UvA-DARE (Digital Academic Repository)

The cyclicalities of R&D investment revisited

van Ophem, H.; van Giersbergen, N.; van Garderen, K.J.; Bun, M.

DOI

[10.1002/jae.2667](https://doi.org/10.1002/jae.2667)

Publication date

2019

Document Version

Submitted manuscript

Published in

Journal of Applied Econometrics

License

Unspecified

[Link to publication](#)

Citation for published version (APA):

van Ophem, H., van Giersbergen, N., van Garderen, K. J., & Bun, M. (2019). The cyclicalities of R&D investment revisited. *Journal of Applied Econometrics*, 34(2), 315-324.
<https://doi.org/10.1002/jae.2667>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

The cyclicalities of R&D investment revisited

Hans van Ophem^{*}
Noud van Giersbergen[†]
Kees Jan van Garderen[‡]
and
Maurice Bun[§]

June 2018[¶]

Abstract

In Fabrizio and Tsoolmon (2014) and Barlevy (2007) it is concluded that R&D investments are procyclical. Fabrizio and Tsoolmon utilize a model based on Barlevy (2007), but differs in some respects and allows for more heterogeneity. However, we doubt whether their implied trends are intended. Fabrizio and Tsoolmon also set missing values for R&D equal to zero leading to unrealistic jumps in investment and its first differences. We reconcile and replicate both the Fabrizio and Tsoolmon and Barlevy papers by considering extensions that encompass both models. Furthermore, we treat missing values more appropriately and consider some alternative specifications to check the robustness of the results. Procyclicality is confirmed, but we find much less heterogeneity than Fabrizio and Tsoolmon (2014) do. In particular obsolescence and patent effectiveness are no longer important in this last respect but external financing is.

Keywords: R&D expenditures, innovation, cyclicity, linear panel data estimation

^{*}Amsterdam School of Economics, University of Amsterdam and Tinbergen Institute. Corresponding author. Address: Amsterdam School of Economics, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands. Email: j.c.m.vanophem@uva.nl.

[†]Amsterdam School of Economics, University of Amsterdam.

[‡]Amsterdam School of Economics, University of Amsterdam.

[§]Dutch Central Bank and Amsterdam School of Economics, University of Amsterdam.

[¶]We are grateful to two referees for their useful remarks.

1 Introduction

The relation between research & development investments (R&D) and the business cycle has been studied in a large number of papers. From a theoretical point of view, Schumpeter (1939) argues that investments in innovation are countercyclical since the opportunity costs in times of a recession are lower than in an upswing. Empirical evidence on the relation between R&D and economic growth, however, suggests procyclicality. Fabrizio and Tzolmon (2014), henceforth denoted by FT, list many references that conclude that there is a positive relation between R&D and industry output growth. From a theoretical point of view, Francois and Lloyd-Ellis (2009), reconcile Schumpeter's theory with the empirical findings by pointing out that innovation is more than only R&D and that other innovative activities might exhibit different business cycle properties. On the other hand, if we look at the empirical literature the evidence is much more heterogeneous than pointing only at procyclicality: e.g. Aghion et al. (2012) allow for heterogeneous impacts of the business cycle and find mixed results, whereas e.g. Kraiczy et al. (2015) find no relation between R&D and the business cycle. At closer inspection, FT also find mixed effects: although for most firms procyclical R&D investments are estimated, for some firms countercyclical R&D investments are found. Because in FT the marginal effects depend on the levels of patent protection, the rate of product obsolescence and the degree of external financing, cyclicalities are heterogeneous.

There are several reasons why such ambiguous results are found. In general it can be stated that none of the studies above use the same model. Differences can be found in the definition of the dependent variable (R&D, $\log(R\&D)$, R&D yes or no etc.), differences in the covariates (none, only lags, lags and leads, firm characteristics etc) and functional form (linear, nonlinear). Furthermore, the character of the data (macro, cross section, panel) and the estimation method (random effects, fixed effects, ignoring the panel character) can differ. In this investigation we will compare the methods used in two related papers to establish whether the conclusions of the papers hold in different settings. We will analyze the research of FT deeper and compare it with Barlevy (2007). FT starts from the model used by Barlevy (2007) and allows the effect of output growth to be different across the levels of patent protection, product obsolescence and external financing the industry that the firm belongs to experiences. Apart from using a different data set, FT also use a different, but related, estimation technique and use controls that are different in one important aspect: they use first differenced variables instead of level variables in their specification. To make the analysis comparable we will estimate the model of Barlevy (2007), extend it according to the suggestions of FT using the same data and the same estimation method. The results are in section 2. In section 3, we will treat missing values on R&D expenditures differently than FT do, and reestimate the models to see whether the conclusions changes. In section 4 we discuss some related models and specifications and section 5 concludes.

