Do Spillovers Matter When Estimating Private Returns to R&D?

Markus Eberhardt† Christian Helmers† Hubert Strauss♯

† University of Oxford
♯ European Investment Bank

18th March 2010
COINVEST Conference, Lisbon

Overview — Returns to R&D

- Firms **invest in R&D** to achieve productivity gains through innovations
- Investment in R&D depends on **expected return** (absolute and relative)
- Particular **characteristics of knowledge**:
  - Non-excludability
  - Non-exhaustability
- **Private and social returns** to R&D differ
- Substantial number of **empirical studies** assessing the private and social returns to R&D at the country, regional, industry and firm-level

Overview — Spillovers

- Most widely used approach is the **knowledge production function** originally proposed by Griliches (1979)
- In Griliches knowledge production function framework **spillovers are omitted** from the specification
- Vast economic literature on **importance of R&D spillovers for growth** (Romer, 1990; Grossman and Helpman, 1991)
- Research question: Can we identify the direct effect of R&D on productivity and its direct returns while omitting spillovers?

Overview — Cross-sectional Dependence

- **Presence of spillovers** can lead to correlation between cross-sectional units
- Spillovers can be regarded as **omitted unobserved factors** in the R&D variable as well as the error term
- If these unobserved factors are correlated, the resulting estimates of private returns to R&D are **biased and inconsistent**
- Account for spillovers through **spatial econometric approach** — but need for specific structure on cross-sectional dependence
- Instead adopt more agnostic **Common Factor approach**
- Preview of preliminary **findings**:
  - Strong evidence for presence of cross-sectional dependence
  - When spillovers in the form of cross-sectional dependence are accounted for, private returns to R&D are at best modest and possibly negative$\Rightarrow$ it's all in the spillovers!?!

Empirical Approach — Knowledge Production Function

- **Cobb-Douglas production function** (Griliches, 1979)

$$Y = AL^{\alpha}K^{\beta}R^{\gamma}e^{\lambda t+\varepsilon} \tag{1}$$

where Y is value-added, L labor, K tangible capital, R knowledge capital (both: stock variables), A is TFP-level, t a time index, e the base of the natural logarithm and ε a stochastic error term

- **Knowledge capital *stock*** is a function of current and past levels of R&D *investments*

$$R = G[W(B)R\&D] \tag{2}$$

where $W(B)$ is a lag polynomial with B being the lag operator.

Empirical Approach — Knowledge Production Function

- Combine (1) and (2), take logarithms and use subscripts i and t to denote cross-sectional units (sectors) and time respectively

$$y_{it} = \psi + \alpha l_{it} + \beta k_{it} + \gamma r_{it} + \lambda_t + \mu_i + e_{it} \tag{3}$$

- r_{it} measure for R&D capital *stock*
- $\gamma = \frac{\partial Y}{\partial R}\frac{R}{Y}$ indicates the elasticity of R&D capital
- $\star\ \rho^G = \gamma\frac{Y}{R}$ gross private rate of return
- $\star\ \rho^N = \rho^G - \delta$ net rate of return where δ is the depreciation rate of R&D capital
- λ_t vs. λt (latter standard)

$\star$ Thus far not completed these computations… for reasons that will become clear.

Data — Overview

- **12 OECD countries** (Czech Republic, Denmark, Finland, Germany, Italy, Japan, Netherlands, Portugal, Slovenia, Sweden, United Kingdom, and the US);
- **12 manufacturing industries**;
- **26 years** spanning over 1980-2005;
- **Sources**: Main source EU KLEMS, R&D expenditure from OECD, GDP deflators from Eurostat and OECD;
- **Output**: Value added (double-deflated by subtracting real input from real output);
- **Labour input**: number of hours worked by persons engaged;
- **Capital input**: tangible assets by book value recorded annually;
- **R&D stocks**: see later slide…

Data — Sectors

SIC	Description: Manufacture of	No. Obs.
15, 16	Food, beverages, tobacco	234
17, 18, 19	Textiles, textile products, leather and leather products	234
20	Wood and products of wood and cork	232
21, 22	Pulp, paper, paper products, printing and publishing	232
24	Chemicals and chemical products	234
25	Rubber and plastic products	223
26	Other non-metallic mineral products	234
27, 28	Basic metals and fabricated metal products	234
29	Machinery and equipment n.e.c.	234
30, 31, 32, 33	Electrical and optical equipment	234
34, 35	Transport equipment	234
36, 37	Manufacturing n.e.c.	234
Total		2,793

