Foreign Linkages, Innovation & Productivity: Evidence from Enterprise Surveys

Anmol Kaur Grewal
Department of International Economics
The Geneva Graduate Institute

Working Paper
Version: September 2023

Foreign Linkages, Innovation & Productivity: Evidence from Enterprise Surveys

Anmol Kaur Grewal[1,2]

Abstract

This paper estimates a three-stage structural model of how foreign linkages affect innovation which in turn affects firm productivity. Using harmonized firm-level data from the World Bank Enterprise Surveys, I construct a panel dataset for 47 developing countries spanning 2003 to 2019. I distinguish between four types of foreign linkages (exports, imports, inward foreign direct investment (FDI) and the use of foreign-licensed technology), and two types of innovation (product innovation and process innovation). To mitigate concerns regarding sample selection and endogeneity biases, I employ advanced panel data methods. In the first stage, I find that being an exporter, using foreign inputs, and using foreign-licensed technology makes firms more likely to invest in R&D, relative to other firms. I also find evidence of sample selection bias which is corrected by using a two-step Heckman selection model. In the second stage, I find that while increases in the R&D intensity increase the probability of product innovation, they have no statistically significant effect on the likelihood of process innovation. In the third stage, I find that product or process innovation is associated with increases in firm-level productivity. These results remain robust across alternative measures of innovation and firm productivity.

[1]Geneva Graduate Institute of International and Development Studies; anmol.grewal@graduateinstitute.ch

[2]I would like to thank my co-supervisors Richard Baldwin and Marcelo Olarreaga for their valuable comments and feedback on this paper. All remaining errors are my own. I am also grateful to the Swiss National Science Foundation for funding this research as a part of the project "Global Value Chains and Firm Performance in Developing Countries" (SNSF Grant number: 100018-159346).

Contents

1 Introduction

This paper aims to contribute to the growing literature on the firm-level dynamics of foreign linkages, innovation and productivity. In an augmented version of the analytical framework originally proposed by Crépon, Duguet, and Mairesse (1998), I use survey data to estimate a structural model of how foreign linkages affect innovation and, in turn, productivity at the firm level.

In the words of Schumpeter, "carrying out innovations is the only function which is fundamental in history". Innovation i.e. the creation of new knowledge is said to be a key driver of economic development (Romer 1990). Based on the works of Griliches (1979) and Pakes and Griliches (1980), the early literature on innovation and productivity adopted the knowledge production function approach with the underlying assumption that innovation inputs determine innovation outputs which then affect productivity. This literature points to a number of channels through which innovation may raise productivity. First, innovation may improve the efficiency with which existing resources are utilised (Hall 2011). Second, innovation may result in structural change in the form of creation of novel sectors which then improves specialisation and drives productivity growth (Katz 2006). Third, innovation via spending on research and development (R&D) activities may also enhance the absorptive capacity of the firm, in turn, facilitating the adoption of new technology and productivity catch-up among weaker firms (Crespi and Zuniga 2012).

Despite the multitude of studies focusing on innovation and productivity, there are still a few key gaps in the literature. First, most of the related research is based on aggregated country-level or sector-level analyses. Among the studies that do use firm-level data, nearly all of them use cross-section data which does not allow one to capture unobserved firm-level heterogeneity (that also goes unaccounted for in the case of more aggregate analyses at the country level). Second, the majority of this literature focuses on advanced economies and so, the firm-level evidence on developing countries is still relatively limited. This is in part due to the scarcity of comparable data on innovation for developing countries where firm-level surveys tend to be relatively costly and cumbersome to carry out. Additionally, in the case of developing economies, innovation is typically more incremental and less radical compared to advanced economies, making it even harder to measure and monitor firms' efforts to innovate.

This paper makes a threefold contribution to the existing literature by filling the aforementioned gaps as follows. First, I use country-wise World

Bank Enterprise Surveys (WBES) to construct a firm-level panel dataset spanning 47 countries over a period of 14 years between 2003 and 2019 [3]. To the best of my knowledge, the dataset so constructed is the most comprehensive in terms of its geographical coverage of developing economies for a study of this type.

Second, the uniquely available panel structure of the World Bank Enterprise Surveys allows me to take into account unobserved firm-level heterogeneity which may be crucial in explaining firms' differences in innovation and productivity across countries. By using advanced panel data methods, I am able to address concerns regarding potential selection bias and endogeneity bias.

Third, this paper contributes to the limited body of empirical work that focuses on firm-level innovation as the mechanism through which foreign linkages may affect productivity. Furthermore, this paper distinguishes between four forms of foreign linkages namely- exports to foreign markets, the use of foreign inputs, inward foreign direct investment (FDI) and the use of foreign-licensed technology. By broadening the definition of foreign linkages, this paper aims to compare the varying degrees to which each type of foreign linkage may affect firm-level innovation and productivity.

By investigating the link between foreign linkages, innovation and productivity, this paper aims to advance the understanding of how emerging markets and developing economies can maximise their potential gains from foreign linkages. Specifically, this paper aims to answer the following research questions: Are firms with foreign linkages more likely to invest in innovation relative to those that don't? Do firms that invest more in R&D tend to produce more innovation output than those that don't? Do firms that produce more innovation output tend to be more productive than those that don't?

The empirical analysis of this paper takes place in three stages, the findings of which can be summarised as follows. In the first stage, I find that being an exporter, using foreign inputs and using foreign-licensed technology makes firms more likely to invest in R&D, relative to other firms. Conditional on the firm's decision to invest, inward FDI has no statistically significant impact on the intensity of R&D investment. I also find evidence of sample selection bias which is corrected by using a two-step Heckman selection model. In the second stage, I find that while increases in the

[3] Following the International Monetary Fund's 2021 classification of countries, 41 of the countries in my sample are classified as "emerging markets and developing economies" while the remaining 6 countries are classified as "advanced economies".

R&D intensity tend to increase the likelihood of introducing a product innovation, they have no statistically significant effect on the likelihood of introducing a process innovation. In the third stage, I find that being a product innovator or a process innovator is associated with increases in firm-level productivity, relative to other firms. These results are robust across a number of specifications and different measures of productivity and innovation.

The rest of the paper is organised as follows. Section 2 discusses the relevant literature and details the contribution of this paper. The theoretical framework and empirical strategy of this paper are presented in Section 3. Section 4 describes the data, Section 5 explains the empirical findings and Section 6 concludes.

2 Literature review

Innovation, typically modelled as the generation of new ideas or blueprints for production, has long been studied as an important driver of productivity growth. Starting with the seminal work by Pakes and Griliches (1980), the early literature on innovation and productivity adopted the knowledge production function approach with the underlying assumption that innovation inputs determine innovation outputs which then affect productivity. The knowledge production function approach gave rise to a new strand of empirical literature that estimates structural models of the relationship between innovation and productivity using survey data. One of the earliest contributions to this strand of the literature was made by Crépon et al. (1998) who proposed a structural framework that explains productivity by innovation output, which in turn is explained by innovation input i.e.e research investment. Using survey data on French manufacturing firms, they found evidence of a positive correlation between firm productivity and higher innovation output, even after controlling for the skill composition of labour and physical capital intensity. Section 3 explains in detail the theoretical underpinnings of the now well-known CDM framework, named after Crépon, Duguet, and Mairesse of Crépon et al. (ibid.).

Building on the CDM model, a number of empirical studies have corroborated the finding of a positive correlation between R&D, innovation and productivity (Crespi and Zuniga 2012; Griffith et al. 2006; Mohnen, Mairesse, et al. 2006; Siedschlag and Zhang 2015). Hall (2011) presents a survey of 25 research papers based on the CDM model and concludes that there are substantial positive impacts of *product* innovation on revenue productivity, however, in the case of *process* innovation, its impact on productivity is more ambiguous. There also exist a number of studies that find no evidence of a relationship in some countries; for example, Benavente (2006) in the case of Chile and Lööf et al. (2001) in the case of Finland.

The mixed evidence from this empirical body of literature is in part due to the differences in the studies' design and methodology (Mohnen and Hall 2013). Typically, country-specific studies based on the CDM model using survey data vary in terms of the sampling methodology, questionnaire design, measures of innovation (e.g. binary or continuous measures, product innovation versus process innovation, etc.) and indicators of economic performance (e.g. labour productivity, multi-factor productivity, sales or profit margins). This restricts the comparability of findings across different studies. To address this issue, Crespi and Zuniga (2012) used harmonized micro-data from innovation surveys and carried out a comparative study across six Latin American countries (Argentina, Chile, Colombia, Costa Rica, Panama, and Uruguay). They used the same specification and identification strategy across all countries and found fairly consistent results. They showed that in all six countries, firms that invest in knowledge are more likely to introduce new technological innovation and those that innovate tend to have greater labour productivity than those that do not.

While the majority of this literature focuses on advanced economies, the firm-level research on developing countries is still relatively limited and rather inconclusive. Cirera and Muzi (2016) offer a comprehensive explanation for why we know so little about the dynamics of firm-level innovation and productivity in developing countries. Firstly, firms in developing economies tend to be much further below the global technological frontier with fairly weak incentives to invest in innovation (Acemoglu et al. 2006). Secondly, implementing innovation surveys is a costly exercise, especially in developing countries where the limited use of internet-based questionnaires implies a heavy reliance on face-to-face interviews for collecting data. Thirdly, the nature of innovation in developing countries tends to be different from that in advanced economies (Cirera and Muzi 2016). Relative to advanced economies, innovation in developing countries tends to be less radical and more incremental in nature, often taking the form of imitation or reverse engineering (Bell and Pavitt 1997). As such, incremental innovation becomes even more subjective and harder to measure and report in developing countries. This lack of harmonized cross-country firm-level innovation surveys leaves a crucial gap in our understanding of the link between innovation and productivity in developing countries.

This paper makes a threefold contribution to the existing literature by filling the aforementioned gap as follows.

First, using country-wise World Bank Enterprise Surveys (WBES), I construct a firm-level panel dataset with the widest possible coverage of emerging markets and developing economies. My sample spans a total of 47 countries over a period of 14 years between 2003 and 2019. To the best of my knowledge, the dataset so constructed is the most comprehensive in terms of its geographical coverage of developing economies for a study of this type. The WBES serves as a source of comparable cross-country innovation data based on a homogeneous set of survey questions asked uniformly to all the firms across different countries.

Second, the uniquely available panel structure of the WBES allows me to take into account unobserved firm-level heterogeneity which is crucial in explaining firms' differences in innovation and productivity across countries. Previous empirical works that have also estimated the CDM model using WBES data include Morris (2018) and Crespi, Tacsir, et al. (2016) among others. This paper contributes to the existing literature by using advanced panel data methods to address concerns about endogeneity and selection bias. This allows us to move beyond correlations and derive causal interpretations from the findings of this study.

Third, this paper contributes to the limited body of empirical work that focuses on firm-level innovation as the mechanism through which foreign linkages may affect productivity. Given that most studies related to this topic are more aggregate country-level or sector-level analyses, they do not typically distinguish between firms that cater to the domestic market and firms that have international linkages. Of the few studies that do so, Castellani and Zanfei (2007) use Italian firm-level data and show that firms with manufacturing activities abroad are characterised by greater R&D efforts, higher propensity to innovate, better innovation performance and the highest productivity premia, relative to other firms. However, they use cross-sectional data and are unable to control for firm-specific unobserved heterogeneity. So, a causal interpretation cannot be drawn from these findings. Criscuolo et al. (2010) adopt the knowledge production function approach in the context of UK manufacturing firms and are able to address concerns regarding potential endogeneity through panel estimation techniques. They find that multinational exporting firms generate more innovation outputs relative to purely domestic firms. In a similar vein, Siedschlag and Zhang (2015) use innovation survey data for Irish firms and showed that firms engaging in international activities (measured by exports and inward FDI) are more likely to invest in innovation, more likely to produce innovation outputs and tend to have higher labour productivity. My research contributes to this body of literature by explicitly modelling foreign linkages into the firm's decision to invest in innovation and its effect on productivity, in turn. In a departure from existing studies, this paper broadens the definition of foreign linkages to include four different measures namely- exports to foreign markets, the use of foreign inputs, inward FDI and the use of foreign-licensed technology. By distinguishing between four forms of foreign linkages, this paper aims to compare the varying degrees to which each type of foreign linkage may affect firm-level innovation and productivity.