2 The models of Fabrizio and Tzolmon (2014) and Barlevy (2007)

The starting point of FT in their analysis of the cyclicalities of R&D is the first-differenced model of R&D used by Barlevy (2007). It reads:

$$\Delta R\&D_{kt} = \beta_0 + \beta_1 \Delta X_{it} + \beta_2 \Delta M_{kt} + \beta_3 \Delta M_{kt-1} + \tau_t + \omega_{kt}, \quad (1)$$

where Δ is the first differencing operator, $R\&D_{kt}$ is the natural log of real R&D investments of firm k in year t , X_{it} is the log industry output relating to the industry i to which firm k belongs in year t and M_{kt} firm-level controls, contemporary and one period lagged. $\alpha\tau_t$ are time fixed effects. This model is extended with some interaction terms that will be discussed later and then estimated by pooled OLS. Note that the addition of the constant β_0 indicates that FT allow for the possibility that there is a trend in R&D expenditures. The model of Barlevy (2007) actually differs in some important aspects. First of all, Barlevy estimates the relation between the growth in R&D and the growth in output using a fixed effects or within estimator. Furthermore, Barlevy uses related but different controls. His model reads:

$$\Delta R\&D_{kt} = (\beta_0 +) \beta_1 \Delta X_{it} + \beta_2 M_{kt} + \beta_3 M_{kt-1} + \alpha_k + \tau_t + \xi_{kt}, \quad (2)$$

where α_k denote firm specific effects. The presence of the constant β_0 is questionable, but this bears no significance because it is not identified in a fixed effects estimation as employed by Barlevy. This means that if firm specific effects exist and if they correlate with one or more of the regressors, estimating eq. (1) with pooled OLS, i.e. what is actually done by FT, will yield biased estimates, whereas estimation of (2) with the fixed effects estimator still gives unbiased results. Finally, note that Barlevy uses level controls and not first differenced controls in his specification.

Cyclicity of R&D expenditure is represented by a nonzero effect of industry output (growth): $\beta_1 \neq 0$. Procyclicality would be represented by $\beta_1 > 0$ and countercyclicality by $\beta_1 < 0$.

As revealed in footnote 4 of FT, the first differenced model is only introduced because the model in levels displays autocorrelation and heteroskedasticity. In fact, it also removes possible endogeneity due to higher-order fixed effects. To show this, rewrite eq. (1) in level variables:

$$R\&D_{kt} = \theta_0 + \beta_0 t + \beta_1 X_{it} + \beta_2 M_{kt} + \beta_3 M_{kt-1} + \eta_k + \theta_t + \varsigma_{kt}. \quad (3)$$

First differencing this model gives the model in eq. (1) where we need to define: $\tau_t = \theta_t - \theta_{t-1}$ and $\omega_{kt} = \varsigma_{kt} - \varsigma_{kt-1}$. Note that first differencing takes care of the potential endogeneity due to the correlation between the firm specific effect η_k and X_{it} and/or M_{kt} , and that η_k for sure creates autocorrelation and quite likely creates heteroskedasticity. If we compare the models in eqs. (1) to (3) they are clearly related and model (2) seems to be a special mix of models (1) and (3). It allows for what is known as incidental trends whereas model (1) does not.

From rewriting the model in level variables it is also clear that the constant added to the model by FT, their β_0 in the first differenced specification, represents a time trend in R&D. FT then proceed by adding their most important variables to the model: measures of industry-level obsolescence (Ob_i), patent effectiveness (PE_i) and external financing (EF_i) both as interactions with growth in industry output and as level variables, although these last variables are not explicitly stated in the equation listed on page 666 (second column) of FT. However, their introduction is clear from the regression output in Table 5 of FT and from the supplementary material provided by the authors. As such adding the level variables is sound econometrics but, as we will see in a moment, it has some unintended consequences for the starting model that uses R&D in levels. The most elaborate model of FT is:

$$\begin{aligned} \Delta R\&D_{kt} = & \beta_0 + \alpha_t + \beta_1 \Delta X_{it} + \beta_2 \Delta M_{kt} + \beta_3 \Delta M_{kt-1} + \\ & + \beta_4 \Delta X_{it} \cdot Ob_i + \beta_5 \Delta X_{it} \cdot PE_i + \beta_6 \Delta X_{it} \cdot EF_i + \delta_4 Ob_i + \delta_5 PE_i + \delta_6 EF_i + \varepsilon_{kt}. \end{aligned} \quad (4)$$

What FT do is allowing for heterogeneous effects of industry output growth on the growth in R&D expenditures where the heterogeneity depends in Ob , PE and EF . FT expect $\beta_4 > 0$, i.e. with higher degree of obsolescence firms will be more sensitive to changes in demand, and $\beta_5 < 0$, i.e. with higher degree of patent protection firms will be less sensitive to changes in demand. No discussion of the expected signs of β_6 , δ_4 , δ_5 and δ_6 is provided, in fact they are not mentioned at all by FT except in the estimation output. As a consequence of the interaction effects, the effect of the cyclical growth of R&D expenditure has become heterogeneous. The marginal effect of industry output growth on R&D growth is now:

$$\frac{\partial E(\Delta R\&D_{kt})}{\partial \Delta X_{it}} = \beta_1 + \beta_4 \cdot Ob_i + \beta_5 \cdot PE_i + \beta_6 \cdot EF_i$$

and depending on the sign of the parameters and the level of the covariates it can indicate procyclicality for some firms and countercyclicality for others.