Data — R&D stocks

- **Construct R&D stocks** applying perpetual inventory method

$$R_{it} = (1-\delta)R_{it-1} + R\&D_{it} \tag{4}$$

R&D denotes R&D *flows* and *R* the corresponding *stock*;

- Compute the **initial capital stock**, i.e.,

$$R_{i1} = R\&D_{i0}\sum_{t=0}^{\infty}\left[\frac{1-\delta}{1+g_i}\right]^t = \frac{R\&D_{i1}}{\delta+g_i} \tag{5}$$

g_i denotes the sector-specific growth rate of R&D investments;

- **Compute g_i** using the first seven years for which R&D expenditure is observed; as long as g_i and δ do not change dramatically within sectors over time, they will be captured by sector-specific effects. Set $\delta = .12$;
- Our **contribution**: create series for Portugal & Slovenia; extended series for all other countries to 2004-5.

Empirical Approach — Spillovers

Motivation

We do not deal with one closed industry, but with a whole array of firms and industries which borrow different amounts of knowledge from different sources according to their economic and technological distance from them (Griliches, 1979:103)

Empirical Approach — Spillovers

Standard approach to model spillovers

- Model TFP as

$$\text{TFP}_{it} = g\left(R_{it}, \sum_k R_{kt}\right) \tag{6}$$

  where R_{it} denotes the R&D stock within sector i and R_{kt} for $k \neq i$ denotes spillovers received from all other sectors

- Estimate Equation (6) as

$$\text{tfp}_{it} = \psi_i + \gamma r_{it} + \chi \sum_k r_{kt} + e_{it} \tag{7}$$

- Equation (7) assumes that spillovers affect TFP linearly (and homogeneously) as captured by the corresponding parameter χ.

Empirical Approach — Spillovers & Cross-sectional Dependence

- **Common factor approach** regards cross-sectional dependence as the result of unobserved, time-varying omitted common shocks that affect each cross-sectional unit differently
- Cross-sectional dependence leads to **inconsistent estimates** if regressors are correlated with the unaccounted common shocks
- Error term is regarded as a linear combination of common time-specific effects with **heterogeneous factor loadings** and an i.i.d. error term
- In addition can incorporate **local spillovers** (within-country; neighbours) with ‘weak’ common factors

Empirical Approach — Cross-sectional Dependence

- Rewrite the **knowledge production function** in Equation (3) denoting the vector of inputs including R&D stock as $\mathbf{X}$

$$y_{it} = \boldsymbol{\beta}'\mathbf{X}_{it} + u_{it} \tag{8}$$

- **Structure of cross-sectional dependence** is described as

$$u_{it} = \boldsymbol{\varphi}_i'\boldsymbol{f}_t + \psi_i + \varepsilon_{it} \tag{9a}$$

$$X_{mit} = \boldsymbol{\chi}_{mi}'\boldsymbol{f}_t + \boldsymbol{\pi}_{mi}'\boldsymbol{g}_t + e_{mit} \tag{9b}$$

where $e_{mit} \sim iid(0, \sigma_{e,i}^2)$ and $\varepsilon_{it} \sim iid(0, \sigma_{\varepsilon,i}^2)$

- $\boldsymbol{f}_t$ contains a fixed number of unobserved common factors ('strong') and influences both regressors and the error term; infinite number of 'weak' common factors (Chudik et al, 2009)
- Estimating (8) without accounting for $\boldsymbol{f}_t$ produces **biased and inconsistent estimates**.

Empirical Approach — Accounting for Cross-sectional Dependence

- Account for cross-sectional dependence by **using cross-section averages** of the dependent and independent variables (Pesaran, 2006)
- Cross-section averages are defined as $\bar{y}_t = N^{-1} \sum_{i=1}^{N} y_{it}$ and $\bar{\mathbf{X}}_t = N^{-1} \sum_{i=1}^{N} \mathbf{X}_{it} \; \forall t$
- Consider (8) in cross-section averages (note that $\bar{\varepsilon}_t = 0$):

$$\bar{y}_t = \bar{\psi} + \bar{\boldsymbol{\beta}}' \bar{\mathbf{X}}_t + \bar{\boldsymbol{\varphi}}' \bar{\boldsymbol{f}}_t \tag{10}$$

which can be expressed as

$$\bar{\boldsymbol{f}}_t = \bar{\boldsymbol{\varphi}}^{-1} (\bar{y}_t - \bar{\psi} - \bar{\boldsymbol{\beta}}' \bar{\mathbf{X}}_t) \tag{11}$$

- If $\bar{\boldsymbol{\varphi}} \neq 0$, the unobserved common factors are captured by cross-sectional means of y and $\mathbf{X}$ since $\bar{\boldsymbol{f}}_t \xrightarrow{p} \boldsymbol{f}_t$.

