Dynamic Models of R&D, Innovation and Productivity: Panel Data Evidence for Dutch and French Manufacturing

Wladimir Raymond[*], Jacques Mairesse[†], Pierre Mohnen[‡] and Franz Palm[§]

December 15, 2012

Abstract

This paper introduces dynamics in the R&D to innovation and innovation to productivity relationships, which have mostly been estimated on cross-sectional data. It estimates by full information maximum likelihood four nonlinear dynamic simultaneous equations models that differ in the way that innovation enters the conditional mean of labor productivity: through an observed binary indicator, an observed intensity variable or through the continuous latent variables that correspond to the observed occurrence or intensity. The estimation controls for individual effects using two unbalanced panels of Dutch and French manufacturing firms from three waves of the Community Innovation Survey. The results provide evidence of robust unidirectional causality from innovation to productivity and of stronger persistence in productivity than in innovation.

Keywords: Panel data, Dynamics, Simultaneous equations, R&D, Innovation, Productivity

JEL classification: C33, C34, C35, L60, O31, O32

[*]Institut National de la Statistique et des Études Économiques (STATEC), EPR2 Unit, Rue Erasme BP304 L-2013 Luxembourg; Tel.: (+352) 247 84 351 wladimir.raymond@statec.etat.lu.

[†]CREST-INSEE, UNU-MERIT, Maastricht University and NBER; mairesse@ensae.fr

[‡]UNU-MERIT, Maastricht University and CIRANO, p.mohnen@maastrichtuniversity.nl

[§]Maastricht University and CESifo fellow, f.palm@maastrichtuniversity.nl

1 Introduction

For decades, R&D and innovation have been recognized by scholars and policy makers as major drivers of country, industry and firm economic performance. Many of the early studies, following the lead of Griliches (1979), have used an augmented production function with R&D capital to estimate the returns to R&D at the firm level. More recently, many studies have relied on innovation survey indicators and on the so-called CDM framework to analyze simultaneously a knowledge production function relating innovation output to R&D, and an augmented production function linking productivity to innovation output (Crépon et al., 1998; Mairesse et al., 2005; Griffith et al., 2006). Both the effects of R&D on innovation output and of innovation output on productivity are usually found to be positive and significant in these studies. Most of them, however, are based on cross-sectional data and cannot account for the dynamic linkages between innovation and economic performance nor control for unobserved firm heterogeneity. This is where the present study comes into play.[1] More specifically, using data from three waves of the Community Innovation Survey (CIS) for France and the Netherlands, and controlling for unobserved individual effects, we examine whether there is evidence of persistence in firm innovation and productivity and of bidirectional causality between them.

There are several reasons why one should introduce dynamics in the interrelationships between R&D, innovation and productivity. Firstly, the time lag between a firm's decision to invest in R&D, the associated R&D outlays and the resulting innovation success may be substantial because of 'time to build', opportunity cost and uncertainty inherent to the innovation process (Majd and Pindyck, 1987). For example, the studies of knowledge production function on firm panel data, where patents proxy for knowledge, specify a relation of patents to distributed lags of R&D (Pakes and Griliches, 1980; Hall et al., 1986). Secondly, scholars argue that a successfully innovative firm is more likely than a non-innovating firm to experience innovation success in the future, in other words, that 'success breeds success'. Several papers have investigated the persistence of innovation success, measured by the number of granted patents (Geroski et al., 1997), the introduction of new or significantly improved products (Peters, 2009) or production methods (Flaig and Stadler, 1994), or the share in total sales accounted for by sales of these products (henceforth the share of innovative sales) (Raymond et al., 2010). Thirdly, it is also argued that the economic performance of a firm, especially of a repeatedly innovating firm, is likely to exhibit persistence. For instance, Bailey et al. (1992), Bartelsman and Dhrymes (1998), and Fariñas and Ruano (2005) find strong evidence of persistence of firm level productivity differentials using transition probabilities on the

[1]See also Parisi et al. (2006) and Huergo and Moreno (2011) for two different attempts to go in this direction.

quintiles or deciles of the distribution of these differentials over time, or using kernel techniques to estimate the conditional distribution of firm level productivity at period t given productivity at period $t-1$. Finally, because of information asymmetry, firms may be more willing to rely on retained earnings rather than to seek external funding for their future innovations (Bhattacharya and Ritter, 1983), implying a feedback effect from productivity to innovation.

To investigate these dynamic aspects, we study four nonlinear dynamic simultaneous equations models that differ in the way that innovation enters the conditional mean of labor productivity: through an observed binary indicator, an observed intensity variable or through the continuous latent variables that correspond to the observed occurrence or intensity. We describe these models in detail in Section 2.

We show in Section 3 how to derive the full information maximum likelihood estimator, henceforth FIML, assuming random effects that are correlated with (sufficiently) time-varying explanatory variables. More specifically, we take care of the initial conditions problem due to the autoregressive structure of the models and the presence of firm effects using Wooldridge's (2005) 'simple solutions' approach, and we handle multiple integration due to the correlations of firm effects and idiosyncratic errors across equations using Gauss-Hermite quadrature sequentially along the lines of Raymond (2007, chapter 6).

In Section 4, we explain the data on which we base our estimations and provide some descriptive statistics. These data come from three waves of the Dutch and the French Manufacturing Community Innovation Surveys (CIS) for 1994-1996, 1998-2000 and 2002-2004, supplemented by a few firm accounting variables. We work with an unbalanced panel to have a larger sample and thus to weaken possible survivorship biases and to obtain more accurate estimates.

In Section 5 we present our results. For both countries they reveal strong persistence in productivity but weaker persistence in innovation, and they indicate a unidirectional causality running from innovation to labor productivity. Whereas past innovation matters to productivity, the most productive enterprises are not more successful in introducing new or significantly improved products and do not attain larger shares of innovative sales than the least productive ones.

2 Models

Our models consist of a knowledge production function (KPF) and an augmented production function (APF) relating respectively innovation output to R&D and other relevant innovation factors, and productivity to innovation output and other relevant production factors. Four variables of innovation output are considered in the analysis. The first is an *observed* binary variable taking the value one if an enterprise is a product innovator, and zero otherwise. In the innovation survey, an enterprise is asked whether it has introduced at least one new or improved product on the market in the last three years. A product innovator is an enterprise that has responded positively to this question. The second variable is the observed share of innovative sales, or observed innovation intensity. This variable is directly reported by the enterprise when filling out the questionnaire of the innovation survey. The share of innovative sales is taken with respect to sales reported in the last year of the three-year period. Finally we consider the two continuous *latent* innovation output variables that underly respectively the propensity to introduce new or improved products on the market and the potential share of innovative sales.

2.1 Knowledge production function

Let y^*_{1it} denote a latent variable underlying firm i's $(i = 1, ..., N)$ propensity to achieve product innovations at period t $(t = 0_i, ..., T_i)$ given past observed incidence of product innovations $y_{1i,t-1}$, past labor productivity $y_{3i,t-1}$, past R&D and other firm- and market-specific characteristics $\mathbf{x}_{1it}$, and unobserved firm heterogeneity α_{1i}.[2] Formally

$$y^*_{1it} = \vartheta_{11} y_{1i,t-1} + \vartheta_{13} y_{3i,t-1} + \boldsymbol{\beta}'_1 \mathbf{x}_{1it} + \alpha_{1i} + \varepsilon_{1it}, \tag{2.1}$$

where ϑ_{11} and ϑ_{13} capture the effect of past product innovation incidence and past productivity on the propensity to innovate, $\boldsymbol{\beta}'_1$ captures the effects of past R&D and other explanatory variables and ε_{1it} denotes idiosyncratic errors encompassing other time-varying unobserved variables that affect y^*_{1it}. The observed dependent variable, y_{1it}, corresponding to y^*_{1it} is defined as

$$y_{1it} = \mathbf{1}[y^*_{1it} > 0], \tag{2.2}$$

where $\mathbf{1}[\ ]$ denotes the indicator function taking the value one if the condition between squared brackets is satisfied, and zero otherwise.

Let y^*_{2it} denote the firm's latent share of innovative sales, or potential innovation intensity, given past observed innovation intensity $y_{2i,t-1}$, past labor productivity $y_{3i,t-1}$, past R&D and other firm- and market-specific characteristics $\mathbf{x}_{2it}$, and firm-specific effects α_{2i}. Formally

$$y^*_{2it} = \vartheta_{22} y_{2i,t-1} + \vartheta_{23} y_{3i,t-1} + \boldsymbol{\beta}'_2 \mathbf{x}_{2it} + \alpha_{2i} + \varepsilon_{2it}, \tag{2.3}$$

[2] By letting t vary from 0_i to T_i, we allow firms to enter and exit the sample at different periods. 0_i denotes the first observation in the sample, which varies across enterprises. Similarly, T_i denotes the last observation in the sample, which also varies across enterprises.

where the coefficients ϑ_{22} and ϑ_{23} capture the effect of past observed share of innovative sales and past labor productivity on the potential innovation intensity, $\boldsymbol{\beta}_2'$ captures the effect of past R&D and other explanatory variables and ε_{2it} denotes idiosyncratic errors. The observed counterpart to y_{2it}^* is defined as

$$y_{2it} = \mathbf{1}[y_{1it}^* > 0]y_{2it}^*. \tag{2.4}$$

In other words, the share of innovative sales of firm i is observed to be positive in period t if its innovation propensity is sufficiently large in that period. If not, the share of innovative sales is set equal to zero.

The product innovation indicator and the share of innovative sales variables are taken from the innovation survey of the two countries. Since the share of innovative sales lies within the unit interval, we use a logit transformation in the estimation in order to normalize it over the entire set of real numbers.[3]

The set of other explanatory variables includes the log R&D per employee, the log market share, and size, industry and time dummy variables.[4] Due to the lengthy and risky nature of research and innovation activities, we use lagged R&D to explain innovation incidence and innovation intensity. Since we cannot construct a stock measure of R&D, we rely instead on lagged R&D expenditures of continuous R&D performers. We also include a lagged dummy variable for non-continuous R&D performers to compensate for the fact that we use positive values of R&D only for continuous R&D performers. Market share is used at the three digit industry level as a measure of relative size that can reflect market power. It is lagged in order to avoid possible endogeneity concerns (due for example to measurement errors in firm sales which would affect both our market share and productivity variables).[5] Employment is used as a measure of firm absolute size. Since the relation between innovation and size is often found to be nonlinear in the empirical literature, we use four class indicators of small enterprises (# employees ≤ 50), medium-sized enterprises ($50 <$ # employees ≤ 250), large enterprises ($250 <$ # employees ≤ 500) and very large enterprises ($500 <$ # employees), the fourth class being considered as the reference.[6] We control for industry effects, according to the OECD (2007) technology-based classification of high-tech, medium-high-tech, medium-low-tech, and low-tech industries, using three dummy variables for the first three

[3] Zero values of the share of innovative sales are replaced by a positive value τ_1 smaller than the minimum positive observed value of that variable, and values one are replaced by a positive value τ_2 higher than the second largest observed value. These choices have a negligible effect on our estimates.

[4] In some specifications, we have also three indicators of the distance to the productivity frontier. We find, however, that they are not statistically significant (see Appendix C).

[5] The market share of a firm is defined as the ratio of its sales over the total sales of the three digit industry it belongs to. The latter is obtained by adding up the sales of all firms in our sample that belong to that industry after multiplying them by the appropriate raising factor.

[6] Size class indicators are less informative than the number of employees. Hence, their use in the innovation output equations weakens any endogeneity of employment.

industry categories and taking low-tech industries as the reference. Such industry-specific effects capture differences in technological opportunities (it is easier to innovate in certain industries than in others) and in intensity of competition (which is expected to be higher in high-tech than in low-tech industries). Since our panel consists only of three periods and we need one for the lagged variables, we need only include a time dummy variable for the period 1998-2000, with 2002-2004 being the reference. This time dummy controls for macroeconomic shocks and for inflation.

2.2 Augmented production function

As in the great majority of studies, we assume a Cobb-Douglas APF written in terms of a log linear productivity equation relating labor productivity to labor (i.e., we do not assume a constant scale elasticity), physical capital per employee, where capital is proxied by physical investment (due to unavailability of a stock measure), and innovation output. We consider four specifications where we explain productivity by latent innovation (i.e. the propensity to achieve product innovations or potential innovation intensity) or by observed innovation (i.e. innovation incidence or observed innovation intensity). In all cases we also condition current labor productivity on its past values and control for unobserved heterogeneity through firm effects. Thus, we can write

$$y_{3it} = \vartheta_{33} y_{3i,t-1} + \boldsymbol{\beta}_3' \mathbf{x}_{3it} + \gamma_j y_{jit}^* + \alpha_{3i} + \varepsilon_{3it}, \tag{2.5a}$$

$$y_{3it} = \vartheta_{33} y_{3i,t-1} + \boldsymbol{\beta}_3' \mathbf{x}_{3it} + \gamma_j y_{jit} + \alpha_{3i} + \varepsilon_{3it}, \tag{2.5b}$$

with $j = 1$ or 2 where innovation propensity (y_{1it}^*) or potential innovation intensity (y_{2it}^*) explains labor productivity in equation (2.5a), and innovation incidence (y_{1it}) or observed innovation intensity (y_{2it}) explains labor productivity in equation (2.5b). The coefficient ϑ_{33} captures the effect of past labor productivity on current labor productivity, $\boldsymbol{\beta}_3'$ captures the effect of standard input variables, i.e. employment and physical investment per employee, γ_j captures the effect of innovation output on labor productivity, and α_{3i} and ε_{3it} denote time-invariant firm effects and idiosyncratic errors. We also control for industry and time effects as in the KPF equations.

3 FIML estimation

We shall now explain how to derive the FIML estimator: how to take care of the initial conditions problem due to the autoregressive structure of the models and the presence of firm effects, how to write the likelihood function, and how to handle the multiple integration due to the correlations of firm effects and idiosyncratic errors across equations.