The meaning of δ_4 , δ_5 and δ_6 becomes clear if we reformulate model (4) in level variables. Just like in the case of the added constant, adding the time constant variables to a first difference specification, gives rise to a trend in the level variable. In the present case, there is not only a pure time trend $\beta_0 \cdot t$ but also three heterogeneous time trends: $\delta_4 \cdot Ob_i \cdot t$, $\delta_5 \cdot PE_i \cdot t$ and $\delta_6 \cdot EF_i \cdot t$. As such this can be a correct specification, but FT do not provide theoretical or empirical arguments why adding these variables to the specification is a good idea. On top of that, the effect of these variables are only identified if no incidental trends are present. If we would actually follow the estimation method of Barlevy (2007), i.e. using a fixed effects estimation, we would allow for firm specific effects in eq. (1) which might potentially be correlated with the explanatory variables of eq. (1). If there is indeed a correlation, then pooled OLS is not the correct estimation method to use, but a fixed effects or first difference estimation will not identify the constant and the parameters of Ob , PE and EF . Consequently the estimation results of FT are only correct under an additional assumption that Barlevy (2007) does not make.

We will now proceed with a replication of the estimation results of FT. In the next section we present estimation results of some alternative models that are less deviant from Barlevy (2007) to investigate whether the results of FT hold under less restrictive assumptions and alternative specifications.

Data¹

FT combine four different US data sources. The Compustat data provides annual information on individual firms over the period 1975-2002. This information is matched to industry level information from the NBER Manufacturing and Productivity Data and the Carnegie Mellon survey. The Carnegie Mellon survey is a cross section from 1994 and provides indicators of the explanatory variables patent effectiveness and obsolescence.

The relation between the growth in R&D expenditures of a firm, measured by $\Delta R\&D_{kt}$, are related to the growth of the output of the industry the firm belongs to and a large number of controls: among others they include growth rates of cash flow, total assets, liabilities, debt and capital stock. Very similar specifications are used in e.g. Barlevy (2007) and Ouyang (2011). The contribution of FT is to allow for additional heterogeneity of the effect of industrial output growth on R&D-expenditures. They

¹The data are taken from the Review of Economics and Statistics replication database.

argue that the cyclicity of R&D investments is influenced by some characteristics of the industry: the degree of patent effectiveness, obsolescence and external financing. Patent effectiveness is measured as the average (across industry) survey response on the question on the percentage of products and process innovations for which patenting was effective in protecting the firm's competitive advantage. The obsolescence indicator is distilled from the results to the question in the Carnegie Mellon survey on the speed with which new product and process innovations are introduced in the focal industry. The external financing indicator is taken from the Compustat data by summing the annual external funds needed to finance the investments of each firm over the sample years 1975-2002 and dividing it by the firm's total capital expenditures. This is then transformed to a industry level variable by using the median across the industry. The growth of the industry output is calculated from the NBER Manufacturing and Productivity Data by calculating the first difference in the natural logarithm of the real gross output of the industry.

Replications

We were able to reproduce the 'B. R&D Estimation'-panel of Table 2 of FT, containing descriptive statistics, exactly. The results of the pooled OLS-estimation of eq. (4), are given in Table 1, columns (1) to (3). They correspond to columns (2), (5) and (6) of Table 5 of FT.

The results presented in these columns are not an exact reproduction of the results presented in Table 5 of FT but they are both in the size of the estimated coefficients and their significance very similar.² All conclusions drawn by FT remain valid. In most cases R&D expenditures are procyclical, depending on patent effectiveness, obsolescence and the external financing within the industry. As noted before the inclusion of these variables introduces heterogeneity and for some firms this will result in a negative marginal effect, indicating countercyclicity. The average marginal effects indicate that for most model specifications and most firms R&D expenditures are procyclical. Patent effectiveness and external financing do not appear to have an effect through growth in output but do have a positive significant themselves. Note that, if anything, patent effectiveness has a negative indirect effect in combination with output growth. This is in line with the procyclicality of R&D investments, as argued by FT. Also confirming the hypothesis of procyclicality of R&D expenditures, obsolescence has both a direct positive impact as an indirect effect through output growth. The average marginal effect of output growth is about 16%, when allowing for interaction effects.