Empirical Approach — Common Correlated Effects Pooled estimators

- **CCEP**: Augment pooled FE model with cross-section averages of the dependent and indep. variables (Pesaran, 2006)

$$y_{it} = \psi_i + \boldsymbol{\beta}'\mathbf{X}_{it} + \sum_{j=1}^{N} \theta_{1i}(\bar{y}_t D_j) + \sum_{k=1}^{m} \sum_{j=1}^{N} \boldsymbol{\theta}_{2ki}(\bar{X}_{kt} D_j) + e_{it} \quad (12)$$

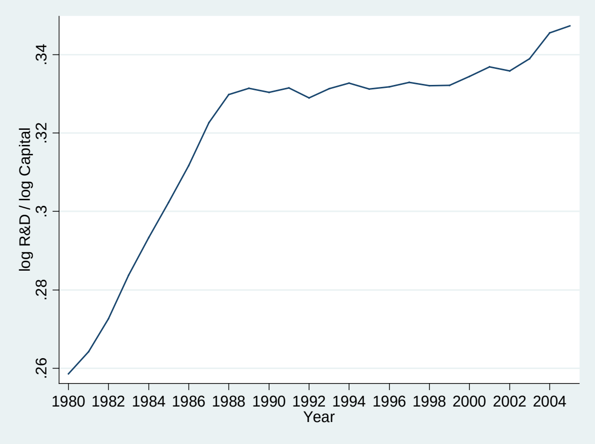
  where $e_{it} \sim iid(0, \sigma^2)$, $\bar{y}_t$ and $\bar{X}_{kt}$ are cross-sectional means for each t of y and the m covariates making up $\mathbf{X}$

- **Mean Group** variant of this estimator (CCEMG), where a simple regression equation for each panel member is augmented with the cross-section averages for y and $\mathbf{X}$, then estimates are averaged across i. Here: limited $T \rightarrow$ imprecise estimator.
- Estimators robust to breakdown in cointegrating relationship.

Estimation

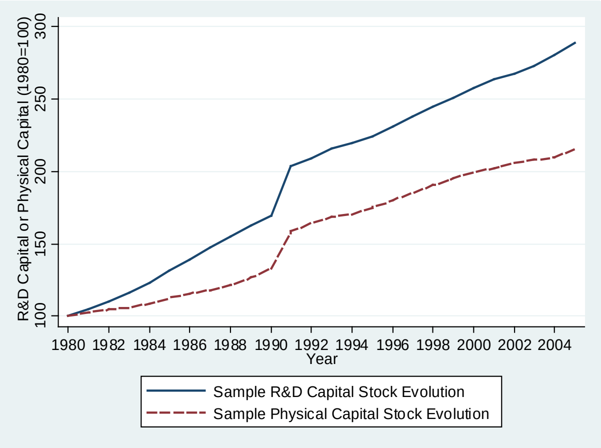
- **Static Pooled** Specification:
  - OLS (POLS)
  - OLS w/ country-sector dummies (FE and 2FE)
  - FE w/ data in deviation from cross-sectional avgs (CDFE)
  - OLS with data in first difference (FD)
  - Pesaran (2006) CCE Pooled estimator (CCEP)
- **Dynamic Pooled** ARDL(1,1) Specification:
  - as above but drop FD estimators (require instrumentation)
  - Arellano & Bond, 1991 (AB) and Blundell & Bond, 1998 (BB)
- **Static Heterogeneous** Specification:
  - Pesaran & Smith (1995) Mean Group Estimator (MG)
  - MG w/ data in deviation from cross-sectional avgs (CDMG)
  - Pesaran (2006) CCE Mean Group estimator (CCEMG)

