European Commission

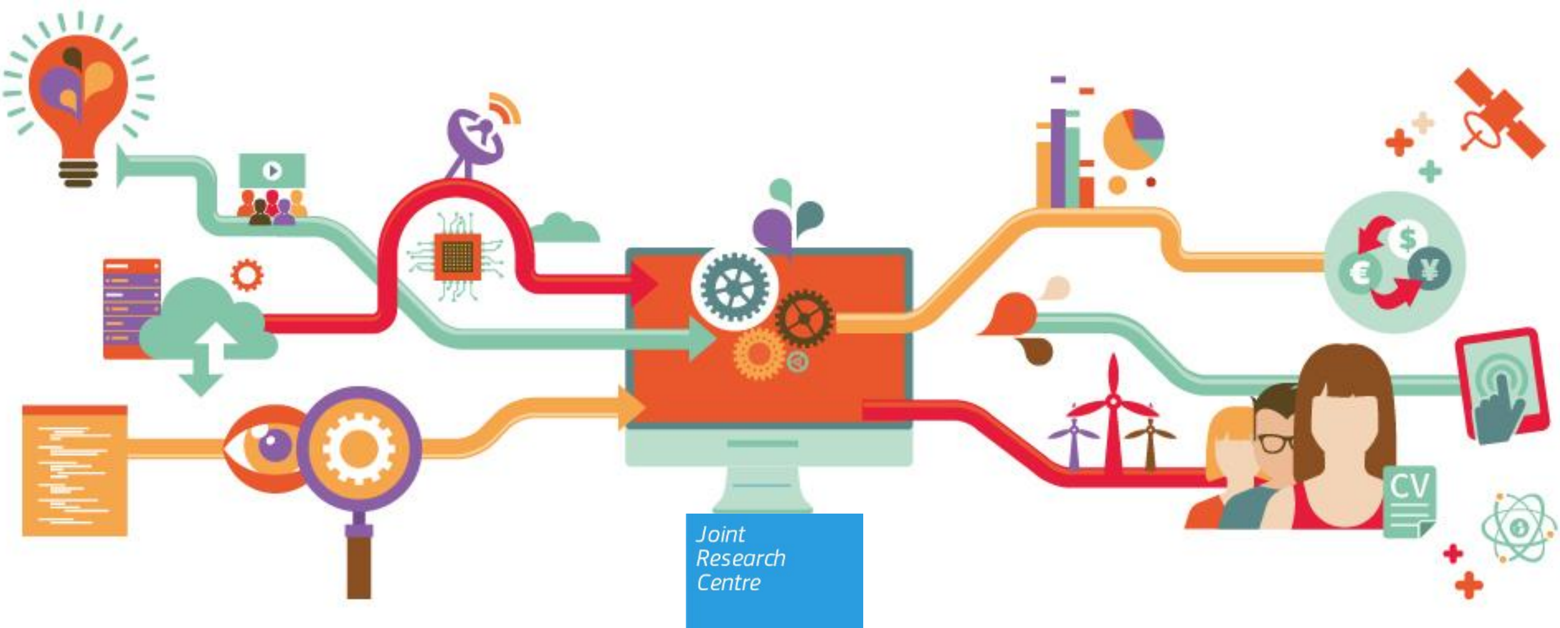
JRC TECHNICAL REPORTS

Who's doing who? Growth of sales, employment, assets, profits and R&D entangled in a curious five-way love triangle

IPTS Working Papers on Corporate R&D and Innovation No 03/2016

Alex COAD, Nicola GRASSANO

2016



Joint Research Centre

European Commission
Joint Research Centre
Institute for Prospective Technological Studies

Contact information
Fernando Hervás Soriano
Address: Edificio Expo. c/ Inca Garcilaso, 3. E-41092 Seville (Spain)
E-mail: jrc-ipts-kfg-secretariat@ec.europa.eu
Tel.: +34 954488463
Fax: +34 954488316

JRC Science Hub
https://ec.europa.eu/jrc

JRC98745

ISSN 1831-9408(online)

How to cite: Coad, A. and Grassano, N. (2016). Who's doing who? Growth of sales, employment, assets, profits and R&D entangled in a curious five-way love triangle, JRC-IPTS Working Papers on Corporate R&D and Innovation No 03/2016, Institute of Prospective Technological Studies, Joint Research Centre.

The JRC-IPTS Working Papers on Corporate R&D and Innovation addresses economic and policy issues related to industrial research and innovation and to the competitiveness of the European industry. Mainly addressed to policy analysts and the academic community, these are policy relevant early-stage scientific articles highlighting policy implications. These working papers are meant to communicate to a broad audience preliminary research findings, generate discussion and attract critical comments for further improvements. All papers have undergone a peer review process.

Who's doing who? Growth of sales, employment, assets, profits and R&D entangled in a curious five-way love triangle[1],[2]

Alex Coad, Nicola Grassano

European Commission, Joint Research Centre, Seville, Spain

Abstract

Understanding causal relationships among key economic variables is crucial for policy makers, who wish to e.g. stimulate private R&D growth. . To this end, we applied a technique recently imported from the Machine Learning community (Structural Vector Autoregressions (SVARs) identified using Independent Components Analysis (ICA)) to a set of the world's largest R&D investors. Our analysis highlights the key role of sales growth, rather than profits growth, in stimulating R&D growth. R&D growth appears at the end of the causal ordering of the growth process. Our results suggest that policies to increase private R&D would do better to target sales rather than profits.

Keywords: R&D investment, firm growth, SVAR, sales growth, industrial dynamics

JEL Classification: L25, O30

[1] We are grateful to Andries Brandsma, Michele Cincera, Sven-Olov Daunfeldt, Paul Desruelle, Fernando Hervás, Dominik Janzing, Koen Jonkers, Roberto Martino, Pierre Mohnen, Edward Woolley, Daniel Vertesy, Antonio Vezzani, Marco Vivarelli, and participants at the European Commission's 6th IRIMA workshop on "R&D Investment and Firm dynamics" (Brussels) and at JRC-IPTS (Seville). Any remaining errors are ours alone.

[2] This Working Paper is issued in the context of the **Industrial Research and Innovation Monitoring and Analysis (IRIMA)** activities that are jointly carried out by the European Commission's Joint Research Centre (JRC) – Institute for Prospective Technological Studies (IPTS) and the Directorate General Research and Innovation - Directorate A, Policy Development and Coordination.

1. Introduction

A fundamental target of innovation policy is to boost the R&D expenditures of private firms. The Lisbon Strategy of the European Commission put forward, back in 2000, the target of having 3% of GDP invested in R&D expenditures, and this 3% target has been reaffirmed by the Europe 2020 Strategy (published in 2010) and subsequent policy documents. Despite the policy interest in encouraging firms to invest more in R&D, however, controversy still surrounds our knowledge of the determinants of private firm's R&D expenditure.

One influential theory on firm growth and R&D investment is the 'demand-pull' approach, which broadly suggests that the growth of demand (measured either at the level of industries or firms) provides firms with the incentives to engage in innovation activities and to introduce new innovations (Schmookler, 1966; Scherer, 1982; Kleinknecht and Verspagen, 1990; Piva and Vivarelli, 2007). The demand-pull approach would predict, in our context, that growth of R&D investment is caused by previous sales growth.

In contrast, the technology-push hypothesis posits that it is the availability of technological opportunities, and the introduction and commercialization of inventions, that stimulates the evolution of firms and industries (see e.g. Dosi, 1988; Breschi et al., 2000; and Dosi and Nelson 2013). With regards to our research context, one interpretation of the technology-push approach would be that that the growth of firms (e.g. in terms of sales, employment, capital, and perhaps also profits) is driven by the successful innovations brought about by prior investments in R&D.

A different perspective holds that it is profits that are the driver of investments in R&D. According to Schumpeterian intuitions, financial performance can be expected to play a major role in driving R&D expenditures (Himmelberg and Petersen, 1994). Scherer (2001) observes how profits and R&D expenditure coevolve in the pharmaceutical sector, and suggests that profits are a major driver of R&D investment.

