

A Service of

ZBW

Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

Raymond, Wladimir; Mairesse, Jacques; Mohnen, Pierre; Palm, Franz

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

CESifo Working Paper, No. 4290

Provided in Cooperation with: Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Raymond, Wladimir; Mairesse, Jacques; Mohnen, Pierre; Palm, Franz (2013) : Dynamic Models of R&D, Innovation and Productivity: Panel Data Evidence for Dutch and French Manufacturing, CESifo Working Paper, No. 4290, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at: https://hdl.handle.net/10419/77668

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



WWW.ECONSTOR.EU

The international platform of Ludwig-Maximilians University's Center for Economic Studies and the Ifo Institute





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

Wladimir Raymond Jacques Mairesse Pierre Mohnen Franz Palm

CESIFO WORKING PAPER NO. 4290 CATEGORY 11: INDUSTRIAL ORGANISATION JUNE 2013

An electronic version of the paper may be downloaded from the SSRN website: www.SSRN.com
from the RePEc website: www.RePEc.org www.RePEc.org • from the CESifo website: www.CESifo-group.org/wp

CESifo Center for Economic Studies & Ifo Institute

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

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 considers four nonlinear dynamic simultaneous equations models that include individual effects and idiosyncratic errors correlated across equations and that differ in the way 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. It estimates these models by full information maximum likelihood 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.

JEL-Code: C330, C340, C350, L600, O310, O320.

Keywords: R&D, innovation, productivity, panel data, dynamics, simultaneous equations.

Wladimir Raymond STATEC / Luxembourg wladimir.raymond@statec.etat.lu Jacques Mairesse Maastricht University / UNU-MERIT & CREST-INSEE, France mairesse@ensae.fr

Pierre Mohnen UNU-MERIT, Maastricht University & CIRANO P.O. Box 616 The Netherlands – 6200 MD Maastricht p.mohnen@maastrichtuniversity.nl Franz Palm Faculty of Economics & Business Administration / Maastricht University P.O.Box 616 The Netherlands – 6200 MD Maastricht f.palm@maastrichtuniversity.nl

May 10, 2013

The authors wish to thank the participants of the various seminars where this paper has been presented for their comments and suggestions, in particular Stéphane Robin, Ulya Ulku and Adrian Wood.

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 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 take into account the dynamic linkages between innovation and economic performance nor 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, 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 quintiles or deciles of the distribution of these differentials over time, or using kernel techniques

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

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 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 Model specifications

Our models consist of a knowledge production function and an augmented production function 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 (KPF)

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 occurrence 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

$$y_{1it}^* = \vartheta_{11}y_{1i,t-1} + \vartheta_{13}y_{3i,t-1} + \beta_1' \mathbf{x}_{1it} + \alpha_{1i} + \varepsilon_{1it}, \qquad (2.1)$$

where ϑ_{11} and ϑ_{13} capture the effect of past product innovation occurrence and past productivity on the propensity to innovate, β'_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

$$y_{2it}^* = \vartheta_{22} y_{2i,t-1} + \vartheta_{23} y_{3i,t-1} + \beta_2' \mathbf{x}_{2it} + \alpha_{2i} + \varepsilon_{2it},$$
(2.3)

where the coefficients ϑ_{22} and ϑ_{23} capture the effect of past observed share of innovative sales and

²By 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 of firm i in the unbalanced panel data sample and T_i its last observation.

past labor productivity on the potential innovation intensity, β'_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 nature of research and innovation activities, we use lagged R&D to explain innovation occurrence and innovation intensity. Since we cannot construct a stock measure of R&D, we restrict ourselves to R&D expenditures of continuous R&D performers. We 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 in particular to measurement errors in firm sales which would affect both our market share and productivity variables).⁵ We take employment as our measure of firm absolute size, and since the relation between innovation and size may be nonlinear, we use four size class indicators: small enterprises (# employees ≤ 50), medium-sized enterprises (50 < # employees ≤ 250), large enterprises $(250 < \# \text{ employees} \le 500)$ and very large enterprises (500 < # employees), the fourth class being considered as the reference. 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 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

³Zero 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 D).

 $^{{}^{5}}$ 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.

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 (APF)

As in the great majority of studies, we assume a Cobb-Douglas APF that we write 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 (proxied here by physical investment due to the 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 occurrence 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

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

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

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 occurrence (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, β'_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 Full information maximum likelihood estimation (FIML)

We shall now explain how to derive the FIML estimator, that is 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 singleequation 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 explanatory variables. Formally

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

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

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

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 β_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 \mathbf{b}_{k2} and β_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:

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

$$\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} = \begin{pmatrix} 1 & & \\ \rho_{\varepsilon_{1}\varepsilon_{2}}\sigma_{\varepsilon_{2}} & \sigma_{\varepsilon_{2}}^{2} & \\ \rho_{\varepsilon_{1}\varepsilon_{3}}\sigma_{\varepsilon_{3}} & \rho_{\varepsilon_{2}\varepsilon_{3}}\sigma_{\varepsilon_{2}}\sigma_{\varepsilon_{3}} & \sigma_{\varepsilon_{3}}^{2} \end{pmatrix}, \ \boldsymbol{\Sigma}_{\boldsymbol{a}} = \begin{pmatrix} \sigma_{a_{1}}^{2} & & \\ \rho_{a_{1}a_{2}}\sigma_{a_{1}}\sigma_{a_{2}} & \sigma_{a_{2}}^{2} & \\ \rho_{a_{1}a_{3}}\sigma_{a_{1}}\sigma_{a_{3}} & \rho_{a_{2}a_{3}}\sigma_{a_{2}}\sigma_{a_{3}} & \sigma_{a_{3}}^{2} \end{pmatrix}$$
(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

$$y_{3it} = \vartheta_{33}y_{3i,t-1} + \beta'_{3}\mathbf{x}_{3it} + \gamma_{1} \left[\vartheta_{11}y_{1i,t-1} + \vartheta_{13}y_{3i,t-1} + \beta'_{1}\mathbf{x}_{1it}\right] + \underbrace{\gamma_{1}\alpha_{1i} + \alpha_{3i}}_{\underline{\alpha_{3i}}} + \underbrace{\gamma_{1}\varepsilon_{1it} + \varepsilon_{3it}}_{\underline{\epsilon_{3it}}}, \quad (3.5)$$

where y_{1it}^* has been replaced by its right-hand side expression of equation (2.1).⁶ The individual effects and the idiosyncratic errors of the reduced form are given by $\underline{\alpha}_i = (\alpha_{1i}, \alpha_{2i}, \underline{\alpha_{3i}})'$ and $\underline{\varepsilon}_{it} = (\varepsilon_{1it}, \varepsilon_{2it}, \underline{\varepsilon}_{3it})'$, 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 $\underline{a}_i = (a_{1i}, a_{2i}, \underline{a}_{3i})'$ 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

⁶In 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.