We now turn to estimating the model of Barlevy (2007) on the data of FT. The specification of the model is given in eq. (2) and the model is estimated using a fixed effects estimator. The results can be found in Table 2, columns (1) to (4). In the fixed effects estimation the significance of output growth is lost if the control variables are added to the specification. Barlevy (2007) found strongly significant positive results pointing at procyclicality of the growth in R&D expenditures. We find no relation with the business cycle if we control for firm characteristics. Although the data come from partly the same source (Compustat) there are differences in the data: here we use about twice the number of observations, the time period considered is slightly different and a different number of industries are distinguished. However, the definitions of the variables are the same. It should be noted that the Barlevy estimation results are quite heterogeneous: his estimates of the effect of output growth on growth in real R&D, where the data used vary across industry codes, included time period and the

²In their Table 5, FT have the entry $\text{Output} \times \dots$ (3 times), but this should read $\Delta \text{Output} \times \dots$

way output is measured, ranges from 0.21 to 0.92.

As concluded before, the estimations of FT are based on stronger assumptions than those of Barlevy. Before estimating both specifications using the same estimation technique and investigating some alternative models, we will first correct an important mistake in the preparation of the data.

3 Treating missing observations differently

A simple diagnostic scatter graph of the dependent variable and the fitted values reveals that there is an interesting problem. The scatter plots in Figure 1 displays the dependent variable in the regressions in Table 1 (x-axis) and the fitted values from the regression of model (1) (y-axis). In the left figure, on both the y- and x-axis there is an unusual concentrations of points. Closer inspection of the files supplied by FT in the data archive of the Review of Economics and Statistics reveals that, missing observations on R&D expenditures and some other variables are replaced by 0's. This results in a large number of zero R&D observations (38.2% of the 71264 observations) and this explains the concentration of points on the y-axis. Another consequence of putting R&D expenditures to 0 is hugely negative and positive changes in the R&D growth rates: if for a certain firm R&D is not observed in a certain year and it is observed in the next year the growth rate is calculated to be huge if R&D is set to 0. The other way around generates very negative growth rates. We corrected this by leaving the missing observations untouched and this resulted in 40922 observations that are plotted in the right hand scatter in Figure 1. Although the growth rates are quite large in absolute value in some cases, the problem appears to have been solved to a large extent. We will use this restricted data set in the remaining part of this paper and indicate it by OGGB in Tables 2 and 3.

Columns (4) to (6) in Table 1 and columns (5) to (8) in Table 2 contain the same estimations as in the earlier columns but using the corrected data. If we only look at the estimated signs of the coefficients, by and large the conclusions with respect to the FT-specification do not change. The estimated coefficients are usually closer to zero and the significance appears to have reduced somewhat despite of the calculated standard errors being smaller. We also find smaller R^2 s. Obsolescence and patent effectiveness remain to have a significant impact. If we concentrate on the full model as displayed in column (6), actually only obsolescence and patent effectiveness themselves have a significant impact. The significance of the interaction term is lost. Note that none of the variables related to output growth are significant. For most firms there is procyclicality of the R&D expenditures and this is stronger if the corrected data are used and if we judge this by the number of firms with procyclical R&D. However the average marginal effect of output growth has reduced to 9%. Table 2, columns (4)-(8) displays larger and more significant results than Table 2, columns (1)-(4). Like in Barlevy (2007) also the estimates of the effect of output growth while controlling for balance sheet variables are now strongly significant.

4 Reconciliation and alternative specifications

Despite claiming to base their model on Barlevy (2007), the model specified and estimated by FT differs in three respects:

- FT use additional controls relating to patent effectiveness, obsolescence and external financing.

This is the key contribution of FT: they allow for heterogeneous effects of output growth.

- Both models relate the growth of R&D to the growth in output but use different estimation methods. FT apply OLS and the resulting estimates are in fact a result of a first difference estimation on R&D.³ By first differencing they remove the firm specific effect and according to FT (their footnote 4) also autocorrelation and heteroskedasticity. Barlevy estimates using a fixed effects estimation. Consequently he allows for second-order fixed effects or incidental trends.
- FT use first differenced balance sheet controls in the growth in R&D equation, whereas Barlevy uses level controls.