Other authors are more sceptical of the alleged role of financial performance in driving R&D investment, however. Kamien and Schwartz (1975, page 26) survey the evidence and conclude: "In sum, the empirical evidence that either liquidity or profitability are conducive to innovative effort appears slim." Relatedly, Himmelberg and Petersen (1994, page 38) write that "Since Schumpeter, economists have argued that internal finance should be an important determinant of R&D expenditures … almost without exception, previous empirical studies have not found evidence of such a relationship"

Thompson (1999) puts forward an alternative model of R&D investment. In this view, firms are boundedly rational, and are far from being able to compute the infinite-horizon profit-maximizing R&D investment levels, because of uncertainties in the business environment, and uncertainties about the outcomes of R&D investments. Instead, firms follow rules of thumb, such as investing a fixed proportion of their sales into R&D.[3]

[3] Ralph Gomory, former senior vice-president of IBM and former member of the US President's Council of Advisers on Science and Technology writes: "You have a product. The product is selling. That gives you a certain stream of revenue. You can take that stream of revenue and put some of it into R&D for the next round. Some of it has to be reserved for manufacturing, some of it for profits.

These conflicting views on the drivers of R&D have highlighted how firm growth and R&D expenditure are part of a complex, endogenous, co-evolving system. Sales growth may result in growth of profits, that may be reinvested in R&D. However, previous investments in employment and assets may be the drivers of subsequent sales growth, if firms need to prepare and plan ahead before realizing their growth projects (Penrose, 1959). R&D growth may lead to employment growth (if R&D projects and new products require more workers) or job cuts[4] (e.g. in the case of labour-saving process innovations), as well as increasing sales and profits, which may have further knock-on effects on other variables. In short, to better understand the causal relations that belie a firm's growth dynamics, we seek an econometric model that has several explananda or dependent variables – we seek to explain the determinants of R&D growth, sales growth, profits growth, etc, as well as how each of these variables affects each of the others. The appropriate econometric model, in this case, would be a vector autoregression model, where a vector of variables is regressed on (current and) lagged values of itself, with each variable taking its turn as dependent variable (Stock and Watson, 2001).

Coad and Rao (2010) applied a (reduced-form) vector autoregression model to Compustat data on large US firms, to observe the coevolution of growth of sales, employment, profits, and R&D expenditure. However, their econometric estimates were associations rather than causal effects, and hence the possibilities for informing policy interventions were limited.

Moneta et al., (2013) introduced a new technique into the econometrics literature, imported from the machine learning community (VAR-LiNGAM; Linear Non-Gaussian Acyclic Model), that was capable of obtaining causal estimates in a structural VAR framework. Moneta et al. (2013) tested this technique on the firm-level data in Coad and Rao (2010), and observed that sales growth played a major role in driving the process of overall firm growth and R&D investment. However, the audience for their paper consisted of econometric theorists, and they did not explore the robustness of their findings to any great extent.

The contribution of this paper is to apply the Moneta et al. (2013) SVAR model to rich data on firm-level R&D investments, controlling for possible confounding influences, to provide new evidence on the determinants and consequences of R&D investments for our sample of the world's largest R&D investors. We contribute by including growth of assets (i.e. capital expenditures) as a fifth dimension of the growth process, to better characterise the evolution of innovative growing firms and the dynamics of their investments in capital assets. Indeed, firms that grow (in terms of R&D, employment, etc) will need to support their expansion by adjusting their capital stocks and physical infrastructure, thus sparking interest in the sequential ordering of capital

---

Now, if you are on an upward swing and your product is succeeding, you have a flow back of money to invest in R&D; and if it isn't, you don't. And in my experience, and the experience of many other people, oddly enough, R&D is determined, more or less, as a percent of sales. It is not an independent variable. Let me say once more. **R&D is often a fixed percent of sales**. Now I exaggerate to make my point. Ten percent is a very reasonable sort of number in a high-tech industry… It may be that, in the correlation, which has often been remarked on, between R&D spending and industrial success, it is the industrial success which causes the R&D spending, not the other way around.'' Gomory (1992, p392), cited in Thompson (1999 p323, emphasis added.)

[4] For a recent contribution on the R&D-employment relationship see Bogliacino and Vivarelli (2012) and Vivarelli (2013).

expenditure growth in the unfolding growth process. We explore the robustness of our findings in a number of ways, including using a relatively new approach to exploring the frequencies of alternative causal orderings. Our results will also be a valuable addition to the literature for the reason that little is known about the performance of VAR-LiNGAM on economic datasets.

2. Methodology

2.1 Background

Policy interventions cannot be based on mere statistical associations or partial correlations between variables, but require an understanding of the causal relations underlying the system (Pearl, 2009). However, research into R&D investment and industrial dynamics generally produces estimates of associations rather than causal effects, because it is difficult to set up laboratory experiments involving firms, and some designs for obtaining causal estimates from observational data (e.g. instrumental variables, regression discontinuity design) are difficult to apply to data on industrial dynamics.

We apply a new technique, imported from the Machine Learning community (Computer Science), to gain new insights into the co-evolution of key variables in the growth process of innovative firms, and in particular to estimate the causal relations between these variables. This technique exploits the statistical information in the (non-gaussian) distributions of growth rate variables to infer directions of causality.

We use the VAR-LiNGAM method in Shimizu et al., (2006) and Moneta et al., (2013). Independent Components Analysis (ICA) (Hyvarinen et al., 2001; Stone, 2004) is used to extract the latent independent components in the SVAR series. These independent components are then arranged in order to produce the most likely causal relationships between variables.

2.2 Vector autoregressions

Consider the following vector autoregression model where the vector consists of two variables, sales growth ($SALES_{it}$) and R&D growth (RD_{it}), for firm i at time t. For simplicity, other variables are omitted and only one lag is included.

$$\begin{aligned} SALES_{it} &= b_{02}RD_{it} + b_{11}SALES_{i,t-1} + b_{12}RD_{i,t-1} + e1_{it} \\ RD_{it} &= b_{03}SALES_{it} + b_{13}SALES_{i,t-1} + b_{14}RD_{i,t-1} + e2_{it} \end{aligned} \quad (1)$$

Given that the vector autoregression (VAR) model consists of the vector $Y_{it} = (SALES_{it}, RD_{it})$, the model can be rewritten as:

$$Y_{it} = B_0(Y_{it}) + B_1(Y_{i,t-1}) + e_{it} \quad (2)$$

If we do not attempt to estimate the matrix of instantaneous effects B_0, then we estimate a reduced-form VAR model as follows (by rearranging (2)):

$$Y_{it} = \{[1-B_0].B_1\}.(Y_{i,t-1}) + [1-B_0]e_{it} \qquad (3)$$

Note that the reduced-form VAR model in (3) does not allow us to estimate the matrix of instantaneous causal effects B_0, nor to be able to properly estimate the matrix of lagged causal effects B_1, because to estimate this latter we need to separate it from the term $[1-B_0]$. Instead, the reduced-form VAR model in (3) can only describe the intertemporal associations between elements of Y_{it}. However, by estimating the matrix of instantaneous causal effects B_0, we can also correctly estimate the matrix of lagged causal effects B_1.

The matrix of instantaneous effects B_0 can be written, with reference to equation (1), as:

$$B_0 = \begin{bmatrix} 0 & b02 \\ b03 & 0 \end{bmatrix} \qquad (4)$$

Y_{it} is therefore a vector of variables regressed upon contemporaneous and lagged values of itself. The instantaneous causal effect of RD_{it} on $SALES_{it}$ is represented by b_{02}, and the instantaneous causal effect of RD_{it} on $SALES_{it}$ is represented by b_{03}. If we assume that the model is acyclic (i.e. no instantaneous feedback loops), then we impose that B_0 is lower-triangular matrix (or can be rearranged or 'row-permuted' to become lower-triangular), and that either b_{02} or b_{03} must be equal to zero.

2.3 Addressing endogeneity

The textbook definition of endogeneity states that, for a regression equation of the form $y_{it} = ax_{it} + e_{it}$, the residuals e_{it} are correlated with the explanatory variable x_{it} (see e.g. Wooldridge 2002, p50). If however x_{it} is uncorrelated with the residuals e_{it}, then x_{it} is said to be exogenous, and hence the causal channel runs from x to y.

Further refinement of the concept of causality in statistics has suggested that x must not only be uncorrelated with e, but fully statistically independent of e, because lack of correlation is a flawed indicator of statistical independence (Mooij et al., 2009).

A key problem affecting causal inference in social science, however, is that "everything correlates to some extent with everything else" - this has been dubbed the 'crud factor' affecting social sciences (Meehl, 1990, p204).