 $\underline{\Sigma_{\varepsilon}}$ and $\underline{\Sigma_{a}}$ given by

$$\underline{\Sigma}_{\underline{\varepsilon}} = \begin{pmatrix} 1 & & \\ \rho_{\varepsilon_1 \varepsilon_2} \sigma_{\varepsilon_2} & \sigma_{\varepsilon_2}^2 & \\ \rho_{\varepsilon_1 \varepsilon_3} \sigma_{\varepsilon_3} & \underline{\rho}_{\varepsilon_2 \varepsilon_3} \sigma_{\varepsilon_2} \sigma_{\varepsilon_3} & \underline{\sigma}_{\varepsilon_3}^2 \end{pmatrix}, \quad \underline{\Sigma}_{\underline{a}} = \begin{pmatrix} \sigma_{a_1}^2 & & \\ \rho_{a_1 a_2} \sigma_{a_1} \sigma_{a_2} & \sigma_{a_2}^2 & \\ \underline{\rho_{a_1 a_3}} \sigma_{a_1} \underline{\sigma}_{a_3} & \underline{\rho}_{a_2 a_3} \sigma_{a_2} \underline{\sigma}_{a_3} & \underline{\sigma}_{a_3}^2 \end{pmatrix}, \quad (3.6)$$

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

$$\underline{\sigma_{\varepsilon_3}^2} = \gamma_1^2 + \sigma_{\varepsilon_3}^2 + 2\gamma_1 \rho_{\varepsilon_1 \varepsilon_3} \sigma_{\varepsilon_3}, \qquad \underline{\sigma_{a_3}^2} = \gamma_1^2 \sigma_{a_1}^2 + \sigma_{a_3}^2 + 2\gamma_1 \rho_{a_1 a_3} \sigma_{a_1} \sigma_{a_3}, \qquad (3.7a)$$

$$\frac{\rho_{\varepsilon_{1}\varepsilon_{3}}}{\left(\gamma_{1}^{2}+\sigma_{\varepsilon_{3}}^{2}+2\gamma_{1}\rho_{\varepsilon_{1}\varepsilon_{3}}\sigma_{\varepsilon_{3}}\right)^{\frac{1}{2}}}, \qquad \frac{\rho_{a_{1}a_{3}}}{\left(\gamma_{1}^{2}\sigma_{a_{1}}^{2}+\sigma_{a_{3}}^{2}+2\gamma_{1}\rho_{a_{1}a_{3}}\sigma_{a_{1}}\sigma_{a_{3}}\right)^{\frac{1}{2}}}, \qquad (3.7b)$$

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

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

$$l_{1i|\underline{a}_{i}} = \prod_{t=0_{i}+1}^{T_{i}} \left[\int_{-\infty}^{-(A_{1it}+a_{1i})} \int_{-\infty}^{\infty} h_{3}(\varepsilon_{1it}, \varepsilon_{2it}, y_{3it}) d\varepsilon_{1it} d\varepsilon_{2it} \right]^{1-y_{1it}}$$

$$\left[\int_{-(A_{1it}+a_{1i})}^{\infty} h_{3}(\varepsilon_{1it}, y_{2it}, y_{3it}) d\varepsilon_{1it} \right]^{y_{1it}},$$

$$(3.8)$$

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

$$A_{1it} \equiv \vartheta_{11}y_{1i,t-1} + \vartheta_{13}y_{3i,t-1} + \beta'_1 \mathbf{x}_{1it} + b_{10} + \mathbf{b}'_{11}\mathbf{y}_{1i0_i} + \mathbf{b}'_{12}\mathbf{x}_{1i}.$$
(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 $h_1(y_{3it}) \int_{-\infty}^{-(A_{1it}+a_{1i})} h_1(\varepsilon_{1it}|y_{3it}) d\varepsilon_{1it}$.⁷ 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.,

 $^{^{7}\}int_{-\infty}^{\infty}h_{3}(\epsilon_{1it},\epsilon_{2it},y_{3it})d\epsilon_{2it} = h_{2}(\epsilon_{1it},y_{3it}) \text{ where } h_{2} \text{ denotes the density of the bivariate normal distribution,}$ and $h_{2}(\epsilon_{1it},y_{3it}) = h_{1}(y_{3it})h_{1}(\epsilon_{1it}|y_{3it}) \text{ where } h_{1} \text{ denotes the density of the univariate normal distribution.}$

2000)

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

$$\Phi_{1}\left(\frac{A_{1it}+a_{1i}+\underline{\rho_{12.3}}\sigma_{\varepsilon_{2}}^{-1}(y_{2it}-A_{2it}-a_{2i})+\underline{\rho_{13.2}}}{\sqrt{1-\underline{R_{1.23}^{2}}}}\frac{\sigma_{\varepsilon_{3}}^{-1}(y_{3it}-A_{3it}-\gamma_{1}A_{1it}-\underline{a_{3i}})}{\sqrt{1-\underline{R_{1.23}^{2}}}}\right),\quad(3.10b)$$

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

$$A_{2it} \equiv \vartheta_{22} y_{2i,t-1} + \vartheta_{23} y_{3i,t-1} + \beta_2' \mathbf{x}_{2it} + b_{20} + \mathbf{b}_{21}' \mathbf{y}_{2i0_i} + \mathbf{b}_{22}' \mathbf{x}_{2i},$$
(3.11a)

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

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

$$\underline{\rho_{12.3}} \equiv \frac{\rho_{\varepsilon_1 \varepsilon_2} - \underline{\rho_{\varepsilon_1 \varepsilon_3}}}{1 - \underline{\rho_{\varepsilon_2 \varepsilon_3}^2}}, \ \underline{\rho_{13.2}} \equiv \frac{\rho_{\varepsilon_1 \varepsilon_3} - \rho_{\varepsilon_1 \varepsilon_2}}{1 - \underline{\rho_{\varepsilon_2 \varepsilon_3}^2}}, \ \underline{R_{1.23}^2} \equiv \frac{\rho_{\varepsilon_1 \varepsilon_2}^2 + \underline{\rho_{\varepsilon_1 \varepsilon_3}^2} - 2\rho_{\varepsilon_1 \varepsilon_2}}{1 - \underline{\rho_{\varepsilon_1 \varepsilon_3}^2}}, \ \underline{\rho_{\varepsilon_2 \varepsilon_3}}.$$
(3.12)

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

$$l_{1i|\underline{a}_{i}} = \prod_{t=0_{i}+1}^{T_{i}} \frac{1}{\sigma_{\varepsilon_{3}}} \phi_{1} \left(\frac{y_{3it} - A_{3it} - \gamma_{1}A_{1it} - \underline{a}_{3i}}{\underline{\sigma_{\varepsilon_{3}}}} \right) \left[\Phi_{1} \left(\frac{-A_{1it} - a_{1i} - \underline{\rho_{\varepsilon_{1}\varepsilon_{3}}}}{\sqrt{1 - \underline{\rho_{\varepsilon_{1}\varepsilon_{3}}}^{2}}} \frac{\sigma_{\varepsilon_{3}}^{-1}(y_{3it} - A_{3it} - \gamma_{1}A_{1it} - \underline{a}_{3i})}{\sqrt{1 - \underline{\rho_{\varepsilon_{1}\varepsilon_{3}}}^{2}}} \right) \right]^{1-y_{1it}} \\ \left[\Phi_{1} \left(\frac{A_{1it} + a_{1i} + \underline{\rho_{12.3}} \sigma_{\varepsilon_{2}}^{-1}(y_{2it} - A_{2it} - a_{2i}) + \underline{\rho_{13.2}}}{\sqrt{1 - \underline{R_{1.23}}^{2}}} \frac{\sigma_{\varepsilon_{3}}^{-1}(y_{3it} - A_{3it} - \gamma_{1}A_{1it} - \underline{a}_{3i})}{\sqrt{1 - \underline{R_{1.23}}^{2}}}} \right) \right]^{1-y_{1it}} \\ \frac{1}{\sigma_{\varepsilon_{2}}\sqrt{1 - \underline{\rho_{\varepsilon_{2}\varepsilon_{3}}^{2}}}} \phi_{1} \left(\frac{y_{2it} - A_{2it} - a_{2i} - \frac{\overline{\rho_{\varepsilon_{2}\varepsilon_{3}}}\sigma_{\varepsilon_{2}}}{\sigma_{\varepsilon_{3}}}(y_{3it} - A_{3it} - \gamma_{1}A_{1it} - \underline{a}_{3i})}{\sigma_{\varepsilon_{2}}\sqrt{1 - \underline{\rho_{\varepsilon_{2}\varepsilon_{3}}^{2}}}}} \right) \right]^{y_{1it}} .$$

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.⁸

⁸As 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 and the maximization of the likelihood function of the reduced form is unfeasible.