3 Empirical methodology

This paper focuses on three key testable hypotheses:

1. Do firms with foreign linkages tend to undertake more innovation effort, relative to firms without foreign linkages?

2. Do firms with higher innovation effort tend to produce more innovation output, relative to other firms?

3. Do firms that have higher innovation output tend to be more productive, relative to those that don't?

Using the structural framework originally proposed by Crépon, Duguet, and Mairesse (1998), this paper aims to address the concerns about endogeneity and selection bias that may arise when testing the above hypotheses. Building upon Siedschlag and Zhang (2015), this paper estimates an augmented version of the three-stage CDM framework which models firm behaviour as follows:

1. In the first stage, firms decide whether or not to invest in R&D activities and how much to invest (i.e. the innovation input).

2. In the second stage, firms produce knowledge (i.e. the innovation output) as a result of the investment in the first stage.

3. In the third stage, final output is produced using the innovation output from the second stage along with other inputs.

3.1 First stage: Innovation input

The first equation models the decision of the firm i at time t to invest in innovation:

$$y^*_{it} = x'_{1it}\beta + \mu_t + \epsilon_{it} \tag{3.1}$$

where y^*_{it} is an unobserved latent variable measuring the predicted utility of investing in innovation, μ_t is a vector of time-fixed effects and ϵ_{it} is an error term with the usual properties. $x_{1it}^{'}$ is a vector of firm-level characteristics that affect innovation effort. These include firm age, size, human capital intensity, obstacles to innovation and the key variables of interest i.e. the four measures of foreign linkages.

In the literature, innovation effort is usually proxied by the natural log of expenditure on R&D per worker. This measure of innovation effort is a latent variable on account of the fact that decisions to invest in innovation are often hard to monitor, measure and report. Managers of small firms, especially those in developing economies, choosing to invest in innovation may do so in very small increments over time. Thus, the actual innovation effort tends to go unreported (or under-reported) unless it exceeds a certain threshold. This means that very few firms end up reporting positive investment in R&D activities at any time (Griffith et al. 2006). This raises concerns about a selection bias. To account for the fact that only the firms' reported innovation effort is observed and to address the resulting problem of selection bias, I assume the following selection equation which describes the propensity of firms to invest in innovation:

$$y_{it} = \begin{cases} 1 & \text{if} \quad y_{it}^{*} = x_{1it}^{'}\beta + \mu_t + \epsilon_{it} > c \\ 0 & \text{if} \quad y_{it}^{*} = x_{it}^{'}\beta + \mu_t + \epsilon_{it} \leq c \end{cases} \tag{3.2}$$

where y_{it} is a binary variable equal to 1 for firms that report investment in innovation input and 0 for firms that report no such investment. Correspondingly, y_{it}^{*} is a latent variable such that firms decide to invest in innovation effort (and report the same) if y_{it}^{*} exceeds a threshold level c.

Conditional on a firm investing in innovation, the actual amount of effort invested in innovation is then observed as follows:

$$w_{it} = \begin{cases} y_{it}^{*} = x_{2it}^{'}\beta + \mu_t + \varphi_{it} & \text{if} \quad y_{it} = 1 \\ 0 & \text{if} \quad y_{it} = 0 \end{cases} \tag{3.3}$$

where w_{it} is the observed innovation expenditure intensity and $x_{2it}^{'}$ is another vector of firm-level characteristics that affect w_{it}. In the empirical analysis that follows, y_{it} is measured by a dummy variable that takes the value 1 if the firm reports a non-zero investment in R&D activities; and 0 otherwise. w_{it} is measured by the log of R&D expenditure per worker. Assuming that ϵ_{it} and φ_{it} are bivariate normal with zero mean, variances equal to one and the correlation coefficient $\rho_{\epsilon\varphi}$, equations (3.2) and (3.3) are estimated jointly using a Heckman two-step selection model for all firms (not only for the subsample of firms that report non-zero R&D intensity).

For identification in the estimation of the first stage equations (3.2) and

(3.3), it is required that there exist at least one variable that affects the decision to invest in innovation but not the intensity of the innovation effort. As is conventional in the literature, I use firm size as the exclusion restriction[4]. Conditional on investing in R&D, the intensity of innovation effort (measured by R&D expenditure per worker) is scaled for firm size and is therefore, less likely to be correlated with the size of the firm. Cohen and Klepper (1996) and Griffith et al. (2006) have shown that larger firms may be more likely to invest in innovation, however, conditional on investing in R&D, large firms do not necessarily invest proportionally more.

Note that the first step of the Heckman selection model is the estimation of equation (3.2) using a Probit regression. Ideally, one should include firm-level fixed effects to control for unobserved firm-specific heterogeneity and mitigate potential endogeneity concerns. However, the so-called "fixed effects Probit" estimator treats the unobserved firm-level heterogeneity θ_i as parameters to be estimated along with β. Not only is this estimator computationally difficult to estimate, but it also results in an incidental parameters problem, unlike in the linear case (Wooldridge 2010). With a small T and large N sample, the estimates of β are inconsistent. Since unobserved firm-level heterogeneity may have an important influence on firm decisions to spend on R&D, I address endogeneity concerns by using the following alternative estimation strategy. The selection equation (3.2) is estimated using Wooldridge's Correlated Random Effects (CRE) Probit model (Wooldridge 2019). The outcome equation (3.3) is estimated using a pooled OLS including year dummies and a set of T Inverse Mills Ratios as regressors (Wooldridge 1995). Both steps are described below.

In the estimation of the selection equation (3.2), the Correlated Random Effects (CRE) Probit model allows for correlation between x'_{1it} and the unobserved firm-level heterogeneity by including within-firm means of time-varying regressors (Mundlak 1978; Wooldridge 2010). In the CRE Probit model specification of equation (3.2), estimated coefficients on the original regressors are interpreted as the estimate of *within-firm* effects whereas the estimates of the within-firm averages of regressors are interpreted as the estimated difference between the *within-firm* and the *between-firm* effects (Schunck 2013).

To estimate the outcome equation (3.3), I first estimate T different probit models of equation (3.2), one for each year in my sample (Mundlak 1978).

[4]Note that even in the absence of a valid exclusion restriction, the outcome equation is identified due to the Inverse Mills Ratio being a non-linear function of the explanatory variables in the selection equation (Wooldridge 2010). The use of firm size as an exclusion restriction here allows the model to be over-identified.

I obtain a set of T Inverse Mills Ratios which are then used as regressors (along with year dummies) in a pooled OLS regression of the outcome equation (3.3). Proposed by Wooldridge (1995), this is a consistent estimator to address sample selection bias in a panel data framework, as an alternative approach to the Heckman selection model.

3.2 Second stage: Innovation output

Next, the second stage of the firm's behaviour is modeled using the following knowledge production function:

$$z_{it} = \widehat{w_{it}}\gamma + x_{3it}^{'}\beta + \mu_t + \omega_{it} \tag{3.4}$$

where z_{it} is a measure of innovation output, $\widehat{w_{it}}$ is the predicted intensity of expenditure on innovation, obtained from the estimation of equations (3.2) and (3.3) for all firms. $x_{3it}^{'}$ is a vector of firm-level characteristics that affect z_{it}, μ_t captures time-fixed effects and ω_{it} is the error term. In order to differentiate between two types of innovation output, I use two dummy variables namely: (i) product innovation which takes the value 1 if the firm introduced any new or significantly improved products (and 0 otherwise); (ii) process innovation which takes the value 1 if the firm introduced any new or significantly improved production processes including methods of supplying services and ways of delivering products (and 0 otherwise).

Given the possibility that the subset of firms reporting investment in innovation may be non-random, there may be concerns about a selection bias. To address this, I follow Griffith et al. (2006) and estimate equation (3.4) for all firms (rather than only for the firms that innovate). To address endogeneity concerns about innovation effort and innovation output being simultaneously determined, I include predicted innovation input ($\widehat{w_{it}}$) (estimated from equation (3.3)) as a regressor in the estimation of equation (3.4). The use of this predicted regressor to proxy for firm's actual R&D intensity offers three advantages. First, it alleviates the endogeneity concerns due to the potential simultaneity between R&D spending and the firm's innovation output, conditional on the exogeneity of the explanatory variables in the Heckman model (Griffith et al. 2006; Hall et al. 2009). Second, by using the predicted levels of R&D intensity, the number of observations in my analysis is expanded to include all firms in the sample, not just the ones that report non-zero R&D spending. Third, this approach allows me to address the issue of selection by reflecting that firms may undertake a certain amount of research effort/spending, even if they do not actually report it (Griffith et al. 2006; Polder et al. 2009). Note that standard errors in the estimation of equation (3.4) are corrected via bootstrapping because of the inclusion of a predicted regressor.

The baseline estimation of equation (3.4) is done using a Linear Probability Model (LPM). Additionally, I estimate equation (3.4) using two alternative specifications: the Pooled Probit model and the Correlated Random Effects (CRE) model (Wooldridge 2019). The rationale behind selecting these specifications is explained below.

Due to its simplicity and ease of interpretation, the LPM is used as the baseline specification to understand the general relationship between the binary outcome z_{it} and the regressors $\widehat{w_{it}}$ and $x^{'}_{3it}$. As noted by Hall (2011), not all innovation effort may be captured by R&D expenditure per worker. Since innovation effort tends to be firm-specific and relatively hard to monitor, it may be the case that unobserved firm-specific factors simultaneously affect both innovation effort and output. This may result in endogeneity unless the unobserved firm-specific factors are controlled for. Therefore, my LPM specification includes firm-level fixed effects θ_i to control for unobservable firm-specific heterogeneity and time-invariant omitted variables that may affect firms' decision to adopt 4IR technologies. The LPM specification also includes industry- and year-fixed effects to control for industry-specific and year-specific factors respectively.

Due to limitations of the LPM model in handling potentially non-linear relationships, I also estimate equation (3.4) using a Pooled Probit model. The Pooled Probit model inherently accounts for nonlinearities between the binary outcome and the predictors. Following Wooldridge (2010), the choice of the Pooled Probit is motivated by two other reasons. First, pooled methods are generally more robust because they do not restrict dependence over time. Second, average partial effects are generally identified by pooled estimation methods and are relatively simple to compute (Wooldridge 2019). Clustering the standard errors in pooled maximum likelihood estimation allows for general serial correlation and consistency of estimates holds (Wooldridge 2010). The same set of industry- and year-fixed effects are also included in the Pooled Probit specification. However, as noted in Section 3.1, the use of firm-level fixed effects is not feasible in the Probit case because of an incidental parameters problem[5](ibid.). Given these considerations, I use the Correlated Random Effects model (Wooldridge 2010, 2019) as an alternative specification.

[5]Alternatively, the conditional logit fixed effects model may be used, however, this estimator is inconsistent in the presence of serial correlation and heteroskedasticity. The assumption of conditional serial independence upon which the conditional logit estimator is based is likely to be violated in the current setting since firms' innovation output may be serially correlated over time (Kwak et al. 2021).

As described earlier, the Correlated Random Effects model allows for correlation between x'_{3it} and the unobserved firm-level heterogeneity by including within-firm means of time-varying regressors (Wooldridge 2010). Assuming that $\theta_i = \pi\overline{x}_i + \nu_i$, equation (3.4) becomes:

$$z_{it} = \widehat{w_{it}}\gamma + x'_{3it}\beta + \pi\overline{x}'_i + \nu_i + \mu_t + \omega_{it} \tag{3.5}$$

where $\overline{x}'_i$ is the vector of within-firm means of regressors[6] over time, which picks up any correlation between the regressors and the error term θ_i.

In the CRE model specification of equation (3.5), β is interpreted as the estimate of *within-firm* fixed-effects whereas π is the estimated difference between the *within-firm* and the *between-firm* effects (Mundlak 1978; Wooldridge 2010). Given this interpretation, the *within-firm* fixed-effects estimates of β from the LPM model of equation (3.4) are comparable with those of the CRE model in equation (3.5) in that both specifications control for unobserved firm-level heterogeneity. Additionally, in equation (3.5), π reflects a comparison of the *within-firm* and *between-firm* effects i.e. the degree to which time-invariant unobserved firm-specific heterogeneity explains observable differences between the dependent variable and the explanatory variables (Schunck 2013).

Both the LPM and the CRE specifications allow me to control for unobserved firm-level heterogeneity (which may be an important source of endogeneity in my model). Additionally, the CRE Probit specification accounts for the binary nature of the dependent variable (i.e. innovation output z_{it}), so I refer to the CRE Probit model as my preferred specification.