3.1 Initial conditions

The initial conditions problem stems from the fact that the first observed value of the lagged dependent variables is correlated with the individual effects. Ignoring or inadequately accounting for this correlation results in a bias of the effect of the lagged dependent variables. Several solutions have been proposed in the econometric literature. We follow the one suggested by Wooldridge (2005).

Wooldridge's 'simple solutions' have been originally applied to autoregressive nonlinear single-equation models with individual effects. We adapt the approach to a model with multiple equations. In other words, we project in each equation the individual effects on the first observation of the corresponding dependent variables and on the observed history of the other sufficiently *time-varying* explanatory variables. Formally

$$\alpha_{1i} = b_{10} + \boldsymbol{b}_{11}' \boldsymbol{y}_{1i0_i} + \boldsymbol{b}_{12}' \mathbf{x}_{1i} + a_{1i}, \tag{3.1}$$

$$\alpha_{2i} = b_{20} + \boldsymbol{b}_{21}' \boldsymbol{y}_{2i0_i} + \boldsymbol{b}_{22}' \mathbf{x}_{2i} + a_{2i}, \tag{3.2}$$

$$\alpha_{3i} = b_{30} + \boldsymbol{b}_{31}' \boldsymbol{y}_{3i0_i} + \boldsymbol{b}_{32}' \mathbf{x}_{3i} + a_{3i}, \tag{3.3}$$

where $\boldsymbol{y}_{ki0_i}$ $(k = 1, 2, 3)$ represents the initial values of the dependent variables, $\mathbf{x}_{ki} = (\mathbf{x}_{ki0_i+1}, ...,$ $\mathbf{x}_{kiT_i})'$ represents the history of (in principle all) the observations of the time-varying explanatory variables, and $\boldsymbol{a}_i = (a_{1i}, a_{2i}, a_{3i})'$ denotes the vector of projection errors assumed orthogonal to $\boldsymbol{y}_{ki0_i}$, $\mathbf{x}_{ki}$ and $\boldsymbol{\varepsilon}_{it} = (\varepsilon_{1it}, \varepsilon_{2it}, \varepsilon_{3it})'$. The ancillary parameters b_{k0}, $\boldsymbol{b}_{k1}$ and $\boldsymbol{b}_{k2}$ are to be estimated alongside the parameters of interest.

Three important remarks are in order regarding equations (3.1)-(3.3). Firstly, if the coefficient vectors $\boldsymbol{\beta}_k$ contain intercepts, only the sums of those intercepts and b_{k0} are identified. Secondly, if the explanatory variables are time-invariant or do not show sufficient *within* variation, then the coefficients $\boldsymbol{b}_{k2}$ and $\boldsymbol{\beta}_k$ cannot be separately identified. As a result, only the sufficiently time-varying explanatory variables enter equations (3.1)-(3.3). Thirdly, in order to discriminate between the effect of the lagged dependent variables and that of the initial values, given the unbalancedness of the panel, we actually have to include in equations (3.1)-(3.3) two types of initial values with different coefficients for firms present in all three waves and for those present only in two waves. Following Wooldridge (2005) we make the following distributional assumptions:

$\varepsilon_{it|\boldsymbol{y}_{i,t-1},\mathbf{x}_{it},\boldsymbol{\alpha}_i} \overset{iid}{\sim} \boldsymbol{Normal}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})$; $\boldsymbol{a}_{i|\boldsymbol{y}_{i0_i},\mathbf{x}_i} \overset{iid}{\sim} \boldsymbol{Normal}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{a}})$ where $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{a}}$ are given by

$$\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} = \begin{pmatrix} 1 & & \\ \rho_{\varepsilon_1\varepsilon_2}\sigma_{\varepsilon_2} & \sigma^2_{\varepsilon_2} & \\ \rho_{\varepsilon_1\varepsilon_3}\sigma_{\varepsilon_3} & \rho_{\varepsilon_2\varepsilon_3}\sigma_{\varepsilon_2}\sigma_{\varepsilon_3} & \sigma^2_{\varepsilon_3} \end{pmatrix}, \ \boldsymbol{\Sigma}_{\boldsymbol{a}} = \begin{pmatrix} \sigma^2_{a_1} & & \\ \rho_{a_1a_2}\sigma_{a_1}\sigma_{a_2} & \sigma^2_{a_2} & \\ \rho_{a_1a_3}\sigma_{a_1}\sigma_{a_3} & \rho_{a_2a_3}\sigma_{a_2}\sigma_{a_3} & \sigma^2_{a_3} \end{pmatrix} \tag{3.4}$$

and are also to be estimated.

3.2 Likelihood

We now derive the likelihood functions. For simplicity, we provide the expressions explicitly only for the specifications where y^*_{1it} or y_{1it} (respectively the latent propensity to achieve product innovations and the observed indicator of innovation occurrence) enters the augmented production function. Those with y^*_{2it} or y_{2it} are presented in Appendix B.

Model with latent innovation propensity

The model with latent innovation propensity as a predictor of labor productivity consists of equations (2.1)-(2.4) and (2.5a) with $j = 1$ in equation (2.5a). These equations constitute the *structural* form of the model. Since y^*_{1it} is unobserved, we cannot, unlike in simultaneous equations models with observed explanatory variables, derive the likelihood function directly where the dependent variable is included as a regressor. As a result, FIML estimates can be obtained only through the likelihood function of the *reduced* form of the model. The reduced-form equations are given by equations (2.1)-(2.4) and

$$y_{3it} = \vartheta_{33}y_{3i,t-1} + \boldsymbol{\beta}'_3\mathbf{x}_{3it} + \gamma_1\left[\vartheta_{11}y_{1i,t-1} + \vartheta_{13}y_{3i,t-1} + \boldsymbol{\beta}'_1\mathbf{x}_{1it}\right] + \underbrace{\gamma_1\alpha_{1i} + \alpha_{3i}}_{\underline{\alpha_{3i}}} + \underbrace{\gamma_1\varepsilon_{1it} + \varepsilon_{3it}}_{\underline{\epsilon_{3it}}}, \tag{3.5}$$

where y^*_{1it} has been replaced by its right-hand side expression of equation (2.1).[7] The individual effects and the idiosyncratic errors of the reduced form are given by $\underline{\boldsymbol{\alpha}_i} = (\alpha_{1i}, \alpha_{2i}, \underline{\alpha_{3i}})'$ and $\underline{\boldsymbol{\varepsilon}_{it}} = (\varepsilon_{1it}, \varepsilon_{2it}, \underline{\varepsilon_{3it}})'$, where $\underline{\alpha_{3i}}$ and $\underline{\varepsilon_{3it}}$ are defined in equation (3.5). After replacing α_{1i}, α_{2i} and α_{3i} by their expressions (3.1) to (3.3) into equations (2.1), (2.3) and (3.5), we obtain the projection errors of the reduced form as $\underline{\boldsymbol{a}_i} = (a_{1i}, a_{2i}, \underline{a_{3i}})'$ with $\underline{a_{3i}} = \gamma_1 a_{1i} + a_{3i}$. Since the structural form idiosyncratic errors and projection errors are both normally distributed, their reduced-form counterparts are also normally distributed with means zero and covariance matrices

[7] In the econometric literature on simultaneous equations models, equation (3.5) is referred to as *restricted* reduced form when written with *all* the parameters of the structural form and *unrestricted* reduced form when written with the underlined parameters. In the latter case, $\gamma_1\vartheta_{11}$ would constitute a new coefficient, say $\underline{\vartheta_{11}}$. The restricted reduced form is of interest in our analysis.

$\underline{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}}$ and $\underline{\boldsymbol{\Sigma}_{\boldsymbol{a}}}$ given by

$$\underline{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}} = \begin{pmatrix} 1 & & \\ \rho_{\varepsilon_1\varepsilon_2}\sigma_{\varepsilon_2} & \sigma^2_{\varepsilon_2} & \\ \underline{\rho_{\varepsilon_1\varepsilon_3}\sigma_{\varepsilon_3}} & \underline{\rho_{\varepsilon_2\varepsilon_3}}\sigma_{\varepsilon_2}\underline{\sigma_{\varepsilon_3}} & \underline{\sigma^2_{\varepsilon_3}} \end{pmatrix}, \quad \underline{\boldsymbol{\Sigma}_{\boldsymbol{a}}} = \begin{pmatrix} \sigma^2_{a_1} & & \\ \rho_{a_1a_2}\sigma_{a_1}\sigma_{a_2} & \sigma^2_{a_2} & \\ \underline{\rho_{a_1a_3}}\sigma_{a_1}\underline{\sigma_{a_3}} & \underline{\rho_{a_2a_3}}\sigma_{a_2}\underline{\sigma_{a_3}} & \underline{\sigma^2_{a_3}} \end{pmatrix}, \tag{3.6}$$

where the underlined components of $\underline{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}}$ and $\underline{\boldsymbol{\Sigma}_{\boldsymbol{a}}}$ are nonlinear functions of their structural form counterparts and are given by

$$\underline{\sigma^2_{\varepsilon_3}} = \gamma_1^2 + \sigma^2_{\varepsilon_3} + 2\gamma_1\rho_{\varepsilon_1\varepsilon_3}\sigma_{\varepsilon_3}, \qquad \underline{\sigma^2_{a_3}} = \gamma_1^2\sigma^2_{a_1} + \sigma^2_{a_3} + 2\gamma_1\rho_{a_1a_3}\sigma_{a_1}\sigma_{a_3}, \tag{3.7a}$$

$$\underline{\rho_{\varepsilon_1\varepsilon_3}} = \frac{\gamma_1 + \rho_{\varepsilon_1\varepsilon_3}\sigma_{\varepsilon_3}}{\left(\gamma_1^2 + \sigma^2_{\varepsilon_3} + 2\gamma_1\rho_{\varepsilon_1\varepsilon_3}\sigma_{\varepsilon_3}\right)^{\frac{1}{2}}}, \qquad \underline{\rho_{a_1a_3}} = \frac{\gamma_1\sigma_{a_1} + \rho_{a_1a_3}\sigma_{a_3}}{\left(\gamma_1^2\sigma^2_{a_1} + \sigma^2_{a_3} + 2\gamma_1\rho_{a_1a_3}\sigma_{a_1}\sigma_{a_3}\right)^{\frac{1}{2}}}, \tag{3.7b}$$

$$\underline{\rho_{\varepsilon_2\varepsilon_3}} = \frac{\gamma_1\rho_{\varepsilon_1\varepsilon_2} + \rho_{\varepsilon_2\varepsilon_3}\sigma_{\varepsilon_3}}{\left(\gamma_1^2 + \sigma^2_{\varepsilon_3} + 2\gamma_1\rho_{\varepsilon_1\varepsilon_3}\sigma_{\varepsilon_3}\right)^{\frac{1}{2}}}, \qquad \underline{\rho_{a_2a_3}} = \frac{\gamma_1\rho_{a_1a_2}\sigma_{a_1} + \rho_{a_2a_3}\sigma_{a_3}}{\left(\gamma_1^2\sigma^2_{a_1} + \sigma^2_{a_3} + 2\gamma_1\rho_{a_1a_3}\sigma_{a_1}\sigma_{a_3}\right)^{\frac{1}{2}}}. \tag{3.7c}$$

The individual likelihood function of the reduced form conditional on $\underline{\boldsymbol{a}_i}$, denoted by $l_{1i|\underline{\boldsymbol{a}_i}}$, is given by

$$l_{1i|\underline{\boldsymbol{a}_i}} = \prod_{t=0_i+1}^{T_i} \left[\int_{-\infty}^{-(A_{1it}+a_{1i})} \int_{-\infty}^{\infty} h_3(\varepsilon_{1it}, \varepsilon_{2it}, y_{3it}) d\varepsilon_{1it} d\varepsilon_{2it}\right]^{1-y_{1it}} \left[\int_{-(A_{1it}+a_{1i})}^{\infty} h_3(\varepsilon_{1it}, y_{2it}, y_{3it}) d\varepsilon_{1it}\right]^{y_{1it}}, \tag{3.8}$$

where h_3 denotes the density function of the trivariate normal distribution and A_{1it} is defined as

$$A_{1it} \equiv \vartheta_{11}y_{1i,t-1} + \vartheta_{13}y_{3i,t-1} + \boldsymbol{\beta}_1'\mathbf{x}_{1it} + b_{10} + \boldsymbol{b}_{11}'\boldsymbol{y}_{1i0_i} + \boldsymbol{b}_{12}'\mathbf{x}_{1i}. \tag{3.9}$$

The first product in equation (3.8) represents the contribution of a non-product innovator to the likelihood function and can be rewritten as $h_1(y_{3it})\int_{-\infty}^{-(A_{1it}+a_{1i})} h_1(\varepsilon_{1it}|y_{3it})d\varepsilon_{1it}$.[8] The second product represents the contribution of a product innovator and is equal to $h_1(\,y_{2it}\,|\,y_{3it}\,)\,h_1(\,y_{3it}\,)$ $\int_{-(A_{1it}+a_{1i})}^{\infty} h_1(\varepsilon_{1it}|y_{2it}, y_{3it})d\varepsilon_{1it}$. These single integrals are univariate cumulative distribution functions (CDFs) of the normal distribution, and are shown to be respectively (see Kotz et al.,

[8] $\int_{-\infty}^{\infty} h_3(\epsilon_{1it}, \epsilon_{2it}, y_{3it})d\epsilon_{2it} = h_2(\epsilon_{1it}, y_{3it})$ where h_2 denotes the density of the bivariate normal distribution, and $h_2(\epsilon_{1it}, y_{3it}) = h_1(y_{3it})h_1(\epsilon_{1it}|y_{3it})$ where h_1 denotes the density of the univariate normal distribution.