Our approach to unravelling the directions of causality is to apply a data-driven SVAR, more specifically VAR-LiNGAM (where LiNGAM stands for Linear Non-Gaussian Acyclic Model, see Shimizu et al., 2006 and Hyvarinen et al., 2008). The identification strategy is based on independent component analysis, is therefore is to recover the SVAR residuals e that are statistically independent of the explanatory variables.

The approach we take is to seek out the components of the variables in our SVAR system that are maximally statistically independent, to avoid the endogeneity that arises when all variables are correlated with each other. More specifically, we apply independent components analysis to uncover the SVAR residuals that are maximally independent of each other - not just uncorrelated, but fully statistically independent. The independent components correspond to the SVAR residuals that are then associated with specific regression equations.[5] With reference to equation (1), we apply independent components analysis to obtain estimates of $e1_{it}$ and $e2_{it}$ that are statistically independent of each other, and then these estimates of the SVAR residuals $e1_{it}$ and $e2_{it}$ are then plugged in to the two equations in (1) such that either b02 or b03 are equal to zero.

The validity of the VAR-LiNGAM estimator depends on several assumptions. First, the SVAR residuals e_{it} should be non-Gaussian. This assumption cannot be tested directly, although we will investigate this assumption by inspecting the distributions of the related (reduced form) VAR residuals. In our context of firm growth rates, there is a large literature on firm growth that suggests that the annual growth rates of firms are highly non-Gaussian (Coad, 2009). A second assumption is that the causal structure is acyclic – that there is one main direction of causality between variables, and that instantaneous feedback loops are not predominant (e.g. within the same time period t, A(t) does not cause B(t) while B(t) simultaneously causes A(t)). However, feedback loops with lags are permitted (e.g. if A affects subsequent values of B, while B affects subsequent values of A). In practical terms, the assumption of acyclicity is satisfied by rearranging the matrix B_0 such that the major causal directions within any time period are given more importance, while relatively minor causal channels are pruned to zero. Another assumption of the VAR-LiNGAM estimator is the usual assumption of linear regression models (such as OLS) that there is a linear relation between the explanatory variables and the dependent variable.

Further details on the VAR-LiNGAM algorithm can be found in Moneta et al., (2013, p715).

3. Database

Our data come from the EU industrial R&D investment Scoreboard (Hernandez et al., 2015), which was compiled by Bureau van Dijk. Taken together, the firms in this database represent about 90% of the total expenditure by business firms on R&D worldwide (Hernandez et al., 2015). The purpose of this database is to facilitate the monitoring of the world's largest R&D investing companies, and to provide evidence to inform European innovation policy. Previous related work on this dataset includes Amoroso (2015), Montresor and Vezzani (2015), Garcia-Manjona and Romero-Merino (2012) and Cincera and Ravet (2010).

[5] A modified version of the microphone analogy (e.g. Stone, 2004) can be helpful. Consider the case of two microphones, one which records voice A, and the other which records A and B. ICA would lead to identify two independent components from the signals recorded by the microphones: the signal of voice A; as well as the independent component corresponding to voice B which is a function of the recorded message on the second microphone, adjusted to remove the signals coming from voice A (such that it is independent of voice A and also the signal recorded by the first microphone). In the case of the first microphone, the recorded signal corresponds to one of the two extracted independent components. Note that our assumption of acyclicity rules out that both microphones record both voices.

The Scoreboard dataset focuses on the world's largest R&D investors. Companies are ranked according to their investment in R&D in the past year and the list is cut at 2500. Moreover, only companies with publicly available annual reports and accounts are considered. As a result of these criteria of inclusion, small and young firms are under-represented. There may exist selection bias on small firms, in the sense that high-R&D small firms will be included, whereas low-R&D small firms will be excluded. However, we make no interpretation of our results in terms of whether small firms are more innovative than large firms. We don't claim to provide results for small innovative firms in particular, but instead we interpret our results in terms of understanding the phenomenon of firm growth and R&D expenditure in our sample of highly innovative firms (i.e. leading R&D investor firms).

Our main variables of interest are the annual growth rates of sales, employment, R&D expenditures, operating profits, and capital expenditures. Appendix 4 provides details on these variables. Growth rates are calculated in the usual way by taking log-differences (Tornqvist et al., 1985; Coad, 2009).[6] These 5 variables each take their turn as dependent variable, and they each appear as explanatory variables for each of the others. Hence, we explore the (causal) relations between each variable on each other variable.

In addition to our main variables of interest, control variables are included, in a first stage, to remove the possibly confounding role of these other possible influences on firm growth. More specifically, we control for the influences of industry-specific growth regimes, year-specific macroeconomy-wide effects, and country-specific effects, by including dummies for sector, year and country. Since our main variables are expressed in growth rates (i.e. differences) rather than size levels, time-invariant firm-specific 'fixed effects' are not included.

We pre-process our SVAR series Y_{it} to remove the influence of control variables X_{it} (consisting of sector dummies, year dummies and country dummies), by estimating median regressions on equation (5), and taking the residuals. We therefore perform our (reduced-form) VAR and SVAR estimations on the 'cleaned' data in vector y_{it}.

$$Y_{it} = a + c.\, X_{it} + y_{it} \qquad (5)$$

y_{it} is then used for our (reduced-form) VAR and SVAR estimations. Our (reduced-form) VAR regression equation is:

$$y_{it} = \alpha + \sum_{\tau=t-s}^{t-1} \beta i\tau \; yi\tau + \xi it \qquad (6)$$

Note that our VAR equation does not investigate instantaneous causal effects, only lagged associations.

Our SVAR regression equation (where the matrix B_0 of instantaneous acyclic causal effects is identified using the VAR-LiNGAM algorithm) is written as:

[6] Note that, by taking logarithms, we lose a small fraction of observations that correspond to non-positive values of operating profits.

$$y_{it} = B_0\, y_{it} + \sum_{\tau=t-s}^{t-1} B i\tau\; y i\tau \quad + \varepsilon_{it} \tag{7}$$

Our SVAR series (the 'dependent variables') are growth rates, rather than size levels, and hence any time-invariant firm-specific components are assumed to have been removed by taking log-differences. Hence, and in keeping with previous applications (e.g. Moneta et al., 2013) we do not control for any time-invariant firm-specific components that might affect growth rates.

4. Analysis

The dataset contains 13,755 observations from 2107 firms, for the years 2003-2013. Summary statistics are not presented, because the SVAR series are normalized to having mean 0.000 and standard deviation 1.000.[7] The correlation matrix in Table 1 shows that the five variables are highly correlated.

Table 1: Correlation matrix. 13,755 observations. All correlations are statistically significant at the 1% level.

	R&D gr	Sales gr	Cap. Ex. gr	Op. Prof. gr.	Empl. gr.
R&D gr	1.0000				
Sales gr	0.2958	1.0000			
Cap. Ex. gr	0.1654	0.303	1.0000		
Op. Prof. gr.	0.0744	0.3528	0.0848	1.0000	
Empl. gr.	0.1427	0.2207	0.1417	0.0363	1.0000

4.1 Reduced-form VAR estimates

Table 2 contains the results for the reduced-form VAR. Growth of R&D is the most strongly related to lagged growth of sales and employment, and the magnitudes of these two associations are roughly similar (compare 0.0687 with 0.0688). Sales growth and employment growth are positively associated with growth of sales and employment in the previous period. Lagged values of sales growth and profits growth are positively associated with subsequent growth of capital expenditures. Growth of profits displays strong negative autocorrelation (the coefficient is -0.212 for the first lag), although it is positively associated with lagged growth of sales. Many associations remain significant at the second lag also.

[7] Normalizing the variables in this way is in keeping with previous applications (Coad and Binder, 2014). One advantage is that it makes the effect sizes comparable across variables.

Although reduced-form VARs can provide interesting intertemporal associations between variables, to describe how the system evolves over time, nevertheless they are unable to provide answers to questions regarding which variables should be manipulated in order to have the desired effects on other variables in subsequent time periods (or even within the same time period). In order to investigate the causal relations between these variables, we now turn to SVAR estimations.

4.2 SVAR estimates

Our assumption of non-Gaussian shocks cannot be investigated directly, because the true SVAR shocks remain unknown (although we do attempt to estimate them). Nevertheless, the plausibility of our assumption of non-Gaussian shocks can be investigated by examining the distribution of our dependent variables. Figure 1 presents quantile-quantile plots of our SVAR variables (growth of R&D, sales, capital expenditures, operating profits, and employment), and suggests that our variables are strongly non-Gaussian (e.g. because the datapoints are not neatly lined up along the diagonals of the qq-plots), which accords with previous work on firm growth variables (Moneta et al., 2013).

