SEEDS
Sustainability Environmental Economics and Dynamics Studies

Working Paper Series

Estimating a non-neutral production function: a heterogeneous treatment effect approach

by

Davide Antonioli, Georgios Gioldasis, Antonio Musolesi

06/2018

SEEDS Working Paper 06/2018
March 2018
Davide Antonioli, Georgios Gioldasis, Antonio Musolesi

Estimating a non-neutral production function: a heterogeneous treatment effect approach

Davide ANTONIOLI
Department of Management and Business Administration,
University of Chieti-Pescara and SEEDS

Georgios GIOLDASIS
Department of Economics and Management (DEM),
University of Ferrara and SEEDS

Antonio MUSOLESI
Department of Economics and Management (DEM),
University of Ferrara and SEEDS

March 29, 2018

Abstract

This paper addresses the issue of estimating a production function that allows us to depart from the standard hypothesis of Hicks neutrality while also coping with the endogeneity of a dummy innovation variable. We consider specifications that relax Hicks neutrality, and we derive the testable conditions for common parametric approximations under which Hicks neutrality holds. The model is estimated through instrumental variables methods, allowing for a heterogeneous effect of innovation on the production process. The econometric analysis rejects Hicks neutrality and highlights three main features: i) a capital-saving technology of innovative with respect to non-innovative firms, ii) a locally progressive technical change and iii) fully heterogeneous technologies when comparing innovative to non-innovative firms.

Keywords: Biased technical change; Hicks neutrality; Innovation; Productivity; Knowledge production function; CDM model; Instrumental variables; heterogeneous treatment effect
JEL classification: C26; C31; D24; O33

Total word count: 8916

1 Introduction

Assessing the effect of technical change (TC) on the production process, both theoretically and empirically and at all levels of aggregation, is one of the major concerns amongst economists because TC is largely recognized as one of the fundamental drivers of economic development. In neoclassical production theory, one of the most commonly used classifications of TC dates to the seminal work of Hicks (1963). In particular, according to Hicks, *"we can classify inventions according as their initial effects are to increase, leave unchanged or diminish the ratio of the marginal product of capital to that of labor"* (Hicks, 1963, p.121). The above types of inventions are referred to as *labour-saving*, *neutral* or *capital-saving* inventions, respectively.

Formally, *Hicks neutrality* (HN) requires that the marginal rate of technical substitution (MRTS) between each pair of inputs be independent of technical change. However, this definition of neutrality has received different interpretations; Hicks' requirement that the firm be in a state of *"internal equilibrium"* (Hicks, 1963, p.113, 234, 236) does not specify explicitly whether the effect of TC should be measured along the firm's expansion path or at the optimal factor proportions, which has resulted in controversy (see, e.g., Antle and Capalbo, 1988).

To clarify this ambiguity, Blackorby et al. (1976) consider three different definitions. The first was suggested by Hicks and is already referred to as HN. The second is *implicit Hicks neutrality* (IHN) and accounts for a ray preserving TC. They also introduce a third definition that they term *extended Hicks neutrality* (EHN). While HN translates equivalently to a weak separability assumption between TC and the inputs (Morimoto, 1974), EHN implies TC that is strongly separable from the inputs in a production function. The three different definitions considered in Blackorby et al. (1976), namely HN, IHN and EHN, are not equivalent in general and coincide only if the production function is input homogeneous.

While HN has received significant attention, several streams of literature question its plausibility. Jones (1965) describes a two-sector economy in which HN is unlikely, although HN may appear in each industry. Further, Steedman (1985) presents a model of interconnected industries in which, under weak alternative sufficient conditions, HN is impossible, ultimately concluding that the compatibility of HN with the other assumptions of a model should be examined. Acemoglu (2015) also mentions that in Atkinson and Stiglitz's (1969) seminal paper, it is implied that when TC is localized to specific factor proportions, then it is biased. Chambers (1988, p. 206) describes other fairly common types of TC that depart from being HN. In particular, if TC is locally progressive or regressive, its effect on productivity results in isoquants of the production function that intercept the old ones. Obviously, such TC does not follow HN, IHN or EHN.