2000)

$$\Phi_1\left(\frac{-A_{1it}-a_{1i}-\underline{\rho_{\varepsilon_1\varepsilon_3}}\;\underline{\sigma_{\varepsilon_3}^{-1}}(y_{3it}-A_{3it}-\gamma_1A_{1it}-\underline{a_{3i}})}{\sqrt{1-\underline{\rho^2_{\varepsilon_1\varepsilon_3}}}}\right), \tag{3.10a}$$

$$\Phi_1\left(\frac{A_{1it}+a_{1i}+\underline{\rho_{12.3}}\sigma_{\varepsilon_2}^{-1}\,(y_{2it}-A_{2it}-a_{2i})+\underline{\rho_{13.2}}\;\underline{\sigma_{\varepsilon_3}^{-1}}(y_{3it}-A_{3it}-\gamma_1A_{1it}-\underline{a_{3i}})}{\sqrt{1-\underline{R^2_{1.23}}}}\right), \tag{3.10b}$$

where A_{1it} is given in equation (3.9), A_{2it} and A_{3it} are given by

$$A_{2it}\equiv\vartheta_{22}y_{2i,t-1}+\vartheta_{23}y_{3i,t-1}+\boldsymbol{\beta}_2'\mathbf{x}_{2it}+b_{20}+\boldsymbol{b}_{21}'\boldsymbol{y}_{2i0_i}+\boldsymbol{b}_{22}'\mathbf{x}_{2i}, \tag{3.11a}$$

$$A_{3it}\equiv\vartheta_{33}y_{3i,t-1}+\boldsymbol{\beta}_3'\mathbf{x}_{3it}+b_{30}+\boldsymbol{b}_{31}'\boldsymbol{y}_{3i0_i}+\boldsymbol{b}_{32}'\mathbf{x}_{3i}, \tag{3.11b}$$

and $\underline{\rho_{12.3}}$, $\underline{\rho_{13.2}}$, and $\underline{R^2_{1.23}}$ are given by

$$\underline{\rho_{12.3}}\equiv\frac{\rho_{\varepsilon_1\varepsilon_2}-\underline{\rho_{\varepsilon_1\varepsilon_3}}\;\underline{\rho_{\varepsilon_2\varepsilon_3}}}{1-\underline{\rho^2_{\varepsilon_2\varepsilon_3}}},\quad \underline{\rho_{13.2}}\equiv\frac{\rho_{\varepsilon_1\varepsilon_3}-\rho_{\varepsilon_1\varepsilon_2}\;\underline{\rho_{\varepsilon_2\varepsilon_3}}}{1-\underline{\rho^2_{\varepsilon_2\varepsilon_3}}},\quad \underline{R^2_{1.23}}\equiv\frac{\rho^2_{\varepsilon_1\varepsilon_2}+\underline{\rho^2_{\varepsilon_1\varepsilon_3}}-2\rho_{\varepsilon_1\varepsilon_2}\;\underline{\rho_{\varepsilon_1\varepsilon_3}}\;\underline{\rho_{\varepsilon_2\varepsilon_3}}}{1-\underline{\rho^2_{\varepsilon_2\varepsilon_3}}}. \tag{3.12}$$

The final expression of $l_{1i|\underline{\boldsymbol{a}_i}}$ is given by

$$\begin{aligned}
l_{1i|\underline{\boldsymbol{a}_i}}=&\prod_{t=0_i+1}^{T_i}\frac{1}{\underline{\sigma_{\varepsilon_3}}}\phi_1\left(\frac{y_{3it}-A_{3it}-\gamma_1A_{1it}-\underline{a_{3i}}}{\underline{\sigma_{\varepsilon_3}}}\right)\left[\Phi_1\left(\frac{-A_{1it}-a_{1i}-\underline{\rho_{\varepsilon_1\varepsilon_3}}\;\underline{\sigma_{\varepsilon_3}^{-1}}(y_{3it}-A_{3it}-\gamma_1A_{1it}-\underline{a_{3i}})}{\sqrt{1-\underline{\rho^2_{\varepsilon_1\varepsilon_3}}}}\right)\right]^{1-y_{1it}}\\
&\left[\Phi_1\left(\frac{A_{1it}+a_{1i}+\underline{\rho_{12.3}}\sigma_{\varepsilon_2}^{-1}\,(y_{2it}-A_{2it}-a_{2i})+\underline{\rho_{13.2}}\;\underline{\sigma_{\varepsilon_3}^{-1}}(y_{3it}-A_{3it}-\gamma_1A_{1it}-\underline{a_{3i}})}{\sqrt{1-\underline{R^2_{1.23}}}}\right)\right.\\
&\left.\frac{1}{\sigma_{\varepsilon_2}\sqrt{1-\underline{\rho^2_{\varepsilon_2\varepsilon_3}}}}\phi_1\left(\frac{y_{2it}-A_{2it}-a_{2i}-\frac{\underline{\rho_{\varepsilon_2\varepsilon_3}}\sigma_{\varepsilon_2}}{\underline{\sigma_{\varepsilon_3}}}(y_{3it}-A_{3it}-\gamma_1A_{1it}-\underline{a_{3i}})}{\sigma_{\varepsilon_2}\sqrt{1-\underline{\rho^2_{\varepsilon_2\varepsilon_3}}}}\right)\right]^{y_{1it}}.
\end{aligned} \tag{3.13}$$

Model with observed innovation incidence

The model with the observed innovation indicator as a predictor of labor productivity consists of equations (2.1)-(2.4) and (2.5b) with $j=1$ in equation (2.5b). Unlike in the previous model, we insert directly the observed innovation indicator in the likelihood function.[9]

[9]As a matter of fact, adopting this approach is recommended in this case. Indeed, the indicator function that relates the observed dependent variable, y_{1it}, to the regressors, which would be used in the likelihood function of the reduced form, is discontinuous. Thus, the maximization of the likelihood function of the reduced form is unfeasible in this case.

The individual likelihood function of the structural form of this model, conditional on $\boldsymbol{a}_i$ and denoted by $l_{2i|\boldsymbol{a}_i}$, has a similar expression to $l_{1i|\underline{\boldsymbol{a}_i}}$. It is given by

$$
\begin{aligned}
l_{2i|\boldsymbol{a}_i} = \prod_{t=0_i+1}^{T_i} \frac{1}{\sigma_{\varepsilon_3}} \phi_1\Bigg(\frac{y_{3it}-A_{3it}-\gamma_1 y_{1it}-a_{3i}}{\sigma_{\varepsilon_3}}\Bigg)\Bigg[\Phi_1\Bigg(\frac{-A_{1it}-a_{1i}-\rho_{\varepsilon_1\varepsilon_3}\sigma_{\varepsilon_3}^{-1}(y_{3it}-A_{3it}-\gamma_1 y_{1it}-a_{3i})}{\sqrt{1-\rho^2_{\varepsilon_1\varepsilon_3}}}\Bigg)\Bigg]^{1-y_{1it}} \\
\Bigg[\Phi_1\Bigg(\frac{A_{1it}+a_{1i}+\rho_{12.3}\sigma_{\varepsilon_2}^{-1}(y_{2it}-A_{2it}-a_{2i})+\rho_{13.2}\ \sigma_{\varepsilon_3}^{-1}(y_{3it}-A_{3it}-\gamma_1 y_{1it}-a_{3i})}{\sqrt{1-R^2_{1.23}}}\Bigg) \\
\frac{1}{\sigma_{\varepsilon_2}\sqrt{1-\rho^2_{\varepsilon_2\varepsilon_3}}}\phi_1\left(\frac{y_{2it}-A_{2it}-a_{2i}-\frac{\rho_{\varepsilon_2\varepsilon_3}\sigma_{\varepsilon_2}}{\sigma_{\varepsilon_3}}(y_{3it}-A_{3it}-\gamma_1 y_{1it}-a_{3i})}{\sigma_{\varepsilon_2}\sqrt{1-\rho^2_{\varepsilon_2\varepsilon_3}}}\right)\Bigg]^{y_{1it}},
\end{aligned}
\tag{3.14}
$$

where $\rho_{12.3}$, $\rho_{13.2}$, and $R^2_{1.23}$, the structural form counterparts of $\underline{\rho_{12.3}}$, $\underline{\rho_{13.2}}$, and $\underline{R^2_{1.23}}$, are derived straightforwardly from equation (3.12) by replacing the underlined correlations by their structural form counterparts.

3.3 Numerical evaluation

The next step consists in obtaining the unconditional counterparts to $l_{1i|\underline{\boldsymbol{a}_i}}$ and $l_{2i|\boldsymbol{a}_i}$, which are obtained by integrating out respectively $\underline{\boldsymbol{a}_i}$ and $\boldsymbol{a}_i$ with respect to their normal distribution. Formally,

$$
l_1 = \prod_{i=1}^{N} \int_{a_{1i}} \int_{a_{2i}} \int_{\underline{a_{3i}}} l_{1i|\underline{\boldsymbol{a}_i}} h_3(a_{1i}, a_{2i}, \underline{a_{3i}}|...) da_{1i} da_{2i} d\underline{a_{3i}}, \tag{3.15}
$$

and

$$
l_2 = \prod_{i=1}^{N} \int_{a_{1i}} \int_{a_{2i}} \int_{a_{3i}} l_{2i|\boldsymbol{a}_i} h_3(a_{1i}, a_{2i}, a_{3i}|...) da_{1i} da_{2i} da_{3i}. \tag{3.16}
$$

Evidently, l_1 and l_2 cannot be derived analytically. Hence, we use Gauss-Hermite quadrature sequentially, along the lines of Raymond (2007), to evaluate the triple integrals.[10] The Gauss-Hermite quadrature states that

$$
\int_{-\infty}^{\infty} e^{-r^2} f(r) dr \simeq \sum_{m=1}^{M} w_m f(a_m), \tag{3.17}
$$

where w_m and a_m are respectively the weights and abscissae of the quadrature with M being the total number of integration points.[11] Numerical tables with values of w_m and a_m are formulated in mathematical textbooks (Abramowitz and Stegun, 1964). The larger M, the more accurate the approximation. Using the results of Appendix A, the unconditional likelihood, l_1, is derived as

[10] The use of this numerical method is well documented in the econometric literature in the context of panel data single-equation models (see e.g. Butler and Moffitt, 1982; Rabe-Hesketh et al., 2005). However, its use in the context of panel data models with multiple equations remains to date limited. A few exceptions are Raymond (2007, chapter 3) who studies the performance of the method in two types of dynamic sample selection models, and Raymond et al. (2010) who apply the method to estimate the persistence of innovation incidence and innovation intensity.

[11] The abscissae of the quadrature, a_m, should not be confused with the projections errors a_{1i}, a_{2i} and a_{3i}.

$$
\begin{aligned}
l_1 \simeq & \prod_{i=1}^{N} \underline{\Delta} \pi^{\frac{-3}{2}} \left[(1-\rho_{a_1 a_2}^2)(1-\underline{\rho_{a_1 a_3}^2})(1-\underline{\rho_{a_1 a_2}^2}) \right]^{\frac{-1}{2}} \sum_{m_3=1}^{M_3} w_{m_3} \prod_{t=0_i+1}^{T_i} \frac{1}{\underline{\sigma_{\varepsilon_3}}} \phi_1 \left(\frac{y_{3it} - A_{3it} - \gamma_1 A_{1it} - a_{m_3}[...]}{\underline{\sigma_{\varepsilon_3}}} \right) \\
& \sum_{m_2=1}^{M_2} w_{m_2} e^{\frac{-2\underline{\Lambda_{23}} a_{m_2} a_{m_3}}{\sqrt{\underline{\Lambda_{22}}\,\underline{\Lambda_{33}}}}} \prod_{t=0_i+1}^{T_i} \left[\frac{1}{\sigma_{\varepsilon_2} \sqrt{1-\underline{\rho_{\varepsilon_2 \varepsilon_3}^2}}} \phi_1 \left(\frac{y_{2it} - A_{2it} - a_{m_2}[...] - \frac{\underline{\underline{\rho_{\varepsilon_2 \varepsilon_3}}\sigma_{\varepsilon_2}}}{\underline{\sigma_{\varepsilon_3}}} (y_{3it} - A_{3it} - \gamma_1 A_{1it} - a_{m_3}[...])}{\sigma_{\varepsilon_2} \sqrt{1-\underline{\rho_{\varepsilon_2 \varepsilon_3}^2}}} \right) \right]^{y_{1it}} \\
& \sum_{m_1=1}^{M_1} w_{m_1} e^{\frac{-2 a_{m_1}}{\sqrt{\underline{\Lambda_{11}}}} \left(\frac{a_{m_2} \underline{\underline{\Lambda_{12}}}}{\sqrt{\underline{\underline{\Lambda_{22}}}}} + \frac{a_{m_3} \underline{\underline{\Lambda_{13}}}}{\sqrt{\underline{\underline{\Lambda_{33}}}}} \right)} \prod_{t=0_i+1}^{T_i} \left[\Phi_1 \left(\frac{-A_{1it} - a_{m_1}[...] - \underline{\rho_{\varepsilon_1 \varepsilon_3}}\ \underline{\sigma_{\varepsilon_3}^{-1}} (y_{3it} - A_{3it} - \gamma_1 A_{1it} - a_{m_3}[...])}{\sqrt{1-\underline{\rho_{\varepsilon_1 \varepsilon_3}^2}}} \right) \right]^{1-y_{1it}} \\
& \left[\Phi_1 \left(\frac{A_{1it} + a_{m_1}[...] + \underline{\rho_{12.3}} \sigma_{\varepsilon_2}^{-1} (y_{2it} - A_{2it} - a_{m_2}[...]) + \underline{\rho_{13.2}}\ \underline{\sigma_{\varepsilon_3}^{-1}} (y_{3it} - A_{3it} - \gamma_1 A_{1it} - a_{m_3}[...])}{\sqrt{1-\underline{R_{1.23}^2}}} \right) \right]^{y_{1it}},
\end{aligned}
\tag{3.18}
$$

where w_{m_k}, a_{m_k} and M_k $(k = 1, 2, 3)$ are respectively the weights, abscissae and total number of points of the quadrature in each stage, $a_{m_k}[...] = \frac{a_{m_k} \sigma_{a_k} \sqrt{2}}{\sqrt{\Lambda_{kk}}}$, and the expressions of Λ_{kl} $(k, l = 1, 2, 3;$ $\Lambda_{kl} = \Lambda_{lk})$ and $\underline{\Delta}$ are given in Appendix A. The FIML estimates of the structural parameters of the model where y_{1it}^* enters the APF are obtained by maximizing $\ln l_1$ subject to the constraints defined in equations (3.7a)-(3.7c).