3.3 Third stage: Productivity

Finally, to study the impact of innovation output on firm productivity, an augmented Cobb-Douglas production function (with constant returns to scale) is estimated as follows:

$$\pi_{it} = \widehat{z_{it}}\alpha + x'_{4it}\delta + \theta_i + \mu_t + v_{it} \tag{3.6}$$

where π_{it} is a measure of productivity $\widehat{z_{it}}$ is the predicted innovation output estimated using equation (3.5). x'_{4it} is a vector of firm-specific characteristics that affect productivity and v_{it} is the error term. For the dependent variable π_{it} in the third stage equation (3.6), I use (revenue) total factor productivity (TFP) estimated using the Levinsohn and Petrin (2003)

[6]Note that $\overline{x}'_i$ also includes the within-firm mean of predicted R&D intensity estimated in the first stage.

method. As a robustness check, I also use two alternative measures of productivity namely the Olley and Pakes (1996) measure of TFP and labour productivity (i.e. log sales per worker).

In the third stage estimation of equation (3.6), predicted innovation output $\widehat{z_{it}}$ is included as regressor to alleviate potential endogeneity concerns about innovation output and productivity being simultaneously determined, conditional on the exogeneity of the regressors in the second stage. Additionally, I include firm-fixed effects to account for unobserved firm-specific characteristics that may jointly affect innovation output and firm productivity. Industry- and year-fixed effects are also included. Just like in the second stage, standard errors need to be corrected because of the use of a predicted regressor. Therefore, bootstrapping of standard errors is performed in the estimation of equation (3.6).

In sum, the three-stage augmented CDM model outlined above is estimated as a sequential system of equations (3.2) - (3.6)[7]. In the first stage, equations (3.2) and (3.3) are estimated using a Heckman two-step selection model for all firms. The predicted level of innovation intensity is then used to estimate equation (3.5) using a Correlated Random Effects (Probit) model in the second stage. Finally, the predicted value of innovation output is then used to estimate equation (3.6) using a fixed-effects regression in the third stage.

4 Data

For the empirical estimation of the aforementioned model, I use country-wise World Bank Enterprise Surveys (WBES) to construct a firm-level panel dataset spanning 47 countries across 14 years between 2003-2019. See Tables A1 for the distribution of firms by country and year. The WBES are firm-level surveys of a representative sample of an economy's private sector. Other than firm characteristics, the surveys cover a broad range of business environment topics including access to finance, corruption, infrastructure, innovation and technology, competition, and performance measures[8]. WBES data is available for 180,000 firms in 154 countries. However, for

[7]The model does not allow for feedback effects between the equations.

[8]Since 2005-06, the WBES have followed a "Global Methodology" to ensure harmonization of the surveys across different countries. Formal (registered) companies from both manufacturing and services sectors with five or more employees are surveyed, excluding those firms that are wholly government-owned. The survey is answered by business owners and top managers. The sampling methodology for Enterprise Surveys is stratified random sampling. For more information, please visit `https://www.enterprisesurveys.org/en/methodology`.

the purpose of this study, the sample is limited on the basis of two requirements. First, I include only those countries in which firms were surveyed at least twice so that a dataset with a panel dimension could be constructed. Second, since this study focuses on innovation, I include only those countries in which the innovation component of the survey was carried out. The resulting panel dataset covers a sample of 47 countries, 41 of which are classified as emerging markets and developing economies. To the best of my knowledge, this dataset has the largest cross-country coverage for a study of this type.

Table 4.1 summarises the variables of interest for this analysis. Full definitions and associated survey questions for each variable are listed in Table A2. Innovation effort is measured by two variables: (i) a dummy variable that captures the firm's decision to report investment in R&D activities; (ii) the amount of R&D expenditure per worker. Innovation output is measured by two dummy variables to distinguish between product innovation and process innovation. Using the information in the innovation module of the WBES, a firm is defined as an "innovator" if it introduced a new or significantly improved product or production process in the last three years[9] . In this sample, 44 percent of the firms are reported to be product innovators and 36 percent of the firms are reported to be process innovators. 53 percent of the firms in this sample are either product innovators or process innovators. One of the concerns regarding the use of a self-reported dummy variable to measure innovation output is that it does not fully capture the heterogeneity between different kinds of innovation introduced by the firm. The constraints of the survey data available so far prevent the use of an alternative variable to measure innovation. In the absence of more objective measures of innovation, I exploit country variations using the cross-country panel dataset, thereby minimizing to some extent the potential measurement error in the innovation variables.

As explained earlier, this paper uses four different measures of firms' foreign linkages. These include dummy variables that capture firms' exports to foreign markets, imports of foreign inputs, inward foreign direct investment and the use of foreign technology via licensing. In the third stage of the model, three measures of productivity are used: total factor productivity (revenue TFP) estimated from the Levinsohn-Petrin method, TFP estimated from the Olley-Pakes method and labour productivity (measured

[9]The exact questions posed in the survey are as follows: "Over the last three fiscal years, (i) Did this establishment introduce onto the market any new or significantly improved products?; (ii) Has this establishment introduced any new or significantly improved production processes including methods of supplying services and ways of delivering products?"

by log sales per employee). Other explanatory variables used in this analysis include age and size of the firm, managerial experience as a proxy for the firm's level of human capital, access to a line of credit, capital intensity and the cost of material inputs [10].

Table 4.1: Summary statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
(0/1) Firm invested in R&D	30580	.19	.39	0	1
Log R&D expenditure per worker (2010 USD)	5806	5.58	1.76	.02	14.38
(0/1) Product Innovation	30300	.44	.5	0	1
(0/1) Process Innovation	27754	.36	.48	0	1
(0/1) Product or Process Innovation	27474	.53	.5	0	1
Firm age	30580	30.24	18.77	3	221
Firm size	30580	106.98	379.29	5	21955
Share of skilled workers	29519	42.56	28.36	0	100
(0/1) Firm has 10% or more of foreign ownership	30580	.09	.29	0	1
(0/1) More than 10% of sales are direct exports	30580	.23	.42	0	1
(0/1) More than 10% of inputs are imported	30580	.57	.49	0	1
(0/1) Foreign-licensed technology	30580	.15	.36	0	1
Manager's experience	30580	19.85	11.58	0	54
(0/1) Line of credit	30580	.46	.5	0	1
Capacity utilisation (%)	30580	72.92	21.91	0	100
Log labour productivity i.e. log sales per worker (2010 USD)	25538	9.95	1.87	0	21.05
Log of net book value of fixed assets (2010USD)	15617	12.08	2.77	.04	28.97
Log of cost of raw materials & intermediate goods (2010USD)	22566	12.12	2.79	.15	22.37

Table 4.2 shows key indicators related to firms' foreign linkages and innovation activity by country. One of the noteworthy insights from this table is that for almost all countries, a far greater share of firms introduced a new product relative to those that report any expenditure on R&D. This points to the possibility that some firms may not report their expenditure on R&D if it falls below a certain level (Griffith et al. 2006). Moreover, some firms may undertake innovation even if they report zero R&D spending. It is precisely this under-reporting of innovation effort that may result in a selection bias. To remedy this, I use a Heckman selection model, as described in the Section 3.1.

Another point to note from Table 4.2 is that in nearly all the countries, the share of firms that are product innovators tends to be higher than the share of firms that are process innovators. Also, the share of firms that use imported inputs tends to be higher than the share of firms that export to foreign markets. Finally, only a small proportion of firms tend to use foreign-licensed technology, so the scope of disembodied transfer of technology seems somewhat limited in this sample of countries. This may be due to the lack of complementary factors like adequate human capital and supporting infrastructure that may facilitate the use of foreign technology in emerging markets and developing economies.

[10] All monetary variables originally recorded in local currency units are converted into standard 2010 USD terms by using the consumer price index from the World Development Indicators and annual averaged exchange rate from the Penn World Tables.

Table 4.2: Firm characteristics by country

Country	(1) Invest in R&D	(2) Product Innovation	(3) Process Innovation	(4) Both Product & Process Innovation	(5) Product or Process Innovation	(6) FDI	(7) Exporter	(8) Foreign inputs	(9) Foreign-licensed Technology
Albania	0.08	0.38	0.26	0.16	0.49	0.16	0.43	0.82	0.28
Argentina	0.37	0.63	0.52	0.42	0.73	0.10	0.26	0.60	0.14
Armenia	0.12	0.41	0.22	0.14	0.28	0.10	0.20	0.79	0.28
Azerbaijan	0.02	0.24	0.02	0.01	0.04	0.09	0.08	0.28	0.22
Bangladesh	0.32	0.29	0.33	0.19	0.44	0.02	0.22	0.49	0.09
Belarus	0.15	0.54	0.50	0.29	0.64	0.17	0.25	0.71	0.08
Bolivia	0.39	0.71	0.64	0.53	0.82	0.09	0.15	0.80	0.19
Bulgaria	0.10	0.25	0.18	0.10	0.33	0.07	0.34	0.52	0.16
Chile	0.34	0.62	0.55	0.46	0.72	0.10	0.17	0.71	0.15
Colombia	0.46	0.66	0.61	0.48	0.80	0.06	0.17	0.61	0.11
Croatia	0.15	0.50	0.27	0.18	0.59	0.11	0.51	0.73	0.13
Czech Republic	0.32	0.48	0.35	0.21	0.61	0.21	0.60	0.74	0.16
Dominican Republic	0.18	0.44	0.27	0.19	0.53	0.19	0.25	0.65	0.23
Ecuador	0.41	0.61	0.51	0.41	0.71	0.14	0.12	0.68	0.12
El Salvador	0.23	0.52	0.45	0.36	0.60	0.12	0.31	0.68	0.15
Estonia	0.16	0.48	0.29	0.17	0.55	0.26	0.62	0.81	0.19
Ethiopia	0.05	0.40	0.36	0.27	0.49	0.12	0.09	0.44	0.20
Georgia	0.07	0.38	0.18	0.12	0.43	0.11	0.21	0.62	0.19
Guatemala	0.35	0.48	0.45	0.27	0.67	0.10	0.27	0.62	0.17
Honduras	0.17	0.56	0.51	0.40	0.67	0.11	0.14	0.68	0.17
Hungary	0.12	0.38	0.25	0.11	0.39	0.20	0.36	0.47	0.11
Jordan	0.03	0.20	0.11	0.06	0.24	0.10	0.40	0.66	0.37
Kazakhstan	0.08	0.29	0.14	0.08	0.31	0.05	0.05	0.56	0.11
Kenya	0.14	0.52	0.42	0.31	0.62	0.15	0.25	0.51	0.18
Kyrgyz Republic	0.09	0.45	0.35	0.28	0.52	0.17	0.19	0.61	0.21
Latvia	0.18	0.50	0.38	0.24	0.55	0.21	0.54	0.71	0.26
Lebanon	0.11	0.33	0.20	0.13	0.40	0.03	0.37	0.63	0.10
Lithuania	0.09	0.43	0.21	0.11	0.45	0.17	0.53	0.72	0.20
Mexico	0.28	0.42	0.37	0.29	0.50	0.09	0.14	0.44	0.13
Montenegro	0.11	0.35	0.20	0.10	0.30	0.09	0.22	0.80	0.15
Morocco	0.04	0.14	0.11	0.07	0.18	0.17	0.28	0.45	0.15
Myanmar	0.02	0.24	0.30	0.16	0.38	0.07	0.13	0.30	0.06
Nicaragua	0.18	0.44	0.46	0.24	0.63	0.08	0.12	0.60	0.09
Panama	0.15	0.39	0.38	0.28	0.49	0.10	0.09	0.62	0.13
Paraguay	0.28	0.64	0.54	0.46	0.73	0.11	0.16	0.77	0.12
Peru	0.44	0.70	0.63	0.51	0.83	0.11	0.28	0.72	0.12
Poland	0.05	0.25	0.10	0.06	0.26	0.05	0.22	0.42	0.11
Romania	0.17	0.49	0.55	0.34	0.70	0.17	0.32	0.65	0.18
Russian Federation	0.10	0.34	0.21	0.13	0.37	0.04	0.07	0.54	0.12
Slovakia	0.13	0.37	0.19	0.12	0.35	0.15	0.45	0.59	0.39
Slovenia	0.34	0.69	0.48	0.42	0.75	0.18	0.73	0.81	0.18
Tajikistan	0.02	0.43	0.22	0.14	0.42	0.12	0.14	0.53	0.22
Turkey	0.05	0.16	0.07	0.03	0.14	0.03	0.33	0.36	0.22
Ukraine	0.09	0.37	0.19	0.14	0.39	0.07	0.19	0.50	0.16
Uruguay	0.32	0.64	0.54	0.43	0.76	0.10	0.23	0.87	0.10
Uzbekistan	0.05	0.31	0.16	0.11	0.36	0.16	0.14	0.30	0.24
Zimbabwe	0.05	0.30	0.30	0.18	0.42	0.17	0.15	0.60	0.15

Values in the table indicate the share of firms column-wise corresponding to each characteristic by country

Despite the many advantages of the WBES dataset that make it well-suited for this analysis, there are a few limitations one must be aware of. First, the WBES methodology does not cover informal firms. Second, there is a relatively low representation of firms from the services sector in this sample. This is because services firms were excluded in some of the key modules of past survey questionnaires, especially the innovation module which is crucial for this study. So, one must note that the findings of this paper are based on the analysis of formal-sector firms predominantly in the manufacturing sector.