Empirically, at a microeconomic level, since the seminal works by Pakes and

Griliches (1984) and Griliches (1998), the question concerning the effect of TC on the production process has often been addressed within the framework of the so-called knowledge production function (KPF), where innovation activity is the main source of technical and knowledge improvements. The KPF is a conceptual framework that suggests a possible causal relationship between unobservable knowledge capital and related observable variables such as innovation inputs (e.g., research and development, R&D), innovation outputs (e.g., patents) and firms' performance. The KPF provided the basis for econometric analyses connecting different and relevant aspects of innovative activities. In their influential article, Crépon et al. (1998) propose a multiple-equation econometric model – commonly labeled CDM – that has a similar structure to Griliches' original conception and uses appropriate estimation methods for taking into account both the potential endogeneity of some of the explanatory variables and the particular nature of the dependent variables. In recent years, the availability of survey data, such as those obtained from the Community Innovation Surveys, has allowed the use of a direct and binary measure of the innovative output that is then introduced into a production function framework as an endogenous dummy variable that accounts for TC (see, e.g., Mairesse and Mohnen, 2010).

Recently, developments and generalizations of the CDM approach have been accomplished, such as the introduction of dynamics into the model, the assessment of measurement errors or the consideration of a Schumpeterian perspective (Lööf et al., 2016). The extremely vast literature clearly indicates a positive and significant effect of innovation on productivity (Mohnen and Hall, 2013) and sheds light on many other relevant relationships among variables.

However, a crucial maintained assumption in this literature is HN. In standard econometric models, innovation is additively introduced into the production function specification. Additivity is equivalent to a multiplicative decomposability of the production function into a function of innovation and a function of the inputs. Uzawa and Watanabe (1961) prove the equivalence of HN and the decomposability of the production function. Therefore, standard econometric models usually impose a strict condition, specifically HN, when assessing the effect of innovation on productivity.

The present study contributes to the literature on the econometrics of productivity and TC in two ways. First, we estimate a model that allows us to relax the neutrality assumption. Second, we extend the analysis by Blackorby et al. (1976) and provide testable conditions, for common parametric specifications, namely Cobb-Douglas (CD) and translog (TL), under which HN, IHN or EHN hold.

The econometric analysis developed here exploits recent advances in the econometric theory of instrumental variables (IVs) with cross-sectional data and a binary endogenous regressor. In the empirical literature, the standard approach considers a parametric approximation of the production function – typically CD – in which innovation enters additively, and the main focus is on the endogeneity of innovation and its binary nature (Musolesi and Huiban, 2010; Mohnen and Hall, 2013). We depart from this literature by allowing for an heterogeneous effect of the dummy endogenous innovation variable by adopting the approach proposed by Wooldridge (2003, 2010). This allows the technology parameters to differ between innovative and non-innovative firms, finally permitting us to address endogeneity while relaxing the assumption of neutrality.

The remainder of the paper is organized as follows. Section 2 describes the theory of TC and presents HN, IHN and EHN, along with their relationships. It also introduces empirical specifications that allow non-neutrality and provides simple conditions under which each of the different definitions of neutrality holds. The data set is described in section 3. This section also presents the results and comments. Finally, section 4 concludes the paper. In supplementary appendices, we present the relationships among HN, IHN and EHN (Appendix A), and we also provide conditions under which neutrality holds within a TL framework (Appendix B). Appendix C provides robustness checks.

2 Hicks neutrality: theory and econometrics

2.1 A selective review of theory

Consider a production function $f : \mathbb{R}_{+}^{k+1} \rightarrow \mathbb{R}_{+}^{1}$ that is expressed as

$$Y = f(\mathbf{X}, I). \tag{1}$$

$\mathbb{R}_{+}$ denotes the positive real numbers. It is commonly assumed that f is twice-differentiable in $\mathbf{X}$, strictly quasiconcave in $\mathbf{X}$ and non-decreasing in I. Y is a measure of output such as value added. $\mathbf{X} = \{X_1, X_2, \ldots, X_k\}^T$ is a k-dimensional vector of inputs that includes conventional production inputs such as capital and labour. I represents innovation activity, which is regarded in production theory as the main source of TC.