The evaluation of l_2 is done in a similar fashion and yields a similar expression except that the underlined parameters are replaced by their non-underlined equivalents and that $\gamma_1 A_{1it}$ is replaced by $\gamma_1 y_{1it}$. The FIML estimates of the structural parameters of the model where y_{1it} enters the APF are obtained by maximizing $\ln l_2$ without additional constraints.

4 Data

The data used in the analysis stem from three waves of the Dutch and the French CIS pertaining to the manufacturing sector, with the exception of the food industry, for the periods 1994-1996 (CIS 2), 1998-2000 (CIS 3) and 2002-2004 (CIS 4). The Dutch and French CIS data are merged respectively with data from the Production Survey (PS) and the *'Enquête Annuelle d'Entreprise'* (EAE) that provide information regarding employment, sales and investment in tangible goods of the enterprise.[12] For each CIS, the merged PS and EAE variables pertain to the last year of the three-year period. We consider enterprises with at least ten employees and positive sales at the end of each period covered by the innovation survey. Only enterprises with a share of total R&D expenditures (intramural + extramural) in total sales smaller than or equal to 50% are included in the analysis for we believe those with total R&D expenditures greater than 50% are of a different

[12] Both Dutch surveys were carried out by the *'Centraal Bureau voor de Statistiek'* (CBS) for the whole manufacturing sector and the two French surveys by the *'Service des Statistiques Industrielles'* (SESSI) of the French Ministry of Industry for the manufacturing sector excluding the food industry.

nature. Note that one of the particularities of the innovation survey is that, for each period, product innovation incidence pertains to any of the three-year period, while the actual share of innovative sales pertains to the last year of the period. As a result, when assessing persistence of innovation, the lag effect refers to two to four years when the incidence of innovation is considered and to four years when the share of innovative sales is considered.[13]

In the following Tables 1, 2 and 3, we show some simple descriptive statistics, mostly means, to present our samples and main variables. However, it is important to keep in mind that our estimates are not based on the differences of most of such means but on differences within the two countries, four industry categories as well as within wave survey patterns, since we estimate our models separately for France and The Netherlands and we control for firm effects and also include industry and time dummies in all the model equations, as explained in Section 2. Table 1 shows, for both countries, the patterns of enterprises' presence in the unbalanced panel after data cleaning. Because of the dynamic structure of the model, an enterprise must be present in at least two consecutive waves of the merged data in order to be included in the analysis. There are 1920 such enterprises in our sample for France and 1228 for The Netherlands. In both countries about one third of the total number of enterprises are present in the three waves. In France about 25% and 40% of them respectively are present in the first and last two waves, while the proportions are the opposite in The Netherlands.

For each pattern, we report the mean and median employment head counts in the sample and in the population where the head counts in the population are obtained by weighting head counts in the sample using a raising factor obtained after correcting for non-response. Because of the lower cut-off points in The Netherlands and possibly differences in the rates of non-responses in the two countries, the differences in average firm size between the sample and the population are larger for France than for The Netherlands. These differences are, however, smaller in the balanced panel, which is to be expected since firms larger than the cut-off points are all included in the samples and are also more likely to survive during the whole period 1994-2004 (Agarwal and Audretsch, 2001). Using the unbalanced panel allows us to obtain more precise estimates as more observations for broader types of enterprises are used and also to control partly for survivorship biases as enterprises are allowed to enter and exit the sample at any period. Overall French enterprises are significantly

[13]The CIS, PS and EAE data are collected at the enterprise level. A combination of a census and a stratified random sampling is used for each wave of the CIS and the PS. A census is used for the population of Dutch enterprises with at least 50 employees, and a stratified random sampling is used for enterprises with less than 50 employees, where the stratum variables are the enterprise economic activity and employment in head counts. The same cut-off point of 50 employees is applied to each wave of the Dutch CIS and the PS. A similar scheme is used for the French CIS except that a cut-off point of 500 employees is used in CIS 2 and 3, and a cut-off point of 250 employees is used in CIS 4. The use of different cut-off points in the census/sampling scheme may result in differences across countries in the size of enterprises in our sample. As for the EAE, all enterprises with at least 20 employees are surveyed.

larger than the Dutch ones both in the sample and in the population, in the balanced as well as in the unbalanced panel.

The last four rows of Table 1 also reveal some interesting differences regarding innovation and labor productivity performance across countries and, for a given country, between the balanced and the unbalanced panel. In both countries, the balanced panel consists of a significantly larger proportion of product innovators. In the case of The Netherlands, the enterprises in the balanced panel are significantly more productive, in terms of sales per employee. The share of innovative sales is significantly larger on average for the Dutch than for the French enterprises which are, however, significantly more productive than the Dutch enterprises. The average amount of R&D per employee of continuous R&D performers is significantly larger for the French than for the Dutch enterprises.

Table 1: Employment, R&D per employee, innovation and sales per employee in each pattern of the unbalanced panel data for Dutch and French manufacturing: CIS 2, CIS 3, and CIS 4

Variable↓	France				The Netherlands			
Pattern→	110	111	011	Total	110	111	011	Total
# enterprises	504	586	829	1920	506	411	311	1228
% in total	26	31	43	100	41	34	25	100
Employment, sample								
Mean	558	1044	398	691	158	217	335	222
Median	160	663	197	336	75	112	108	96
Employment, population								
Mean	215	726	200	334	111	172	197	155
Median	73	405	74	97	56	93	57	68
Mean variables of interest								
R&D/employee†	6.68	7.79	8.51	7.80	4.22	4.96	4.25	4.57
Product innovator	0.51	0.72	0.50	0.59	0.60	0.62	0.49	0.58
Share of innov. sales‡	0.23	0.23	0.19	0.22	0.32	0.30	0.27	0.30
Sales/employee*	215	210	235	221	150	224	164	184

Patterns refer to presence/absence of firms in the three successive waves. †For continuous R&D performers, in 1000s of euros. ‡For product innovators. *In 1000s of euros.

Table 2 gives the means of the non-transformed dependent and explanatory variables for the unbalanced samples and for the subsamples of product innovators. Comparing first all firms with the innovating firms, it appears that in both countries product innovators do not seem to be more productive on average. However, this corresponds to the existence among non-product innovators of a few firms with very high values of sales per employee, and actually the means of the logarithm of sales per employee variable, which we use in our equations, downweigh outlier values and are significantly higher for product innovators in both countries. We also observe that on average in both countries product innovators are larger in terms of employment and have a larger market share.

Comparing now the two countries, we see that Dutch enterprises, either overall or for product innovators only, have on average much higher physical investments per employee, larger market shares but smaller sales per employee than their French counterparts. We also see that the Dutch innovators have on average a significantly higher share of innovative sales but a significantly lower mean R&D per employee than French innovators. It is finally interesting to note that in France the majority of product innovators and non-innovators are very large in size in contrast to The Netherlands where they are mostly medium-sized enterprises.

Table 2: Means of dependent and explanatory variables: Unbalanced panel from Dutch and French CIS 2, CIS 3 and CIS 4

Variable	France		The Netherlands	
	All enterprises	Product. innov.	All enterprises	Product. innov.
Product innovator	0.59	-	0.58	-
Share of innov. sales	-	0.22	-	0.30
Sales/employee†	220.53	215.04	184.85	180.70
R&D/employee‡	-	8.09	-	4.68
Investment/employee†	7.29	7.36	9.10	9.62
Employment				
# employees	691.36	935.72	222.13	258.78
Size class 1	0.13	0.06	0.20	0.14
Size class 2	0.29	0.21	0.63	0.65
Size class 3	0.20	0.22	0.08	0.11
Size class 4	0.39	0.51	0.09	0.10
Market share (%)	1.52	1.86	1.66	2.01
# observations	4427	2618	2867	1670

†In 1000s of euros. ‡For continuous R&D performers, in 1000s of euros.

Table 3 gives the same statistics as Table 2 but separately for each period of the unbalanced panel. For both countries, we observe a significant decrease in the proportion of product innovators and in the mean share of innovative sales between 1994 and 2004. The marked increase of the share of innovative sales between the last two periods for France is not large enough to offset the large decrease that occurs between the first two periods. We also see, for both countries, a strong increase in the mean nominal sales per employee and in the mean market share between 1994 and 2004, while on average employment decreases for France and increases for The Netherlands. The growth that we observe in the mean R&D and physical investment per employee between 1994 and 2004 is relatively modest (and only statistically significant for The Netherlands).

5 Results

We now turn to the results of the estimation of the models. We shall first quickly comment on the general results before discussing the estimated effects of innovation output on productivity and especially what are the revealed dynamics of, and between, innovation and productivity. Tables 4 and 5 present the estimation results for the model with latent innovation as a predictor of labor

Table 3: Means of dependent and explanatory variables for each CIS of the unbalanced panel, France and The Netherlands

Variable	France			The Netherlands		
	1994-1996	1998-2000	2002-2004	1994-1996	1998-2000	2002-2004
Product innovator	0.65	0.56	0.58	0.66	0.60	0.45
Share of innov. sales†	0.30	0.15	0.23	0.33	0.32	0.21
Sales/employee‡	158.51	230.52	254.75	149.43	170.11	254.91
R&D/employee*	7.59	7.28	8.64	3.96	4.76	5.14
Investment/employee‡	6.44	8.12	6.83	8.58	8.01	11.61
Employment						
# employees	806.37	673.99	626.33	175.36	236.04	257.87
Size class 1	0.12	0.14	0.12	0.18	0.21	0.20
Size class 2	0.26	0.31	0.28	0.68	0.61	0.60
Size class 3	0.15	0.19	0.25	0.08	0.08	0.09
Size class 4	0.47	0.36	0.35	0.06	0.10	0.11
Market share (%)	1.38	1.45	1.73	1.30	1.75	1.97
# enterprises	1091	1920	1416	917	1228	722

†For product innovators. ‡In 1000s of euros. *For continuous R&D performers, in 1000s of euros.

productivity, and Tables 6 and 7 present the results for the model with observed innovation as a predictor of labor productivity.

5.1 The effects of size, market share, R&D and the error terms

It is first of all remarkable and comforting to notice that the results are quite consistent and robust across models. Tables 4 to 7 show that larger French manufacturing firms are more likely to be product innovators while no such evidence is found for Dutch manufacturing. Since the reference size category consists of firms with more than 500 employees, the negative signs for the other categories indicate a positive size effect. Size does not play a big role in explaining differences in innovation intensities: only for France is there a sign of a difference attributable to size between the enterprises with 50 to 250 and those with above 500 employees. For both countries, however, market share plays a positive and significant role not only in the innovation incidence but also in the share of innovative sales: a 10% increase in market share increases the probability to innovate by 0.002 and the share of innovative sales by 1 to 2%. In both countries, enterprises that undertook R&D activities continuously during the previous two to four years are more likely to be product innovators and attain a larger share of innovative sales. Moreover, past R&D intensity increases both the innovation propensity (and the probability of becoming a product innovator) and intensity of the average Dutch enterprise. The estimates show that a 1% increase in R&D intensity corresponds four years later to an increase in the probability of being a product innovator of about 0.05, and to an increase in the share of innovative sales (of product innovators) of about

Table 4: FIML estimates of the model with latent innovation propensity to explain productivity: Unbalanced panel from Dutch and French CIS 2, CIS 3 and CIS 4[‡]

Variable	Coefficient	(Std. Err.)	Coefficient	(Std. Err.)
	France		The Netherlands	
		Innovation incidence$_t$		
Innovation incidence$_{t-1}$	0.181†	(0.103)	0.051	(0.171)
Innovation incidence$_{0_i}$, 3 waves	0.167	(0.111)	0.696**	(0.189)
Innovation incidence$_{0_i}$, 2 waves	0.339**	(0.066)	0.721**	(0.112)
(Sales/employee)$_{t-1}$, in log	0.000	(0.017)	0.010	(0.025)
(R&D/employee)$_{t-1}$, in log	0.025	(0.027)	0.173**	(0.045)
$(\mathrm{D}_{\text{non-continuous R\&D}})_{t-1}$	-0.534**	(0.070)	-0.587**	(0.093)
Size class				
$\mathrm{D}_{\#\text{ employees}\leq 50}$	-0.470**	(0.130)	-0.271	(0.170)
$\mathrm{D}_{50<\#\text{ employees}\leq 250}$	-0.335**	(0.092)	-0.049	(0.144)
$\mathrm{D}_{250<\#\text{ employees}\leq 500}$	-0.095	(0.077)	0.304†	(0.182)
Market share$_{t-1}$, in log	0.074**	(0.024)	0.077**	(0.029)
		Share of innovative sales$_t$, in logit		
Share of innov. sales$_{t-1}$, in logit	0.110*	(0.050)	0.043	(0.044)
Share of innov. sales$_{0_i}$, 3 waves	0.064	(0.046)	0.132**	(0.040)
Share of innov. sales$_{0_i}$, 2 waves	0.185**	(0.027)	0.174**	(0.030)
(Sales/employee)$_{t-1}$, in log	0.002	(0.034)	0.001	(0.029)
(R&D/employee)$_{t-1}$, in log	0.105†	(0.057)	0.258**	(0.065)
$(\mathrm{D}_{\text{non-continuous R\&D}})_{t-1}$	-1.093**	(0.154)	-0.697**	(0.144)
Size class				
$\mathrm{D}_{\#\text{ employees}\leq 50}$	-0.349	(0.311)	0.022	(0.276)
$\mathrm{D}_{50<\#\text{ employees}\leq 250}$	-0.361†	(0.206)	0.065	(0.222)
$\mathrm{D}_{250<\#\text{ employees}\leq 500}$	0.118	(0.169)	0.156	(0.268)
Market share$_{t-1}$, in log	0.152**	(0.055)	0.093*	(0.047)
		Labor productivity$_t$: sales/employee, in log		
(Sales/employee)$_{t-1}$, in log	0.531**	(0.056)	0.319**	(0.066)
(Sales/employee)$_{0_i}$, 3 waves	0.336**	(0.056)	0.282**	(0.066)
(Sales/employee)$_{0_i}$, 2 waves	0.856**	(0.012)	0.584**	(0.024)
Latent innovation propensity$_t$	0.074**	(0.020)	0.121**	(0.029)
(Investment/employee)$_t$, in log	0.065**	(0.006)	0.119**	(0.012)
Employment$_t$, in log	-0.027**	(0.008)	-0.082**	(0.018)
		Covariance matrix		
Individual effects				
σ_{a_1}	0.259†	(0.148)	0.470**	(0.138)
σ_{a_2}	0.745**	(0.179)	0.680**	(0.193)
σ_{a_3}	0.096**	(0.025)	0.160**	(0.060)
$\rho_{a_1 a_2}$	0.514**	(0.152)	0.540**	(0.145)
$\rho_{a_1 a_3}$	-0.090	(0.408)	-0.221	(0.336)
$\rho_{a_2 a_3}$	0.158	(0.279)	0.064	(0.331)
Idiosyncratic errors				
σ_{ε_2}	2.469**	(0.076)	1.780**	(0.096)
σ_{ε_3}	0.313**	(0.009)	0.587**	(0.019)
$\rho_{\varepsilon_1 \varepsilon_2}$	0.99, fixed after grid search		0.95, fixed after grid search	
$\rho_{\varepsilon_1 \varepsilon_3}$	-0.206**	(0.071)	-0.237**	(0.067)
$\rho_{\varepsilon_2 \varepsilon_3}$	-0.208**	(0.070)	-0.229**	(0.061)
# observations	2505		1639	
Log-likelihood	-5048.917		-3920.898	