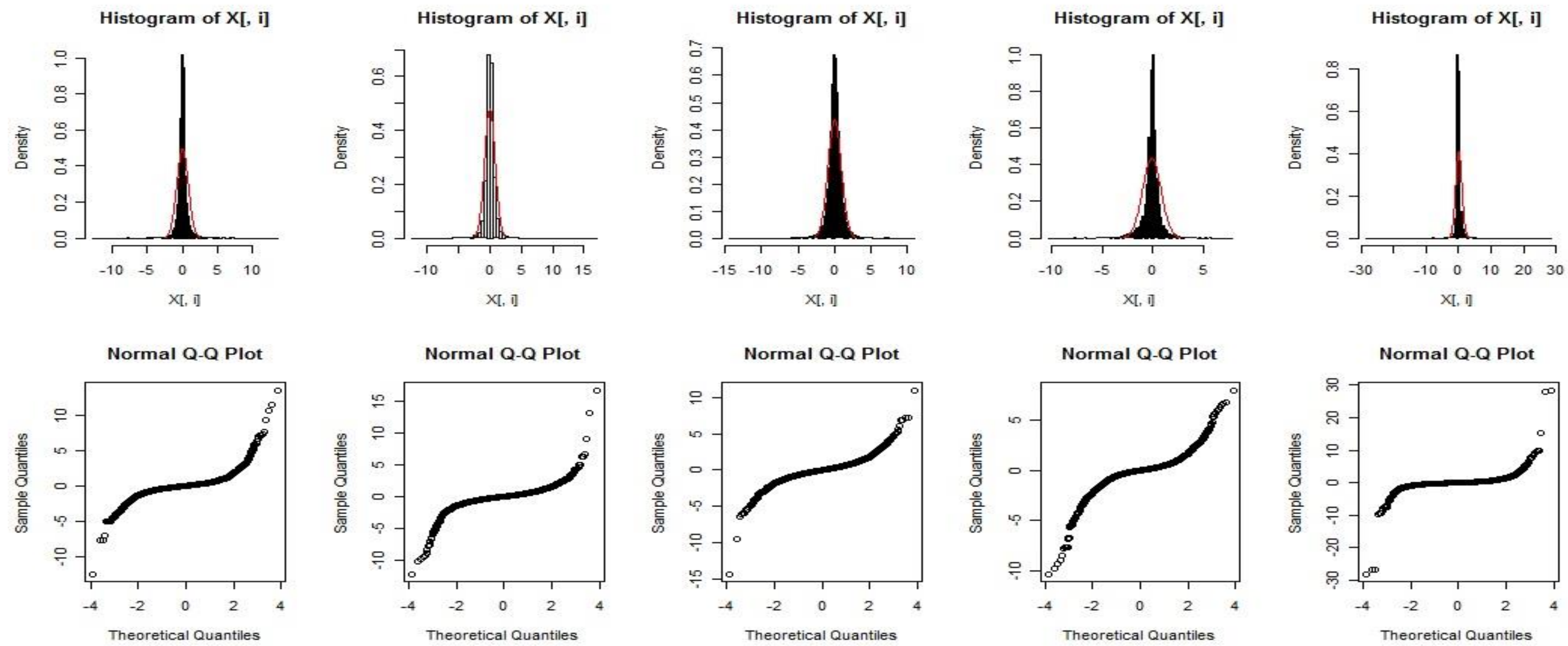
Table 3 contains our SVAR results, both instantaneous effects and lagged effects. We begin by commenting the instantaneous effects. Table 3 shows that sales growth comes first in the causal ordering, having large positive causal effects on the other variables. Sales growth thus appears to kick-start the growth process, with other growth in other dimensions co-evolving as a consequence of the initial sales growth stimulus. Growth of capital expenditures follows suit, having a positive influence on growth of R&D and employment (but no significant influence on growth of profits). R&D growth comes third, having a positive effect on employment growth as well as a negative causal effect on growth of profits (presumably because expenditures on R&D appear in the firms accounts as a cost, thereby diminishing profits).

Table 2: Reduced-form Vector Autoregression, with two lags. Regressions are presented with dependent variables at the top of each column, and explanatory variables in the rows below.

	(1)	(2)	(3)	(4)	(5)
	R&D gr	Sales gr	Cap. Ex. gr	Op. Prof. gr.	Empl. gr.
R&D gr (t-1)	0.0244**	0.0192*	0.00526	-0.00408	0.0117**
	(0.0115)	(0.0114)	(0.00942)	(0.00504)	(0.00527)
Sales gr (t-1)	0.0687***	0.102***	0.126***	0.0954***	0.0504***
	(0.0154)	(0.0210)	(0.0167)	(0.0128)	(0.0115)
Cap. Ex. gr (t-1)	0.00891	0.000852	-0.227***	-0.0121**	0.0175***
	(0.00642)	(0.00990)	(0.0152)	(0.00617)	(0.00614)
Op. Prof. gr (t-1)	0.0294***	-0.00292	0.0855***	-0.212***	0.00562
	(0.00847)	(0.0106)	(0.0127)	(0.0174)	(0.00394)
Empl. gr (t-1)	0.0688***	0.102***	0.0487***	0.0130	0.0503***
	(0.0152)	(0.0200)	(0.0189)	(0.00849)	(0.0131)
R&D gr (t-2)	0.00336	0.0206***	0.00575	-0.00369	0.00133
	(0.00969)	(0.00739)	(0.00853)	(0.00825)	(0.00487)
Sales gr (t-2)	0.0378***	-0.00504	0.0450***	0.000661	0.0349***
	(0.0128)	(0.00854)	(0.0138)	(0.00670)	(0.00849)
Cap. Ex. gr (t-2)	0.00336	0.00493	-0.144***	-0.0102	0.0145***
	(0.00706)	(0.00745)	(0.0152)	(0.00703)	(0.00534)
Op. Prof. gr (t-2)	-0.00464	-0.0185**	0.0426***	-0.137***	-0.00213
	(0.00557)	(0.00763)	(0.0106)	(0.0125)	(0.00412)
Empl. gr (t-2)	0.00164	0.0229***	-0.0113	-0.00147	0.0110**
	(0.0118)	(0.00851)	(0.0117)	(0.00428)	(0.00550)
Constant	-0.0770***	-0.0623***	-0.0313***	0.0160***	-0.0759***
	(0.00502)	(0.00602)	(0.00801)	(0.00557)	(0.00351)
Observations	8,813	8,813	8,813	8,813	8,813
Pseudo-R2	0.0233	0.0245	0.0414	0.0276	0.0190

Note: *The vector of 5 variables is spread across the five columns, and regressed on two lags of itself. Estimations performed using median regression (i.e. 50% quantile regression) with 100 bootstrap replications. Standard errors in parentheses. Key to significance levels: *** p<0.01, ** p<0.05, * p<0.1*

Figure 1: quantile-quantile plot.



***Note:** The five variables, in order, are growth of R&D, growth of sales, growth of capital expenditures, growth of operating profits, and growth of employment*

Table 3: SVAR results. The five dependent variables are listed in the first column, and their determinants should be read across the rows.