In a production function with multiple inputs, i.e., for $k > 1$, technical change might differently affect the marginal productivity of each input. Therefore, it can be classified as *biased* or *neutral*, according to whether the ratios between marginal products are changed. Hicks (1963) is the first to distinguish inventions according to the above observation. Particularly, he focuses on cases with two input factors, namely capital and labour. Under the requirement that the firm remains in a state of *"internal equilibrium"*, he defines an invention as *labour-saving*, *neutral* or *capital-saving* according to whether its initial effect is to increase, leave unchanged or decrease the ratio of the marginal product of capital to that of labour, respectively. Uzawa and Watanabe (1961) generalize Hicks' classification to cases of more than two factors of production.

Hicks' classification has resulted in controversial interpretations concerning whether the effect described above should be observed along the expansion path of the firm or along the optimal factor proportions. In particular, while Blackorby et al. (1976) state that this effect should be considered along the firm's expansion path, Kennedy and Thirlwall (1972, 1977) consider HN along a ray from the origin and argue that Hicks' definition of neutrality does not imply an expansion *path preserving* innovation.

Nevertheless, Blackorby et al. (1976) attempt to clarify this ambiguity by distinguishing three different definitions of neutrality. The first generalizes Hicks' definition for cases of firms with more than two inputs; neutrality holds if the MRTS between each pair of inputs is independent of I. This definition is denoted HN and translates to the following expression:

$$\frac{\partial f(\mathbf{X}, I)/\partial X_r}{\partial f(\mathbf{X}, I)/\partial X_l} = \phi_{rl}(\mathbf{X}), \ \ \forall r \neq l \tag{2}$$

where $\phi_{rl}, r, l = 1, 2, \ldots, k, r \neq l$ are functions of $\mathbf{X}$. Therefore, HN requires that the MRTS for every X_r, $X_l, r \neq l$ component of $\mathbf{X}$ be a function of $\mathbf{X}$ only. Blackorby et al. (1976) refer to innovation as HN if it is expansion path preserving. Further, it is proven (see Morimoto, 1974) that HN holds if and only if the inputs $\mathbf{X}$ are weakly separable from I in f, such that the production function is described by:

$$f(\mathbf{X}, I) = g(h(\mathbf{X}), I) \tag{3}$$

where g and h are real functions. The above implies that HN and weak separability of $\mathbf{X}$ from I impose equivalent restrictions on f.

Blackorby et al. (1976) introduce a second definition of neutrality, namely IHN, that requires the MRTS between each pair of inputs to be independent of I, at constant factor proportions. This definition implies that the MRTS is homogeneous of degree zero in the inputs $\mathbf{X}$. Further, they refer to innovation as IHN if it is *ray preserving* and prove that IHN holds if and only if the transformation function $\tilde{f}(\mathbf{X}, Y, I) = \max_{\lambda}\{\lambda > 0 : f(\lambda^{-1}\mathbf{X}, I) \geq Y\}$ can be written in the following form:

$$\tilde{f}(\mathbf{X}, Y, I) = \tilde{g}(\tilde{h}(\mathbf{X}, Y), Y, I), \tag{4}$$

where $\tilde{g}$ and $\tilde{h}$ are real functions. $\tilde{f}$ is a distance function that uniquely represents the technology f. According to (4), I is IHN if it is separable from $\mathbf{X}$ in $\tilde{f}$.

A third definition of neutrality accounts for cases in which I is strongly separable from $\mathbf{X}$ in the production function f (see Chambers, 1988, p.45). By definition, EHN holds if:

$$\frac{\partial}{\partial X_r}\left[\ln f(\mathbf{X}, I)\right] = \phi_r(\mathbf{X}), \ \forall r = 1, 2, \ldots, k, \tag{5}$$

where $\phi_r, \ r = 1, 2, \ldots, k$ are functions of $\mathbf{X}$. Moreover, it is proven that EHN holds if and only if the production function can be multiplicatively decomposed into a function $\bar{h}$ of inputs only and a function $\bar{g}$ of I only, such that:

$$f(\mathbf{X}, I) = \bar{g}(I)\bar{h}(\mathbf{X}). \tag{6}$$

Obviously, if innovation is EHN, it is also HN, because eq.(6) implies eq.(3).