[‡]Three dummies of category of industry, a time dummy and an intercept are included in each equation.
Significance levels : $\dagger$: 10% $*$: 5% $**$: 1%

0.2%.[14] The estimates are lower for France showing that a 1% increase in R&D intensity does

[14]The marginal effect of a regressor on the probability of the average firm to become a product innovator is obtained in the usual way, (see Greene, 2011, page 689). Furthermore, because of the logit transformation of the share of innovative sales, the R&D elasticity is obtained by multiplying the coefficient of R&D by $1 - y_{2t}$ where y_{2t}

Table 5: FIML estimates of the model with latent innovation intensity to explain productivity: Unbalanced panel from Dutch and French CIS 2, CIS 3 and CIS 4[‡]

Variable	Coefficient	(Std. Err.)	Coefficient	(Std. Err.)
	France		The Netherlands	
	Innovation incidence$_t$			
Innovation incidence$_{t-1}$	0.151	(0.098)	0.027	(0.178)
Innovation incidence$_{0_i}$, 3 waves	0.171†	(0.102)	0.738**	(0.194)
Innovation incidence$_{0_i}$, 2 waves	0.343**	(0.066)	0.719**	(0.116)
(Sales/employee)$_{t-1}$, in log	0.004	(0.018)	0.009	(0.026)
(R&D/employee)$_{t-1}$, in log	0.023	(0.028)	0.173**	(0.047)
$(\mathrm{D}_{\text{non-continuous R\&D}})_{t-1}$	-0.546**	(0.070)	-0.607**	(0.095)
Size class				
$\mathrm{D}_{\#\ \text{employees}\leq 50}$	-0.509**	(0.126)	-0.325†	(0.175)
$\mathrm{D}_{50<\#\ \text{employees}\leq 250}$	-0.353**	(0.091)	-0.101	(0.150)
$\mathrm{D}_{250<\#\ \text{employees}\leq 500}$	-0.111	(0.077)	0.225	(0.187)
Market share$_{t-1}$, in log	0.069**	(0.024)	0.072*	(0.030)
	Share of innovative sales$_t$, in logit			
Share of innov. sales$_{t-1}$, in logit	0.138**	(0.048)	0.055	(0.043)
Share of innov. sales$_{0_i}$, 3 waves	0.044	(0.043)	0.113**	(0.040)
Share of innov. sales$_{0_i}$, 2 waves	0.191**	(0.026)	0.169**	(0.030)
(Sales/employee)$_{t-1}$, in log	0.004	(0.033)	-0.001	(0.029)
(R&D/employee)$_{t-1}$, in log	0.100†	(0.055)	0.261**	(0.062)
$(\mathrm{D}_{\text{non-continuous R\&D}})_{t-1}$	-1.018**	(0.154)	-0.691**	(0.140)
Size class				
$\mathrm{D}_{\#\ \text{employees}\leq 50}$	-0.187	(0.296)	0.096	(0.271)
$\mathrm{D}_{50<\#\ \text{employees}\leq 250}$	-0.366†	(0.198)	0.169	(0.219)
$\mathrm{D}_{250<\#\ \text{employees}\leq 500}$	0.103	(0.162)	0.323	(0.266)
Market share$_{t-1}$, in log	0.179**	(0.052)	0.110*	(0.046)
	Labor productivity$_t$: sales/employee, in log			
(Sales/employee)$_{t-1}$, in log	0.527**	(0.056)	0.320**	(0.066)
(Sales/employee)$_{0_i}$, 3 waves	0.337**	(0.056)	0.282**	(0.066)
(Sales/employee)$_{0_i}$, 2 waves	0.852**	(0.012)	0.583**	(0.024)
Latent share of innovative sales$_t$	0.043**	(0.010)	0.084**	(0.022)
(Investment/employee)$_t$, in log	0.065**	(0.006)	0.120**	(0.012)
Employment$_t$, in log	-0.025**	(0.008)	-0.070**	(0.017)
	Covariance matrix			
Individual effects				
σ_{a_1}	0.322**	(0.092)	0.492**	(0.134)
σ_{a_2}	0.673**	(0.190)	0.642**	(0.213)
σ_{a_3}	0.095**	(0.026)	0.158**	(0.060)
$\rho_{a_1 a_2}$	0.546**	(0.128)	0.544**	(0.146)
$\rho_{a_1 a_3}$	-0.094	(0.315)	-0.121	(0.350)
$\rho_{a_2 a_3}$	-0.108	(0.302)	-0.151	(0.362)
Idiosyncratic errors				
σ_{ε_2}	2.481**	(0.075)	1.791**	(0.098)
σ_{ε_3}	0.320**	(0.011)	0.596**	(0.021)
$\rho_{\varepsilon_1\varepsilon_2}$	0.99, fixed after grid search		0.95, fixed after grid search	
$\rho_{\varepsilon_1\varepsilon_3}$	-0.298**	(0.077)	-0.266**	(0.074)
$\rho_{\varepsilon_2\varepsilon_3}$	-0.301**	(0.075)	-0.288**	(0.073)
# observations	2505		1639	
Log-likelihood	-5045.613		-3922.326	

[‡]Three dummies of category of industry, a time dummy and an intercept are included in each equation.
Significance levels : † : 10% ∗ : 5% ∗∗ : 1%

not significantly affect the probability to innovate four years later, although it corresponds to an increase in the share of innovative sales of about 0.1% (and significant at the 10% level). Overall

denotes the share of innovative sales in level (see Appendix D).

Table 6: FIML estimates of the model with observed innovation indicator to explain productivity: Unbalanced panel from Dutch and French CIS 2, CIS 3 and CIS 4‡

Variable	Coefficient	(Std. Err.)	Coefficient	(Std. Err.)
	France		The Netherlands	
	Innovation incidence$_t$			
Innovation incidence$_{t-1}$	0.155	(0.117)	0.010	(0.182)
Innovation incidence$_{0_i}$, 3 waves	0.174	(0.135)	0.767**	(0.200)
Innovation incidence$_{0_i}$, 2 waves	0.339**	(0.070)	0.736**	(0.116)
(Sales/employee)$_{t-1}$, in log	0.003	(0.017)	0.009	(0.026)
(R&D/employee)$_{t-1}$, in log	0.027	(0.030)	0.177**	(0.047)
$(D_{\text{non-continuous R\&D}})_{t-1}$	-0.567**	(0.071)	-0.621**	(0.096)
Size class				
$D_{\#\ \text{employees} \leq 50}$	-0.577**	(0.134)	-0.323†	(0.174)
$D_{50 < \#\ \text{employees} \leq 250}$	-0.356**	(0.102)	-0.115	(0.148)
$D_{250 < \#\ \text{employees} \leq 500}$	-0.102	(0.081)	0.198	(0.188)
Market share$_{t-1}$, in log	0.054*	(0.024)	0.068*	(0.030)
	Share of innovative sales$_t$, in logit			
Share of innov. sales$_{t-1}$, in logit	0.109*	(0.049)	0.036	(0.044)
Share of innov. sales$_{0_i}$, 3 waves	0.061	(0.045)	0.141**	(0.040)
Share of innov. sales$_{0_i}$, 2 waves	0.180**	(0.028)	0.173**	(0.030)
(Sales/employee)$_{t-1}$, in log	0.005	(0.034)	0.001	(0.029)
(R&D/employee)$_{t-1}$, in log	0.101†	(0.058)	0.256**	(0.065)
$(D_{\text{non-continuous R\&D}})_{t-1}$	-1.119**	(0.159)	-0.710**	(0.145)
Size class				
$D_{\#\ \text{employees} \leq 50}$	-0.502	(0.315)	-0.022	(0.277)
$D_{50 < \#\ \text{employees} \leq 250}$	-0.355†	(0.210)	0.005	(0.223)
$D_{250 < \#\ \text{employees} \leq 500}$	0.108	(0.171)	0.064	(0.269)
Market share$_{t-1}$, in log	0.131*	(0.055)	0.081†	(0.048)
	Labor productivity$_t$: sales/employee, in log			
(Sales/employee)$_{t-1}$, in log	0.532**	(0.056)	0.330**	(0.066)
(Sales/employee)$_{0_i}$, 3 waves	0.341**	(0.056)	0.282**	(0.066)
(Sales/employee)$_{0_i}$, 2 waves	0.861**	(0.012)	0.594**	(0.023)
Observed innovation indicator$_t$	0.056	(0.042)	0.197**	(0.059)
(Investment/employee)$_t$, in log	0.065**	(0.006)	0.121**	(0.012)
Employment$_t$, in log	-0.012	(0.007)	-0.064**	(0.016)
	Covariance matrix			
Individual effects				
σ_{a_1}	0.284	(0.177)	0.492**	(0.143)
σ_{a_2}	0.757**	(0.150)	0.710**	(0.185)
σ_{a_3}	0.094**	(0.025)	0.154**	(0.058)
$\rho_{a_1 a_2}$	0.535**	(0.170)	0.527**	(0.148)
$\rho_{a_1 a_3}$	0.242	(0.362)	0.055	(0.365)
$\rho_{a_2 a_3}$	0.271	(0.314)	0.224	(0.329)
Idiosyncratic errors				
σ_{ε_2}	2.463**	(0.083)	1.758**	(0.096)
σ_{ε_3}	0.308**	(0.008)	0.580**	(0.018)
$\rho_{\varepsilon_1 \varepsilon_2}$	0.99, fixed after grid search		0.95, fixed after grid search	
$\rho_{\varepsilon_1 \varepsilon_3}$	-0.071	(0.114)	-0.223**	(0.072)
$\rho_{\varepsilon_2 \varepsilon_3}$	-0.024	(0.043)	-0.128*	(0.052)
# observations	2505		1639	
Log-likelihood	-5054.238		-3925.706	

‡Three dummies of category of industry, a time dummy and an intercept are included in each equation.
Significance levels : † : 10% ∗ : 5% ∗∗ : 1%

our results support the existence of a lag effect when it comes to innovation.[15]

[15] To have a better appreciation of the time span between R&D investment and innovation success, we would of course need a longer and yearly panel allowing us to estimate a distributed lag model, along the lines for example of Pakes and Griliches (1980) and Hall et al. (1986) as regards R&D and patents.