5 Results

5.1 First Stage: Innovation Input

The first stage of the augmented CDM framework models the firm's decision to invest in R&D. This is the "Innovation Input" stage. To address concerns regarding the endogenous sample selection in this stage, I use a Heckman sample selection model. This model is estimated via the two-step efficient estimator, the results of which are presented in Table 5.1. Column (1) tabulates the results from the outcome equation where the dependent variable is innovation intensity i.e. the log of R&D investment per worker. Column (2) tabulates the results from the selection equation with the dependent variable being a dummy variable whether or not the firm invests in R&D. Starting with the selection equation in column (2), firm age seems to have no statistically significant effect on the firm's likelihood of investing in R&D. Inward FDI has a negative and statistically significant effect on the decision to invest whereas the other three forms of foreign linkages all have a positive and statistically significant impact on the decision to invest. Having inward FDI makes a firm 6.8 percentage points less likely to invest in R&D, relative to other firms. Being an exporter increases the probability of investing in R&D by 29 percentage points, relative to non-exporting firms. Similarly, the use of foreign inputs increases the likelihood of investing in R&D by 34 percentage points, relative to firms that use domestic inputs. The use of foreign-licensed technology increases the likelihood of investing in R&D by 34 percentage points, relative to firms that don't use foreign-licensed technology. Other control variables such as firm size, access to credit and experience of the manager also have a positive and statistically significant effect on the decision to invest in R&D as expected.

The relationships noted above between the firms' foreign linkages and their decision to invest in R&D may be driven by the following mechanisms. Firms receiving inward FDI tend to benefit from knowledge spillovers directly from their foreign parent companies, reducing the firm's incentive to undertake their own investment in R&D. This may be the reason why inward FDI is seen to have a negative effect on the firm's decision to invest in R&D. Exporting firms may have a greater incentive to invest in R&D because of competitive pressures faced from foreign firms and/or because of the potential knowledge spillovers from exporting to foreign markets. The use of foreign inputs may mean greater access to better quality of inputs and/or better technology embodied in those inputs. Both these aspects would result in greater incentives to invest in R&D and undertake innovation either in the form of new products produced using the greater variety of foreign inputs and/or new processes using the technology embodied in those inputs. The use of foreign-licensed technology represents the import of disembodied technology from abroad. This may spur firms to invest in R&D, especially in developing economies where innovation typically takes the form of imitation or reverse engineering of foreign products.

Turning to column (1) of Table 5.1, the first thing to note is that firm size is not included as a regressor here because it is an exclusion restriction. Identification in the first stage requires that there exists a variable that affects the firm's decision to invest in R&D but not the intensity of the firm's innovation effort i.e. the amount of R&D invested per worker. Following the convention in the literature, I use firm size as the exclusion restriction[11]. Larger firms may be more likely to invest in innovation, however, conditional on investing in R&D, large firms do not necessarily invest proportionally more (Cohen and Klepper 1996; Griffith et al. 2006).

Column (1) shows a log-level specification with innovation intensity (i.e. log R&D expenditure per worker) as the dependent variable. Conditional on the firm's decision to invest, inward FDI has no statistically significant impact on the innovation intensity. Exporting firms tend to have 68 percent higher R&D intensity, relative to non-exporting firms. Firms using foreign inputs tend to have 86 percent higher R&D intensity compared to those that use domestic inputs. Firms that use foreign-licensed technology tend to have 78 percent higher R&D intensity compared to those that don't. Again, access to credit and the experience of the manager also have a positive and statistically significant effect on the innovation inten-

[11]As a robustness check, I tried an alternative specification of the Heckman model including firm size as a regressor in both the selection as well as outcome equations (results available upon request). In this case, the outcome equation is said to be just-identified on account of the Inverse Mills Ratio being a non-linear function of the explanatory variables in the selection equation. As expected, firm size turned out to be statistically significant in the selection equation but statistically insignificant in the outcome equation. Taking this as evidence that firm size affects the decision to invest in R&D but not the intensity of R&D spending, I proceed with using firm size as a valid exclusion restriction, allowing the model to be over-identified.

sity. The coefficient on the Inverse Mills Ratio may be interpreted as a test for sample selection since it represents the covariance between the errors in the selection and the outcome equations, under the assumptions of the Heckman model. The estimated Inverse Mills Ratio is found to be statistically significant, indicating the presence of selection bias in this sample.

As described in Section 3.1, since it is not possible to include firm-fixed effects to control for unobserved firm-level heterogeneity in the Heckman model, I use an alternative approach to address the issue of sample selection in a panel data framework. The results are presented in Table 5.2. The selection equation is estimated using a Correlated Random Effects (CRE) Probit while the outcome equation is estimated using the estimator proposed by Wooldridge (1995). Starting with the selection equation in column (2), the estimates are qualitatively the same as those in Table 5.1, in terms of expected signs and statistical significance. However, in terms of the magnitude, the estimates are all much smaller compared to those obtained in the Heckman two-step model of Table 5.1. This can explained by the fact that the CRE model allows regressors to be correlated with unobserved firm-level heterogeneity. Therefore, the first ten rows of estimates in Table 5.2 are interpreted as the *within-firm* effects of regressors on the propensity to invest in R&D, while the next ten rows are interpreted as the *between-firm* effects. In the Heckman model of Table 5.1, however, firm-level heterogeneity is unaccounted for, therefore the estimated coefficients seem to be biased upwards.

Turning to column (1), the estimated *within-firm* effects of regressors in the outcome equation in Table 5.2 are also similar to those in Table 5.1 in terms of signs and statistical significance but smaller in magnitude. Again, I find the Inverse Mills Ratio to be statistically significant, which indicates the presence of selection bias in this sample. I correct this selection bias by including the estimated Inverse Mills Ratio as a regressor in my estimation of the second stage and third stage equations.

Table 5.1: First stage: Heckman selection model

	Outcome equation	Selection equation
	(1)	(2)
	OLS	Probit
VARIABLES	Log R&D intensity	(0/1) Firm invests in R&D
Firm age	-0.0056	-0.00069
	(0.0041)	(0.0015)
Firm age squared	0.000079**	0.000016
	(0.000033)	(0.000012)
FDI	0.12	-0.068**
	(0.090)	(0.032)
Exporter	0.68***	0.29***
	(0.090)	(0.023)
Importer	0.86***	0.34***
	(0.089)	(0.021)
Foreign-licensed technology	0.78***	0.34***
	(0.095)	(0.025)
Log of firm size		0.14***
		(0.0082)
Manager's experience	0.012***	0.0022**
	(0.0027)	(0.00089)
Line of credit	0.76***	0.33***
	(0.089)	(0.020)
Capacity utilisation	-0.0052***	-0.0025***
	(0.0014)	(0.00045)
Inverse Mills ratio	2.71***	
	(0.23)	
Constant	1.25***	-1.60***
	(0.48)	(0.076)
Observations	30,580	30,580
Year FE	Yes	Yes
Industry FE	Yes	Yes

Estimates reported are marginal effects.
Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 5.2: First stage: Alternative specification

VARIABLES	Outcome equation (1) Wooldridge (1995) estimator Log R&D intensity	Selection equation (2) CRE Probit (0/1) Invest in R&D
Firm age	-0.0092***	0.0017
	(0.0035)	(0.0013)
Firm age squared	0.000089***	-0.000013
	(0.000030)	(0.000010)
FDI	0.088	0.014
	(0.074)	(0.027)
Exporter	0.31***	0.035*
	(0.061)	(0.018)
Importer	0.49***	0.061***
	(0.060)	(0.013)
Foreign-licensed technology	0.45***	0.091***
	(0.065)	(0.017)
Log of firm size		0.055***
		(0.010)
Manager's experience	0.0098***	0.00014
	(0.0020)	(0.00056)
Line of credit	0.34***	0.041***
	(0.060)	(0.013)
Capacity utilisation	-0.0031***	-0.00023
	(0.0012)	(0.00027)
Firm age (average)		-0.0018
		(0.0014)
Firm age squared (average)		0.000018*
		(0.000011)
FDI (average)		-0.042
		(0.028)
Exporter (average)		0.025
		(0.019)
Importer (average)		0.032**
		(0.014)
Foreign-licensed technology (average)		0.011
		(0.017)
Log of firm size (average)		-0.0071
		(0.010)
Capacity utilisation (average)		0.00054
		(0.00059)
Line of credit (average)		0.032**
		(0.014)
Capacity utilisation (average)		-0.00028
		(0.00030)
Inverse Mills Ratio	1.48***	
	(0.13)	
Constant	3.51***	
	(0.31)	
Observations	5,804	34,323
R-squared	0.092	
Year FE	Yes	Yes
Industry FE	Yes	Yes
Firm FE	No	No

Estimates reported are marginal effects.
*** p<0.01, ** p<0.05, * p<0.1
Standard errors (in parentheses) are clustered at the firm level.

5.2 Second Stage: Innovation Output

In the second stage of the CDM framework, I use the predicted R&D intensity from the first stage to model how firms produce innovation output as a result of the innovation effort undertaken in the first stage. This is done by estimating three alternative models: a Linear Probability Model, a Pooled Probit model, and a Correlated Random Effects (CRE) Probit model. Standard errors are bootstrapped due to the use of a predicted regressor i.e. predicted log R&D expenditure as a share of sales. Additionally, standard errors are clustered at the firm level to allow for arbitrary serial correlation in errors within the same firm over time[12]. Table 5.3 presents the results from the second stage using product innovation as the dependent variable while Table 5.4 presents the second stage results with process innovation as the dependent variable.

The LPM specification includes firm-fixed effects to account for unobserved firm-level heterogeneity and the underlying sample contains only those firms that have been surveyed twice (i.e. no singletons are included in this sample). That is why the number of observations in column (1) is much smaller than those in columns (2) and (3). In the LPM specification i.e. column (1) of Table 5.3, I find that predicted R&D intensity has no statistically significant effect on the likelihood of introducing a product innovation. Having inward FDI, using foreign inputs and using foreign technology seem to have no statistically significant effect on the likelihood of introducing a product innovation. However, exporting firms are 12 percentage points more likely to introduce product innovations, relative to non-exporting firms.

The results of the Pooled Probit and CRE Probit models in column (2) and (3) account for the binary nature of the dependent variable and are comparable in terms of estimated marginal effects. From columns (2) and (3), I find that a 100 percent increase in the predicted R&D intensity tends to increase the probability of product innovation by 17-18 percentage points. A 100 percent increase in the size of the firm tends to increase the probability of product innovation by 4.2-6.6 percentage points. All four of the foreign linkages variables seem to have no statistically significant *within-firm* effect on the likelihood of product innovation in the CRE Probit specification.

Turning to the results for process innovation in 5.4, again, the LPM specification shows that predicted R&D intensity has no statistically significant

[12]These results are also robust to clustering at the country level, allowing errors to be correlated across firms within the same country. These results are available upon request.

effect on the likelihood of introducing a process innovation. The only determinant of process innovation in the LPM specification seems to be access to a line of credit, which makes firms 15 percentage points more likely to introduce process innovation.

Contrary to the results for product innovation, in columns (2) and (3), I find that a 100 percent increase in the predicted R&D intensity makes firms less likely to introduce a process innovation. This difference may be explained as follows. Process innovation involves improving existing internal operations and means of production and delivery. These improvements often occur through informal, tacit knowledge spillovers that may not necessarily require increases in formal R&D spending. Product innovation, on the other hand, involves the development of new or improved products or services which often requires substantial investment in formal R&D activities, including research laboratories, dedicated experts, and the acquisition of new technologies or knowledge.