In general, the three definitions of neutrality are not equivalent. Antle and Capalbo (1988, p.38) note that innovation that results in a renumbering of the isoquants is also neutral in terms of the MRTS at points on the expansion path but may not be neutral in terms of optimal factor proportions. Moreover, a priori, neither HN nor IHN is sufficient for EHN. Nevertheless, there are conditions under which HN, IHN and EHN coincide. First, under the assumption that the production function is input homogeneous, the three definitions are equivalent (see Morimoto, 1974; Blackorby et al., 1976). IHN and HN are equivalent if and only if the production function is input homothetic, while the equivalence of IHN and EHN implies the homotheticity of f. A detailed presentation of the relationships among HN, IHN and EHN is provided in Appendix A.

2.2 Econometric specification: relaxing and testing Hicks neutrality

The above definitions of neutrality can be assessed using a suitable econometric framework. We assume that the inputs of production are the conventional factors capital (K) and labour (L). The innovation activity of the firm is described by a binary variable $I \in \{0, 1\}$, as in many previous works (see, e.g., Mairesse and Mohnen, 2010; Mohnen and Hall, 2013). We consider a specification that allows us to relax the neutrality assumption. This is achieved by focusing on a model in which innovation not only produces a shift in the production technology – as is usually imposed in econometric analyses – but is also interacted with labour and capital inputs, thus allowing the technology parameters to differ between innovative and non-innovative firms. For simplicity, comparability with previous studies and data congruence, the main analysis is conducted assuming a CD technology, while in Appendix B, we derive testable conditions concerning the TL form and its relationship with HN.

Empirical studies usually consider a CD production function in which innovation enters additively. Given a sample of n observations, the econometric model is described by:

$$\ln Y_i = \ln f^{\mathsf{CD}}(K_i, L_i, I_i, \epsilon_i) = \alpha + \alpha_I I_i + \alpha_K \ln K_i + \alpha_L \ln L_i + \epsilon_i, \tag{7}$$

where i denotes the i-th observation, and ϵ is the error term. Then, the MRTS[1] between

[1]Computation of the MRTS involves deriving the marginal products of the inputs. The marginal product of $X_r, r = 1, 2, \ldots, k$ at $(Y_i, \mathbf{X}_i, I_i), i = 1, 2, \ldots, n$ is given by $\partial E(Y_i|\mathbf{X}_i, I_i)/\partial X_{r,i}$, which assumes the exogeneity of all variables, that is, $E(\epsilon_i|\mathbf{X}_i, I_i) = 0, i = 1, 2, \ldots, n$. (see Verbeek, 2008, for a discussion of marginal effects in the linear model) This assumption is considered in this section only

L and K is given by:

$$MRTS_i^{\texttt{CD}} = -\frac{\alpha_L}{\alpha_K}\frac{K_i}{L_i}, \tag{8}$$

which, being independent of I, implies that HN holds. Moreover, IHN is also imposed because for constant factor proportions, the MRTS is both independent of I and constant. Finally, I is EHN because (5) holds, that is:

$$\frac{\partial}{\partial K_i}E\left[\ln Y_i|K_i, L_i, I_i\right] = \alpha_K \frac{1}{K_i} \quad \& \quad \frac{\partial}{\partial L_i}E\left[\ln Y_i|K_i, L_i, I_i\right] = \alpha_L \frac{1}{L_i}.$$

Alternatively, it suffices to observe that I is strongly separable from the inputs in $f_{\texttt{CD}}$. In summary, in the case of a CD production function with added innovation as described by (7), all definitions of neutrality described above are satisfied. This equivalence is also implied by the input homogeneity of $f^{\texttt{CD}}$.