Table 7: FIML estimates of the model with observed innovation intensity to explain productivity: Unbalanced panel from Dutch and French CIS 2, CIS 3 and CIS 4‡

Variable	Coefficient	(Std. Err.)	Coefficient	(Std. Err.)
	France		The Netherlands	
		Innovation incidence$_t$		
Innovation incidence$_{t-1}$	0.245**	(0.088)	0.014	(0.176)
Innovation incidence$_{0_i}$, 3 waves	0.114	(0.092)	0.760**	(0.194)
Innovation incidence$_{0_i}$, 2 waves	0.343**	(0.062)	0.735**	(0.115)
(Sales/employee)$_{t-1}$, in log	0.000	(0.016)	0.008	(0.026)
(R&D/employee)$_{t-1}$, in log	0.020	(0.026)	0.177**	(0.047)
$(D_{\text{non-continuous R\&D}})_{t-1}$	-0.433**	(0.063)	-0.608**	(0.095)
Size class				
$D_{\#\text{ employees}\leq 50}$	-0.514**	(0.115)	-0.319†	(0.173)
$D_{50<\#\text{ employees}\leq 250}$	-0.354**	(0.082)	-0.102	(0.148)
$D_{250<\#\text{ employees}\leq 500}$	-0.129†	(0.069)	0.220	(0.187)
Market share$_{t-1}$, in log	0.069**	(0.021)	0.071*	(0.030)
		Share of innovative sales$_t$, in logit		
Share of innov. sales$_{t-1}$, in logit	0.139**	(0.041)	0.044	(0.044)
Share of innov. sales$_{0_i}$, 3 waves	0.030	(0.036)	0.130**	(0.040)
Share of innov. sales$_{0_i}$, 2 waves	0.177**	(0.026)	0.172**	(0.030)
(Sales/employee)$_{t-1}$, in log	0.002	(0.031)	-0.001	(0.029)
(R&D/employee)$_{t-1}$, in log	0.076†	(0.045)	0.262**	(0.064)
$(D_{\text{non-continuous R\&D}})_{t-1}$	-0.708**	(0.127)	-0.690**	(0.144)
Size class				
$D_{\#\text{ employees}\leq 50}$	-0.213	(0.258)	0.029	(0.275)
$D_{50<\#\text{ employees}\leq 250}$	-0.389*	(0.170)	0.070	(0.221)
$D_{250<\#\text{ employees}\leq 500}$	0.001	(0.136)	0.178	(0.268)
Market share$_{t-1}$, in log	0.191**	(0.044)	0.098*	(0.047)
		Labor productivity$_t$: sales/employee, in log		
(Sales/employee)$_{t-1}$, in log	0.420**	(0.054)	0.324**	(0.067)
(Sales/employee)$_{0_i}$, 3 waves	0.427**	(0.053)	0.285**	(0.067)
(Sales/employee)$_{0_i}$, 2 waves	0.836**	(0.012)	0.590**	(0.023)
Observed share of innovative sales$_t$	0.107**	(0.009)	0.045**	(0.012)
(Investment/employee)$_t$, in log	0.066**	(0.006)	0.120**	(0.012)
Employment$_t$, in log	-0.037**	(0.007)	-0.071**	(0.017)
		Covariance matrix		
Individual effects				
σ_{a_1}	0.154	(0.106)	0.494**	(0.132)
σ_{a_2}	0.511**	(0.140)	0.649**	(0.207)
σ_{a_3}	0.133**	(0.019)	0.157**	(0.060)
$\rho_{a_1a_2}$	0.557**	(0.124)	0.550**	(0.139)
$\rho_{a_1a_3}$	0.011	(0.230)	-0.002	(0.361)
$\rho_{a_2a_3}$	0.117	(0.195)	0.120	(0.382)
Idiosyncratic errors				
σ_{ε_2}	2.459**	(0.065)	1.795**	(0.098)
σ_{ε_3}	0.357**	(0.012)	0.586**	(0.019)
$\rho_{\varepsilon_1\varepsilon_2}$	0.99, fixed after grid search		0.95, fixed after grid search	
$\rho_{\varepsilon_1\varepsilon_3}$	-0.570**	(0.042)	-0.274**	(0.076)
$\rho_{\varepsilon_2\varepsilon_3}$	-0.691**	(0.038)	-0.259**	(0.069)
# observations	2505		1639	
Log-likelihood	-5021.290		-3923.897	

‡Three dummies of category of industry, a time dummy and an intercept are included in each equation.
Significance levels : † : 10% $*$: 5% $**$: 1%

We observe a negative and statistically significant correlation, at time t, in both countries and in all specifications between the error terms in the innovation and the labor productivity equations.

This can be explained by the fact that, in order to innovate, enterprises may need to increase their personnel, which in the short run may lead to a decrease in labor productivity because of adjustment costs and time to learn. As for the correlation between the two innovation equations, $\rho_{\epsilon_1\epsilon_2}$, the log-likelihood values tended to be the highest for values around 0.95 for The Netherlands and 1 for France, which lead us to fix these values in the estimation.

5.2 The effects of innovation output on labor productivity

Table 8 compares the four sets of estimated elasticities and semi-elasticities of labor productivity with respect to innovation from Tables 4 to 7, and presents tests on their equality across the two countries and on whether the models with latent and observed innovation output as a predictor of labor productivity are equally close to the 'true' unknown model. All these elasticities are positive and highly statistically significant except in the case of the observed innovation indicator for France.[16] We can make more precisely the following remarks. First, they are statistically and significantly different across countries in the specification with observed innovation. For instance, *ceteris paribus*, becoming an innovator increases labor productivity by 20% on average in Dutch manufacturing but only by 6% in French manufacturing. Furthermore, in France, a 1% increase in the share of innovative sales increases labor productivity by 0.12% compared to 0.05% in Dutch manufacturing. Secondly, using Vuong's (1989) LR test for non-nested hypotheses, we conclude that the models, with respectively latent and observed innovation output as a predictor of labor productivity, are equally close to the 'true' unknown model. This result contrasts sharply with that of Duguet (2006) who used a similar test by Davidson and MacKinnon (1981) to conclude that observed innovation is a better predictor of TFP growth than latent innovation. Mairesse et al. (2005) report that the estimates of innovation output in the productivity equation are significant only when the endogeneity of innovation is accounted for, which appears to be also the case in this paper. For instance, the semi-elasticity of labor productivity with respect to becoming an innovator drops from 20% to 4% and becomes statistically insignificant in Dutch manufacturing, and the elasticity of labor productivity with respect to the observed share of innovative sales decreases from 0.12% to 0.01% in French manufacturing. The endogeneity of the innovation output regressor in the labor productivity equation operates through the covariance matrices of the individual effects on the one hand and the idiosyncratic errors on the other hand. The correlations between innovation output and labor productivity, $\rho_{\varepsilon_1\varepsilon_3}$ and $\rho_{\varepsilon_2\varepsilon_3}$, are in general statistically significant, although

[16] Note that the estimated elasticities of productivity with respect to labor and physical capital shown in Tables 4 to 7 are also all statistically significant and their orders of magnitude are not unreasonable. In other words, we find decreasing elasticities of scale by slightly less than 5% in French manufacturing and slightly more in Dutch manufacturing, and elasticities for physical capital on the low side, especially for France, which could be expected since capital is proxied by investment.

surprisingly negative. The null hypothesis of exogeneity of the innovation output regressor in the labor productivity equation is clearly rejected at any conventional significance level using a Wald or an LR test.

Table 8: Elasticities and semi-elasticities of labor productivity with respect to innovation†

I) Measures of innovation outpu	France		The Netherlands		Test of equality	
	Slope	(Std. Err.)	Slope	(Std. Err.)	|z|	p-value
	Latent innovation to explain labor productivity					
1) Latent innovation propensity	0.074**	(0.020)	0.121**	(0.029)	1.316	0.188
2) Latent share of innov. sales	0.049**	(0.011)	0.099**	(0.026)	1.800	0.072
	Observed innovation to explain labor productivity					
1) Observed innovation indicator	0.056	(0.042)	0.197**	(0.059)	1.956*	0.050
2) Observed share of innov. sales	0.122**	(0.010)	0.054**	(0.014)	3.993**	0.000

II) Vuong's LR test	France		The Netherlands	
	|z|	p-value	|z|	p-value
Latent 1) vs observed 1)	1.252	0.211	1.391	0.164
Latent 2) vs observed 2)	0.838	0.402	0.323	0.746

†I) The z-statistic is computed as the ratio of the difference of the elasticities across countries, assuming independence between them, over the standard error of that difference. II) The non-nested null hypothesis of the test is H0: both models are as close to the 'true' model. The resulting z-statistic is computed as $z{=}[\ln l_1{-}\ln l_2 - \ln obs(k_1{-}k_2)/2]/[obs{\times}var(\ln l_{1i}{-}\ln l_{2i})]^{\frac{1}{2}}$, where obs is the number of observations, k_j $(j{=}1{,}2)$ the number of parameters and $var()$ is the sample variance of the difference in the pointwise log-likelihoods of both models.

Significance levels : $*$: 5% $**$: 1%.

The effect of product innovation on labor productivity can be ascribed partly to a differential of productivity in the production of old and new products, and partly to a differential in the pricing of old and new products. As shown in Harrison et al. (2008), there is little evidence that the productivities in producing old and new products are statistically different. Let us then concentrate on the price differentials and see how much of a differential there is given our estimation results. Labor productivity can be written as $y_3 = pg$, where g is labor productivity in volume and p is average price, which by definition is equal to $p = s_q p_n + (1 - s_q)p_o$, where $s_q = \frac{q_n}{q_o}$ is the physical ratio of new to old products (assuming the two can be added up), p_n is the price of new products and p_o denotes the price of old products. After some manipulations it can be shown that $\nu = \frac{gp_o}{1-y_2\left(1-\frac{1}{r}\right)}$, where $r = \frac{p_n}{p}$, and the elasticity of labor productivity with respect to the share of sales due to new products is equal to $El_{y_3,y_2} = \frac{1}{\frac{1}{y_2\left(1-\frac{1}{r}\right)}-1}$. If we equate this expression with the one we have derived from our model in Appendix D, $El_{y_3,y_2} = \frac{\gamma_2}{1-y_2}$, use the estimated values for γ_2 and the observed sample averages for y_2, we can solve for r. Depending on whether we take the estimates obtained using the observed innovation intensity or those obtained using the latent innovation intensity, we obtain for France respectively a ratio of 2.21 and 1.31, and for The Netherlands a ratio of 1.25 and 1.55. These ratios may be somewhat overestimated if there is also a difference in the productivity of old and new products, since new products are normally produced with new technologies.

5.3 Dynamics

In both countries, there is clear evidence of a unidirectional causality running from innovation to labor productivity during the period under study. In other words, four-year lagged R&D has a positive and significant effect on current innovation output, which itself is positively and significantly affected by past innovation output, mostly through innovation output in the initial period, and has a positive and significant effect on labor productivity. The lagged feedback effect of labor productivity on innovation is not economically nor statistically significant. This result suggests that the most productive enterprises at period $t-1$ do not necessarily invest more in R&D at period t.[17] Such finding of unidirectional causality seems to be new in the empirical literature. It is robust across the four specifications and for the two countries.

In order to assess true persistence, as defined in the econometric literature on panel data, it is important to take care of individual effects and initial conditions (Hsiao, 2003). In our case, given that our panels are unbalanced with two or three consecutive observations by firm, we have an additional difficulty disentangling the effect of the lagged dependent variable and of the initial conditions. Indeed the initial values of the dependent variable on which we must project the individual effects corresponds to a two-period lag for enterprises that are present in all three periods but only to a one period lag for those that are present in two adjacent periods. For the latter, the lagged dependent variable is the same as the initial value and therefore the associated coefficient picks up the sum of the two effects. Without imposing this constraint, we notice that, as expected, the coefficient of the initial value for the 2-wave unbalanced panel is practically equal to the sum of the initial value for the 3-wave balanced panel and the one period lagged effect.

As can be seen in Tables 4 to 7, our results show no evidence of true persistence in product innovation in Dutch manufacturing neither for the incidence indicator nor the intensity variable. In other words, once individual effects and the incidence and the intensity of innovation at the initial period are controlled for, achieving successful innovations in Dutch manufacturing and generating innovative sales are no longer time dependent. In contrast, our estimates for French manufacturing (at least in the innovation intensity) support the 'success breeds success' idea. This evidence remains, however, weak as the one-period lagged innovation is less influential than the fixed effects as projected on innovations in the initial period.

For both countries we find very strong evidence of true persistence of labor productivity. Even after controlling for individual effects and initial productivity, one period lagged productivity conditions current productivity. Although fairly robust given the limitations of our panel data, our

[17] Estimation results from regressions explaining R&D at period t by productivity at period $t-1$ are not reported but can be obtained upon request.

differing results for persistence in innovation and productivity could be due to errors of specification in our models, such as large random errors in the innovation measures or important missing factors in the productivity equation, for example skills, management practices and organizational characteristics.[18]

As already stated, we find no evidence of a lagged effect of labor productivity on innovation. In other words, becoming more productive does not stimulate enterprises to become more innovative. Furthermore, as can be seen from the estimates documented in Appendix C, enterprises that were four years earlier closer to the technological frontier, defined in terms of labor productivity, are not more successful in achieving innovations and do not attain a larger share of innovative sales. This confirms the absence of a feedback effect of labor productivity on innovation. Indeed, in the presence of such a feedback effect, the distance to the frontier dummies should capture indirect effects of labor productivity, and hardly any of them is significant.

In order to capture enterprises' unobserved ability to be innovative and productive we account for individual effects in each equation of the model. Likelihood ratio (LR) tests suggest that they have indeed to be taken into account as the specifications assuming their absence in the model are rejected at the 1% significance level.[19]

6 Conclusion

We have in this study examined whether French and Dutch manufacturing firms display persistence in innovation and productivity, whether innovation Granger causes productivity or whether the reverse holds, whether the dynamics in the R&D-innovation-productivity relationship differs between French and Dutch manufacturing firms, and finally whether models with observed or latent, qualitative or quantitative, innovation indicators yield different estimation results. To do so, we have used unbalanced panels of French and Dutch manufacturing firms resulting from three waves of the respective Community Innovation Surveys. With few exceptions, the results we obtain are not very different for the two countries and are robust to various specifications of the innovation-productivity relationship. As in many related studies based on cross-sectional firm data, we find that R&D activities undertaken continuously during the previous two to four years, and the intensity of such activities, affect significantly the incidence and the intensity of product innovations. We find weak, if any, evidence of persistence in product innovation, but strong evidence of persistence in labor productivity levels. Both the incidence and the intensity of product innovation play

[18] Since we rely on a revenue measure of productivity because of unavailable output price information at the firm level, the persistence in productivity can also reflect persistence in firm market power.

[19] To save space, the results of the LR tests are not reported but can be obtained upon request.

an important role in enhancing firm labor productivity. Past productivity does not, however, affect product innovation significantly. Thus, our results provide evidence of a unidirectional causality running from innovation to productivity, without a feedback effect, and of a strong persistence in productivity but not in innovation. Our results are robust to different ways of modeling and estimating and hold for both countries.

In order to assess the generality of the result, it would be interesting to estimate the same model on more country data and longer periods, which will become possible with additional waves of innovation surveys in many countries. With the decision, at least in the European Union, to hold innovation surveys every two years, it would be worthwhile in the future to re-estimate this model with shorter lags (two years instead of four) and see whether the conclusions regarding the dynamics of innovation still hold. Productivity could also be due to process innovation. The introduction of process (and possibly other forms of) innovation would require one or more additional equations, a challenging but not impossible task. Finally, instead of examining the levels of productivity one could also consider estimating the link between innovation and productivity growth.