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

$$\begin{split} l_{2i|a_{i}} &= \prod_{t=0_{i}+1}^{T_{i}} \frac{1}{\sigma_{\varepsilon_{3}}} \phi_{1} \left(\frac{y_{3it} - A_{3it} - \gamma_{1} y_{1it} - a_{3i}}{\sigma_{\varepsilon_{3}}} \right) \left[\Phi_{1} \left(\frac{-A_{1it} - a_{1i} - \rho_{\varepsilon_{1}\varepsilon_{3}} \sigma_{\varepsilon_{3}}^{-1} (y_{3it} - A_{3it} - \gamma_{1} y_{1it} - a_{3i})}{\sqrt{1 - \rho_{\varepsilon_{1}\varepsilon_{3}}^{2}}} \right) \right]^{1-y_{1it}} \\ &\left[\Phi_{1} \left(\frac{A_{1it} + a_{1i} + \rho_{12.3} \sigma_{\varepsilon_{2}}^{-1} (y_{2it} - A_{2it} - a_{2i}) + \rho_{13.2} \sigma_{\varepsilon_{3}}^{-1} (y_{3it} - A_{3it} - \gamma_{1} y_{1it} - a_{3i})}{\sqrt{1 - R_{1.23}^{2}}} \right) \right]^{1-y_{1it}} \\ &\left[\frac{1}{\sigma_{\varepsilon_{2}} \sqrt{1 - \rho_{\varepsilon_{2}\varepsilon_{3}}^{2}}} \phi_{1} \left(\frac{y_{2it} - A_{2it} - a_{2i} - \frac{\rho_{\varepsilon_{2}\varepsilon_{3}} \sigma_{\varepsilon_{2}}}{\sigma_{\varepsilon_{3}}} (y_{3it} - A_{3it} - \gamma_{1} y_{1it} - a_{3i})}{\sigma_{\varepsilon_{2}} \sqrt{1 - \rho_{\varepsilon_{2}\varepsilon_{3}}^{2}}} \right) \right]^{y_{1it}}, \end{split}$$

where $\rho_{12.3}$, $\rho_{13.2}$, and $R_{1.23}^2$ are derived straightforwardly from equation (3.12) by replacing the underlined correlations by their structural form counterparts, that is

$$\rho_{12.3} \equiv \frac{\rho_{\varepsilon_1 \varepsilon_2} - \rho_{\varepsilon_1 \varepsilon_3} \ \rho_{\varepsilon_2 \varepsilon_3}}{1 - \rho_{\varepsilon_2 \varepsilon_3}^2}, \ \rho_{13.2} \equiv \frac{\rho_{\varepsilon_1 \varepsilon_3} - \rho_{\varepsilon_1 \varepsilon_2} \ \rho_{\varepsilon_1 \varepsilon_2} \ \rho_{\varepsilon_2 \varepsilon_3}}{1 - \rho_{\varepsilon_2 \varepsilon_3}^2}, \ R_{1.23}^2 \equiv \frac{\rho_{\varepsilon_1 \varepsilon_2}^2 + \rho_{\varepsilon_1 \varepsilon_3}^2 \ - 2\rho_{\varepsilon_1 \varepsilon_2} \ \rho_{\varepsilon_1 \varepsilon_3} \ \rho_{\varepsilon_2 \varepsilon_3}}{1 - \rho_{\varepsilon_2 \varepsilon_3}^2}.$$

$$(3.15)$$

3.3 Numerical evaluation

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

$$l_1 = \prod_{i=1}^N \int_{a_{1i}} \int_{a_{2i}} \int_{\underline{a_{3i}}} l_{1i|\underline{a_i}} h_3(a_{1i}, a_{2i}, \underline{a_{3i}}|...) da_{1i} da_{2i} d\underline{a_{3i}},$$
(3.16)

and

$$l_2 = \prod_{i=1}^{N} \int_{a_{1i}} \int_{a_{2i}} \int_{a_{3i}} l_{2i|a_i} h_3(a_{1i}, a_{2i}, a_{3i}|...) da_{1i} da_{2i} da_{3i}.$$
(3.17)

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.⁹ The Gauss-Hermite quadrature states that

$$\int_{-\infty}^{\infty} e^{-r^2} f(r) dr \simeq \sum_{m=1}^{M} w_m f(a_m), \qquad (3.18)$$

where w_m and a_m are respectively the weights and abscissae of the quadrature with M being the total number of integration points.¹⁰ 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

⁹The 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.

¹⁰The abscissae of the quadrature, a_m , should not be confused with the projections errors a_{1i} , a_{2i} and a_{3i} .

approximation. Using the results of Appendix A, the unconditional likelihood, l_1 , is derived as

$$l_{1} \simeq \prod_{i=1}^{N} \Delta \pi^{\frac{-3}{2}} \Big[(1 - \rho_{a_{1}a_{2}}^{2}) (1 - \underline{\rho_{a_{1}a_{3}}^{2}}) (1 - \underline{\rho_{a_{1}a_{2}}^{2}}) \Big]^{\frac{-1}{2}} \sum_{m_{3}=1}^{M_{3}} w_{m_{3}} \prod_{t=0_{i}+1}^{T_{i}} \frac{1}{\sigma_{\varepsilon_{3}}} \phi_{1} \left(\frac{y_{3it} - A_{3it} - \gamma_{1}A_{1it} - a_{m_{3}}[...]}{\sigma_{\varepsilon_{3}}} \right) \\ \sum_{m_{2}=1}^{M_{2}} w_{m_{2}} \frac{e^{2\Delta_{23}a_{m_{2}}a_{m_{3}}}}{\sigma_{\sqrt{\Delta_{22}}\Delta_{33}}} \prod_{t=0_{i}+1}^{T_{i}} \left[\frac{1}{\sigma_{\varepsilon_{2}}\sqrt{1 - \underline{\rho_{\varepsilon_{2}\varepsilon_{3}}^{2}}}} \phi_{1} \left(\frac{y_{2it} - A_{2it} - a_{m_{2}}[...] - \frac{\rho_{\varepsilon_{2}\varepsilon_{3}}\sigma_{\varepsilon_{2}}}{\sigma_{\varepsilon_{3}}} (y_{3it} - A_{3it} - \gamma_{1}A_{1it} - a_{m_{3}}[...])}{\sigma_{\varepsilon_{2}}\sqrt{1 - \underline{\rho_{\varepsilon_{2}\varepsilon_{3}}^{2}}}}} \right) \Big]^{y_{1it}}$$

$$(3.19)$$

$$\begin{split} &\sum_{m_{1}=1}^{M_{1}} w_{m_{1}} e^{\frac{-2a_{m_{1}}}{\sqrt{\Delta_{11}}} \left(\frac{a_{m_{2}}\Delta_{12}}{\sqrt{\Delta_{22}}} + \frac{a_{m_{3}}\Delta_{13}}{\sqrt{\Delta_{33}}} \right)} \prod_{t=0_{i}+1}^{T_{i}} \left[\Phi_{1} \left(\frac{-A_{1it} - a_{m_{1}}[...] - \underline{\rho_{\varepsilon_{1}\varepsilon_{3}}}}{\sqrt{1 - \underline{\rho_{\varepsilon_{1}\varepsilon_{3}}}}} \frac{\sigma_{\varepsilon_{3}}^{-1}(y_{3it} - A_{3it} - \gamma_{1}A_{1it} - a_{m_{3}}[...])}{\sqrt{1 - \underline{\rho_{\varepsilon_{1}\varepsilon_{3}}}}} \right) \right]^{1-y_{1it}} \\ & \left[\Phi_{1} \left(\frac{A_{1it} + a_{m_{1}}[...] + \underline{\rho_{12,3}} \sigma_{\varepsilon_{2}}^{-1}(y_{2it} - A_{2it} - a_{m_{2}}[...]) + \underline{\rho_{13,2}}}{\sqrt{1 - \underline{R_{1,23}}^{2}}} \frac{\sigma_{\varepsilon_{3}}^{-1}(y_{3it} - A_{3it} - \gamma_{1}A_{1it} - a_{m_{3}}[...])}{\sqrt{1 - \underline{R_{1,23}^{2}}}} \right) \right]^{y_{1it}}, \end{split}$$

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 and descriptive statistics

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.¹¹ 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.¹² Note that one of the particularities of the innovation survey is that, for

¹¹Both 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.