In (7), innovation additively enters the production function. To relax HN, we consider a CD production function in which innovation also interacts with the inputs, as given by:

$$\ln Y_i = \ln f_{\texttt{CDh}}\left(K_i, L_i, I_i, \epsilon_i\right) = \alpha + \alpha_I I_i + \alpha_K \ln K_i + \alpha_{KI} I_i \ln K_i + \alpha_L \ln L_i + \alpha_{LI} I_i \ln L_i + \epsilon_i. \tag{9}$$

Then, the MRTS between L and K becomes:

$$MRTS_i^{\texttt{CDh}} = -\frac{\alpha_L + \alpha_{LI} I_i}{\alpha_K + \alpha_{KI} I_i}\frac{K_i}{L_i}. \tag{10}$$

In this case, neither HN nor IHN hold, unless the following condition holds:

$$\alpha_{KI}\alpha_L = \alpha_{LI}\alpha_K. \tag{11}$$

Moreover, the definition of EHN does not hold, generally, because:

$$\frac{\partial}{\partial K_i}E\left[\ln Y_i|K_i, L_i, I_i\right] = (\alpha_K + \alpha_{KI} I_i)\frac{1}{K_i} \quad \& \quad \frac{\partial}{\partial L_i}E\left[\ln Y_i|K_i, L_i, I_i\right] = (\alpha_L + \alpha_{LI} I_i)\frac{1}{L_i}.$$

Alternatively, it suffices to show that I is not strongly separable from the inputs in $f^{\texttt{CDh}}$, unless the following condition holds:

$$\alpha_{LI} = \alpha_{KI} = 0. \tag{12}$$

In summary, in the case of a CD production function with interactions as described in (9), the definitions of HN, IHN and EHN are not satisfied, unless particular conditions are met.

to simplify the presentation without losing any relevant information.

2.3 Estimation methods

The particular structure of models in which a dummy endogenous innovation additively enters the production function (7) is typically referred to as the dummy endogenous variable model. This specification, which is standard in the econometric literature focusing on the effect of innovation on productivity (Mohnen and Hall, 2013), can be consistently estimated using, among other methods, the standard IV estimator (IV-2SLS, hereafter; see, e.g., Wooldridge, 2010; Kelejian, 1971; Angrist and Krueger, 2001) and selecting the instruments within the determinants of the innovation function (Musolesi and Huiban, 2010). However, in a model with interactions between the endogenous dummy and the explanatory variables, the adoption of the IV-2SLS is more problematic. The main problem arises because each interaction term IX_r will be also endogenous. Therefore, estimating such a model by standard IV-2SLS would require finding instruments for all the endogenous variables $I, IX_r, r = 1, 2, \ldots, k$. If $\mathbf{Z}$ is a set of ρ valid instruments for I, then a natural set of instruments for IX_r is $\{Z_j X_r : Z_j \in \mathbf{Z}\}$. This approach results in a total of $(k + 1)\rho$ IVs, while it is well known that the estimation by standard IV-2SLS in the presence of many instruments exhibits substantial bias and makes inference inaccurate (see Hansen et al., 2008)

Consequently, we use an alternative IV approach proposed by Wooldridge (2010) (IV-W), which is more efficient than IV-2SLS and has a number of other interesting features. The implementation of IV-2SLS requires the *zero correlation assumption*, i.e., $\mathbf{E}(\epsilon) = \mathbf{E}(\epsilon\mathbf{X}) = \mathbf{E}(\epsilon\mathbf{Z})$ =0, whereas to use IV-W, the error term ϵ should have *zero conditional mean* – $\mathbf{E}(\epsilon \mid \mathbf{X}, \mathbf{Z}) = 0$ – which is a stronger exogeneity assumption that ensures that $\mathbf{E}(\epsilon) = \mathbf{E}(\epsilon\mathbf{X}) = \mathbf{E}(\epsilon\mathbf{Z})$ =0 but also implies that ε is uncorrelated with any function of $\mathbf{X}$ and $\mathbf{Z}$. Under the *zero conditional mean assumption*, the two-step approach proposed by Wooldridge for a model in which the endogenous dummy enters additively, as in (7), consists in estimating $P(I = 1 \mid \mathbf{X}, \mathbf{Z}) = F(\mathbf{X}, \mathbf{Z};\gamma)$ by (probit) maximum likelihood (ML) and then estimating the structural equation by IV-2SLS using 1, $\mathbf{X}$ and the estimated conditional probability $\hat{P}$ as instruments.