We will now proceed with estimating some alternative models to make the analyses more comparable. We start with rewriting the model estimated by FT in levels and then proceed estimating this equation with a fixed effects estimation. A crucial observation here is that FT introduce a constant and three time constant explanatory variables (Patent Effectiveness, Obsolescence and External Financing) in the model they estimate.⁴ If we rewrite the model actually estimated by FT, i.e. the model in eq. (4), in level variables we obtain:

$$R\&D_{kt} = \beta_0 t + \beta_1 X_{it} + \beta_2 M_{kt} + \beta_3 M_{kt-1} + \beta_4 X_{it} \cdot Ob_i + \beta_5 X_{it} \cdot PE_i + \beta_6 X_{it} \cdot EF_i + \delta_4 Ob_i \cdot t + \delta_5 PE_i \cdot t + \delta_6 EF_i \cdot t + \eta_k + \theta_t + \varsigma_{kt}. \quad (5)$$

As such this model can make perfect sense, but in their discussion of R&D investments, indeed as a level variable - the growth in R&D investment is only introduced when discussing their empirical model, FT do not discuss time trends in general and time trend related to patent effectiveness, obsolescence and external financing in particular. If we review the results in Table 1, we can conclude that there is in general a negative trend in R&D expenditures but that this negative trend is less negative or even positive for firms with relatively high degrees of patent effectiveness, obsolescence and/or external financing.

The model in eq. (5) is, as far as we are aware, never used in the literature on R&D investments. Furthermore, as becomes clear from footnote 4 of FT, the authors are actually interested in estimating R&D in levels but resort to using first-differenced variables to mitigate the problems of autocorrelation and heteroskedasticity.⁵⁶ If we estimate eq (5) with a first difference estimation, the results of FT as they presented in Table 1 will be found, apart from the standard errors that are calculated somewhat

³We will disregard the difference between a fixed effects estimation and the first difference estimation. As discussed in Baltagi (2013) both estimation methods yield consistent estimates under the same assumptions, but the fixed effects estimator is more efficient.

⁴A constant is present in the model FT written down on page 666, the three time invariant variables are not although they are presented in Table 5 on page 672.

⁵The authors do not actually mention the potential existence of firm specific effects in footnote 4 as a reason for using first differences. However, note that the autocorrelation and heteroskedasticity experienced by FT might be due to these firm specific effects.

⁶One of the referees suggested that testing for unit roots might give an argument to use levels or first differences. We tested for the stationarity of the dependent and independent variables using the Levin, Lin and Chu-test as discussed in Baltagi (2013, Chapter 12). All series were measures in real terms using the GDP deflator as in FT. For all series the null of nonstationarity was rejected.

differently. Now consider a model which is in line with the discussion of FT:⁷

$$R\&D_{kt} = \beta_1 X_{it} + \beta_2 M_{kt} + \beta_3 M_{kt-1} + \beta_4 X_{it} \cdot Ob_i + \beta_5 X_{it} \cdot PE_i + \beta_6 X_{it} \cdot EF_i \\ + \delta_4 Ob_i + \delta_5 PE_i + \delta_6 EF_i + \eta_k + \theta_t + \varsigma_{kt}, \quad (6)$$

the difference being that we deleted all variables related to the trend. We estimate this model using the fixed effects estimator. Due to using this estimator the effects of the time constant explanatory variables cannot be estimated.⁸ The relevant estimation results are presented in Table 3, column (1). Note that the estimations now are performed on 45061 observations. The conclusions we can draw from the estimation results in Table 3 column (1), are quite different from the conclusions from Table 1 column (6). The effect of industry output growth remains insignificant, although the sign has changed. The effect of patent effectiveness and obsolescence on the impact of industry output growth also changed: using a fixed effect estimation reverses the sign and maintains insignificance. The signs are now contrary to the theoretical predictions of FT. This can signal two things. Either the model estimated is not the correct model and the model estimated by FT represent reality better or there is a tendency in the direction of acyclical R&D expenditures for firms in industries with higher patent effectiveness and with lower levels of obsolescence. Finally, the effect of external financing remains positive but has become strongly significant. With respect to heterogeneity we can conclude from the average marginal effects of output growth on R&D expenditures that the procyclicality of R&D is even stronger in this alternative model: procyclicality is found for 952% of the firms whereas it was 88% in Table 1.

We now turn to using the same controls as Barlevy (2007). Instead of controls specified as first differences, Barlevy uses level variables. The results, using pooled OLS like FT do, are presented in Table 3, column (2). The results are far from significant, apart from the interaction term relating to external financing (6). The effect of obsolescence in combination with output growth is again corresponding to the theoretical expectations of FT, although far from significant. The average marginal effect of output growth for this model is quite large, almost 24%, and the number of firm with countercyclical R&D is very small (2.5%).

If we now estimate the same model using a fixed effects estimation, this is the estimation method used by Barlevy (2007) except that three interactions are used as additional regressors, we find Table 3, column (3). All significance is lost and incidental trends appear to be nonexistent. Note that the R^2 is in line with the previous results, indicating that there is significance of the control variables.

The same conclusion follows if we estimate the model of FT as in Table 1 and therefore use controls in differences, but not applying pooled OLS but the fixed effects estimator. The results are in Table 3, column (4). As explained before, the effects of the time constant explanatory variables cannot be estimated.