From the CRE Probit specification in column (3), I find that being an exporter increases the likelihood of process innovation by 9 percentage points, relative to non-exporting firms. Using foreign inputs and foreign-licensed technology both increase the likelihood of process innovation by 11 percentage points each. Exposure to competitors in foreign markets as well as access to a greater variety of (potentially cheaper and/or higher-quality) inputs from abroad may result in positive spillovers in terms of improved production processes. Similarly, the use of foreign-licensed technology may facilitate more advanced production techniques, that may not be available domestically or may be prohibitively expensive to develop internally. Access to credit and the years of experience of the manager also affect the likelihood of process innovation positively.

Table 5.3: Second stage: Product Innovation

VARIABLES	(1) LPM-FE (0/1) Product Innovation	(2) Pooled Probit (0/1) Product Innovation	(3) CRE Probit (0/1) Product Innovation
Predicted R&D intensity	0.024	0.18***	0.17***
	(0.14)	(0.037)	(0.042)
Firm age	0.0046*	0.0016***	0.0051***
	(0.0024)	(0.00048)	(0.0017)
Firm age squared	-0.000039**	-0.000012**	-0.000038***
	(0.000020)	(4.6e-06)	(0.000013)
FDI	0.018	-0.033***	-0.037
	(0.058)	(0.013)	(0.032)
Exporter	0.12*	-0.027	0.032
	(0.070)	(0.018)	(0.033)
Importer	0.091	0.065***	0.0032
	(0.084)	(0.022)	(0.032)
Foreign-licensed technology	0.079	0.017	0.0014
	(0.076)	(0.021)	(0.028)
Log of firm size	0.059	0.042***	0.066***
	(0.036)	(0.0083)	(0.015)
Manager's experience	0.0012	-0.0013***	-0.00060
	(0.0018)	(0.00045)	(0.00091)
Line of credit	0.11	0.032	0.013
	(0.073)	(0.020)	(0.028)
Inverse Mills Ratio	0.18	0.25***	0.26***
	(0.28)	(0.070)	(0.082)
Predicted R&D intensity (average)			0.016**
			(0.0065)
Firm age (average)			-0.0037**
			(0.0017)
Age squared (average)			0.000027*
			(0.000014)
FDI (average)			0.0030
			(0.031)
Exporter (average)			-0.065**
			(0.028)
Importer (average)			0.066***
			(0.023)
Foreign-licensed technology (average)			0.017
			(0.027)
Log firm size (average)			-0.025*
			(0.014)
Manager's experience (average)			-0.00074
			(0.00087)
Line of credit (average)			0.019
			(0.022)
Constant	-0.33		
	(0.70)		
Observations	7,268	30,580	30,300
R-squared	0.644		
Year FE	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes
Firm FE	Yes	No	No

Estimates reported are marginal effects.
Bootstrapped standard errors (in parentheses) are clustered at the firm level.
*** p<0.01, ** p<0.05, * p<0.1

In Tables 5.5 and 5.6, I repeat the second stage estimation using interaction terms as additional regressors. Specifically, I interact the level of predicted R&D intensity with the four foreign linkages variables: *FDI, Exporter, Im-*

Table 5.4: Second stage: Process Innovation

VARIABLES	(1) LPM-FE (0/1) Process Innovation	(2) Pooled Probit (0/1) Process Innovation	(3) CRE Probit (0/1) Process Innovation
Predicted R&D intensity	-0.020	-0.059**	-0.072**
	(0.19)	(0.026)	(0.035)
Firm age	0.00094	0.00050	0.0021
	(0.0025)	(0.00046)	(0.0019)
Firm age squared	-5.4e-07	-2.7e-06	-5.9e-06
	(0.000023)	(4.1e-06)	(0.000016)
FDI	-0.011	0.019	0.028
	(0.063)	(0.012)	(0.039)
Exporter	0.097	0.058***	0.090***
	(0.083)	(0.016)	(0.031)
Importer	0.15	0.13***	0.11***
	(0.10)	(0.017)	(0.031)
Foreign-licensed technology	0.16	0.12***	0.11***
	(0.096)	(0.016)	(0.030)
Log of firm size	0.070	0.011*	0.026
	(0.047)	(0.0059)	(0.017)
Manager's experience	0.0027	0.00078**	0.0020**
	(0.0021)	(0.00038)	(0.00093)
Line of credit	0.15*	0.13***	0.10***
	(0.088)	(0.015)	(0.030)
Inverse Mills Ratio	0.27	-0.084*	-0.075
	(0.36)	(0.050)	(0.069)
Predicted R&D intensity (average)			0.020***
			(0.0068)
Firm age (average)			-0.0016
			(0.0020)
Age squared (average)			2.7e-06
			(0.000016)
FDI (average)			-0.010
			(0.039)
Exporter (average)			-0.036
			(0.025)
Importer (average)			0.018
			(0.019)
Foreign-licensed technology (average)			0.0081
			(0.021)
Log firm size (average)			-0.014
			(0.013)
Manager's experience (average)			-0.0014*
			(0.00074)
Line of credit (average)			0.027
			(0.021)
Constant	-0.52		
	(0.93)		
Observations	6,463	27,754	27,754
R-squared	0.606		
Year FE	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes
Firm FE	Yes	No	No

Estimates reported are marginal effects.
Bootstrapped standard errors (in parentheses) are clustered at the firm level.
*** p<0.01, ** p<0.05, * p<0.1

porter and Foreign-licensed Technology. The use of these interaction terms allows me to check whether the impact of R&D spending on the likelihood of product or process innovation varies depending on the firm's foreign linkages.

The results for product innovation in Table 5.5 are very similar to those in Table 5.3. A 100 percent increase in the predicted R&D intensity increases the likelihood of product innovation by 19.8 percentage points in the CRE Probit specification. For the LPM specification, the estimated coefficients of the interaction terms are presented in Table 5.3 whereas for the Pooled Probit and CRE Probit models, the marginal effects are plotted in Figure 2. Most of the interaction terms are not found to be jointly statistically significant with the predicted R&D intensity variable, except for a few. For example in panel (c) of Figure 2, results from the CRE Probit specification indicate that the marginal effect of R&D intensity on product innovation is smaller for firms receiving inward FDI, compared to those without FDI. This is consistent with the hypothesis that firms receiving inward FDI tend to benefit from knowledge spillovers directly from their foreign parent companies, reducing the firm's incentive to undertake their own innovation efforts. I also find that the marginal effect of R&D intensity on product innovation is larger for exporting firms, compared to non-exporting ones.

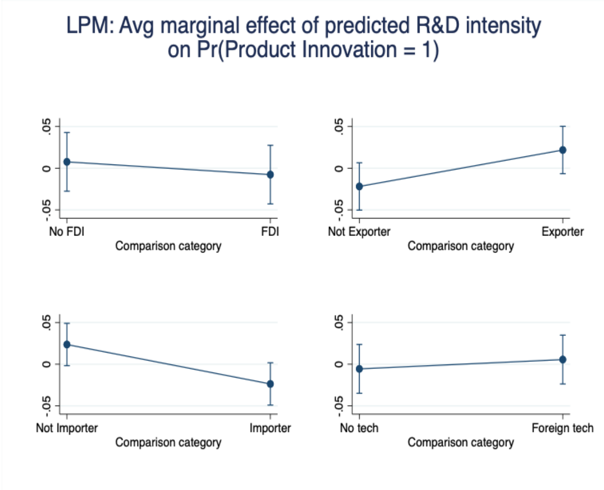
Table 5.6 presents the results for process innovation with interaction effects. Upon including interaction terms, I find that predicted R&D intensity has no statistically significant effect on the probability of process innovation. The other estimated marginal effects are similar in magnitude to those in Table 5.4. Again, most of the interaction terms are not found to be jointly statistically significant with the predicted R&D intensity variable, except for the FDI interaction term. The marginal effects of interaction terms are plotted in Figure 3. In the CRE Probit specification (panel (c) of Figure 3), I find that the marginal effect of R&D spending on the likelihood of process innovation is smaller for firms with inward FDI, compared to those without FDI.

Table 5.5: Second stage: Product Innovation (with interaction effects)

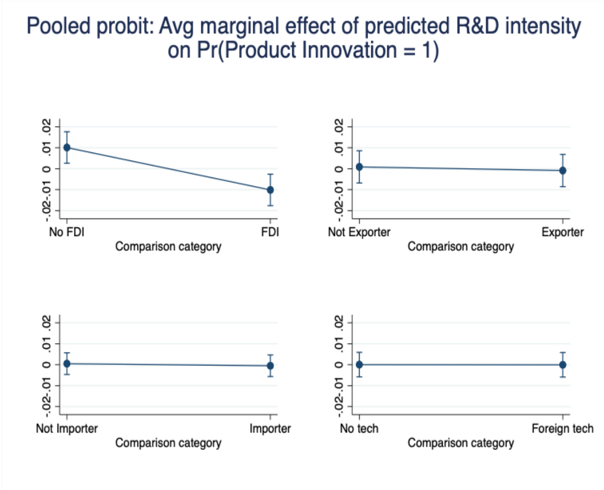
VARIABLES	(1) LPM-FE (0/1) Product Innovation	(2) Pooled Probit (0/1) Product Innovation	(3) CRE Probit (0/1) Product Innovation
Predicted R&D intensity	0.0442	0.213***	0.198***
	(0.153)	(0.0420)	(0.0442)
Firm age	0.00457*	0.00175***	0.00525***
	(0.00261)	(0.000424)	(0.00197)
Firm age squared	-0.0000388*	-0.0000135***	-0.0000391**
	(0.0000204)	(0.00000399)	(0.0000173)
FDI	0.0265	-0.0390***	-0.0445
	(0.0626)	(0.0114)	(0.0351)
Exporter	0.0779	-0.0308	0.0269
	(0.0712)	(0.0209)	(0.0335)
Importer	0.118	0.0579**	-0.000943
	(0.0856)	(0.0245)	(0.0314)
Foreign technology	0.0673	0.0164	0.00126
	(0.0741)	(0.0226)	(0.0326)
Log of firm size	0.0633	0.0520***	0.0758***
	(0.0393)	(0.00931)	(0.0191)
Manager's experience	0.000998	-0.00153***	-0.000873
	(0.00174)	(0.000503)	(0.000617)
Line of credit	0.103	0.0294	0.0104
	(0.0647)	(0.0217)	(0.0297)
Inverse Mills Ratio	0.201	0.345***	0.342***
	(0.311)	(0.0773)	(0.101)
FDI × Predicted R&D intensity	-0.00765		
	(0.0179)		
Exporter × Predicted R&D intensity	0.0219		
	(0.0145)		
Importer × Predicted R&D intensity	-0.0237*		
	(0.0130)		
Foreign technology × Predicted R&D intensity	0.00559		
	(0.0149)		
Predicted R&D intensity (average)			0.0162**
			(0.00761)
Firm age (average)			-0.00371*
			(0.00202)
Age squared (average)			0.0000272
			(0.0000179)
FDI (average)			0.00582
			(0.0355)
Exporter (average)			-0.0643**
			(0.0308)
Importer (average)			0.0640***
			(0.0210)
Foreign-licensed technology (average)			0.0168
			(0.0266)
Log firm size (average)			-0.0252*
			(0.0142)
Manager's experience (average)			-0.000710
			(0.000617)
Line of credit (average)			0.0202
			(0.0209)
Constant	-0.395		
	(0.801)		
Observations	7268	30300	30300
R^2	0.645		
Year FE	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes
Firm FE	Yes	No	No

Estimates reported are marginal effects.
Bootstrapped standard errors (in parentheses) are clustered at the firm level.
*** p<0.01, ** p<0.05, * p<0.1