Appendix A   Three-stage Gauss-Hermite quadrature

The trivariate normal density function of the structural form projection errors, a_{1i}, a_{2i} and a_{3i}, denoted by $h_3(a_{1i}, a_{2i}, a_{3i}|...)$, is written as

$$h_3(a_{1i}, a_{2i}, a_{3i}|...) = \Gamma e^{\frac{-1}{2}\left(\Lambda_{11}\frac{a_{1i}^2}{\sigma_{a_1}^2}+2\Lambda_{12}\frac{a_{1i}}{\sigma_{a_1}}\frac{a_{2i}}{\sigma_{a_2}}+2\Lambda_{13}\frac{a_{1i}}{\sigma_{a_1}}\frac{a_{3i}}{\sigma_{a_3}}+2\Lambda_{23}\frac{a_{2i}}{\sigma_{a_2}}\frac{a_{3i}}{\sigma_{a_3}}+\Lambda_{22}\frac{a_{2i}^2}{\sigma_{a_2}^2}+\Lambda_{33}\frac{a_{3i}^2}{\sigma_{a_3}^2}\right)}, \tag{A.1}$$

where

$$\Gamma = \left(\sigma_{a_1}\sigma_{a_2}\sigma_{a_3}\right)^{-1}(2\pi)^{\frac{-3}{2}}(\Delta)^{\frac{-1}{2}}, \tag{A.2}$$

and the expressions of Δ and Λ_{kl} $(k, l = 1, 2, 3;\ \Lambda_{kl} = \Lambda_{lk})$ are given by

$$\begin{aligned}
\Delta &= 1-\rho_{a_1a_2}^2-\rho_{a_1a_3}^2-\rho_{a_2a_3}^2+2\rho_{a_1a_2}\rho_{a_1a_3}\rho_{a_2a_3},\\
\Lambda_{11} &= \frac{1-\rho_{a_2a_3}^2}{\Delta},\ \Lambda_{12} = \frac{\rho_{a_1a_3}\rho_{a_2a_3}-\rho_{a_1a_2}}{\Delta},\\
\Lambda_{22} &= \frac{1-\rho_{a_1a_3}^2}{\Delta},\ \Lambda_{23} = \frac{\rho_{a_1a_2}\rho_{a_1a_3}-\rho_{a_2a_3}}{\Delta},\\
\Lambda_{33} &= \frac{1-\rho_{a_1a_2}^2}{\Delta},\ \Lambda_{13} = \frac{\rho_{a_1a_2}\rho_{a_2a_3}-\rho_{a_1a_3}}{\Delta}.
\end{aligned} \tag{A.3}$$

The trivariate density expression of the reduced-form projection errors is written straightforwardly by replacing a_{3i}, σ_{a_3}, $\rho_{a_1a_3}$ and $\rho_{a_2a_3}$ by their underlined counterparts, hence the expressions of $\underline{\Delta}$ and $\underline{\Lambda_{kl}}$ $(\underline{\Lambda_{kl}} = \underline{\Lambda_{lk}})$.

Let us rewrite l_1 (eq. (3.15)) as

$$l_1 = \prod_{i=1}^{N}\int_{\underline{a_{3i}}} G_1(\underline{a_{3i}}|...)\prod_{t=0_i+1}^{T_i}\frac{1}{\underline{\sigma_{\varepsilon_3}}}\phi_1\left(\frac{y_{3it}-A_{3it}-\gamma_1A_{1it}-\underline{a_{3i}}}{\underline{\sigma_{\varepsilon_3}}}\right)H_1(\underline{a_{3i}}|...)d\underline{a_{3i}}, \tag{A.4}$$

where $G_1(\underline{a_{3i}}|...)$ and $H_1(\underline{a_{3i}}|...)$ are functions of the sole projection error $\underline{a_{3i}}$. $G_1(\underline{a_{3i}}|...)$ is derived from the trivariate density of the reduced-form projection errors and is equal to $e^{\frac{-1}{2}\left(\underline{\Lambda_{33}}\,\underline{a_{3i}}^2\,\underline{\sigma_{a_3}}^{-2}\right)}$ with $\underline{\Lambda_{kl}}$ obtained from equation (A.3), and $H_1(\underline{a_{3i}}|...)$ is given by

$$\begin{aligned}
H_1(\underline{a_{3i}}|...) = \int_{a_{2i}} G_2(a_{2i}, \underline{a_{3i}}|...)\prod_{t=0_i+1}^{T_i}\left[\sigma_{\varepsilon_2}^{-1}\left(1-\underline{\rho_{\varepsilon_2\varepsilon_3}}^2\right)^{\frac{-1}{2}}\right]&\\
\phi_1\left(\frac{y_{2it}-A_{2it}-a_{2i}-\frac{\underline{\rho_{\varepsilon_2\varepsilon_3}}\sigma_{\varepsilon_2}}{\underline{\sigma_{\varepsilon_3}}}(y_{3it}-A_{3it}-\gamma_1A_{1it}-\underline{a_{3i}})}{\sigma_{\varepsilon_2}\sqrt{1-\underline{\rho_{\varepsilon_2\varepsilon_3}}^2}}\right)H_2(a_{2i}, \underline{a_{3i}}|...)da_{2i}&.
\end{aligned} \tag{A.5}$$

$G_2(a_{2i}, \underline{a_{3i}}|...)$ is also derived from the trivariate density of the reduced-form projection errors. It is equal to $e^{\frac{-1}{2}\left(\underline{\Lambda_{22}}a_{2i}^2\sigma_{a_2}^{-2}\right)-\underline{\Lambda_{23}}a_{2i}\sigma_{a_2}^{-1}\underline{a_{3i}}\sigma_{a_3}^{-1}}$, and $H_2(a_{2i},\underline{a_{3i}}|...)$ is written as

$$H_2(a_{2i},\underline{a_{3i}}|...)=\int_{a_{1i}} G_3(a_{1i},a_{2i},\underline{a_{3i}}|...)\prod_{t=0_i+1}^{T_i}\left[\Phi_1\left(\frac{-A_{1it}-a_{1i}-\underline{\rho_{\varepsilon_1\varepsilon_3}}\ \underline{\sigma_{\varepsilon_3}^{-1}}(y_{3it}-A_{3it}-\gamma_1 A_{1it}-\underline{a_{3i}})}{\sqrt{1-\underline{\rho_{\varepsilon_1\varepsilon_3}^2}}}\right)\right]^{1-y_{1it}}$$
$$\left[\Phi_1\left(\frac{A_{1it}+a_{1i}+\underline{\rho_{12.3}}\sigma_{\varepsilon_2}^{-1}\,(y_{2it}-A_{2it}-a_{2i})+\underline{\rho_{13.2}}\ \underline{\sigma_{\varepsilon_3}^{-1}}(y_{3it}-A_{3it}-\gamma_1 A_{1it}-\underline{a_{3i}})}{\sqrt{1-\underline{R_{1.23}^2}}}\right)\right]^{y_{1it}} da_{1i}, \tag{A.6}$$

with $G_3(a_{1i},a_{2i},\underline{a_{3i}}|...)$ equal to $e^{\frac{-1}{2}\left(\underline{\Lambda_{11}}a_{1i}^2\sigma_{a_1}^{-2}\right)-\underline{\Lambda_{12}}a_{1i}\sigma_{a_1}^{-1}a_{2i}\sigma_{a_2}^{-1}-\underline{\Lambda_{13}}a_{1i}\sigma_{a_1}^{-1}\underline{a_{3i}\sigma_{a_3}^{-1}}}$. The three-stage quadrature approach consists in approximating the single integral in equation (A.6) using the formula of equation (3.17) after making an appropriate variable change. Then, $H_2(a_{2i},\underline{a_{3i}}|...)$ is replaced by the resulting approximated expression into equation (A.5). A second approximation is carried out for the single integral of equation (A.5) using the same formula. We then plug the resulting expression of $H_1(\underline{a_{3i}}|...)$ into equation (A.4) and apply again the quadrature formula, hence the final expression of l_1 (see equation (3.18)).

The performance of the Gauss-Hermite quadrature is worth mentioning. It is known to be inaccurate if the panel size, T_i, or intraclass correlation, also known as equicorrelation, is large.[20] For instance, Rabe-Hesketh et al. (2005) show that, in the context of a random-effect probit, the quadrature yields biased estimates when $T_i = 10$ with an equicorrelation of 0.9, or for any equicorrelation greater than or equal to 0.45 when $T_i = 100$ (see also Lee, 2000). However, Raymond (2007, chapter 3) shows that the quadrature works very well in the context of dynamic sample selection models with a panel of small size ($T_i = 4$) and equicorrelation between 0.3 and 0.5, the latter range being that of the equicorrelation when these models are estimated on the Dutch innovation survey data. Thus, we expect the quadrature to produce accurate estimates.[21] Evidently, we would need to carry out Monte Carlo analyses that use our nonlinear dynamic simultaneous equations models as a benchmark in order to assess the extent of the accuracy of the quadrature in these models. This is beyond the scope of our analysis and is left for future research.

Appendix B Models with y_{2it}^* or y_{2it} as a predictor of labor productivity

Model with latent share of innovative sales

[20] In the context of panel data, the intraclass correlation is a special form of serial correlation. It is defined as $\frac{\sigma_{a_j}^2}{\sigma_{a_j}^2+\sigma_{\epsilon_j}^2}$ $(j = 1, 2, 3)$ with $\sigma_{\epsilon_1} = 1$.

[21] The equicorrelation is about 0.1 when the dynamic sample selection models are estimated on the French innovation survey data. Thus, the poor performance of the Gauss-Hermite quadrature is even less of an issue.

The model with latent share of innovative sales as a predictor of labor productivity consists of equations (2.1)-(2.4) and (2.5a) with $j=2$ in equation (2.5a). These equations constitute the structural form of the model. The reduced-form equations are given by equations (2.1)-(2.4) and

$$y_{3it} = \vartheta_{33} y_{3i,t-1} + \boldsymbol{\beta}_3' \mathbf{x}_{3it} + \gamma_2 \left[\vartheta_{22} y_{2i,t-1} + \vartheta_{23} y_{3i,t-1} + \boldsymbol{\beta}_2' \mathbf{x}_{2it} \right] + \underbrace{\gamma_2 \alpha_{2i} + \alpha_{3i}}_{\underline{\alpha_{3i}}} + \underbrace{\gamma_2 \varepsilon_{2it} + \varepsilon_{3it}}_{\underline{\epsilon_{3it}}}, \tag{B.1}$$

where y_{2it}^* has been replaced by its right-hand side expression of equation (2.3). The relations between the underlined components of $\underline{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}}$ and $\underline{\boldsymbol{\Sigma}_{\boldsymbol{a}}}$ and the structural counterparts become

$$\underline{\sigma_{\varepsilon_3}^2} = \gamma_2^2 \sigma_{\varepsilon_2}^2 + \sigma_{\varepsilon_3}^2 + 2\gamma_2 \rho_{\varepsilon_2 \varepsilon_3} \sigma_{\varepsilon_2} \sigma_{\varepsilon_3}, \qquad \underline{\sigma_{a_3}^2} = \gamma_2^2 \sigma_{a_2}^2 + \sigma_{a_3}^2 + 2\gamma_2 \rho_{a_2 a_3} \sigma_{a_2} \sigma_{a_3}, \tag{B.2a}$$

$$\underline{\rho_{\varepsilon_1 \varepsilon_3}} = \frac{\gamma_2 \rho_{\varepsilon_1 \varepsilon_2} \sigma_{\varepsilon_2} + \rho_{\varepsilon_1 \varepsilon_3} \sigma_{\varepsilon_3}}{\left(\gamma_2^2 \sigma_{\varepsilon_2}^2 + \sigma_{\varepsilon_3}^2 + 2\gamma_2 \rho_{\varepsilon_2 \varepsilon_3} \sigma_{\varepsilon_2} \sigma_{\varepsilon_3} \right)^{\frac{1}{2}}}, \qquad \underline{\rho_{a_1 a_3}} = \frac{\gamma_2 \rho_{a_1 a_2} \sigma_{a_2} + \rho_{a_1 a_3} \sigma_{a_3}}{\left(\gamma_2^2 \sigma_{a_2}^2 + \sigma_{a_3}^2 + 2\gamma_2 \rho_{a_2 a_3} \sigma_{a_2} \sigma_{a_3} \right)^{\frac{1}{2}}}, \tag{B.2b}$$

$$\underline{\rho_{\varepsilon_2 \varepsilon_3}} = \frac{\gamma_2 \sigma_{\varepsilon_2} + \rho_{\varepsilon_2 \varepsilon_3} \sigma_{\varepsilon_3}}{\left(\gamma_2^2 \sigma_{\varepsilon_2}^2 + \sigma_{\varepsilon_3}^2 + 2\gamma_2 \rho_{\varepsilon_2 \varepsilon_3} \sigma_{\varepsilon_2} \sigma_{\varepsilon_3} \right)^{\frac{1}{2}}}, \qquad \underline{\rho_{a_2 a_3}} = \frac{\gamma_2 \sigma_{a_2} + \rho_{a_2 a_3} \sigma_{a_3}}{\left(\gamma_2^2 \sigma_{a_2}^2 + \sigma_{a_3}^2 + 2\gamma_2 \rho_{a_2 a_3} \sigma_{a_2} \sigma_{a_3} \right)^{\frac{1}{2}}}. \tag{B.2c}$$

The likelihood function of this model is similar to l_1 except that $\gamma_1 A_{1it}$ is replaced by $\gamma_2 A_{2it}$ where A_{2it} is defined in equation (3.11a). The FIML estimates of the structural parameters are obtained by maximizing the log-likelihood subject to the constraints (B.2a)-(B.2c) in lieu of (3.7a)-(3.7c).

Model with observed share of innovative sales

The model with observed share of innovative sales as a predictor of labor productivity consists of equations (2.1)-(2.4) and (2.5b) with $j=2$ in equation (2.5b). The likelihood function of this model is similar to l_1 except that the underlined parameters are replaced by their non-underlined equivalents and that $\gamma_1 A_{1it}$ is replaced by $\gamma_2 y_{2it}$. The FIML estimates of the structural parameters of this model are obtained by maximizing the log-likelihood with no additional constraints.