Due to the lack of significance, we conclude that it is not necessary to allow for second-order

⁷We added Ob, PE and EF as well, although not discussed in the theoretical part of FT, they are included in their estimations.

⁸This deviates with estimation on first differences employed by FT. However, the fixed effects estimator is more efficient and more observations are actually used in the estimations. Note that also the first difference estimator does not make it possible to identify time constant effects. The differences in the estimation results however are marginal. Another advantage of the fixed effects estimator is that it is the more natural approach if the data contain gaps in time like in the present case.

fixed effects as suggested by Barlevy (2007). If we concentrate only on the signs of the estimated coefficients, the conclusions are very similar to the ones following from Table 1. There is almost a one-to-one correspondence of the sign (14 out of 16 cases). However, as noted, almost all significance is lost. With respect to heterogeneity, only the external financing cross term is significant but this is precisely the variable that is not emphasized in FT. In fact it is hardly discussed by them apart as entries in regression tables.

Although not very well known in economics⁹, it is actually possible to estimate the effect of time constant explanatory variables while allowing for fixed effects. In a complete linear setting this was first discussed by Mundlak (1978). He proved that if the fixed effects are linearly related to the time averages of the time varying explanatory variables, adding these time averages to the model and estimating with either random effects estimation or the pooled regression estimation will yield the fixed effect estimates. These are the estimated effects of the time varying explanatory variables. What appears to be less well known, is that Mundlak’s assumption of linearity is superfluous. The result actually holds without specifying the relation between the fixed effects and the explanatory variables. Furthermore, even the standard errors calculated in the random effects estimation are correct i.e. are the same as the standard errors calculated if the fixed effects estimator would have been used.¹⁰ Because we can actually use a random effects estimator to calculate the fixed effects estimator, we can also add time invariant explanatory variables and their effects are indeed identified. However, for consistency we need to make the assumption that these time invariant explanatory variables are not correlated with the fixed effects. Note that this assumption is implicitly also made by FT. The model in the Mundlak specification reads:

$$R\&D_{kt} = \beta_1 X_{it} + \beta_2 \Delta M_{kt} + \beta_3 \Delta M_{kt-1} + \beta_4 X_{it} \cdot Ob_i + \beta_5 X_{it} \cdot PE_i + \beta_6 X_{it} \cdot EF_i + \delta_2 Ob_i + \delta_3 PE_i + \delta_4 EF_i + \beta_1^* \bar{X}_i + \beta_2^* \bar{X}_i \cdot Ob_i + \beta_3^* \bar{X}_i \cdot PE_i + \beta_4^* \bar{X}_i \cdot EF_i + \eta_k + \theta_t + \varsigma_{kt}, \quad (7)$$

The random effect estimator gives the following effects of time constant variables Obsolescence, Patent Effectiveness and External Financing on R&D: 8.993 (se = 1.498), -0.562 (se = 0.427) and -0.282 (se = 0.241). Only the effect of Patent Effectiveness is significant and has the same sign as reported in FT.

Our conclusion is that if the same explanatory variables and appropriate estimation methods are used, the heterogeneity of cyclicity of R&D investment found by FT is reduced considerably. Heterogeneity appears to be dependent especially on the external financing of the firm. Just like FT, we do find procyclicality for the vast majority of firms.

Barlevy (2007) and Fabrizio and Tsolmon (2014) both relate R&D to industry output. Although this reduces firm specific idiosyncrasies, using firm specific output appears to be the more direct measure and this is more in line with the hypotheses of Schumpeter (1939). In Table 4, columns (1)-(4), we therefore present, estimates of the same specifications as in Table 3 but replacing industry by firm output. The estimation results are very similar: the vast majority of firms appears to display procyclical R&D investments and there is heterogeneity but mainly with respect to external financing.

⁹No mention of this more general results can be found in e.g. Baltagi (2013). In e.g. sociology and biology, adding the time averages of the time varying explanatory variables is quite common, see e.g. Allison (2009); Neuhaus and Kalbfleisch (1998); Phillips (2006).

¹⁰A proof of these claims is available on request.

Compared to Table 3, the sign of this variable has changed indicating that depending more on external financing tends to make R&D expenditures less procyclical.

From Table 2 of FT it is clear that the growth in R&D attains some unrealistically high values in absolute sense: the minimum of the growth rate is 1169% and the maximum 1294%.¹¹ Apart from our doubt whether these figures can be correct there is also another problem. The difference on the year to year log of R&D only measures the growth rate for relatively small changes. To investigate this we restrict the sample to absolute growth rates of 50% or smaller.¹² This reduces the number of observations from about 45000 to about 37000. The estimation results, again for the same specification as in Table 3, are in columns (5)-(8) of Table 4. The overall conclusions do not change: more than 90% of the firms display procyclical R&D expenditures and there is heterogeneity but again due to external financing.