 $^{^{12}}$ We delete enterprises with a share of total R&D expenditures (intramural + extramural) in total sales greater than 50% since they are likely to specialize in doing and trading R&D, hence should be classified in R&D services

each period, product innovation occurrence relates to the introduction of a new product over a 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 occurrence of innovation is considered and to four years when the share of innovative sales is considered.¹³ In this paper we consider as innovators only those firms that have introduced a product new to the firm, but not necessarily new to the market.

In the following Tables 1, 2 and 3, we show some simple descriptive statistics, mostly means, to present our samples and main variables. 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.

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

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 enterprises with product innovators, it appears that in both countries product innovators do not seem to

not in manufacturing.

¹³The 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 but a cut-off point of 500 employees is used in CIS 2 and 3, and one of 250 employees in CIS 4. The use of different cut-off points in the census/sampling scheme may result in differences across countries in the the French sampling we would have lost too many observations for the Netherlands. In the EAE, all enterprises with at least 20 employees are surveyed.

Variable↓		France				1	The Ne	therlar	nds
$Pattern \rightarrow$	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

Table 1: Employment in the sample and in the population for each pattern of the unbalanced panel data samples of Dutch and French manufacturing enterprises: CIS 2, CIS 3, and CIS 4

Pattern refers to the presence/absence of firms in the three successive waves.

be more productive on average. This is due to the existence among non-product innovators of a few firms with very high values of sales per employee. When we take the logarithms of sales per employee, as we do in our estimating equations, we downweigh outlier values and obtain significantly higher productivities 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 contrast to The Netherlands where they are mostly medium-sized enterprises.

Variable	Fr	ance	The Netherlands		
	All enterprises	Product. innov.	All enterprises	Product. innov.	
Product innovator	0.59	-	0.58	-	
Share of innov. sales	-	0.22	-	0.30	
$Sales/employee^{\dagger}$	220.53	215.04	184.85	180.70	
$R\&D/employee^{\ddagger}$	-	8.09	-	4.68	
$Investment/employee^{\dagger}$	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	

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

[†]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 our 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). Although the differences between the means of the main variables across countries are informative, 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 3: Means of dependent and explanatory variables for each CIS of the unbalanced panel data samples for 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 ^{\dagger}	0.30	0.15	0.23	0.33	0.32	0.21
$\mathrm{Sales}/\mathrm{employee}^\ddagger$	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
$Investment/employee^{\ddagger}$	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.

5 Estimation results

We now turn to the results of the estimation of the models. We shall first quickly comment on the general results before discussing particularly the estimated effects of innovation output on productivity and the dynamic interrelations between innovation and productivity. Tables 4 and 5 present the estimation results for the model with latent innovation as a predictor of labor productivity, and Tables 6 and 7 present the results for the model with observed innovation as a

predictor of labor productivity.