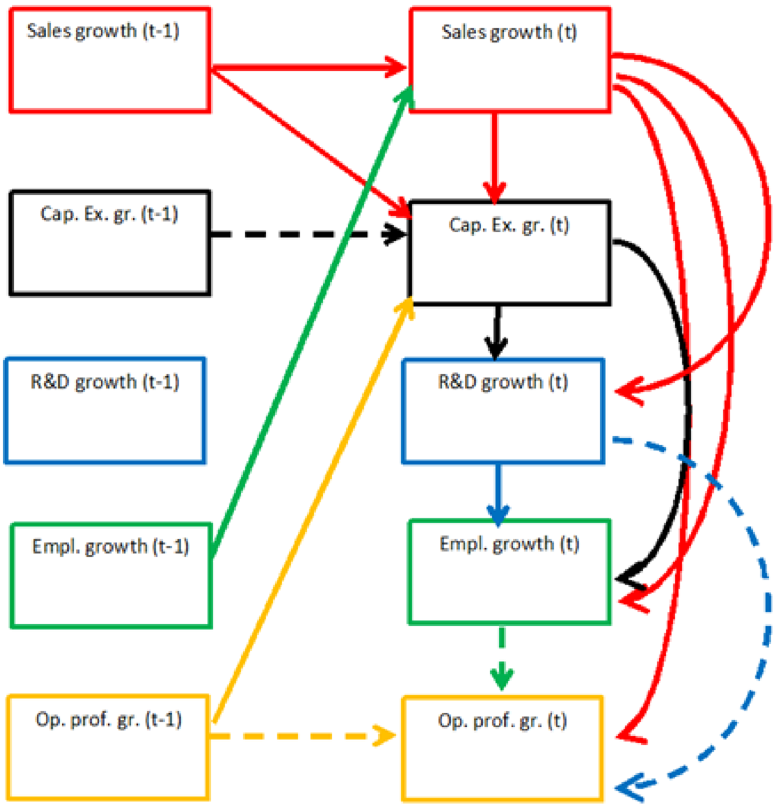
	B0					B1					B2				
	R&D gr	Sales gr	Cap. Ex.	Op.Prof.gr.	Empl.gr.	R&D gr	Sales gr	Cap. Ex.	Op.Prof.gr.	Empl.gr.	R&D gr	Sales gr	Cap. Ex.	Op.Prof.gr.	Empl.gr.
R&D gr	.	**0.2539**	**0.0682**	.	.	0.0175	**0.0375**	**0.0229**	**0.0226**	**0.0364**	-0.0020	**0.0317**	0.0113	-0.0002	-0.0092
	.	0.0185	0.0106	.	.	0.0117	0.0130	0.0055	0.0084	0.0126	0.0078	0.0089	0.0059	0.0049	0.0082
Sales gr	.	.	.	.	.	**0.0200**	**0.0933**	-0.0019	-0.0057	**0.1029**	0.0143	-0.0036	0.0078	**-0.0183**	**0.0223**
	.	.	.	.	.	0.0095	0.0182	0.0098	0.0086	0.0184	0.0081	0.0081	0.0080	0.0071	0.0085
Cap.Ex.gr	.	**0.3300**	.	.	.	-0.0068	**0.0872**	**-0.2012**	**0.0753**	0.0118	0.0006	**0.0451**	**-0.1326**	**0.0331**	-0.0140
	.	0.0129	.	.	.	0.0086	0.0145	0.0114	0.0092	0.0145	0.0069	0.0110	0.0130	0.0089	0.0095
Op.Prof.gr.	**-0.0694**	**0.5436**	-0.0088	.	**-0.0571**	**-0.0123**	**0.0318**	-0.0129	**-0.1786**	**-0.0348**	-0.0121	0.0076	**-0.0124**	**-0.1038**	**-0.0132**
	0.0150	0.0276	0.0088	.	0.0194	0.0054	0.0112	0.0066	0.0178	0.0093	0.0066	0.0054	0.0058	0.0116	0.0039
Empl.gr.	**0.1285**	**0.2197**	**0.0733**	.	.	0.0058	0.0038	**0.0321**	-0.0005	0.0116	-0.0032	**0.0228**	**0.0218**	0.0021	0.0056
	0.0185	0.0262	0.0112	.	.	0.0053	0.0089	0.0053	0.0042	0.0097	0.0047	0.0084	0.0054	0.0037	0.0057

Notes: coefficients and standard errors are reported. Coefficients significant at the 1% level appear in bold ink.

Employment growth comes next, also having a negative effect on growth of profits (because the direct effect of the wage bill on profits is negative; although there are many indirect effects via e.g. subsequent increases in sales). Growth of operating profits comes at the end of the causal ordering.

The lagged effects are similar to those observed for the instantaneous case. Ignoring for now the autocorrelation coefficients (shown along the diagonals), we see that sales growth has significant positive effects on subsequent growth of R&D, capital expenditures, and profits. Growth of capital expenditures, in turn, has positive significant effects on R&D growth and employment growth, but a negative effect on growth of profits. R&D growth has a significant positive effect on subsequent sales growth (even after one year, which is perhaps surprisingly fast) and a negative effect on profits. Employment growth boosts subsequent growth of R&D and sales, but has a negative direct effect on growth of operating profits at both the first and second lag. Finally, growth of operating profits has a small but statistically significant positive effect on growth of R&D (at the first lag only) and also a significant effect on capital expenditures. Figure 2 provides a graphical representation of the SVAR results in Table 3.

Figure 2: graphical representation of the SVAR results for the full sample

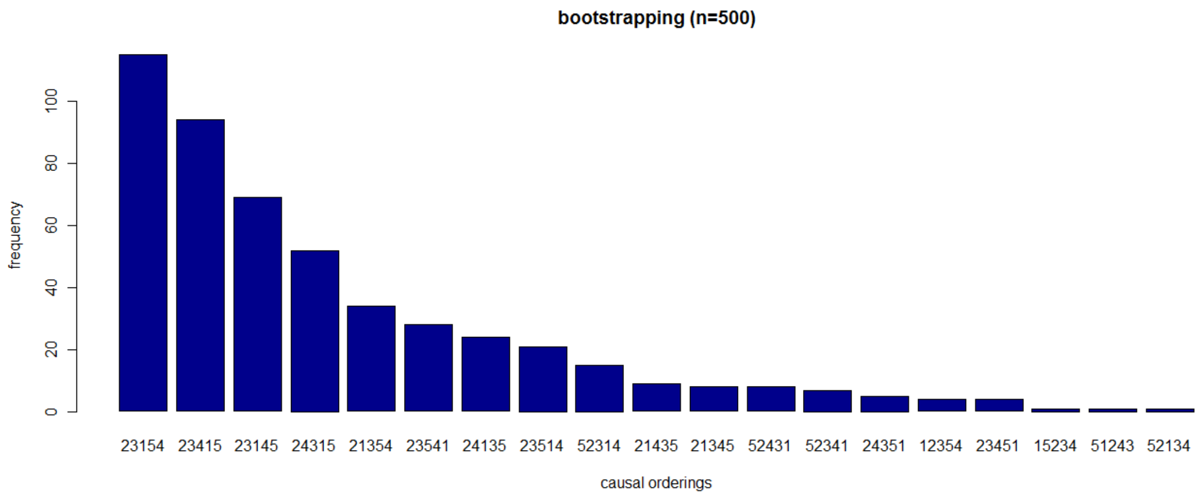


Note: Solid arrows indicate positive effects, dashed arrows negative ones. Coefficients not significant at the 95% level are not shown. Coefficients between -0.05 and 0.05 are not shown.
For simplicity, this figure does not show the second lag. Rows have been permuted to reflect the empirically observed causal ordering: sales growth comes first, followed by capital expenditures growth, R&D growth, then growth of employment and operating profits.

Figure 3 analyses the robustness of the SVAR results, by using the bootstrap analysis in Duschl and Brenner (2013). Although many different causal orderings are observed, nevertheless the most commonly-observed causal orderings account for most of the cases.

It is sales growth ("2") that generally occurs first in the causal ordering, thus taking the role of the *primus motor* in the growth process. Growth of employment ("5") generally occurs at the end of the growth process, being instead influenced by growth of the other variables. Growth of capital expenditures ("3") generally occurs in the first half of the causal ordering. Growth of R&D expenditure generally occurs in the second half of the causal ordering.

Figure 3: robustness analysis: causal pathways in bootstrapped samples.



Note: 500 bootstrap replications. Key to the variable numbers: 1 = R&D growth, 2 = sales growth, 3 = capital expenditures growth, 4 = operating profits growth, 5 = employment growth.

4.3 Further robustness analysis

In contradistinction to standard regressions, where robustness is generally evaluated in terms of whether statistical significance is observed in different subsamples and specifications, we examine the robustness of our results in two directions. First, we examine the robustness of the estimated causal structure (i.e. does B cause A, or does A cause B?) Second, we examine the robustness of our results in terms of whether the results remain statistically significant across subsamples and specifications (does the effect of B on A remain statistically significant?).

Appendices 1 and 2 present our results from alternative specifications, where only four out of the five SVAR series are included in each set of estimations. The reason for this is that, if the SVAR series are all highly correlated between them, then this might lead to difficulties in identifying the distinct roles played by the variables. Tables A1.1 and A2.1 confirm that it is sales growth that drives the process of firm growth in both of these specifications. Figures A1.1 and A2.1 show that this role of sales growth as the primary causal factor is robust. Appendices 1 and 2 also confirm that growth of R&D investment generally appears at the end of the causal ordering of growth.