R&D-Physical Capital Ratio 1980-2005



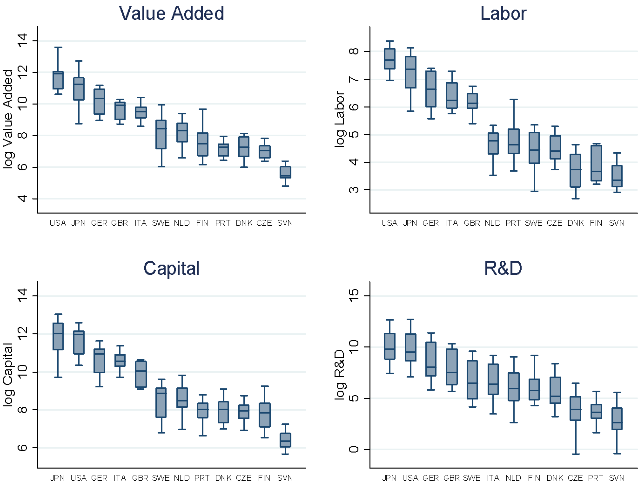
Notes: Aggregate evolution for R/K across 12 countries.

Evolution of R&D and Physical Capital Stock (1980 = 100)

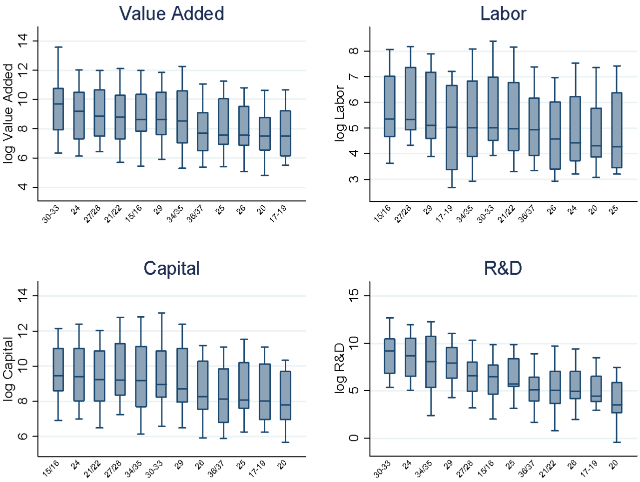


Notes: The kink in 1991 is caused by Germany, a major innovator, entering our sample.

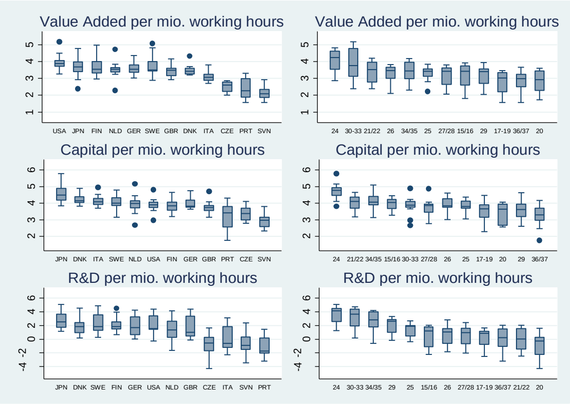
Variation in Inputs across Countries for SIC 2-digit Industries (2005)



Variation in Inputs across SIC 2-digit Industries (2005)



Labour-Deflated Input Variation across Countries & Industries



Notes: The data is transformed into mio. Euros per mio. working hours (in logs) and plotted in order of median value. The left column plots variation by country, the right column by SIC 2-digit industry. All data presented in this graph are for 2005.

Possible factor structure of the data

Principle Component Analysis

Share of variation in the data explained by the first two PCs.

	Panel A: Variables in Levels				
	$\ln Y$	$\ln L$	$\ln K$	$\ln R$	all four
$T = 9, N = 119$					
Variance expl. by 1st & 2nd Component	0.760	0.836	0.948	0.930	0.862
$T = 24, N = 84$					
Variance expl. by 1st & 2nd Component	0.797	0.730	0.922	0.946	0.838

	Panel B: Variables in First Diff.				
	$\Delta\ln Y$	$\Delta\ln L$	$\Delta\ln K$	$\Delta\ln R$	all four
$T = 8, N = 119$					
Variance expl. by 1st & 2nd Component	0.472	0.563	0.691	0.648	0.568
$T = 23, N = 84$					
Variance expl. by 1st & 2nd Component	0.349	0.371	0.441	0.509	0.342