$\begin{tabular}{ c c c c c } \hline France & The Netherlands \\ \hline Innovation occurrence_{t-1} & Innovation occurrence_{t} \\ \hline Innovation occurrence_{0,1} 3 waves & 0.167 & (0.111) & 0.666^{**} & (0.189) \\ Innovation occurrence_{0,2} 2 waves & 0.339^{**} & (0.066) & 0.721^{**} & (0.112) \\ (Sales/employee)_{t-1,1} in log & 0.000 & (0.017) & 0.010 & (0.025) \\ (R&D/employee)_{t-1,1} in log & 0.025 & (0.027) & 0.173^{**} & (0.045) \\ (Dnon-continuous R&D)_{t-1} & -0.534^{**} & (0.070) & -0.587^{**} & (0.093) \\ Size class & & & & & & & & & & & & & & & & & & $	Variable	Coefficient (Std. Err.)		Coefficient (Std. Err.)		
$\begin{tabular}{ c c c c c c } \hline Innovation occurrence_t & Innovation occurrence_{t-1} & 0.181^{\dagger} & (0.103) & 0.051 & (0.171) \\ Innovation occurrence_{0,1} 3 waves & 0.167 & (0.111) & 0.696^{**} & (0.189) \\ Innovation occurrence_{0,1} 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) \\ (Dnon-continuous R&D)_{t-1} & -0.534^{**} & (0.070) & -0.587^{**} & (0.093) \\ Size class & & & & & & & & & & & & & & & & & & $		France		The Netherlands		
$\begin{array}{c c c c c c c c c c c c c c c c c c c $		Innovatio		on occurrence _t		
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	Innovation occurrence $_{t-1}$	0.181^{\dagger}	(0.103)	0.051	(0.171)	
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	Innovation occurrence _{0_i} , 3 waves	0.167	(0.111)	0.696^{**}	(0.189)	
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	Innovation occurrence _{0_i} , 2 waves	0.339^{**}	(0.066)	0.721^{**}	(0.112)	
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	$(Sales/employee)_{t-1}$, in log	0.000	(0.017)	0.010	(0.025)	
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	$(R\&D/employee)_{t-1}$, in log	0.025	(0.027)	0.173^{**}	(0.045)	
Size class $D_{\#} employees \leq 50$ -0.470^{**} (0.130) -0.271 (0.170) $D_{50<\#} employees \leq 500$ -0.095 (0.092) -0.049 (0.144) $D_{250<\#} employees \leq 500$ -0.095 (0.077) 0.304^{\dagger} (0.182) Market share _{t-1} , in log 0.074^{**} (0.024) 0.077^{**} (0.029) Market share _{t-1} , in log 0.110^* (0.024) 0.077^{**} (0.029) Share of innov. sales _{0,1} , 3 waves 0.064 (0.046) 0.132^{**} (0.040) Share of innov. sales _{0,1} , 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^{\dagger} (0.57) 0.258^{**} (0.065) (Dnon-continuous R&D) _{t-1} -1.093^{**} (0.154) -0.697^{**} (0.144) Size class $D_{\#}$ employees ≤ 50 -0.349 (0.311) 0.022 (0.276) $D_{50<\#}$ employees ≤ 500 0.118 (0.169) 0.156 (0.28)	$(D_{non-continuous R\&D})_{t-1}$	-0.534**	(0.070)	-0.587**	(0.093)	
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	Size class					
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$D_{\# \text{ employees}} < 50$	-0.470**	(0.130)	-0.271	(0.170)	
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$D_{50 < \# \text{ employees} < 250}$	-0.335**	(0.092)	-0.049	(0.144)	
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	$D_{250 < \# \text{ employees} < 500}$	-0.095	(0.077)	0.304^{\dagger}	(0.182)	
$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	Market share $t-1$, in log	0.074^{**}	(0.024)	0.077^{**}	(0.029)	
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$			Share of innov	vative sales _t , in \log	çit	
$\begin{array}{rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr$	Share of innov. sales $t-1$, in logit	0.110*	(0.050)	0.043	(0.044)	
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	Share of innov. sales 0_i , 3 waves	0.064	(0.046)	0.132^{**}	(0.040)	
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	Share of innov. sales 0_i , 2 waves	0.185^{**}	(0.027)	0.174^{**}	(0.030)	
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$(Sales/employee)_{t-1}$, in log	0.002	(0.034)	0.001	(0.029)	
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$(R\&D/employee)_{t-1}$, in log	0.105^{\dagger}	(0.057)	0.258^{**}	(0.065)	
Size class 0 -0.349 (0.311) 0.022 (0.276) D50<# employees<250	$(D_{non-continuous B&D})_{t=1}$	-1.093**	(0.154)	-0.697**	(0.144)	
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	Size class				~ /	
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$D_{\# employees} \leq 50$	-0.349	(0.311)	0.022	(0.276)	
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$D_{50<\#} \text{ employees} \leq 250$	-0.361^{\dagger}	(0.206)	0.065	(0.222)	
Market share_{t-1}, in log 0.152^{**} (0.055) 0.093^* (0.047) Labor productivity_t: sales/employee, in log 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 0.366^{**} (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) Ga1 0.259^{\dagger} (0.148) 0.470^{**} (0.138) σ_{a_1} 0.259^{\dagger} (0.179) 0.680^{**} (0.193) σ_{a_2} 0.096^{**} (0.025) 0.160^{**} (0.060)	$D_{250} < \# \text{ employees} \le 500$	0.118	(0.169)	0.156	(0.268)	
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	Market share t_{-1} , in log	0.152^{**}	(0.055)	0.093^{*}	(0.047)	
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$		L	abor productivity	<i>t</i> : sales/employee.	in log	
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$(Sales/employee)_{t=1}$, in log	0.531**	(0.056)	0.319**	(0.066)	
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$(Sales/employee)_0$, 3 wayes	0.336^{**}	(0.056)	0.282^{**}	(0.066)	
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	$(Sales/employee)_0$, 2 waves	0.856**	(0.012)	0.584^{**}	(0.024)	
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	Latent innovation propensity t	0.074^{**}	(0.020)	0.121**	(0.029)	
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$(Investment/employee)_t$, in log	0.065^{**}	(0.006)	0.119^{**}	(0.012)	
$\begin{array}{c c} \hline & & & & & \\ \hline & & & & & \\ \hline & & & & \\ \hline & & & &$	Employment, in \log	-0.027**	(0.008)	-0.082**	(0.018)	
Individual effects σ_{a_1} 0.259^{\dagger} (0.148) 0.470^{**} (0.138) σ_{a_2} 0.745^{**} (0.179) 0.680^{**} (0.193) σ_{a_2} 0.096^{**} (0.025) 0.160^{**} (0.060)	<u> </u>		Covar	iance matrix	()	
$ \begin{array}{cccc} \sigma_{a_1} & 0.259^{\dagger} & (0.148) & 0.470^{**} & (0.138) \\ \sigma_{a_2} & 0.745^{**} & (0.179) & 0.680^{**} & (0.193) \\ \sigma_{a_3} & 0.096^{**} & (0.025) & 0.160^{**} & (0.060) \\ \end{array} $	Individual effects		00101			
$ σ_{a_2} $ $ 0.745^{**} $ $ (0.179) $ $ 0.680^{**} $ $ (0.193) $ $ σ_{a_2} $ $ 0.066^{**} $ $ (0.25) $ $ 0.166^{**} $ $ (0.660) $	σ_{a}	0.259^{\dagger}	(0.148)	0.470^{**}	(0.138)	
σ 0.006** (0.025) 0.160** (0.060)	σ_{a_2}	0.745**	(0.179)	0.680**	(0.193)	
0.000 (0.020) (0.100 (0.000))	σ_{a_2}	0.096**	(0.025)	0.160**	(0.060)	
ρ_{q_1,q_2} 0.514 ^{**} (0.152) 0.540 ^{**} (0.145)	$\rho_{a_1a_2}$	0.514^{**}	(0.152)	0.540^{**}	(0.145)	
$\rho_{a_1a_2}$ -0.090 (0.408) -0.221 (0.336)	$\rho_{a_1a_2}$	-0.090	(0.408)	-0.221	(0.336)	
$\rho_{a_1a_3}$ 0.158 (0.279) 0.064 (0.331)	$\rho_{a_2a_2}$	0.158	(0.279)	0.064	(0.331)	
Idiosyncratic errors	Idiosyncratic errors		· · · ·		()	
$\sigma_{\hat{\epsilon}_2}$ 2.469** (0.076) 1.780** (0.096)	σ_{ε_2}	2.469^{**}	(0.076)	1.780^{**}	(0.096)	
σ_{ε_2} 0.313** (0.009) 0.587** (0.019)	$\sigma_{\hat{\varepsilon}_2}$	0.313^{**}	(0.009)	0.587^{**}	(0.019)	
$\rho_{e_{i}e_{2}}$ 0.99, fixed after grid search 0.95, fixed after grid search	ρ_{siso}	0.99, fixed af	ter grid search	0.95, fixed aft	er grid search	
$\rho_{e_1e_2}$ -0.206** (0.071) -0.237** (0.067)	$\rho_{\epsilon_1\epsilon_2}$	-0.206**	(0.071)	-0.237**	(0.067)	
ρ_{sasa} -0.208 ^{**} (0.070) -0.229 ^{**} (0.061)	ρ_{solution}	-0.208**	(0.070)	-0.229**	(0.061)	
# observations 2505 1639	# observations	25	505		1639	
Log-likelihood -5048.917 -3920.898	Log-likelihood	-5	048.917		-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%

Variable	Coefficient (Std. Err.) France Innovation		Coefficient (Std. Err.)		
			The Netherlands		
			on occurrence _t		
Innovation occurrence $t-1$	0.151	(0.098)	0.027	(0.178)	
Innovation occurrence _0_i, 3 waves	0.171^{\dagger}	(0.102)	0.738^{**}	(0.194)	
Innovation occurrence $_{0_i},2$ waves	0.343^{**}	(0.066)	0.719^{**}	(0.116)	
$(\text{Sales/employee})_{t-1}$, 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)	
$(D_{non-continuous R\&D})_{t-1}$ Size class	-0.546**	(0.070)	-0.607**	(0.095)	
$D_{\# \text{ employees}} < 50$	-0.509**	(0.126)	-0.325^{\dagger}	(0.175)	
$D_{50<\# \text{ employees} \le 250}$	-0.353**	(0.091)	-0.101	(0.150)	
$D_{250<\# \text{ employees}} \leq 500$	-0.111	(0.077)	0.225	(0.187)	
Market share $t-1$, in log	0.069^{**}	(0.024)	0.072^{*}	(0.030)	
		Share of innov	vative sales _t , in \log	git	
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^{\dagger}	(0.055)	0.261^{**}	(0.062)	
$(D_{non-continuous R\&D})_{t-1}$ Size class	-1.018**	(0.154)	-0.691**	(0.140)	
$D_{\# \text{ employees}} < 50$	-0.187	(0.296)	0.096	(0.271)	
$D_{50<\# \text{ employees} \leq 250}$	-0.366^{\dagger}	(0.198)	0.169	(0.219)	
$D_{250<\# \text{ employees}} \leq 500$	0.103	(0.162)	0.323	(0.266)	
Market share t_{-1} , in log	0.179^{**}	(0.052)	0.110^{*}	(0.046)	
		Labor productivity	v_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)	
		Covar	iance 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)	
$ ho_{a_1a_2}$	0.546^{**}	(0.128)	0.544^{**}	(0.146)	
$ ho_{a_1a_3}$	-0.094	(0.315)	-0.121	(0.350)	
$\rho_{a_2a_3}$	-0.108	(0.302)	-0.151	(0.362)	
Idiosyncratic errors	0.401**		1 501**		
o_{ε_2}	2.481	(0.075)	1.791	(0.098)	
o_{ε_3}	0.320°	(U.U11)	0.596	(0.021)	
	0.99, fixed a	(0.077)	0.90, fixed aff	(0.074)	
	-0.298	(0.077)	-0.200	(0.074)	
$\mu_{\epsilon_2\epsilon_3}$	-0.301	(0.075)	-0.288	(0.073)	
# observations		2505		1639	
Log-likelihood	-5045.613		-3922.326		

