One referee has noted the relation of the results to the theory of absorptive capacities (Cohen and Levinthal, 1989) – a respective interaction effect could be tested in a slightly alternative specification.

The preferred panel cointegration model with random coefficients and dynamic regressors confirms the positive effect of domestic R&D on TFP but it does not confirm the effect of foreign R&D on TFP.

References

Baltagi, Badi H. 2001. *Econometric analysis of panel data*. 2nd Edition. West Sussex: John Wiley & Sons.

Chiang, Min-Hsien; Chihwa Kao. 2002. *Nonstationary Panel Time Series Using NPT 1.3 – A User Guide*. National Cheng-Kung University and Syracuse University.

Chiang, Min-Hsien; Chihwa Kao. 2002. „Chihwa Kao." <http://web. syr.edu/~cdkao> (31.07.2002).

Cohen, Wesley; Levinthal, Daniel. 1989. "Innovation and Learning: The Two Faces of R&D." *Economic Journal* 99, 569-196.

Coe, David T.; Elhanan Helpman. 1995. "International R&D spillovers." *European Economic Review* 39, 859-887.

Cook, Dennis R.; Stanford Weisberg. 1994. *An introduction to regression graphics*. New York: Wiley.

Dickey D.; W. Fuller. 1979. "Distribution of the Estimators for Autoregressive Time Series with a Unit Root." *Journal of the American Statistical Association* 74.

Dickey D.; W. Fuller. 1981. "Likelihood Ratio Tests for Autoregressive Time Series with Unit Root." *Econometrica* 49.

Edmond, Chris. 2000. "Some panel cointegration models of international R&D spillovers." *Journal of Macroeconomics*.

Engle, Robert F.; Clive W.J. Granger. 1987. "Co-integration and Error Correction: Representation, Estimation and Testing." *Econometrica* 55.

Engle, Robert. F.; B. Yoo. 1987. "Forecasting and Testing in Cointegrated Systems." *Journal of Econometrics* 35.

Granger, C.W.J.; P. Newbold. 1974. "Spurious regressions in econometrics." *Journal of Econometrics* 2, 111-120.

Greene, William H. 2000. *Econometric analysis*. 4th Edition. New Jersey: Prentice Hall.

Grossman, Gene M.; Elhanan Helpman. 1991. "Innovation and growth in the global economy." *MIT Press, Cambridge, MA.*

Gumprecht, Daniela, 2003, „*Ein Panel Kointegrationsmodell für internationale Forschungs- und Entwicklungs Spillovers*", unpublished master thesis at the Department of Statistics and Decision Support Systems, University of Vienna, 2003.

Gumprecht, Nicole. 2003. „*Regression mit zufälligen Koeffizienten: Software Implementierung"*. unpublished master thesis at the Department of Statistics and Decision Support Systems, University of Vienna.

Helpman, Elhanan. 2003 (aktueller Stand). „Professor Elhanan Helpman's Data On The Web." <http://post.economics.harvard.edu/faculty/helpman/data.html> (24.01.03).

Hsiao, Cheng. 1986. *Analysis of panel data*. 1st Edition. New York: Cambridge University Press.

Johnston, John. 1984. *Econometric Methods*. 3rd Edition. New York: McGraw-Hill.

Kao, Chihwa. 1999. "Spurious Regression and Residual-Based Tests for Cointegration in Panel Data." *Journal of Econometrics* 90, 1-44.

Kao, Chihwa; Min-Hsien Chiang. 1997. *On the Estimation and Inference of a Cointegrated Regression in Panel Data*. Syracuse University.

Kao, Chihwa; Min-Hsien Chiang; Bangtian Chen. 1999. „International R&D Spillovers: An application of estimation and inference in panel cointegration." *Oxford Bulletin of Economics and Statistics* 61, 693-711.

Keller, Wolfgang. 1998. “How Trade Patterns and Technology Flows affect Productivity Growth.” *National Bureau of Economic Research* [Working Paper].

Levin, Andrew; Chien-Fu Lin. 1992. *Unit foot tests in panel data: asymptotic and finite-sample properties*. University of California, San Diego, CA [Discussion paper 92-93].

Levin, Andrew; Chien-Fu Lin. 1993. *Unit root tests in panel data: New results*. University of California, San Diego, CA [Discussion paper 93-56].

Lichtenberg, Frand; Bruno van Pottelsberghe de la Potterie. 1998. “International R&D Spillovers: A Comment.” *European Economic Review* 42.

Müller, Werner G.; Michaela Nettekoven. 1999. „A panel data analysis: research and development spillover.“ *Economics Letters* 64, 37-41.

Navaretti, Giorgio B.; David Tarr. “International Knowledge Flows and Economic Performance: A Review of the Evidence.” *The World Bank Economic Review* 14, No. 1: 1-15.

Pedroni, Peter. 1995. *Panel Cointegration: Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the PPP Hypothesis*. Indiana University [Working Paper in Economics No. 95-013].

Pedroni, Peter. 1997. *Panel Cointegration; Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the PPP Hypothesis: New Results*. [Manuscript].

Pedroni, Peter. 1998. *Approximate Critical Values for Cointegration Tests in Heterogeneous Panels with Multiple Regressors.* [Manuscript].

Romer, P.M. 1990. “Endogenous technical change.” *Journal of Political Economy* 98.

Saikkonen, Pentti. “Asymptotically efficient estimation of cointegration regressions.” *Econometric Theory* 7, 1-21.

Stock, J.H. 1987. “Asymptotic properties of least squares estimations of co-integrating vectors.” *Econometrica* 55.

Swamy, P. 1971. Statistical Inference in Random Coefficient Regression Models. New York: Springer-Verlag.