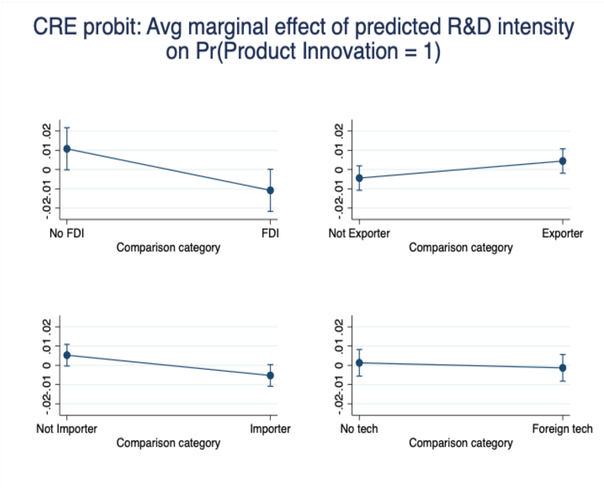
Figure 2: Marginal (interaction) effects of R&D intensity on Product Innovation



(a) LPM



(b) Pooled Probit



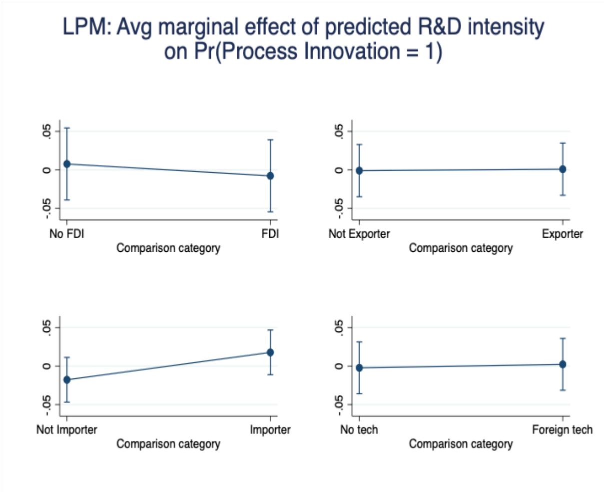
(c) CRE Probit

Table 5.6: Second stage: Process Innovation (with interaction effects)

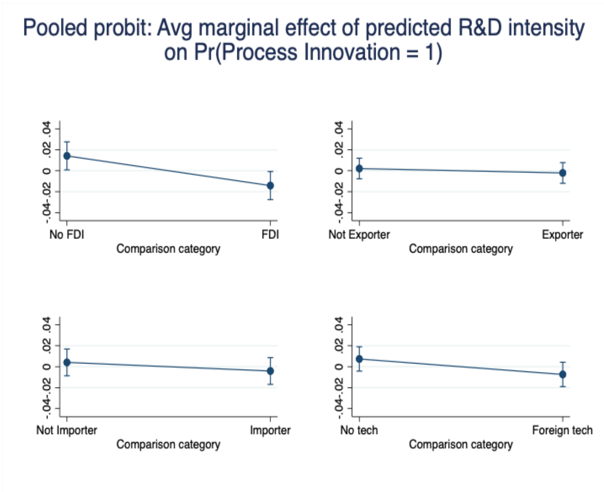
VARIABLES	(1) LPM-FE (0/1) Process Innovation	(2) Pooled Probit (0/1) Process Innovation	(3) CRE Probit (0/1) Process Innovation
Predicted R&D intensity	-0.0959	-0.0411	-0.0552
	(0.197)	(0.0375)	(0.0450)
Firm age	0.000737	0.000580	0.00219
	(0.00252)	(0.000472)	(0.00145)
Firm age squared	0.00000193	-0.00000345	-0.00000662
	(0.0000223)	(0.00000445)	(0.0000142)
FDI	0.0211	0.0198	0.0277
	(0.0747)	(0.0143)	(0.0451)
Exporter	0.104	0.0566***	0.0888***
	(0.0802)	(0.0174)	(0.0277)
Importer	0.131	0.127***	0.108***
	(0.0857)	(0.0209)	(0.0251)
Foreign technology	0.158**	0.123***	0.114***
	(0.0793)	(0.0233)	(0.0245)
Log of firm size	0.0513	0.0167*	0.0308*
	(0.0528)	(0.00935)	(0.0170)
Manager's experience	0.00321	0.000628	0.00191**
	(0.00213)	(0.000423)	(0.000801)
Line of credit	0.160**	0.127***	0.100***
	(0.0795)	(0.0190)	(0.0227)
Inverse Mills Ratio	0.108	-0.0344	-0.0291
	(0.444)	(0.0882)	(0.0901)
FDI × Predicted R&D intensity	-0.00769		
	(0.0239)		
Exporter × Predicted R&D intensity	0.000936		
	(0.0173)		
Importer × Predicted R&D intensity	0.0178		
	(0.0148)		
Foreign technology × Predicted R&D intensity	0.00225		
	(0.0172)		
Predicted R&D intensity (average)			0.0200**
			(0.00881)
Firm age (average)			-0.00164
			(0.00156)
Age squared (average)			0.00000275
			(0.0000148)
FDI (average)			-0.00939
			(0.0407)
Exporter (average)			-0.0358
			(0.0255)
Importer (average)			0.0194
			(0.0224)
Foreign-licensed technology (average)			0.00750
			(0.0207)
Log firm size (average)			-0.0142
			(0.0140)
Manager's experience (average)			-0.00145*
			(0.000798)
Line of credit (average)			0.0270*
			(0.0162)
Constant	-0.0859		
	(1.135)		
Observations	6463	27754	27754
R^2	0.607		
Year FE	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes
Firm FE	Yes	No	No

Estimates reported are marginal effects.
Bootstrapped standard errors (in parentheses) are clustered at the firm level.
*** p<0.01, ** p<0.05, * p<0.1

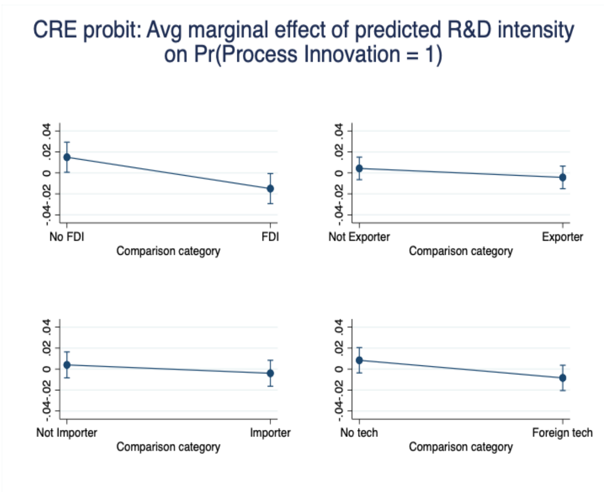
Figure 3: Marginal (interaction) effects of R&D intensity on Process Innovation



(a) LPM



(b) Pooled Probit



(c) CRE Probit

5.3 Third Stage: Productivity

The final stage of the CDM framework models how the innovation output produced in the second stage affects firm-level productivity. For my baseline measure of productivity, I estimate revenue total factor productivity (TFP) using a control function approach to address the simultaneity between input choice and unobserved productivity shocks (Levinsohn and Petrin 2003). I estimate both gross revenue TFP as well as value-added TFP using the Levinsohn and Petrin (ibid.) approach. As a robustness check, I also use two alternative measures of productivity namely the Olley and Pakes (1996) measure of TFP and labour productivity (i.e. log sales per worker).

Levinsohn and Petrin (2003) suggest that intermediate inputs be used to control for the correlation between inputs and unobserved productivity shocks. Since intermediate inputs respond more smoothly (compared to investment) to productivity shocks, they can be used as suitable proxies for such unobservable shocks [13]. Using this control function approach, I obtain firm-level estimates of revenue TFP (in terms of both gross and value-added output) which is then used as the dependent variable in the third stage of the augmented CDM framework. Table 5.7 shows the results from an augmented Cobb-Douglas specification which is estimated using a fixed-effects regression to control for unobserved firm-level heterogeneity. The dependent variable is gross revenue TFP in odd-numbered columns and value-added TFP in even-numbered columns. The explanatory variables of interest are the predicted probabilities of product innovation and process innovation estimated previously in the second stage. Year-fixed effects, industry-fixed effects, and firm-fixed effects have been included in all specifications. Again, the standard errors have been bootstrapped since predicted probabilities of product innovation and process innovation are used as regressors.

The estimated marginal effects in Table 5.7 may be interpreted as follows. An increase in the likelihood of product innovation by 1 percentage point increases the gross TFP by 2.33 percent and the value-added TFP by 3.91 percent. There seems to be no statistically significant effect of process innovation on gross TFP. However, an increase in the likelihood of process innovation by 1 percentage point increases the value-added TFP by 3.39 percent. None of the foreign linkages variables are found to have a statistically significant impact on gross TFP or value-added TFP, other than their

[13] Following Levinsohn and Petrin (2003), log of sales is used as the dependent variable; firm size and proportion of skilled workers are used as freely variable inputs; capital is used as the state variable and estimates of realised material inputs cost are used as the proxy for unobserved productivity.

indirect impact through the predicted product/process innovation. Greater capital stock and larger firm size are found to affect TFP negatively, across all four specifications. This may be on account of diminishing returns to each additional unit of capital, as firms accumulate capital stock beyond a certain amount. As firm size increases, TFP may decrease due to diseconomies of scale and internal inefficiencies[14].

In order to investigate whether the marginal effect of predicted product/process innovation on TFP differs on the basis of the foreign linkages, I include interaction terms as additional regressors. The results are presented in Table 5.8. Upon including interactions between predicted product innovation and the four foreign linkages variables, I find that an increase in the likelihood of product innovation by 1 percentage point increases gross TFP by 3.02 percent and value-added TFP by 4.88 percent. Similarly for process innovation, the corresponding increase in value-added TFP is 4.44 percent. There seems to be no statistically significant effect of process innovation on gross TFP. In column (4), I also find that exporting firms tend to have 9.5 percent higher value-added TFP, compared to non-exporting firms. The marginal effect of process innovation on value-added TFP is found to be slightly lower for exporting firms, compared to non-exporting firms. All other interaction terms are found to be statistically insignificant.

As a robustness check, I repeat the third stage analysis with alternative measures of productivity in Table 5.9. Using the Olley-Pakes measures of TFP and log sales per worker as the alternative dependent variables in the third stage, the results remain qualitatively the same as those in Table 5.7. An increase in the likelihood of product innovation increases gross TFP (Olley-Pakes) by 3.24 percent and value-added TFP (Olley-Pakes) by 3.91 percent. Similarly, an increase in the likelihood of process innovation increases gross TFP (Olley-Pakes) by 2.86 percent and value-added TFP (Olley-Pakes) by 3.39 percent. When labour productivity is used as the dependent variable, the marginal effect of product innovation on productivity is found to be 2.45 percent and that of process innovation is found to be 2.63 percent. This shows that the estimated positive effect of predicted product/process innovation on productivity is robust across different measures of productivity.

[14]This negative relationship between TFP and firm size has also been documented by Morris (2018) using the same WBES dataset but a different estimation approach.

Table 5.7: Third stage: Fixed effects regression

	(1) Log TFP (LP)	(2) Log VA TFP (LP)	(3) Log TFP (LP)	(4) Log VA TFP (LP)
Predicted product innovation	0.0233*	0.0391**		
	(0.0136)	(0.0179)		
Predicted process innovation			0.0145	0.0339**
			(0.0132)	(0.0157)
Log capital stock	-0.163***	-0.504***	-0.198***	-0.524***
	(0.0304)	(0.0388)	(0.0316)	(0.0323)
Log of firm size	-0.216***	-0.356***	-0.235***	-0.334***
	(0.0705)	(0.109)	(0.0688)	(0.110)
FDI	0.0317	0.100	-0.0190	-0.0750
	(0.191)	(0.183)	(0.161)	(0.199)
Exporter	-0.000134	-0.127	0.0450	0.0962
	(0.177)	(0.154)	(0.169)	(0.161)
Importer	0.0329	-0.0628	0.0291	-0.0210
	(0.138)	(0.155)	(0.165)	(0.117)
Foreign-licensed technology	0.106	-0.186	0.0748	-0.103
	(0.156)	(0.170)	(0.142)	(0.141)
Inverse Mills Ratio	0.892	0.815	0.549	0.871*
	(0.653)	(0.583)	(0.522)	(0.509)
Constant	1.324	6.581***	2.819**	7.161***
	(1.931)	(1.684)	(1.398)	(1.558)
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Observations	2090	1948	1945	1830
R^2	0.629	0.816	0.654	0.842

Estimates reported are marginal effects.