Table 5: FIML estimates of the model with latent innovation intensity to explain productivity: Unbalanced panel data samples from Dutch and French CIS 2, CIS 3 and CIS 4^{\ddagger}

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

Variable	Coefficient (Std. Err.)		Coefficient	(Std. Err.)	
	France		The Netherlands		
	Innovatio		on occurrence $_t$		
Innovation occurrence $t-1$	0.155	(0.117)	0.010	(0.182)	
Innovation occurrence _{0i} , 3 waves	0.174	(0.135)	0.767^{**}	(0.200)	
Innovation occurrence 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)	
$(D_{non-continuous R\&D})_{t-1}$	-0.567**	(0.071)	-0.621**	(0.096)	
Size class					
$D_{\text{\# employees}} < 50$	-0.577**	(0.134)	-0.323^{\dagger}	(0.174)	
$D_{50<\# \text{ employees} \le 250}$	-0.356**	(0.102)	-0.115	(0.148)	
$D_{250 < \# \text{ employees} < 500}$	-0.102	(0.081)	0.198	(0.188)	
Market share t_{-1} , in log	0.054^{*}	(0.024)	0.068^{*}	(0.030)	
		Share of innov	vative sales _t , in log	it	
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^{\dagger}	(0.058)	0.256^{**}	(0.065)	
$(D_{non-continuous R&D})_{t-1}$	-1.119**	(0.159)	-0.710**	(0.145)	
Size class					
$D_{\# \text{ employees}} \leq 50$	-0.502	(0.315)	-0.022	(0.277)	
$D_{50<\# employees \le 250}$	-0.355^{\dagger}	(0.210)	0.005	(0.223)	
$D_{250} < \# \text{ employees} \le 500$	0.108	(0.171)	0.064	(0.269)	
Market share t_{-1} , in log	0.131^{*}	(0.055)	0.081^{\dagger}	(0.048)	
	L	abor productivity	t_t : sales/employee,	in log	
$(\text{Sales/employee})_{t=1}$, 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)	
		Covar	iance 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)	
$\rho_{a_1a_2}$	0.535^{**}	(0.170)	0.527^{**}	(0.148)	
$\rho_{a_1a_3}$	0.242	(0.362)	0.055	(0.365)	
$ ho_{a_2a_3}$	0.271	(0.314)	0.224	(0.329)	
Idiosyncratic errors					
$\sigma_{arepsilon_2}$	2.463^{**}	(0.083)	1.758^{**}	(0.096)	
$\sigma_{arepsilon_3}$	0.308^{**}	(0.008)	0.580^{**}	(0.018)	
$\rho_{\varepsilon_1\varepsilon_2}$	0.99, fixed af	ter grid search	0.95, fixed after grid search		
$\rho_{\varepsilon_1\varepsilon_3}$	-0.071	(0.114)	-0.223**	(0.072)	
$\rho_{\epsilon_2\epsilon_3}$	-0.024	(0.043)	-0.128*	(0.052)	
# observations	25	505	1	639	
Log-likelihood	-5054.238		-3925.706		

 $\label{eq:table 6: FIML estimates of the model with observed innovation indicator to explain productivity: Unbalanced panel data samples from Dutch and French CIS 2, CIS 3 and CIS 4^{\ddagger}$

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

Variable	Coefficient (Std. Err.)		Coefficient (Std. Err.)		
	France		The Netherlands		
	Innovatio		n occurrence _t		
Innovation occurrence $t-1$	0.245^{**}	(0.088)	0.014	(0.176)	
Innovation occurrence _{0_i} , 3 waves	0.114	(0.092)	0.760^{**}	(0.194)	
Innovation occurrence _{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)	
$(D_{non-continuous R\&D})_{t-1}$	-0.433**	(0.063)	-0.608**	(0.095)	
Size class					
$D_{\text{\# employees} \leq 50}$	-0.514^{**}	(0.115)	-0.319^{\dagger}	(0.173)	
$D_{50 < \# \text{ employees} \le 250}$	-0.354**	(0.082)	-0.102	(0.148)	
$D_{250 < \# \text{ employees} \le 500}$	-0.129^{\dagger}	(0.069)	0.220	(0.187)	
Market share $t-1$, in log	0.069^{**}	(0.021)	0.071^{*}	(0.030)	
		Share of innova	tive 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)	
$(\text{Sales/employee})_{t-1}$, in log	0.002	(0.031)	-0.001	(0.029)	
$(R\&D/employee)_{t-1}$, in log	0.076^{\dagger}	(0.045)	0.262^{**}	(0.064)	
$(D_{non-continuous R\&D})_{t-1}$	-0.708**	(0.127)	-0.690**	(0.144)	
Size class					
$D_{\text{\# employees} \leq 50}$	-0.213	(0.258)	0.029	(0.275)	
$D_{50 < \# \text{ employees} \le 250}$	-0.389^{*}	(0.170)	0.070	(0.221)	
$D_{250 < \# \text{ employees} \le 500}$	0.001	(0.136)	0.178	(0.268)	
Market share $t-1$, in log	0.191^{**}	(0.044)	0.098^{*}	(0.047)	
	La	abor 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)	
$\operatorname{Employment}_t$, in \log	-0.037**	(0.007)	-0.071**	(0.017)	
		Covaria	nce 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)	
$ ho_{a_1a_2}$	0.557^{**}	(0.124)	0.550^{**}	(0.139)	
$ ho_{a_1a_3}$	0.011	(0.230)	-0.002	(0.361)	
$ ho_{a_2a_3}$	0.117	(0.195)	0.120	(0.382)	
Idiosyncratic errors	at at		-tt-		
$\sigma_{arepsilon_2}$	2.459**	(0.065)	1.795**	(0.098)	
$\sigma_{arepsilon_3}$	0.357**	(0.012)	0.586^{**}	(0.019)	
$\rho_{\varepsilon_1\varepsilon_2}$	0.99, fixed at	tter 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	28	505	1	639	
Log-likelihood	-5021.290		-3923.897		

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

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

5.1 The effects of size, market share, R&D, physical capital 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.¹⁴ Size does not seem to play, however, a significant role in explaining differences in innovation intensities conditional on being an innovator. For both countries, however, market share plays a positive and significant role not only in the innovation occurrence but also in the share of innovative sales: a 10% increase in market share increases the probability to innovate by 0.2% and the share of innovative sales by 1 to 2%.

In France and The Netherlands, 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 5%, and to an increase in the share of innovative sales (of product innovators) of about 0.2%.¹⁵ The estimates are lower for France showing that a 1% increase in R&D intensity does 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 our results are also consistent with a lagged impact of R&D on innovation.¹⁶

We observe a negative and statistically significant correlation, at time t, in both countries and in all specifications between the idiosyncratic error terms in the innovation and the labor productivity equations. This could correspond to 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.¹⁷

Note that the estimated elasticities of productivity with respect to labor and physical capital

 $^{^{14}}$ 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.

¹⁵The 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} denotes the share of innovative sales in level (see Appendix C).

¹⁶To 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.

¹⁷As 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.

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.