This two-step approach has a number of very interesting features since i) although generated instruments are used, the usual IV-2SLS standard errors and test statistics remain asymptotically valid for the second stage; ii) provided that the homoskedasticity assumption $\text{Var}(\epsilon|\mathbf{X}, \mathbf{Z}) = \sigma^2$ holds, the IV-2SLS estimator of the second step is asymptotically the most efficient for the class of estimators in which the instruments are functions of $(\mathbf{X}, \mathbf{Z})$; and, possibly most important, iii) this estimator possesses an important robustness property since the estimator of the structural equation in the second step is consistent, even if the model for $P(I = 1|\mathbf{X}, \mathbf{Z})$ is not correctly specified, while the requirements that are needed for consistency are much weaker (White, 1982). In other words, the innovation function does not have to be correctly specified to obtain consistent estimates for the parameters in the augmented production function.

To estimate a model with interactions such as (9), Wooldridge (2003, 2010) proposes a method (IV-W-H) that is a simple extension of that presented above and provides a solution to the problem of many instruments in the standard IV-2SLS. This is achieved by using the optimal – in terms of efficiency – instruments $P \equiv P(I = 1|\mathbf{X}, \mathbf{Z})$ for I and PX_r for $IX_r, r = 1, 2, \ldots, k$. This approach consists of the following two steps:

a. Estimate $P(I = 1|\mathbf{X}, \mathbf{Z}) = G(\mathbf{X}, \mathbf{Z}; \gamma)$ by ML and obtain the fitted values $\hat{P}$.

b. Estimate the structural equation by IV-2SLS using 1, $\mathbf{X}$, $\hat{P}$ and $\hat{P}\mathbf{X}$ as instruments.

As before, under the *zero conditional mean assumption*, the IV-2SLS of the structural equation is a consistent and asymptotically normal estimator, and again, the binary response model does not need to be correctly specified to achieve consistency. Some additional remarks are in order.

First, note that to estimate (9), we use the same parametrization as in Wooldridge (2003, 2010), where the interaction terms are mean centred, i.e., $I\left(\mathbf{X} - \overline{\mathbf{X}}\right)$. Second, note that $\overline{\mathbf{X}}$ is introduced in $\left(\mathbf{X} - \overline{\mathbf{X}}\right)$ as an estimator of $\mathbf{E}(\mathbf{X})$ and this should be accounted for when computing the standard errors. However, according to Wooldridge (2010), this will not have serious consequences, and heteroskedasticity-robust standard errors could still be employed; alternatively, bootstrapped standard errors are a viable alternative. Third, the parameter associated with innovation, α_I, measures, under weak assumptions, the average treatment effect (ATE), i.e., $\alpha_I = \alpha_I^{ATE}$ for both the additive (7) and the interaction model (9). As the interaction terms are mean centred and using the same notation as in Cerulli (2014), we can define the following:

$$\begin{aligned} ATE(\mathbf{X}) &= \mathbf{E}\left(\ln Y \mid \mathbf{X}, \mathbf{Z}, I = 1\right) - \mathbf{E}\left(\ln Y \mid \mathbf{X}, \mathbf{Z}, I = 0\right) \\ &= \alpha_I + \alpha_{KI}(\ln K - \overline{\ln K}) + \alpha_{LI}\left(\ln L - \overline{\ln L}\right). \end{aligned} \tag{13}$$

While in the additive model, the innovation effect is constant, specifying (9) allows for an heterogeneous effect of innovation across firms, which is a function of the production inputs. This is why, in the heterogeneous case, we will also focus attention on the estimation of the distribution of such an effect. Finally, suppose that the unobservable stochastic part of the model, ϵ, differs between innovative and non-innovative firms, that is, $\epsilon = \epsilon_0 + I(\epsilon_1 - \epsilon_0), \epsilon_1 \neq \epsilon_0$. In such a case, under a fairly weak assumption, the above procedure continues to be consistent. The assumption that is required for consistency is a mean independence assumption $\mathbf{E}[I(\epsilon_1 - \epsilon_0)|\mathbf{X}, \mathbf{Z}] = \mathbf{E}[I(\epsilon_1 - \epsilon_0)]$, which is generally reasonable for continuously distributed responses (see Angrist, 1991; Wooldridge, 2010).