Appendix C FIML estimates with y_{1it}^* in the labor productivity equation and distance to frontier regressors

The notion of technological frontier is mostly used in the macroeconomic literature on growth convergence. Among various testable hypotheses one is that innovation becomes more important as an economy approaches the world technology frontier (see e.g. Acemoglu et al., 2003, 2006). We can identify in each 3-digit industry the enterprise with the largest productivity and then define for each enterprise a technology gap variable as the difference between the largest productivity within each 3-digit industry and the productivity of the enterprise belonging to that industry.

Table 9: FIML estimates of the model with latent innovation propensity to explain productivity and with distance to frontier regressors: Unbalanced panel from Dutch and French CIS 2, CIS 3 and CIS 4[‡]

Variable	Coefficient	(Std. Err.)	Coefficient	(Std. Err.)
	France		The Netherlands	
		Innovation incidence$_t$		
Innovation incidence$_{t-1}$	0.182†	(0.109)	0.063	(0.165)
Innovation incidence$_{0_i}$, 3 waves	0.176	(0.112)	0.710**	(0.186)
Innovation incidence$_{0_i}$, 2 waves	0.352**	(0.062)	0.693**	(0.097)
Distance to frontier				
$(\mathrm{D}_{\mathrm{Q}_2})_{t-1}$	0.050	(0.079)	0.045	(0.108)
$(\mathrm{D}_{\mathrm{Q}_3})_{t-1}$	0.040	(0.082)	0.063	(0.108)
$(\mathrm{D}_{\mathrm{Q}_4})_{t-1}$	0.089	(0.090)	0.126	(0.116)
(R&D/employee)$_{t-1}$, in log	0.029	(0.028)	0.177**	(0.045)
$(\mathrm{D}_{\text{non-continuous R\&D}})_{t-1}$	-0.519**	(0.072)	-0.588**	(0.093)
Size class				
$\mathrm{D}_{\#\text{ employees}\leq 50}$	-0.447**	(0.130)	-0.261	(0.169)
$\mathrm{D}_{50<\#\text{ employees}\leq 250}$	-0.337**	(0.090)	-0.034	(0.143)
$\mathrm{D}_{250<\#\text{ employees}\leq 500}$	-0.099	(0.076)	0.328†	(0.182)
Market share$_{t-1}$, in log	0.085**	(0.027)	0.090**	(0.030)
		Share of innovative sales$_t$, in logit		
Share of innov. sales$_{t-1}$, in logit	0.116*	(0.049)	0.046	(0.043)
Share of innov. sales$_{0_i}$, 3 waves	0.057	(0.046)	0.130**	(0.040)
Share of innov. sales$_{0_i}$, 2 waves	0.185**	(0.026)	0.171**	(0.027)
Distance to frontier				
$(\mathrm{D}_{\mathrm{Q}_2})_{t-1}$	0.376*	(0.176)	0.006	(0.175)
$(\mathrm{D}_{\mathrm{Q}_3})_{t-1}$	0.112	(0.182)	0.019	(0.176)
$(\mathrm{D}_{\mathrm{Q}_4})_{t-1}$	0.200	(0.202)	-0.124	(0.190)
(R&D/employee)$_{t-1}$, in log	0.111†	(0.058)	0.256**	(0.065)
$(\mathrm{D}_{\text{non-continuous R\&D}})_{t-1}$	-1.065**	(0.157)	-0.695**	(0.143)
Size class				
$\mathrm{D}_{\#\text{ employees}\leq 50}$	-0.274	(0.309)	0.036	(0.275)
$\mathrm{D}_{50<\#\text{ employees}\leq 250}$	-0.340†	(0.198)	0.068	(0.222)
$\mathrm{D}_{250<\#\text{ employees}\leq 500}$	0.127	(0.166)	0.146	(0.269)
Market share$_{t-1}$, in log	0.171**	(0.060)	0.085†	(0.050)
		Labor productivity$_t$: sales/employee, in log		
(Sales/employee)$_{t-1}$, in log	0.531**	(0.056)	0.324**	(0.066)
(Sales/employee)$_{0_i}$, 3 waves	0.337**	(0.056)	0.280**	(0.066)
(Sales/employee)$_{0_i}$, 2 waves	0.857**	(0.012)	0.587**	(0.024)
Latent innovation propensity$_t$	0.080**	(0.021)	0.124**	(0.029)
(Investment/employee)$_t$, in log	0.065**	(0.006)	0.119**	(0.012)
Employment$_t$, in log	-0.029**	(0.009)	-0.083**	(0.018)
# observations	2505		1639	
Log-likelihood	-5045.059		-3918.637	

[‡]Three dummies of category of industry, a time dummy and an intercept are included in each equation.
To save space, the covariance matrix of the individual effects and the error terms are not reported.
Significance levels : $\dagger$: 10% $*$: 5% $**$: 1%

Then, looking at the distribution (within each industry) of the technology gap variable, we define three dummy variables $\mathrm{D}_{\mathrm{Q}_2}$, $\mathrm{D}_{\mathrm{Q}_3}$ and $\mathrm{D}_{\mathrm{Q}_4}$ which take the value one if the technology gap lies respectively between the first (>) and second quartile ($\leq$), the second (>) and the third quartile ($\leq$), and above (>) the third quartile. The dummy variable $\mathrm{D}_{\mathrm{Q}_1}$, which takes the value one if the technology gap lies below or at the first quartile, is used as the reference. Firms for which $\mathrm{D}_{\mathrm{Q}_1}$ is equal to one are the closest to the technological frontier. If the above-mentioned hypothesis is satisfied, we expect the effects of $\mathrm{D}_{\mathrm{Q}_2}$, $\mathrm{D}_{\mathrm{Q}_3}$ and $\mathrm{D}_{\mathrm{Q}_4}$ to be negative and statistically significant. We consider the lagged values of the dummy variables in the estimation for the same reason as for the market share regressor. Furthermore, these dummy variables capture not only the distance to technological frontier but also a type of (indirect) feedback effect of productivity on innovation. As a result, in order to avoid multicollinearity problems, whenever these dummy variables are included in the estimation, the above-mentioned feedback effect of productivity is ignored, i.e. we assume $\vartheta_{13} = \vartheta_{23} = 0$.

Since the results with the distance to frontier regressors are very similar across model specifications, we report them only for the model with the innovation propensity as a predictor of labor productivity (see Table 9). In other words, we still observe a unidirectional causality running from innovation to productivity with the lagged distance to frontier dummies being insignificant. The remaining estimation results can be obtained upon request.

Appendix D Elasticity of labor productivity with respect to the share of innovative sales

Let the productivity equation be written as

$$\ln(y_{3t}) = \gamma_2 \operatorname{logit}(y_{2t}) + ... + \varepsilon_{3t}, \tag{D.1}$$

where y_{3t} denotes productivity, y_{2t} denotes the share of innovative sales and $\operatorname{logit}(y_{2t}) = \ln\left(\frac{y_{2t}}{1-y_{2t}}\right)$.[22] The elasticity of productivity with respect to the share of innovative sales, denoted by $El_{y_{3t},y_{2t}}$, is by definition $\partial \ln(y_{3t})/\partial \ln(y_{2t})$ and is derived as

$$El_{y_{3t},y_{2t}} = \frac{\partial \ln(y_{3t})}{\partial \operatorname{logit}(y_{2t})} \frac{\partial \operatorname{logit}(y_{2t})}{\partial \ln(y_{2t})} = \gamma_2 \frac{\partial \operatorname{logit}(y_{2t})}{\partial \ln(y_{2t})}.$$

By making the variable change $v_{2t} = \ln(y_{2t})$ and writing

$$\operatorname{logit}(y_{2t}) = v_{2t} - \ln\left[1 - e^{v_{2t}}\right], \tag{D.2}$$

we can derive $\partial \operatorname{logit}(y_{2t})/\partial \ln(y_{2t})$ as

$$\frac{\partial \operatorname{logit}(y_{2t})}{\partial \ln(y_{2t})} = \frac{\partial \operatorname{logit}(y_{2t})}{\partial v_{2t}} = \frac{1}{1 - e^{v_{2t}}} = \frac{1}{1 - y_{2t}}. \tag{D.3}$$

[22]For simplicity in the notation, we discard the firm subscript i, the other regressors and the individual effects.

The elasticity of productivity with respect to the share of innovative sales is then written as

$$El_{y_{3t},y_{2t}} = \frac{\gamma_2}{1 - y_{2t}} \tag{D.4}$$

and is to be evaluated at values of interest (e.g. sample mean) of the share of innovative sales (in level). When the latent share of innovative sales enters the productivity equation, we evaluate this elasticity at predicted values of interest of the latent share of innovative sales.

Since $El_{y_{3t},y_{2t}}$ is a linear function of γ_2, the standard error of the estimated elasticity is straightforwardly obtained as

$$S.E.(\widehat{El}_{y_{3t},y_{2t}}) = \frac{S.E.(\widehat{\gamma_2})}{1 - y_{2t}}. \tag{D.5}$$

References

Abramowitz, M., Stegun, I. A., 1964. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. National Bureau of Standards Applied Mathematics, US Government Printing Office, Washington, D.C..

Acemoglu, D., Aghion, P., Zilibotti, F., 2003. Vertical integration and distance to frontier. Journal of the European Economic Association 1, 630–638.

Acemoglu, D., Aghion, P., Zilibotti, F., 2006. Distance to frontier, selection, and economic growth. Journal of the European Economic Association 4, 37–74.

Agarwal, R., Audretsch, D. B., 2001. Does entry size matter? The impact of the life cycle and technology on firm survival. Journal of Industrial Economics 49, 21–43.

Bailey, M. N., Hulten, C., Campbell, D., Bresnahan, T., Caves, R. E., 1992. Productivity dynamics in manufacturing plants. Brookings Papers on Economic Activity. Microeconomics, 187–267.

Bartelsman, E. J., Dhrymes, P. J., 1998. Productivity dynamics: U.S. manufacturing plants, 1972-1986. Journal of Productivity Analysis 9, 5–34.

Bhattacharya, S., Ritter, J. R., 1983. Innovation and communication: Signalling with partial disclosure. Review of Economic Studies 50, 331–346.

Butler, J. S., Moffitt, R., 1982. A computationally efficient quadrature procedure for the one-factor multinomial probit model. Econometrica 50, 761–764.

Crépon, B., Duguet, E., Mairesse, J., 1998. Research, innovation and productivity: An econometric analysis at the firm level. Economics of Innovation and New Technology 7, 115–158.

Davidson, R., MacKinnon, J. G., 1981. Several tests for model specification in the presence of alternative hypotheses. Econometrica 49, 781–793.

Duguet, E., 2006. Innovation height, spillovers and TFP growth at the firm level: Evidence from French manufacturing. Economics of Innovation and New Technology 15, 415–442.

Fariñas, J. C., Ruano, S., 2005. Firm productivity, heterogeneity, sunk costs and market selection. International Journal of Industrial Organization 23, 505–534.

Flaig, G., Stadler, M., 1994. Success breeds success. The dynamics of the innovation process. Empirical Economics 19, 55–68.

Geroski, P. A., van Reenen, J., Walters, C. F., 1997. How persistently do firms innovate? Research Policy 26, 33–48.

Greene, W. H., 2011. Econometric Analysis, 7^{th} Edition. Prentice Hall, New Jersey.

Griffith, R., Huergo, E., Mairesse, J., Peters, B., 2006. Innovation and productivity across four European countries. Oxford Review of Economic Policy 22, 483–498.

Griliches, Z., 1979. Issues in assessing the contribution of research and development to productivity growth. Bell Journal of Economics 10, 92–116.

Hall, B. H., Griliches, Z., Hausman, J. A., 1986. Patents and R and D: Is there a lag? International Economic Review 27, 265–283.

Harrison, R., Jaumandreu, J., Mairesse, J., Peters, B., 2008. Does innovation stimulate employment? A firm-level analysis using comparable micro-data from four European countries. NBER Working Paper 14216.

Hsiao, C., 2003. Analysis of Panel Data, 2^{nd} Edition. Cambridge University Press.

Huergo, E., Moreno, L., 2011. Does history matter for the relationship between R&D, innovation and productivity. Industrial and Corporate Change 20, 1335–1368.

Kotz, S., Balakrishnan, N., Johnson, N. L., 2000. Continuous Multivariate Distributions, 2^{nd} Edition. Vol. 1. John Wiley & Sons, INC.

Lee, L.-F., 2000. A numerically stable quadrature procedure for the one-factor random-component discrete choice model. Journal of Econometrics 95, 117–129.

Mairesse, J., Mohnen, P., Kremp, E., 2005. The importance of R&D and innovation for productivity: A reexamination in light of the French innovation survey. Annales d'Économie et de Statistique 79/80, 487–527.

Majd, S., Pindyck, R. S., 1987. Time to build, option value, and investment decisions. Journal of Financial Economics 18, 7–27.

OECD, 2007. OECD Science, Technology and Industry Scoreboard 2007: Innovation and Performance in the Global Economy. OECD; 2007 edition, Paris.

Pakes, A., Griliches, Z., 1980. Patents and R&D at the firm level: A first report. Economics Letters 5, 377–381.

Parisi, M. L., Schiantarelli, F., Sembenelli, A., 2006. Productivity, innovation and R&D: Micro evidence for Italy. European Economic Review 50, 2037–2061.

Peters, B., 2009. Persistence of innovation: Stylized facts and panel data evidence. Journal of Technology Transfer 34, 226–243.

Rabe-Hesketh, S., Skrondal, A., Pickles, A., 2005. Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. Journal of Econometrics 128, 301–323.

Raymond, W., 2007. The dynamics of innovation and firm performance: An econometric panel data analysis. PhD Dissertation, Maastricht University.

Raymond, W., Mohnen, P., Palm, F., Schim van der Loeff, S., 2010. Persistence of innovation in Dutch manufacturing: Is it spurious? Review of Economics and Statistics 92, 495–504.

Vuong, Q. H., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57, 307–333.

Wooldridge, J. M., 2005. Simple solutions to the initial conditions problem in dynamic nonlinear panel data models with unobserved heterogeneity. Journal of Applied Econometrics 20, 39–54.