Appendices 1 and 2 also identify a role for employment growth driving R&D growth, both within-the-period and with a lag. This is at odds with the instantaneous causal relationship observed between R&D growth and employment growth highlighted in our main SVAR results in Table 3 – therefore we remain cautious about whether R&D growth drives employment growth, or vice versa, when considering within-the-year effects.

Appendix 3 contains results from a 1-lag SVAR, to complement our main SVAR results in Table 3 from a 2-lag SVAR. Here we see, again, that it is sales growth that drives the growth process. Again, capital expenditures growth comes second in the causal ordering. The 1 lag model, however, suggests that the instantaneous causal orderings between growth of R&D, growth of profits, and growth of employment are not stable across specifications. We remain cautious about the causal orderings between these latter three variables. Although Table A3.1 shows that growth of operating profits has a causal effect on growth of R&D, nevertheless this effect is not significant. Figure A3.2 shows that a number of alternative causal orderings are observed in the bootstrapping exercise of Duschl and Brenner (2013), although in the vast majority of cases sales growth comes first in the causal ordering, and growth of capital expenditures also appears early on in the causal ordering, with growth of R&D consistently appearing towards the end of the causal ordering.

Taken together, our further robustness analysis has highlighted heterogeneity across firms in terms of their growth patterns. Not all firms grow in the same way. Firms might have different causal orderings between variables as they grow. Nevertheless, our robustness analysis shows that some regularities are found in the vast majority of cases, such as sales growth coming first in the causal ordering.

Another possible avenue for robustness analysis would be to consider that there may be heterogeneity across sectors (e.g. do pharmaceutical firms grow in the same way as automobiles?). We have not investigated this in depth, because to focus on individual sectors would mean having a reduced number of observations in our dataset. We leave for future research these investigations of how heterogeneities in subsamples and possible exceptional cases may belie the broader relationships observed at the aggregate level.

5. Discussion and Conclusion

Our results suggest that it is sales growth that is the key stimulus for R&D investment. This is consistent with the demand-pull theory of industrial evolution (Schmookler, 1966) as well as the behavioural model whereby boundedly-rational firms invest in R&D as a fixed proportion of sales (Thompson, 1999). There is little evidence that firms first need to make profits before they invest in R&D. However, we cannot rule out that *anticipated* profits, in contrast to realized profits, may have a bigger effect on R&D investment – because we have no data on *anticipated* profits.

One possible channel for policy interventions to boost sales growth (and hence R&D growth and employment growth) would be to encourage firms to boost their sales through increased exporting activity. In this context, the European Commission's (2015) Single Market Strategy could play a role in boosting R&D growth via increases in sales. Another possible channel is through the use of procurement policy to generate sales for innovative firms with growth ambitions (e.g. Rolfstam, 2013).

R&D growth appears to be determined at the end of the causal ordering of the growth process, as a consequence of the other variables. We do not observe any effects of R&D investment on subsequent firm performance in terms of sales growth or profits growth. This is not surprising, though, for several reasons. First, there may be long time lags between when a firm invests in R&D and when it capitalizes on its subsequent innovation success – the payback period may be 10 years or more (Grabowski et al., 2002). Second, investment in innovation entails a lot of uncertainty, such that some spectacular successes might be found alongside a large number of relatively unsuccessful outcomes (Grabowski et al., 2002; Coad and Rao, 2008).

Our SVAR analysis has highlighted the key role of sales growth, rather than profits growth, in stimulating R&D growth (and firm growth more generally). One could speculate that investment in R&D is not driven by rational calculation, but the 'animal spirits' of innovation, perhaps tinted with over-optimism, whereby industrialists put aside their elaborate forecasts and use their gut feelings, and their need for achievement, to channel large amounts of available funds into new R&D projects.

References

Amoroso, S. (2015). Profits, R&D and Labour. JRC-IPTS Working Papers on Corporate R&D and Innovation No 10/2015, Institute of Prospective Technological Studies, Joint Research Centre.

Bogliacino, F., & Vivarelli, M. (2012). The job creation effect of R&D expenditure. *Australian Economic Papers*, 51(2), 96-113.

Breschi, S., Malerba, F., Orsenigo, L., (2000). Technological Regimes and Schumpeterian Patterns of Innovation, Economic Journal, 110, 388-410.

Cincera, M., Ravet, J., 2010. Financing constraints and R&D investments of largecorporations in Europe and the US. *Science and Public Policy* 37 (6), 455–466.

Coad A., 2009. "The Growth of Firms: a Survey of Theories and Empirical Evidence." Edward Elgar, Cheltenham, UK and Northampton, MA, USA.

Coad A., Binder M., (2014). Causal linkages between work and life satisfaction and their determinants in a structural VAR approach. *Economics Letters* 124, 263–268.

Coad A., Rao R., (2008). Innovation and firm growth in high-tech sectors: A quantile regression approach*. Research Policy*, 37 (4), 663-648.

Coad, A. and Rao, R. (2010). Firm growth and R&D expenditure. *Economics of Innovation and New Technology*, 19(2): 127-145.

Dosi G., (1988). Sources, Procedures, and Microeconomic Effects of Innovation. Journal of Economic Literature 26 (3), 1120-1171.

Dosi, G., Nelson, R.R., (2013). The Evolution of Technologies: An Assessment of the State-of-the-Art. Eurasian Business Review 3, 3–46.

Duschl M., Brenner T. (2013). Growth dynamics in regional systems of technological activities. Working Papers on Innovation and Space, University of Marburg.

European Commission (2015). A Single Market Strategy for Europe - Analysis and Evidence. http://ec.europa.eu/growth/single-market/index_en.htm

García-Manjóna, J.V., Romero-Merino, M.E., 2012. Research, development, and firm growth. Empirical evidence from European top R&D spending firms. *Research Policy* 41 (6), 1084–1090.

García-Quevedo, J., Pellegrino, G., & Vivarelli, M. (2014). R&D drivers and age: Are young firms different?. *Research Policy*, 43(9), 1544-1556.

Grabowski H., Vernon J., DiMasi J.A., (2002). Returns on Research and Development for 1990s New Drug Introductions. Pharmacoeconomics 20 Suppl 3, 11-29.

Hernandez, H., Hervas Soriano, F., Tubke, A., Vezzani, A., Dosso M., Amoroso A., Grassano N., Coad A., and Gkotsis P. (2015). The 2015 EU industrial R&D investment Scoreboard. Publications Office of the European Union, Luxembourg.

Himmelberg C.P., Petersen B.C., 1994. R&D and Internal Finance: A Panel Study of Small Firms in High-Tech Industries. *Review of Economics and Statistics* 76, 38-51.

Hyvärinen, A., Karhunen, J., and Oja, E. (2001). Independent component analysis, volume 26. Wiley-interscience.

Hyvarinen A., Shimizu S., Hoyer P.0., (2008). Causal modelling combining instantaneous and lagged effects: an identifiable model based on non-Gaussianity. In: Proceedings of the 25th International Conference on Machine Learning (ICML2008), Helsinki Finland, pp. 424–431.

Kamien M.I., Schwartz N.L. (1975). Market Structure and Innovation: A Survey. *Journal of Economic Literature* 13 (1), 1-37

Kleinknecht, A., Verspagen, B., 1990. Demand and Innovation: Schmookler Re-examined. Research Policy 19, 387–394.

Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports*, 66(1), 195-244.

Mooij J., D. Janzing, J. Peters, and B. Scholkopf (2009). "Regression by Dependence Minimization and Its Application to Causal Inference," Proc. 26th Int'l Conf. Machine Learning, pp. 745-752.

Moneta A, Entner D, Hoyer P, Coad A (2013). Causal Inference by Independent Component Analysis: Theory and Applications*. Oxford Bulletin of Economics and Statistics* 75 (5), 705-730.

Montresor, S., & Vezzani, A. (2015). The production function of top R&D investors: Accounting for size and sector heterogeneity with quantile estimations. *Research Policy*, 44(2), 381-393.

Pearl, J. (2009) Causality: Models, Reasoning and Inference, 2nd edition. Cambridge: Cambridge University Press.

Penrose, E.T., (1959). The Theory of the Growth of the Firm. Basil Blackwell: Oxford, UK.

Piva, M., Vivarelli, M., (2007). Is Demand-pulled Innovation Equally Important in Different Groups of Firms? Cambridge Journal of Economics 31, 691–710.