5 Conclusion

In this paper we replicated Fabrizio and Tzolmon (2014) and Barlevy (2007). The results of FT were retrieved almost precisely whereas significance was lost in the Barlevy specification. Significance of the Barlevy estimates was restored after having corrected the data, whereas the overall conclusion of the FT remained the same after this correction. If we use suitable (fixed effects) estimation methods, no significance was found for the key factor of Fabrizio and Tzolmon (2014) obsolescence and patent effectiveness. Heterogeneity of the cyclicity of R&D expenditures remained, however, due to the external financing interaction term. These conclusions were corroborated when we use firm specific output growth in stead of industry specific output growth and when we restricted the sample to more realistic growth numbers of R&D. All in all, overall practicality, corroborating the conclusions of both Barlevy (2007) and Fabrizio and Tzolmon (2014) and the existence of heterogeneity of procyclicality across firms, corroborating the results of Fabrizio and Tzolmon (2014), is confirmed by our replication.

References

- Aghion, P., Askenazy, P., Berman, N., Cetto, G., Eymard, L., 2012. Credit constraints and the cyclicity of R&D investment: evidence from France. *Journal of the European Economic Association* 10, 1001–1024.
- Allison, P., 2009. *Fixed Effects Regression Models*. SAGE.
- Baltagi, B., 2013. *Econometric Analysis of Panel Data*, 5th Edition. John Wiley & Sons, Chichester.
- Barlevy, G., 2007. On the cyclicity of research and development. *The American Economic Review* 97 (4), 1131–1164.
- Fabrizio, K., Tzolmon, U., 2014. An empirical examination of the procyclicality of R&D investment and innovation. *The Review of Economics and Statistics* 96 (4), 662–675.
- Francois, P., Lloyd-Ellis, H., 2009. Schumpeterian cycles with pro-cyclical R&D. *Review of Economic Dynamics* 12, 567–591.

¹¹These large growth rates are partly due to replacing missing observations on R&D by zeros and adding 1 to the logarithm.

¹²We also estimated with a cut off of 25% and 100% but the conclusions remain the same.

- Kraiczy, N., Hack, A., Kellermanns, F., 2015. CEO innovation orientation and r&d intensity in small and medium-sized firms: the moderating role of firm growth. *Journal of Business Economics* 85, 851–872.
- Mundlak, Y., 1978. On the pooling of time series and cross section data. *Econometrica* 46, 69–85.
- Neuhaus, J., Kalbfleisch, J., 1998. Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics* 54, 638–645.
- Ouyang, M., 2011. On the cyclicity of R&D. *The Review of Economics and Statistics* 93(2), 542–533.
- Phillips, J., 2006. Explaining discrepant findings in cross-sectional and longitudinal analysis: an application to U.S. homicide rates. *Social Science Research* 35, 948–974.
- Schumpeter, J., 1939. *Business Cycles: a Theoretical, Historical and Statistical Analysis of the Capitalist Process*. McGraw-Hill.

Table 1: OLS estimates of growth in R&D expenditures (eq. 4)

	(1)	(2)	(3)	(4)	(5)	(6)
ΔOutput	0.272*** (0.034)	-0.777** (0.262)	-0.747* (0.292)	0.143*** (0.022)	-0.125 (0.156)	-0.053 (0.181)
$\Delta\text{Output} \times \text{Patent Effectiveness}$		-0.054 (0.306)	-0.178 (0.314)		-0.279 (0.212)	-0.313 (0.213)
Patent Effectiveness		0.183*** (0.023)	0.101*** (0.026)		0.070*** (0.018)	0.052* (0.021)
$\Delta\text{Output} \times \text{Obsolescence}$		0.341*** (0.079)	0.341*** (0.088)		0.114* (0.048)	0.091 (0.057)
Obsolescence		0.048*** (0.010)	0.033** (0.010)		0.029*** (0.007)	0.026*** (0.008)
$\Delta\text{Output} \times \text{External Financing}$			0.039 (0.063)			0.044 (0.045)
External Financing			0.014*** (0.004)			0.001 (0.003)
Constant	0.038 (0.026)	-0.189*** (0.040)	-0.115** (0.042)	-0.037** (0.014)	-0.156*** (0.028)	-0.142*** (0.030)
Data used	FT	FT	FT	OGGB	OGGB	OGGB
N (observations)	71264	71264	71264	40922	40922	40922
R^2	0.224	0.225	0.226	0.122	0.123	0.123
Average marginal effect output	0.272	0.168	0.157	0.143	0.074	0.089
%Avg marginal effect < 0	0.0	8.1	14.9	0.0	5.0	12.0

Robust standard errors, clustered by firm in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
 Additional controls as in FT, i.e. first differenced, including year fixed effects.