Bootstrapped standard errors (in parentheses) are clustered at the firm level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5.8: Third stage: With interaction terms

	(1) Log TFP (LP)	(2) Log VA TFP (LP)	(3) Log TFP (LP)	(4) Log VA TFP (LP)
Predicted product innovation	0.0302**	0.0488**		
	(0.0135)	(0.0203)		
Predicted process innovation			0.0155	0.0440***
			(0.0152)	(0.0169)
Log capital stock	-0.163***	-0.502***	-0.199***	-0.525***
	(0.0298)	(0.0319)	(0.0237)	(0.0322)
Log of firm size	-0.217***	-0.354***	-0.227***	-0.329***
	(0.0733)	(0.0963)	(0.0709)	(0.0765)
FDI	-0.591	-0.183	-0.633*	-0.559
	(0.571)	(0.645)	(0.378)	(0.549)
Exporter	0.454	0.505	0.572*	0.953***
	(0.523)	(0.607)	(0.347)	(0.363)
Importer	0.259	0.247	0.0173	0.351
	(0.419)	(0.457)	(0.319)	(0.344)
Foreign-licensed technology	0.186	-0.278	0.489	0.566
	(0.435)	(0.543)	(0.374)	(0.436)
Inverse Mills Ratio	1.041*	0.991*	0.581	1.002*
	(0.570)	(0.536)	(0.560)	(0.606)
FDI × Predicted product innovation	0.00991	0.00403		
	(0.00914)	(0.00936)		
Importer × Predicted product innovation	-0.00418	-0.00596		
	(0.00752)	(0.00840)		
Exporter × Predicted product innovation	-0.00698	-0.00996		
	(0.00719)	(0.00835)		
Foreign technology × Predicted product innovation	-0.000956	0.00197		
	(0.00586)	(0.00838)		
FDI × Predicted process innovation			0.0105*	0.00778
			(0.00574)	(0.00812)
Importer × Predicted process innovation			0.000482	-0.00735
			(0.00671)	(0.00693)
Exporter × Predicted process innovation			-0.00849	-0.0138**
			(0.00568)	(0.00569)
Foreign technology × Predicted process innovation			-0.00631	-0.00985
			(0.00605)	(0.00702)
Constant	0.737	5.806***	2.681*	6.474***
	(1.567)	(1.711)	(1.465)	(1.733)
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Observations	2090	1948	1945	1830
R^2	0.631	0.817	0.657	0.845

Estimates reported are marginal effects.
Bootstrapped standard errors (in parentheses) are clustered at the firm level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5.9: Third stage: Robustness check (alternate measures of productivity)

	(1) Log TFP (OP)	(2) Log VA TFP (OP)	(3) Log TFP (OP)	(4) Log VA TFP (OP)	(5) Log labour productivity	(6) Log labour productivity
Predicted product innovation	0.0324***	0.0391**			0.0245**	
	(0.0123)	(0.0164)			(0.0118)	
Predicted process innovation			0.0286**	0.0339**		0.0263**
			(0.0127)	(0.0155)		(0.0109)
Log capital stock	-0.553***	-0.518***	-0.581***	-0.537***	0.136***	0.114***
	(0.0232)	(0.0368)	(0.0308)	(0.0386)	(0.0231)	(0.0278)
Log of firm size	-0.288***	-0.369***	-0.301***	-0.348***	-0.329***	-0.356***
	(0.0954)	(0.105)	(0.0979)	(0.117)	(0.0650)	(0.0647)
FDI	0.212	0.100	0.0798	-0.0750	0.219	0.126
	(0.205)	(0.156)	(0.199)	(0.206)	(0.180)	(0.194)
Exporter	0.00289	-0.127	0.111	0.0962	0.0691	0.0982
	(0.127)	(0.140)	(0.119)	(0.157)	(0.152)	(0.0996)
Importer	-0.0404	-0.0628	0.00588	-0.0210	-0.0136	-0.00388
	(0.121)	(0.148)	(0.108)	(0.143)	(0.121)	(0.0974)
Foreign-licensed technology	-0.212*	-0.186	-0.143	-0.103	-0.174	-0.135
	(0.126)	(0.159)	(0.114)	(0.143)	(0.113)	(0.0950)
Inverse Mills Ratio	0.570	0.815	0.543	0.871	0.417	0.458
	(0.480)	(0.610)	(0.486)	(0.598)	(0.366)	(0.385)
Constant	7.397***	6.581***	8.203***	7.161***	8.112***	8.498***
	(1.410)	(1.787)	(1.376)	(1.780)	(1.128)	(1.075)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2230	1948	2053	1830	2768	2603
R^2	0.857	0.815	0.878	0.842	0.918	0.927

Estimates reported are marginal effects.
Bootstrapped standard errors (in parentheses) are clustered at the firm level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

5.4 Composite measure of innovation

So far, this paper has distinguished between two forms of innovation: product innovation and process innovation. However, these two measures tend to be highly correlated since firms that undertake one form of innovation typically also undertake innovation of the other form. Due to this correlation between product innovation and process innovation, the regression specifications I have estimated so far have included only one measure of innovation at a time. Including both product and process innovation dummies together results in both of them being insignificant. This is well-documented by Mohnen and Hall (2013) who suggest that since product and process innovations often occur together, it is only their joint effect that can be suitably identified. Keeping this in mind, this section repeats the analysis of the second stage and third stage of the CDM model using a composite measure of innovation. I call this measure "innovation" and define it as follows: "innovation" is a dummy variable that takes the value 1 if the firm undertakes either a product innovation or a process innovation; and 0 otherwise. Using this composite measure of innovation, I am able to address any concerns about a potential omitted variable bias that may arise in the previously used specifications that include product and process innovation dummies separately.

When using the composite measure of "innovation" in Table 5.10, the results are qualitatively comparable with my earlier results. In my preferred CRE Probit specification, I find that a 100 percent increase in the R&D intensity increases the likelihood of introducing a product or process innovation by 9.4 percentage points. Inward FDI has no statistically significant effect on the likelihood of introducing product or process innovation. Exporting firms are 7.9 percentage points more likely to introduce a product or process innovation, compared to non-exporting firms. Importing firms are 6.4 percentage points more likely to introduce a product or process innovation, compared to non-importing firms.

Table 5.11 presents the results from the third stage fixed effects regression using the predicted value of "innovation" estimated in the second stage. The estimated marginal effect of predicted "innovation" on TFP is similar in magnitude to that of predicted product innovation and predicted process innovation in Table 5.7. An increase in the probability of innovation by 1 percentage point increases gross TFP by 2.9 percent and value-added TFP by 5.12 percent. Again, none of the foreign linkages variables are found to have a statistically significant effect on TFP, other than their indirect effect through the predicted innovation. These results are found to be robust across other specifications using alternate measures of productivity.

Table 5.10: Second stage: Composite measure of innovation

VARIABLES	(1) LPM-FE Innovation	(2) Pooled Probit Innovation	(3) CRE Probit Innovation
Predicted R&D intensity	0.047	0.11**	0.094***
	(0.14)	(0.044)	(0.034)
Firm age	0.0057**	0.0012**	0.0068***
	(0.0025)	(0.00053)	(0.0019)
Firm age squared	-0.000043**	-8.8e-06*	-0.000050***
	(0.000021)	(5.1e-06)	(0.000017)
FDI	0.0020	-0.016	-0.0099
	(0.060)	(0.014)	(0.044)
Exporter	0.11	0.020	0.079**
	(0.076)	(0.020)	(0.034)
Importer	0.11	0.10***	0.064**
	(0.089)	(0.022)	(0.030)
Foreign-licensed technology	0.085	0.062***	0.044
	(0.082)	(0.021)	(0.037)
Log of firm size	0.060	0.034***	0.054***
	(0.037)	(0.0097)	(0.017)
Manager's experience	0.0013	-0.00065	0.00023
	(0.0018)	(0.00053)	(0.00086)
Line of credit	0.10	0.080***	0.046
	(0.079)	(0.020)	(0.028)
Inverse Mills Ratio	0.24	0.17**	0.17***
	(0.28)	(0.085)	(0.066)
Predicted R&D intensity (average)			0.017**
			(0.0082)
Firm age (average)			-0.0059***
			(0.0021)
Age squared (average)			0.000043**
			(0.000018)
FDI (average)			-0.0071
			(0.043)
Exporter (average)			-0.065**
			(0.029)
Importer (average)			0.040*
			(0.024)
Foreign-licensed technology (average)			0.020
			(0.029)
Log firm size (average)			-0.021
			(0.015)
Manager's experience (average)			-0.00094
			(0.00093)
Line of credit (average)			0.036*
			(0.021)
Constant	-0.41		
	(0.71)		
Observations	6,118	27,474	27,474
R-squared	0.665		
Year FE	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes
Firm FE	Yes	No	No

Estimates reported are marginal effects.
Bootstrapped standard errors (in parentheses) are clustered at the firm level.
*** p<0.01, ** p<0.05, * p<0.1

Table 5.11: Third Stage: Predicted Innovation & TFP

	(1) Log TFP (LP)	(2) Log VA TFP (LP)	(3) Log TFP (OP)	(4) Log VA TFP (OP)	(5) Log labour productivity
Predicted innovation	0.0290**	0.0512***	0.0426***	0.0512***	0.0349***
	(0.0128)	(0.0178)	(0.0136)	(0.0153)	(0.0131)
Log capital stock	-0.196***	-0.519***	-0.575***	-0.533***	0.113***
	(0.0297)	(0.0342)	(0.0366)	(0.0317)	(0.0292)
Log of firm size	-0.205***	-0.327***	-0.306***	-0.341***	-0.353***
	(0.0694)	(0.102)	(0.0749)	(0.101)	(0.0917)
FDI	-0.0141	-0.0398	0.113	-0.0398	0.147
	(0.163)	(0.189)	(0.182)	(0.170)	(0.223)
Exporter	0.0164	-0.0863	-0.0565	-0.0863	0.00186
	(0.161)	(0.166)	(0.127)	(0.163)	(0.136)
Importer	-0.0103	-0.129	-0.118	-0.129	-0.0840
	(0.121)	(0.165)	(0.132)	(0.185)	(0.115)
Foreign-licensed technology	0.131	-0.0730	-0.156	-0.0730	-0.143
	(0.143)	(0.142)	(0.102)	(0.176)	(0.140)
Inverse Mills Ratio	1.121**	1.330**	0.795	1.330**	0.647
	(0.510)	(0.640)	(0.506)	(0.607)	(0.512)
Constant	0.651	4.833**	6.514***	4.833***	7.316***
	(1.577)	(2.148)	(1.851)	(1.809)	(1.748)
Year FE	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
Observations	1690	1585	1794	1585	2274
R^2	0.662	0.847	0.879	0.847	0.933

Estimates reported are marginal effects.
Bootstrapped standard errors (in parentheses) are clustered at the firm level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

6 Conclusion

The objective of this paper is to study how foreign linkages affect firm-level innovation and productivity in the context of emerging markets and developing economies. This paper fills a number of important gaps in the existing literature. First, I use a cross-country firm-level dataset which is the most comprehensive in terms of its geographical coverage of emerging markets and developing economies for a study of this type. Second, the uniquely available panel structure of the World Bank Enterprise Surveys allows me to use advanced panel data methods to address concerns regarding potential selection bias and endogeneity bias. Third, this paper contributes to the limited body of empirical work that focuses on firm-level innovation as the mechanism through which foreign linkages may affect productivity.

Building upon the analytical framework proposed by Crépon et al. (1998), I estimate a three-stage structural model, the results of which are summarised as follows. In the first stage, I find that being an exporter, using foreign inputs and using foreign-licensed technology makes firms more likely to invest in R&D, relative to firms that do not have such foreign linkages. Conditional on the firm's decision to invest, inward FDI has no statistically significant impact on the intensity of R&D investment. I also find evidence of sample selection bias which is corrected by using a two-step Heckman selection model. In the second stage, I find that while increases in the R&D intensity tend to increase the probability of introducing a product innovation, they have no statistically significant effect on the likelihood of introducing a process innovation. In the third stage, I find that being a product innovator or a process innovator is associated with increases in firm-level productivity, relative to other firms. These results are robust across a number of specifications and different measures of productivity and innovation.

The findings of this paper have a number of policy implications relevant to developing countries. First, policies aimed at driving innovation and consequent increases in productivity should be formulated in sync with policies that incentivize and facilitate firms in strengthening linkages with foreign firms. Second, innovation effort may not necessarily result in innovation output (as noted in the results for process innovation in Table 5.6). Especially in the case of developing economies, there is a significant role for complementary factors such as adequate levels of human capital and supporting infrastructure to reap the potential gains from innovation effort. Third, both forms of technology transfer namely embodied transfer (via the use of foreign inputs) and disembodied transfer (via the use of foreign-licensed technology) seem to be important for spurring process innovation in developing economies. This suggests that stronger foreign linkages may help in facilitating technological catch-up for firms in developing economies.