Rolfstam, M. (2013). Public Procurement and Innovation. Edward Elgar Publishing.

Scherer, F.M., (1982). Demand-Pull and Technological Invention: Schmookler Revisited. Journal of Industrial Economics 30, 225–37

Scherer, F.M., (2001). The Link Between Gross Profitability and Pharmaceutical R&D Spending. *Health Affairs* 20, 216-220.

Schmookler, J., (1966). Invention and Economic Growth. Harvard University Press MA.

Shimizu, S., Hoyer, P., Hyvärinen, A., and Kerminen, A. (2006). A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7:2003–2030.

Stock J.H., Watson M.W. (2001). Vector Autoregressions. *Journal of Economic Perspectives* 15 (4), 101-115.

Stone, J. (2004). Independent component analysis: a tutorial introduction. MIT Press, Cambridge, MA.

Thompson P., (1999). Rationality, Rules of Thumb, and R&D. *Structural Change and Economic Dynamics* 10, 321-340

Tornqvist L., Vartia P., Vartia Y.O. (1985). How Should Relative Changes Be Measured? *American Statistician* 39(1), 43-46.

Vivarelli, M. (2013). Technology, Employment and Skills: An Interpretative Framework. Eurasian Business Review, 3(1), 66-89.

Wooldridge, J. M. (2002). Econometric analysis of cross section and panel data. MIT press.

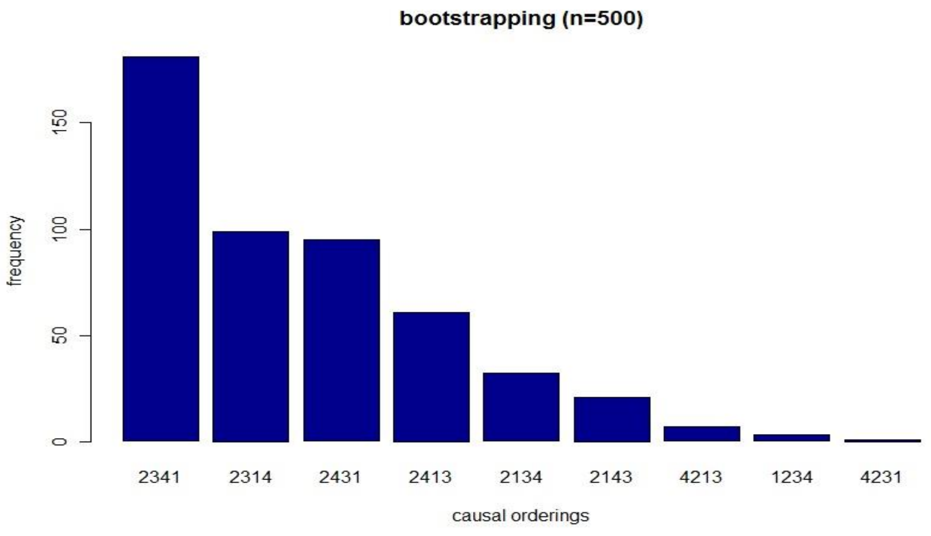
Appendix 1: SVAR on four variables only: growth of R&D, sales, operating profits and employment

Table A1.1: SVAR results: Growth of R&D, Sales, Operating Profits, and Employment.

	Instantaneous effects				Lagged effects			
	R&D gr	Sales gr	Op.Prof.gr.	Empl.gr.	R&D gr	Sales gr	Op.Prof.gr.	Empl.gr.
R&D gr	0	**0.2511**	-0.0122	**0.0943**	0.0155	**0.0665**	**0.0172**	**0.0457**
	0	0.0281	0.0173	0.016	0.0092	0.0139	0.0066	0.0147
Sales gr	0	0	0	0	**0.0178**	**0.1218**	0.0012	**0.0951**
	0	0	0	0	0.0088	0.0193	0.008	0.0193
Op.Prof.gr.	0	**0.4376**	0	0	**-0.0146**	0	**-0.1215**	**-0.0245**
	0	0.0258	0	0	0.0047	0.0118	0.0148	0.009
Empl.gr.	0	**0.2902**	**-0.0527**	0	0.0059	**0.0467**	-0.007	**0.0227**
	0	0.0255	0.0143	0	0.0055	0.0082	0.004	0.0078

Notes: coefficients and standard errors. Coefficients significant at the 1% level appear in bold ink. The 4 SVAR dependent variables are listed in the first column, and their determinants should be read across the rows.

Figure A1.1: Robustness of the causal ordering for the four SVAR



Note: 500 bootstrap replications. Key to the variable numbers: 1 = R&D growth, 2 = sales growth, 3 = operating profits growth, 4 = employment growth. Robustness of the dominant causal pathway: bootstrapping exercise following Duschl and Brenner (2013).

Appendix 2: SVAR on four variables only: growth of R&D, sales, capital expenditures and employment

Table A2.1: SVAR results: Growth of R&D, Sales, capital expenditures, and Employment

	Instantaneous effects				Lagged effects			
	R&D gr	Sales gr	Cap. Ex. gr.	Empl.gr.	R&D gr	Sales gr	Cap. Ex. gr.	Empl.gr.
R&D gr	0	**0.2225**	**0.0595**	**0.0932**	0.0119	**0.0656**	**0.0198**	**0.0486**
	0	0.0243	0.0113	0.0174	0.0082	0.0132	0.0057	0.0145
Sales gr	0	0	0	0	**0.0174**	**0.1203**	0.0051	**0.0938**
	0	0	0	0	0.0087	0.0194	0.0073	0.0198
Cap. Ex. gr.	0	**0.333**	0	0	0.0014	**0.1064**	**-0.1744**	0.0014
	0	0.0137	0	0	0.0116	0.0104	0.0128	0.0099
Empl.gr.	0	**0.2383**	**0.0773**	0	0.0067	**0.0332**	**0.0315**	**0.0245**
	0	0.0241	0.0102	0	0.0049	0.0088	0.005	0.0081

Notes: coefficients and standard errors. Coefficients significant at the 1% level appear in bold ink. The 4 SVAR dependent variables are listed in the first column, and their determinants should be read across the rows.

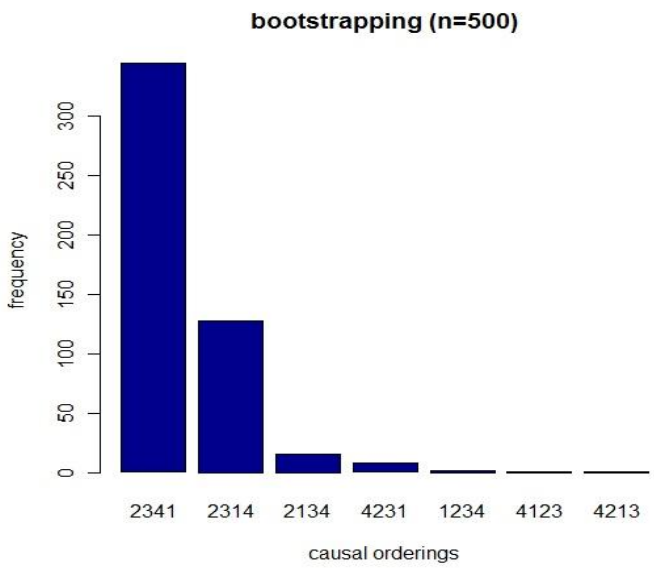
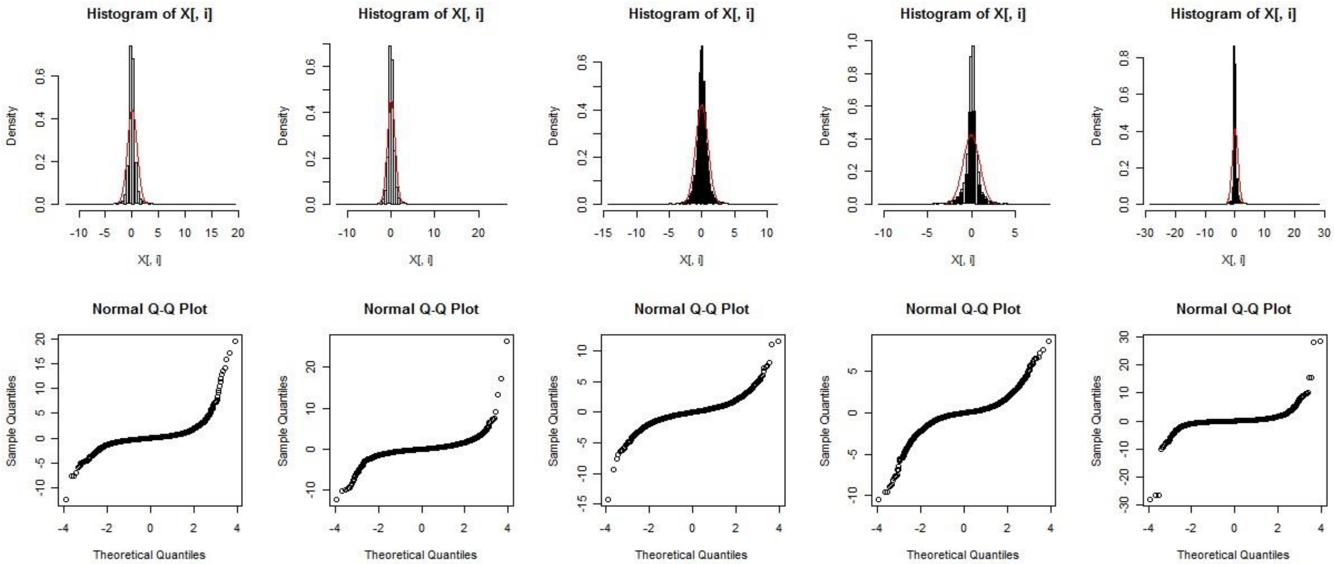