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.¹⁸

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.

We can make more precisely the following remarks. Firstly, these estimates appear statistically different across countries only in the specification with observed innovation. A product innovator has on average a 20% higher labor productivity than a non-innovator in Dutch manufacturing and a 6% higher labor productivity in French manufacturing. By contrast, in French manufacturing a 1% increase in the share of innovative sales raises labor productivity on average by 0.12%, compared to 0.05% in Dutch manufacturing. Labor productivity is more responsive to increases in product innovation in French than in Dutch manufacturing enterprises.

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 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. A likely reason for the difference may be that the two-step estimation procedure used by Duguet (2006) ignores the correlation between the errors in the innovation and productivity equations.

As stressed in Mairesse et al. (2005), the important point is that the estimates of innovation output in the productivity equation are significant only when the endogeneity of innovation is taken into account, and this is also what appears here. When innovation is treated as exogenous

 $^{^{18}\}mathrm{To}$ save space, the results of the LR tests are not reported but can be obtained upon request.

in the productivity equation, 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 (results not tabulated but available upon request). In our paper, the endogeneity of the innovation output regressor in the labor productivity equation operates in both types of models largely 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. 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.

 $Table \ 8: \ Elasticities \ and \ semi-elasticities \ of \ labor \ productivity \ with \ respect \ to \ innovation^{\dagger}$

I) Measures of innovation outpu	France		The N	The Netherlands		equality	
	Slope	(Std. Err.)	Slope	(Std. Err.)	$ \mathbf{z} $	p-value	
		Latent inne	ovation to ex	plain labor pro	ductivity		
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 innovatio			ovation 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	$ \mathbf{z} $	p-va	alue	
Latent 1) vs observed 1)	1.252		0.211	1.391	0.1	164	
Latent 2) vs observed 2)	C).838	0.402	0.323	0.7	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%.

5.3 The dynamics of innovation and productivity

Persistence in innovation

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 occurrence indicator nor the intensity variable. In other words, once individual effects and the occurrence 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.

Persistence in productivity

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 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.¹⁹

Causality

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.²⁰ 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.

¹⁹Since 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.

²⁰Estimation results from regressions explaining R&D at period t by productivity at period t-1 are not reported but can be obtained upon request.

Furthermore, as can be seen from the estimates documented in Appendix D, 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.

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 innovationproductivity 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 occurrence 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 occurrence and the intensity of product innovation play 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.

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

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

where

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

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

$$\Delta = 1 - \rho_{a_1 a_2}^2 - \rho_{a_1 a_3}^2 - \rho_{a_2 a_3}^2 + 2\rho_{a_1 a_2}\rho_{a_1 a_3}\rho_{a_2 a_3},$$

$$\Lambda_{11} = \frac{1 - \rho_{a_2 a_3}^2}{\Delta}, \quad \Lambda_{12} = \frac{\rho_{a_1 a_3}\rho_{a_2 a_3} - \rho_{a_1 a_2}}{\Delta},$$

$$\Lambda_{22} = \frac{1 - \rho_{a_1 a_3}^2}{\Delta}, \quad \Lambda_{23} = \frac{\rho_{a_1 a_2}\rho_{a_1 a_3} - \rho_{a_2 a_3}}{\Delta},$$

$$\Lambda_{33} = \frac{1 - \rho_{a_1 a_2}^2}{\Delta}, \quad \Lambda_{13} = \frac{\rho_{a_1 a_2}\rho_{a_2 a_3} - \rho_{a_1 a_3}}{\Delta}.$$
(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. (3.16)) as

$$l_{1} = \prod_{i=1}^{N} \int_{\underline{a_{3i}}} G_{1}(\underline{a_{3i}}|...) \prod_{t=0_{i}+1}^{T_{i}} \frac{1}{\underline{\sigma_{\varepsilon_{3}}}} \phi_{1}\left(\frac{y_{3it} - A_{3it} - \gamma_{1}A_{1it} - \underline{a_{3i}}}{\underline{\sigma_{\varepsilon_{3}}}}\right) H_{1}(\underline{a_{3i}}|...) d\underline{a_{3i}},$$
(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 $e^{\frac{-1}{2}\left(\underline{\Lambda_{33}} a_{3i}^2 \frac{\sigma_{a3}^{-2}}{a_{a3}}\right)}$ with $\underline{\Lambda_{kl}}$ obtained from equation (A.3), and $H_1(\underline{a_{3i}}|...)$ is given by

$$H_{1}(\underline{a_{3i}}|...) = \int_{a_{2i}} G_{2}(a_{2i}, \underline{a_{3i}}|...) \prod_{t=0_{i}+1}^{T_{i}} \left[\sigma_{\varepsilon_{2}}^{-1} \left(1 - \underline{\rho_{\varepsilon_{2}\varepsilon_{3}}^{2}} \right)^{\frac{-1}{2}} \right] \\ \phi_{1}\left(\frac{y_{2it} - A_{2it} - a_{2i} - \frac{\rho_{\varepsilon_{2}\varepsilon_{3}}\sigma_{\varepsilon_{2}}}{\sigma_{\varepsilon_{3}}} (y_{3it} - A_{3it} - \gamma_{1}A_{1it} - \underline{a_{3i}})}{\sigma_{\varepsilon_{2}}\sqrt{1 - \underline{\rho_{\varepsilon_{2}\varepsilon_{3}}^{2}}}} \right) H_{2}(a_{2i}, \underline{a_{3i}}|...) da_{2i}.$$
(A.5)

 $G_2(a_{2i}, a_{3i}|...)$ is also derived from the trivariate density of the reduced-form projection errors. It

is equal to $e^{\frac{-1}{2}(\underline{\Lambda}_{22}a_{2i}^2\sigma_{a_2}^{-2})-\underline{\Lambda}_{23}a_{2i}\sigma_{a_2}^{-1}\underline{a}_{3i}\sigma_{a_3}^{-1}}$, and $H_2(a_{2i},\underline{a}_{3i}|\dots)$ is written as

$$H_{2}(a_{2i}, \underline{a_{3i}}|...) = \int_{a_{1i}} G_{3}(a_{1i}, a_{2i}, \underline{a_{3i}}|...) \prod_{t=0_{i}+1}^{T_{i}} \left[\Phi_{1} \left(\frac{-A_{1it} - a_{1i} - \underline{\rho_{\varepsilon_{1}\varepsilon_{3}}}}{\sqrt{1 - \underline{\rho_{\varepsilon_{1}\varepsilon_{3}}}}} \frac{\sigma_{\varepsilon_{3}}^{-1}(y_{3it} - A_{3it} - \gamma_{1}A_{1it} - \underline{a_{3i}})}{\sqrt{1 - \underline{\rho_{\varepsilon_{1}\varepsilon_{3}}}}} \right) \right]^{1-y_{1it}} \left[\Phi_{1} \left(\frac{A_{1it} + a_{1i} + \underline{\rho_{12.3}} \sigma_{\varepsilon_{2}}^{-1}(y_{2it} - A_{2it} - a_{2i}) + \underline{\rho_{13.2}}}{\sqrt{1 - \underline{R_{1.23}}^{2}}} \frac{\sigma_{\varepsilon_{3}}^{-1}(y_{3it} - A_{3it} - \gamma_{1}A_{1it} - \underline{a_{3i}})}{\sqrt{1 - \underline{R_{1.23}}^{2}}} \right) \right]^{y_{1it}} da_{1i},$$

$$(A.6)$$

with $G_3(a_{1i}, a_{2i}, a_{3i}|...)$ equal to $e^{\frac{-1}{2}(\underline{\Lambda_{11}}a_{1i}^2\sigma_{a_1}^{-2})-\underline{\Lambda_{12}}a_{1i}\sigma_{a_1}^{-1}a_{2i}\sigma_{a_2}^{-1}-\underline{\Lambda_{13}}a_{1i}\sigma_{a_1}^{-1}\underline{a_{3i}}\sigma_{a_3}^{-1}}{\underline{\Lambda_{13}}a_{1i}\sigma_{a_1}^{-1}\underline{A_{13}}\sigma_{a_2}^{-1}}$. The three-stage quadrature approach consists in approximating the single integral in equation (A.6) using the formula of equation (3.18) after making an appropriate variable change. Then, $H_2(a_{2i}, a_{3i}|...)$ 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 $H_1(a_{3i}|...)$ into equation (A.4) and apply again the quadrature formula, hence the final expression of l_1 (see equation (3.19)).