Figure 1: Growth in R&D expenditures and fitted values

Table 2: Fixed effects estimates of growth in R&D expenditures eq. (2)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ Output	0.241*** (0.044)	0.230*** (0.055)	-0.026 (0.042)	0.019 (0.050)	0.357*** (0.027)	0.263*** (0.034)	0.133*** (0.028)	0.101** (0.033)
Data used	FT	FT	FT	FT	OGGB	OGGB	OGGB	OGGB
Year fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
Balance sheet controls	No	No	Yes	Yes	No	No	Yes	Yes
N (observations)	85604	85604	75807	75807	52870	52870	44144	44144
R^2 (overall)	0.003	0.000	0.237	0.236	0.007	0.004	0.114	0.111

Robust standard errors, clustered by firm in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Balance sheet controls in levels as in Barlevy (2007).

Table 3: Alternative estimates of (growth in) R&D expenditures, OGGB data

	(1) R&D	(2) Δ R&D	(3) Δ R&D	(4) Δ R&D
Output	0.055 (0.148)			
Output $\times$ Patent Effectiveness	0.066 (0.234)			
Output $\times$ Obsolescence	-0.005 (0.053)			
Output $\times$ External Financing	0.105*** (0.029)			
Δ Output		-0.069 (0.160)	-0.098 (0.299)	-0.159 (0.279)
Δ Output $\times$ Patent Effectiveness		0.004 (0.199)	-0.068 (0.279)	-0.112 (0.262)
Δ Output $\times$ Obsolescence		0.076 (0.049)	0.074 (0.090)	0.102 (0.083)
Δ Output $\times$ External Financing		0.142*** (0.032)	0.038 (0.053)	0.042 (0.051)
Constant	-2.452*** (0.356)	-0.118*** (0.023)	-0.526*** (0.068)	-0.025 (0.015)
Estimation method	FE	OLS	FE	FE
Type of controls	Level	Level	Level	Δ
N (observations)	45061	44144	44144	40922
R^2	0.451	0.124	0.092	0.088
Average marginal effect output	0.135	0.238	0.107	0.103
%Avg marginal effect < 0	4.8	2.5	1.2	3.5

Robust standard errors, clustered by firm in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 4: Alternative estimates of (growth in) R&D expenditures using firm output, OGGB data

	(1) R&D	(2) Δ R&D	(3) Δ R&D	(4) Δ R&D	(5) R&D	(6) Δ R&D	(7) Δ R&D	(8) Δ R&D
Output	0.007 (0.112)				0.062 (0.139)			
Output $\times$ Patent Effectiveness	0.048 (0.093)				0.080 (0.114)			
Output $\times$ Obsolescence	0.008 (0.035)				0.010 (0.041)			
Output $\times$ External Financing	-0.021** (0.008)				-0.022* (0.009)			
Δ Output		0.033 (0.091)	0.042 (0.108)	0.053 (0.128)		0.079 (0.049)	0.082 (0.052)	0.023 (0.045)
Δ Output $\times$ Patent Effectiveness		-0.207** (0.077)	-0.086 (0.089)	0.002 (0.109)		-0.078 (0.041)	-0.043 (0.041)	-0.020 (0.037)
Δ Output $\times$ Obsolescence		0.051 (0.030)	0.036 (0.036)	0.024 (0.043)		0.008 (0.016)	0.001 (0.017)	0.008 (0.015)
Δ Output $\times$ External Financing		-0.007 (0.006)	-0.018* (0.007)	-0.029** (0.009)		-0.009** (0.003)	-0.012*** (0.003)	-0.009** (0.003)
Constant	-1.093*** (0.157)	-0.102*** (0.021)	-0.518*** (0.061)	-0.029* (0.014)	-1.142*** (0.190)	-0.061*** (0.010)	-0.288*** (0.030)	0.004 (0.006)
Sample	Full	Full	Full	Full	Restricted	Restricted	Restricted	Restricted
Estimation method	FE	OLS	FE	FE	FE	OLS	FE	FE
Type of controls	Level	Level	Level	Δ	Level	Level	Level	Δ
N (observations)	52891	51690	51690	45840	36678	36678	36678	37589
R^2	0.433	0.135	0.099	0.094	0.504	0.080	0.061	0.081
Average marginal effect output	0.036	0.084	0.096	0.103	0.112	0.061	0.058	0.030
%Avg marginal effect < 0	8.2	0.2	0.0	7.8	0.0	0.0	7.5	7.4

Robust standard errors, clustered by firm in parentheses.

* p<0.05, ** p<0.01, *** p<0.001.

Full: full sample, no restrictions on growth R&D. Restricted: restricted sample, abs(growth R&D) $\leq$ 50 %.