Possible factor structure of the data

Average Correlation Coefficients and Pesaran (2004) CD-test

PANEL A: VARIABLES IN LEVELS				
	$\ln Y$	$\ln L$	$\ln K$	$\ln R$
avg ρ	0.22	0.24	0.51	0.32
avg $\|\rho\|$	0.59	0.59	0.79	0.77
CD	101.93	104.27	198.69	145.50
PANEL B: VARIABLES IN FD				
	$\ln Y$	$\ln L$	$\ln K$	$\ln R$
avg ρ	0.13	0.15	0.16	0.02
avg $\|\rho\|$	0.30	0.31	0.36	0.39
CD	53.17	58.70	62.44	11.57
PANEL C: AR(2) RESIDUALS				
	$\ln Y$	$\ln L$	$\ln K$	$\ln R$
avg ρ	0.10	0.11	0.08	0.04
avg $\|\rho\|$	0.32	0.32	0.33	0.30
CD	43.04	46.21	37.84	13.71

Results — Static Specification

	POLS [1]	FE [2]	2FE [3]	CDFE [4]	CCEP [5]	CCEP [6]	FDOLS [7]	FDOLS [8]
ln L_{it}	**0.412**	**0.382**	**0.608**	**0.612**	**0.600**	**0.610**	**0.544**	**0.641**
	[35.30]**	[13.13]**	[18.46]**	[19.47]**	[17.92]**	[17.41]**	[16.10]**	[18.25]**
ln K_{it}	**0.529**	**0.748**	**0.492**	**0.553**	**0.282**	**0.117**	**0.511**	**0.309**
	[41.36]**	[23.96]**	[11.04]**	[15.74]**	[5.97]**	[2.31]*	[9.68]**	[4.54]**
ln R_{it}	**0.096**	**0.114**	**0.062**	**0.082**	**0.099**	**0.065**	**0.122**	**0.041**
	[22.75]**	[8.35]**	[4.36]**	[7.21]**	[4.54]**	[2.90]**	[5.22]**	[1.56]
Year dummies	Included		Implicit	Implicit		Included		Included
CRS	0.00	0.00	0.00	0.00	0.69	0.00	0.01	0.91
AB AR(1)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
AB AR(2)	0.00	0.00	0.00	0.00	0.78	0.79	0.03	0.01
Residuals	I(1)	I(1)	I(1)	I(1)	I(0)	I(0)	I(0)	I(0)
CD Test (p)	24.37 (.00)	15.34 (.00)	-1.51 (.13)	2.28 (.02)	8.92 (.00)	-0.99 (.32)	34.66 (.00)	-0.11 (.91)
Obs	2793	2793	2793	2793	2793	2793	2650	2650
R-squared	0.97	0.99	0.99	0.99	1.00	1.00	0.24	0.32

Notes: Robust t-statistics in parentheses. CRS tests constant returns to scale (H_0 CRS), AB is the Arellano and Bond (1991) serial correlation test (H_0 no serial correlation), Residuals reports the results from Pesaran (2007) panel unit root tests with I(1) integrated of order 1 and I(0) integrated of order zero (stationary).
Estimators: POLS — pooled OLS, FE — country-sector FE, 2FE — two-way FE, CDFE — cross-sectionally demeaned FE, CCEP — Pooled Common Correlated Effects estimator, FDOLS — pooled OLS with variables in first difference.