Figure A2.1: Robustness of the causal ordering for the four SVAR series

Note: 500 bootstrap replications. Key to the variable numbers: 1 = R&D growth, 2 = sales growth, 3 = capital expenditure growth, 4 = employment growth. Robustness of the dominant causal pathway: bootstrapping exercise following Duschl and Brenner (2013).

Appendix 3: Alternative specification: 1 lag instead of 2.

Figure A3.1: QQ-plot for variables for the 1-lag SVAR.



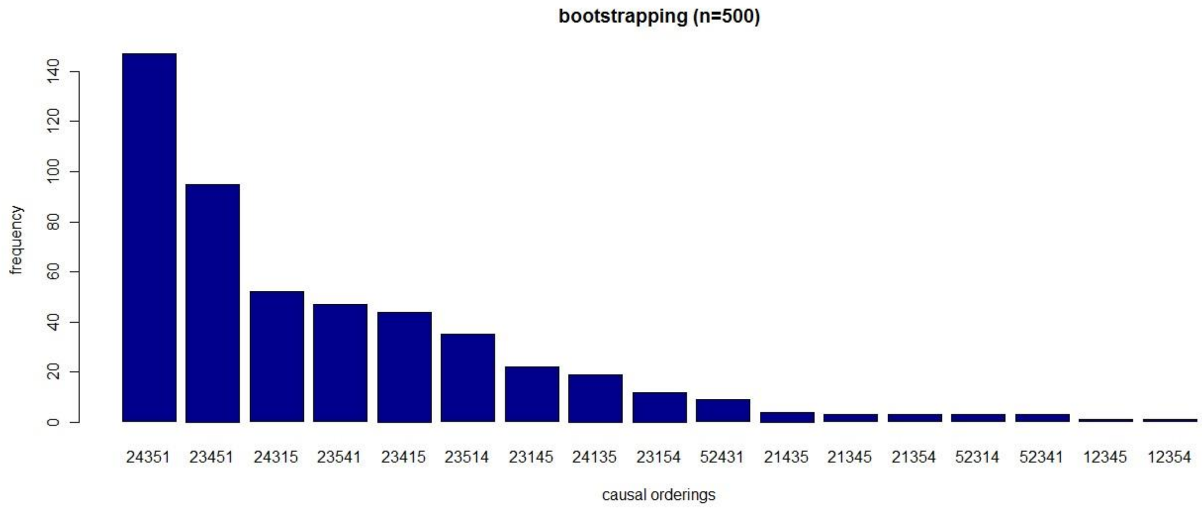
Note: The five variables, in order, are growth of R&D, growth of sales, growth of capital expenditures, growth of operating profits, and growth of employment.

Table A3.1: 5-variable SVAR estimates from a 1-lag model (instead of including 2 lags)

	B0					B1				
	R&D gr	Sales gr	Cap. Ex.	Op.Prof.gr.	Empl.gr.	R&D gr	Sales gr	Cap. Ex.	Op.Prof.gr.	Empl.gr.
R&D gr	0	**0.2313**	**0.0554**	-0.0127	**0.0924**	0.0107	**0.0506**	**0.0187**	0.0139	**0.0389**
	0	0.0260	0.0112	0.0156	0.0177	0.0090	0.0146	0.0056	0.0075	0.0128
Sales gr	0	0	0	0	0	0.0151	**0.1098**	0.0033	0.0022	**0.0911**
	0	0	0	0	0	0.0091	0.0174	0.0078	0.0074	0.0173
Cap. Ex.	0	**0.3295**	0	0	0	0.0055	**0.0801**	**-0.1615**	**0.0660**	-0.0059
	0	0.0138	0	0	0	0.0120	0.0159	0.0109	0.0091	0.0093
Op.Prof.gr.	0	**0.4444**	**-0.0262**	0	0	**-0.0133**	0.0029	**-0.0187**	**-0.1120**	**-0.0229**
	0	0.0253	0.0071	0	0	0.0039	0.0099	0.0049	0.0137	0.0083
Empl.gr.	0	**0.2681**	**0.0742**	**-0.0549**	0	0.0063	**0.0322**	**0.0283**	**-0.0109**	**0.0199**
	0	0.0216	0.0109	0.0111	0	0.0050	0.0095	0.0050	0.0039	0.0073

Notes: *coefficients and standard errors. Coefficients significant at the 1% level appear in bold ink. The 5 SVAR dependent variables are listed in the first column, and their determinants should be read across the rows.*

Figure A3.2: bootstrap analysis for a 1-lag SVAR



***Note:** 500 bootstrap replications. Key to the variable numbers: 1 = R&D growth, 2 = sales growth, 3 = capital expenditures growth, 4 = operating profits growth, 5 = employment growth. Bootstrapping exercise following Duschl and Brenner (2013).*

Appendix 4: Variables definitions (taken from Hernandez et al., 2015)

Research and Development (R&D) investment in the Scoreboard is the cash investment funded by the companies themselves. It excludes R&D undertaken under contract for customers such as governments or other companies. It also excludes the companies' share of any associated company or joint venture R&D investment. Being that disclosed in the annual report and accounts, it is subject to the accounting definitions of R&D. For example, a definition is set out in International Accounting Standard (IAS) 38 "Intangible assets" and is based on the OECD "Frascati" manual. Research is defined as original and planned investigation undertaken with the prospect of gaining new scientific or technical knowledge and understanding. Expenditure on research is recognised as an expense when it is incurred. Development is the application of research findings or other knowledge to a plan or design for the production of new or substantially improved materials, devices, products, processes, systems or services before the start of commercial production or use. Development costs are capitalised when they meet certain criteria and when it can be demonstrated that the asset will generate probable future economic benefits. Where part or all of R&D costs have been capitalised, the additions to the appropriate intangible assets are included to calculate the cash investment and any amortisation eliminated.

Net sales follow the usual accounting definition of sales, excluding sales taxes and shares of sales of joint ventures & associates. For banks, sales are defined as the "Total (operating) income" plus any insurance income. For insurance companies, sales are defined as "Gross premiums written" plus any banking income.

Operating profit is calculated as profit (or loss) before taxation, plus net interest cost (or minus net interest income) minus government grants, less gains (or plus losses) arising from the sale/disposal of businesses or fixed assets.

Capital expenditure (Capex) is expenditure used by a company to acquire or upgrade physical assets such as equipment, property, industrial buildings. In accounts capital expenditure is added to an asset account (i.e. capitalised), thus increasing the asset's base. It is disclosed in accounts as additions to tangible fixed assets.

Number of employees is the total consolidated average employees or year-end employees if average not stated.

JRC Mission

As the Commission's in-house science service, the Joint Research Centre's mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy Directorates-General, the JRC addresses key societal challenges while stimulating innovation through developing new methods, tools and standards, and sharing its know-how with the Member States, the scientific community and international partners.

Serving society
Stimulating innovation
Supporting legislation