As an area of future research, it would be interesting to study if there are any lagged effects of innovation effort on innovation output. This type of analysis would require firm-level longitudinal data over a longer period of time, which is currently limited given the cost of carrying out innovation surveys. Based on the availability of future data, it would also be useful to consider more objective measures of innovation output (as opposed to self-reported binary variables) to better understand how innovation results in productivity gains.

References

Acemoglu, Daron, Philippe Aghion, and Fabrizio Zilibotti (2006). "Distance to frontier, selection, and economic growth". In: *Journal of the European Economic association* 4.1, pp. 37–74.

Bell, Martin and Keith Pavitt (1997). "Technological accumulation and industrial growth: contrasts between developed and developing countries". In: *Technology, globalisation and economic performance* 83137, pp. 83–137.

Benavente, Jose Miguel (2006). "The role of research and innovation in promoting productivity in Chile". In: *Economics of innovation and New Technology* 15.4-5, pp. 301–315.

Castellani, Davide and Antonello Zanfei (2007). "Internationalisation, innovation and productivity: how do firms differ in Italy?" In: *World Economy* 30.1, pp. 156–176.

Cirera, Xavier and Silvia Muzi (2016). *Measuring firm-level innovation using short questionnaires: Evidence from an experiment.* The World Bank.

Cohen, Wesley M and Steven Klepper (1996). "A reprise of size and R & D". In: *The Economic Journal* 106.437, pp. 925–951.

Crépon, Bruno, Emmanuel Duguet, and Jacques Mairesse (1998). "Research, Innovation And Productivity: An Econometric Analysis At The Firm Level". In: *Economics of Innovation and new Technology* 7.2, pp. 115–158.

Crespi, Gustavo, Ezequiel Tacsir, and Fernando Vargas (2016). "Innovation dynamics and productivity: Evidence for Latin America". In: *Firm innovation and productivity in Latin America and the Caribbean*, pp. 37–71.

Crespi, Gustavo and Pluvia Zuniga (2012). "Innovation and productivity: evidence from six Latin American countries". In: *World development* 40.2, pp. 273–290.

Criscuolo, Chiara, Jonathan E Haskel, and Matthew J Slaughter (2010). "Global engagement and the innovation activities of firms". In: *International Journal of Industrial Organization* 28.2, pp. 191–202.

Griffith, Rachel, Elena Huergo, Jacques Mairesse, and Bettina Peters (2006). "Innovation and productivity across four European countries". In: *Oxford review of economic policy* 22.4, pp. 483–498.

Griliches, Zvi (1979). "Issues in assessing the contribution of research and development to productivity growth". In: *Bell Journal of economics* 10.1, pp. 92–116.

Hall, Bronwyn H (2011). *Innovation and productivity.* Tech. rep. National bureau of economic research.

Hall, Bronwyn H, Francesca Lotti, and Jacques Mairesse (2009). "Innovation and productivity in SMEs: empirical evidence for Italy". In: *Small business economics* 33, pp. 13–33.

Katz, Jorge (2006). "Structural change and domestic technological capabilities". In: *Cepal Review*.

Kwak, Do Won, Robert S. Martin, and Jeffrey M. Wooldridge (2021). In: *Journal of Econometric Methods*. DOI: `doi:10.1515/jem-2021-0005`. URL: `https://doi.org/10.1515/jem-2021-0005`.

Levinsohn, James and Amil Petrin (2003). "Estimating production functions using inputs to control for unobservables". In: *The review of economic studies* 70.2, pp. 317–341.

Lööf, Hans, Almas Heshmati, Rita Asplund, and Svein-Olav Nåås (2001). *Innovation and performance in manufacturing industries: A comparison of the Nordic countries*. Tech. rep. SSE/EFI working paper series in economics and finance.

Mohnen, Pierre and Bronwyn H Hall (2013). "Innovation and productivity: An update". In: *Eurasian Business Review* 3.1, pp. 47–65.

Mohnen, Pierre, Jacques Mairesse, and Marcel Dagenais (2006). "Innovativity: A comparison across seven European countries". In: *Economics of Innovation and New Technology* 15.4-5, pp. 391–413.

Morris, Diego M (2018). "Innovation and productivity among heterogeneous firms". In: *Research Policy* 47.10, pp. 1918–1932.

Mundlak, Yair (1978). "On the pooling of time series and cross section data". In: *Econometrica: journal of the Econometric Society*, pp. 69–85.

Olley, G Steven and Ariel Pakes (1996). "The Dynamics of Productivity in the Telecommunications Equipment Industry". In: *Econometrica: Journal of the Econometric Society*, pp. 1263–1297.

Pakes, Ariel and Zvi Griliches (1980). "Patents and R&D at the firm level: A first report". In: *Economics letters* 5.4, pp. 377–381.

Polder, Michael, George van Leeuwen, Pierre Mohnen, and Wladimir Raymond (2009). "Productivity effects of innovation modes". In.

Romer, Paul M (1990). "Endogenous technological change". In: *Journal of political Economy* 98.5, Part 2, S71–S102.

Schunck, Reinhard (2013). "Within and between estimates in random-effects models: Advantages and drawbacks of correlated random effects and hybrid models". In: *The Stata Journal* 13.1, pp. 65–76.

Siedschlag, Iulia and Xiaoheng Zhang (2015). "Internationalisation of firms and their innovation and productivity". In: *Economics of Innovation and New Technology* 24.3, pp. 183–203.

Wooldridge, Jeffrey M (1995). "Selection corrections for panel data models under conditional mean independence assumptions". In: *Journal of econometrics* 68.1, pp. 115–132.

Wooldridge, Jeffrey M (2010). *Econometric analysis of cross section and panel data*. MIT press.

— (2019). "Correlated random effects models with unbalanced panels". In: *Journal of Econometrics* 211.1, pp. 137–150.

Table A1: Country Coverage

Country	2003	2006	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total
Albania	0	0	0	49	0	0	0	64	0	0	0	0	0	126	239
Argentina	0	524	0	0	627	0	0	0	0	0	0	434	0	0	1,585
Armenia	0	0	0	86	0	0	0	92	0	0	0	0	0	0	178
Azerbaijan	0	0	0	86	0	0	0	83	0	0	0	0	0	0	169
Bangladesh	0	0	0	0	0	105	0	766	0	0	0	0	0	0	871
Belarus	0	0	53	0	0	0	0	103	0	0	0	0	0	0	156
Bolivia	0	312	0	0	81	0	0	0	0	0	0	93	0	0	486
Bulgaria	0	0	0	59	0	0	0	90	0	0	0	0	0	358	507
Chile	0	545	0	0	693	0	0	0	0	0	0	0	0	0	1,238
Colombia	0	603	0	0	640	0	0	0	0	0	0	484	0	0	1,727
Croatia	0	0	0	26	0	0	0	111	0	0	0	0	0	139	276
Czech Republic	0	0	0	37	0	0	0	80	0	0	0	0	0	257	374
Dominican Republic	0	0	0	0	98	0	0	0	0	0	73	0	0	0	171
Ecuador	0	167	0	0	59	0	0	0	0	0	0	59	0	0	285
El Salvador	0	369	0	0	115	0	0	0	0	0	368	0	0	0	852
Estonia	0	0	0	46	0	0	0	51	0	0	0	0	0	121	218
Ethiopia	0	0	0	0	0	131	0	0	0	322	0	0	0	0	453
Georgia	0	0	74	0	0	0	0	95	0	0	0	0	0	183	352
Guatemala	426	0	0	0	310	0	0	0	0	0	0	128	0	0	864
Honduras	0	200	0	0	115	0	0	0	0	0	79	0	0	0	394
Hungary	0	0	0	93	0	0	0	64	0	0	0	0	0	0	157
Jordan	0	0	0	0	0	0	0	250	0	0	0	0	0	182	432
Kazakhstan	0	0	0	126	0	0	0	155	0	0	0	0	0	650	931
Kenya	0	0	0	0	0	0	0	196	0	0	0	0	382	0	578
Kyrgyz Republic	0	0	0	67	0	0	0	83	0	0	0	0	0	128	278
Latvia	0	0	0	58	0	0	0	61	0	0	0	0	0	107	226
Lebanon	0	0	0	0	0	0	0	172	0	0	0	0	0	236	408
Lithuania	0	0	0	51	0	0	0	76	0	0	0	0	0	114	241
Mexico	0	943	0	0	1,054	0	0	0	0	0	0	0	0	0	1,997
Montenegro	0	0	0	25	0	0	0	30	0	0	0	0	0	0	55
Morocco	0	0	0	0	0	0	0	129	0	0	0	0	0	252	381
Myanmar	0	0	0	0	0	0	0	0	307	0	337	0	0	0	644
Nicaragua	431	312	0	0	104	0	0	0	0	0	97	0	0	0	944
Panama	0	190	0	0	92	0	0	0	0	0	0	0	0	0	282
Paraguay	0	306	0	0	92	0	0	0	0	0	0	96	0	0	494
Peru	0	330	0	0	685	0	0	0	0	0	0	495	0	0	1,510
Poland	0	0	0	71	0	0	0	103	0	0	0	0	0	610	784
Romania	0	0	0	109	0	0	0	148	0	0	0	0	0	0	257
Russian Federation	0	0	0	299	0	0	1,029	0	0	0	0	0	0	710	2,038
Slovakia	0	0	0	51	0	0	0	72	0	0	0	0	0	0	123
Slovenia	0	0	0	44	0	0	0	73	0	0	0	0	0	139	256
Tajikistan	0	0	84	0	0	0	0	89	0	0	0	0	0	73	246
Turkey	0	0	509	0	0	0	0	769	0	0	0	0	0	768	2,046
Ukraine	0	0	308	0	0	0	0	578	0	0	0	0	0	772	1,658
Uruguay	0	300	0	0	285	0	0	0	0	0	0	80	0	0	665
Uzbekistan	0	0	103	0	0	0	0	120	0	0	0	0	0	656	879
Zimbabwe	0	0	0	0	0	156	0	0	0	0	239	0	0	0	395
Total	857	5,101	1,131	1,383	5,050	392	1,029	4,703	307	322	1,193	1,869	382	6,581	30,300

Table A2: Variable definitions

Variable name	Associated survey question
(0/1) Firm invested in R&D	During the last fiscal year, did this establishment spend on formal research and development (R&D) activities, either in-house or contracted to other companies, excluding market research surveys?
Log R&D per worker (2010 USD)	During the last fiscal year, how much money did this establishment spend on R&D, either in-house or contracted to other companies?
(0/1) Product Innovation	Did this establishment introduce a new or significantly improved product(s) or service(s) over the last three years?
(0/1) Process Innovation	Did this establishment introduce a new or significantly improved process (including manufacturing, logistics, delivery and distribution of goods/services) over the last three years?
Age	In what year did this establishment begin operations in this country?
Full-time employees	What was the number of permanent, full-time workers at the end of last fiscal year?
Share of skilled workers	What was the share of skilled production workers at the end of last fiscal year?
(0/1) Foreign direct investment	Does this establishment have more than 10% ownership by foreign individuals or companies?
(0/1) Direct exports	Does this establishment export more than 10% of its total sales?
(0/1) Foreign inputs	Does this establishment import more than 10% of its material inputs from abroad?
(0/1) Foreign-licensed technology	Does this establishment at present use technology licensed from a foreign-owned company, excluding office software?
Manager's experience	How many years of experience does the top manager have in the type of sector that the establishment presently operates?
(0/1) Line of credit	At this time, does this establishment have a line of credit or a loan from a financial institution?
Capacity utilization (%)	In the last fiscal year, what was this establishment's output produced as a proportion of the maximum output possible if using all the resources available?
Log sales per worker (2010 USD)	In the last fiscal year, what were this establishment's total annual sales for all products and services?
Log net book value of fixed assets (2010 USD)	If this establishment were to purchase machinery, vehicles, and equipment it uses now, in their current condition, how much would they cost, independently of whether they are owned, rented or leased?
Log cost of material inputs (2010 USD)	In the last fiscal year, what was the annual cost of all raw materials, intermediate goods and other inputs used in the production activity?