Results — Dynamic Specification

	POLS	FE	2FE	AB	BB	CDFE	CCEP	CCEP
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
PANEL A: UNRESTRICTED ARDL MODEL COEFFICIENTS — see paper								
PANEL B: LONG-RUN COEFFICIENTS (UNRESTRICTED MODEL)								
Labour	**0.418**	**0.437**	**0.700**			**0.328**		
	[4.31]**	[3.20]**	[3.95]**			[1.92]		
Capital	**0.198**	**0.607**	**0.402**			**0.578**		
	[1.33]	[6.00]**	[2.96]**			[3.66]**		
R&D stock	**0.356**	**0.011**	**-0.023**			**0.160**		
	[4.16]**	[0.19]	[0.38]			[2.29]*		
Year dummies	included		implicit	included	included	implicit		included
COMFAC	0.00	0.00	0.01	0.58	0.18	0.00	0.20	0.16
CRS	0.46	0.62	0.48	0.47	0.43	0.40	0.28	0.00
AB AR(1)	0.00	0.97	0.93	0.00	0.00	0.41	0.01	0.55
AB AR(2)	0.02	0.27	0.54	0.15	0.79	0.46	0.03	0.00
CD-test (p)	-0.62 (.54)	25.76 (.00)	-0.73 (.46)	-0.92 (.36)	-1.47 (.14)	7.31 (.00)	7.76 (.00)	0.37 (.37)
PANEL C: LONG-RUN COEFFICIENTS (COMMON FACTOR RESTRICTIONS IMPOSED)								
Labour				**0.918**	**0.543**		**0.556**	**0.579**
				[6.23]**	[4.38]**		[10.21]**	[10.66]**
Capital				**-0.385**	**0.299**		**0.289**	**0.197**
				[1.70]	[2.21]*		[3.65]**	[2.32]*
R&D stock				**-0.147**	**0.094**		**0.022**	**0.000**
				[1.14]	[1.67]		[0.77]	[0.01]

Notes: COMFAC is a test for common factor restrictions in the ARDL model. Residual tests for all models: I(1)/I(0). Sargan tests rejects instrument validity in AB, BB.

Results — Heterogeneous parameter specification

	Country-sector regressions			Country regressions			Sector regressions		
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
	MG	CDMG	CCEMG	POLS	2FE	CCEP	POLS	2FE	CCEP
ln L_{it}	**0.614**	**0.610**	**0.641**	**0.405**	**0.257**	**0.226**	**0.402**	**0.214**	**0.184**
	[9.89]**	[10.22]**	[11.95]**	[11.34]**	[2.66]*	[2.15]	[12.95]**	[1.84]	[1.65]
ln K_{it}	**-0.049**	**0.343**	**0.178**	**0.478**	**0.693**	**0.776**	**0.421**	**0.353**	**0.430**
	[0.52]	[5.55]**	[1.83]	[24.16]**	[5.59]**	[6.06]**	[8.71]**	[3.34]**	[3.96]**
ln R_{it}	**-0.022**	**0.163**	**0.053**	**0.047**	**0.090**	**0.106**	**0.205**	**0.127**	**0.143**
	[0.40]	[4.63]**	[0.96]	[2.11]	[1.21]	[1.65]	[6.16]**	[1.69]	[1.87]
trend	**0.025**								
	[7.55]**								
CD Test (p)	25.98 (.00)	4.91 (.00)	4.95 (.00)						
Residuals	I(1)/I(0)	I(1)/I(0)	I(0)						
Panel-t ln L_{it}	26.46**	34.10**	23.29**						
Panel-t ln K_{it}	-0.69	26.67**	6.12**						
Panel-t ln R_{it}	-2.82**	20.31**	4.28**						
Panel-t trend	36.99**								
Sum	0.543	1.115	0.872	0.930	1.040	1.108	1.028	0.694	0.757
reject CRS	40%	71%	39%						
Avg n	19.5	19.5	22.2	232.8	232.8	232.8	232.8	232.8	232.8
N	143	143	119	12	12	12	12	12	12

Notes: We present robust means for country-sector, country- and sector estimates in each case. N refers to the panel members from which these averages are calculated: in the sector (country) case estimates for 12 sectors (countries) and in the country-sector case 143 country-sectors. For CCEMG we had to drop Slovenia and Czech Republic due to the short time-series span in their data. Panel-t statistics constructed following Pedroni (2000) are standard normal distributed.

Findings

- **What we set out to do:** Contrasted estimates for a Griliches knowledge production function across empirical specifications with different assumptions about error term independence (lack of spillover effects) as well as technology homogeneity across countries and/or sectors
- **Initial findings:**
  1. Strong indication for presence of spillovers and the indivisibility of R&D from spillovers
  2. Ignoring cross-sectional dependence yields positive and statistically & economically significant private returns to R&D
  3. Once cross-sectional dependence accounted for: no evidence for positive private returns to R&D
- **Our interpretation:** Estimates of elasticity of private returns to R&D ignoring cross-sectional dependence confound direct effect of R&D with combined effect of own R&D and spillovers.

Thank you.

Markus Eberhardt
St Catherine's College, Oxford and CSAE

Christian Helmers
Wolfson College, Oxford and UC Berkeley

Hubert Strauss
European Investment Bank