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.²¹ 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.²² 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.

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

Model with latent share of innovative sales

 $^{^{21}}$ In the context of panel data, the intraclass correlation is a special form of serial correlation. It is defined as $\frac{\sigma_{a_j}^2}{\sigma_{a_j}^2 + \sigma_{e_j}^2}$ (j = 1, 2, 3). ²²The equicorrelation is about 0.1 when the dynamic sample selection models are estimated on the French inno-

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

$$y_{3it} = \vartheta_{33}y_{3i,t-1} + \beta_3' \mathbf{x}_{3it} + \gamma_2 \left[\vartheta_{22}y_{2i,t-1} + \vartheta_{23}y_{3i,t-1} + \beta_2' \mathbf{x}_{2it} \right] + \underbrace{\gamma_2 \alpha_{2i} + \alpha_{3i}}_{\underline{\alpha_{3i}}} + \underbrace{\gamma_2 \varepsilon_{2it} + \varepsilon_{3it}}_{\underline{\epsilon_{3it}}}, \quad (B.1)$$

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

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

$$\frac{\rho_{\varepsilon_1\varepsilon_3}}{\left(\gamma_2^2\sigma_{\varepsilon_2}^2 + \sigma_{\varepsilon_3}^2 + 2\gamma_2\rho_{\varepsilon_2\varepsilon_3}\sigma_{\varepsilon_2}\sigma_{\varepsilon_3}\right)^{\frac{1}{2}}}, \qquad \frac{\rho_{a_1a_3}}{\left(\gamma_2^2\sigma_{a_2}^2 + \sigma_{a_3}^2 + 2\gamma_2\rho_{a_2a_3}\sigma_{a_2}\sigma_{a_3}\right)^{\frac{1}{2}}}, \qquad (B.2b)$$

$$\underline{\rho_{\varepsilon_2\varepsilon_3}} = \frac{\gamma_2\sigma_{\varepsilon_2} + \rho_{\varepsilon_2\varepsilon_3}\sigma_{\varepsilon_3}}{\left(\gamma_2^2\sigma_{\varepsilon_2}^2 + \sigma_{\varepsilon_3}^2 + 2\gamma_2\rho_{\varepsilon_2\varepsilon_3}\sigma_{\varepsilon_2}\sigma_{\varepsilon_3}\right)^{\frac{1}{2}}}, \qquad \underline{\rho_{a_2a_3}} = \frac{\gamma_2\sigma_{a_2} + \rho_{a_2a_3}\sigma_{a_3}}{\left(\gamma_2^2\sigma_{a_2}^2 + \sigma_{a_3}^2 + 2\gamma_2\rho_{a_2a_3}\sigma_{a_2}\sigma_{a_3}\right)^{\frac{1}{2}}}.$$
 (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 Elasticity of labor productivity with respect to the share of innovative sales

Let the productivity equation be written as

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

where y_{3t} denotes productivity, y_{2t} denotes the share of innovative sales and $logit(y_{2t}) = ln\left(\frac{y_{2t}}{1-y_{2t}}\right)^{23}$. The elasticity of productivity with respect to the share of innovative sales, denoted by $El_{y_{3t},y_{2t}}$, is

 $^{^{23}}$ For simplicity in the notation, we discard the firm subscript *i*, the other regressors and the individual effects.

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

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

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

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

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

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

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

$$S.E.(\widehat{El}_{y_{3t},y_{2t}}) = \frac{S.E.(\widehat{\gamma}_2)}{1 - y_{2t}}.$$
(C.5)

Appendix D 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. Accemoglu 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. Then, looking at the distribution (within each industry) of the technology gap variable, we define

Variable	Coefficient	(Std. Err.)	Coefficient	(Std. Err.)		
	France		The Netherlands			
	Innovation occurrence _t					
Innovation occurrence $t-1$	0.182^{\dagger}	(0.109)	0.063	(0.165)		
Innovation occurrence _{0_i} , 3 waves	0.176	(0.112)	0.710^{**}	(0.186)		
Innovation occurrence _{0_i} , 2 waves	0.352^{**}	(0.062)	0.693^{**}	(0.097)		
Distance to frontier						
$(DQ_2)_{t-1}$	0.050	(0.079)	0.045	(0.108)		
$(DQ_3)_{t-1}$	0.040	(0.082)	0.063	(0.108)		
$(D_{\mathbf{Q}_4})_{t-1}$	0.089	(0.090)	0.126	(0.116)		
$(R\&D/employee)_{t-1}$, in log	0.029	(0.028)	0.177^{**}	(0.045)		
$(D_{non-continuous R\&D})_{t-1}$ Size class	-0.519**	(0.072)	-0.588**	(0.093)		
$D_{\# \text{ employees}} < 50$	-0.447**	(0.130)	-0.261	(0.169)		
$D_{50<\# \text{ employees} \le 250}$	-0.337**	(0.090)	-0.034	(0.143)		
$D_{250<\# \text{ employees}} \leq 500$	-0.099	(0.076)	0.328^{\dagger}	(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. $\mathrm{sales}_{0_i},2$ waves	0.185^{**}	(0.026)	0.171^{**}	(0.027)		
Distance to frontier						
$(DQ_2)_{t-1}$	0.376^{*}	(0.176)	0.006	(0.175)		
$(DQ_3)_{t-1}$	0.112	(0.182)	0.019	(0.176)		
$(DQ_4)_{t-1}$	0.200	(0.202)	-0.124	(0.190)		
$(R\&D/employee)_{t-1}$, in log	0.111^{\dagger}	(0.058)	0.256^{**}	(0.065)		
$(D_{non-continuous R\&D})_{t-1}$	-1.065**	(0.157)	-0.695**	(0.143)		
Size class						
$D_{\# \text{ employees} \leq 50}$	-0.274	(0.309)	0.036	(0.275)		
$D_{50 < \# \text{ employees} \le 250}$	-0.340^{\dagger}	(0.198)	0.068	(0.222)		
$D_{250 < \# employees \le 500}$	0.127	(0.166)	0.146	(0.269)		
Market share $t-1$, in log	0.171**	(0.060)	0.085^{\dagger}	(0.050)		
	Labor productivity _t : sales/employee, in log					
$(\text{Sales/employee})_{t-1}$, 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 $\operatorname{propensity}_t$	0.080^{**}	(0.021)	0.124**	(0.029)		
$(Investment/employee)_t$, in log	0.065**	(0.006)	0.119**	(0.012)		
$\operatorname{Employment}_t$, in \log	-0.029**	(0.009)	-0.083**	(0.018)		
# observations	25	05		1639		
Log-likelihood	-5045.059		-3918.637			

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[‡]

[‡]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%

three dummy variables D_{Q_2} , D_{Q_3} and D_{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 D_{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 D_{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 D_{Q_2} , D_{Q_3} and D_{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.

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.
- 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, 7th 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.
- Hsiao, C., 2003. Analysis of Panel Data, 2nd 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, 2nd 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.