3 Data and Results

3.1 Data

The data used for the analysis come from the tenth and last Survey on Manufacturing Firms ("Indagine sulle Imprese Manifatturiere") provided by Unicredit-Capitalia, which is complemented with balance sheets sourced from either AIDA (the Italian Balance Sheet Dataset of the Bureau van Dijk) or from the Chambers of Commerce Registry (UNICREDIT, 2008). The same survey, although in different waves, has been widely used in the economics literature on firms' innovation activities (for example Parisi et al., 2006; Hall et al., 2009). Information on the innovation activity of firms was derived from the survey that was conducted in 2007 and posed questions referring to the three-year period 2004-2006, while the variables derived from balance sheets refer to the year 2006. The initial sample comes from a stratified survey: all firms with more than 500 employees are included, while for the firms with fewer than 500 employees, a sample is extracted and stratified according to the information collected from the company registry for the variables size, value added, geographical location and industry. To estimate the econometric model, the main variables we consider are as follows:

1. The natural logarithm of value added (lnY): the measure refers to 2006 and is reported on the balance sheet.

2. The natural logarithm of the capital stock (lnK): the measure refers to 2006, is calculated by summing the value of fixed assets, and is estimated through a perpetual inventory method considering the usual rate of depreciation of 0.05, including investments. Both measures of fixed assets and investments are available from 1998 to 2006, and both are deflated with the respective aggregate price index (derived from ISTAT, the Italian National Statistical Office).

3. The natural logarithm of labour (lnL): number of employees in 2006 reported on the balance sheet.

4. Innovation, I: an innovation dummy taking value 1 if the firm affirmed having introduced at least one product or one process innovation in the previous three years (2004-2006), 0 otherwise.

5. Sectors: the manufacturing firms are classified by sector according to the two-digit ATECO2002 classification, which derives from the NACE Rev.1.1 Eurostat classification.

The aim of our preliminary data analysis is to identify the outliers in the sample, and the econometric sample is obtained by adopting the cleaning procedure detailed below.

First, we drop observations with missing or inconsistent values, resulting in a sample of 3237 firms. Second, outliers are detected using the boxplot rule (Tukey's method) on the variables under investigation. In so doing, another 232 observations (7% of the total) that exceed the boxplot's outer fences are dropped, resulting in a dataset of 3005 firms and a significant reduction in the range of the variables. We also search for outliers with respect to productivity. Consequently, we define total factor productivity as $TFP = Y/(K^{0.3}L^{0.7})$ and then again apply the boxplot rule to detect another 81 outlying observations. Therefore, the final dataset consists of 2924 firms.

Descriptive statistics of the main inputs and output are presented in Table 1, while the $\mathbf{Z}$ variables are described in Table 4 of Appendix C. Note that the percentage of innovating firms in the final dataset employed in this work is 64%. This value is the same as in Hall et al. (2009), who construct a panel data set starting from different waves of the same survey used in this work. This value is also very close to the percentage of innovating firms obtained using the Italian CIS survey (Hall et al., 2008).

3.2 Main results

First, we focus on the choice of the set of the $\mathbf{Z}$ variables. Importantly, as explained in subsection 2.3, Wooldridge's approach for both the additive and non-additive specifications (IV-W and IV-W-H) is consistent even if the model for $P(I = 1|\mathbf{X}, \mathbf{Z})$ is not correctly specified. This is an important robustness property.

In Table 2, we report the main results, which are obtained by using Wooldridge's approach and selecting the $\mathbf{Z}$ variables using a backward selection procedure, with a threshold of 0.10 for the p-value. The presentation of this single set of results, leaving the remaining to Appendix C, is chosen because of the extreme stability of the results across the different specifications. Appendix C presents many additional results. These results are obtained using alternative definitions of the vector $\mathbf{Z}$ and also adopting the standard IV-2SLS method, where the instruments are selected to be strong and valid.

Examining the results reported in the first two columns of Table 2, which are obtained estimating the additive specification, reveals their consistency with previous empirical work. While the baseline OLS approach does not provide evidence supporting the positive role of innovation in the production process, with $\widehat{\alpha_I} = 0.0189$ and being non-significant, when we use the IV-W method, we find a positive and significant Hicks-neutral effect of innovation, with $\widehat{\alpha_I} = 0.285$, which is in line with the existing literature (Hall, 2011), where this parameter ranges from approximately 0.2 to approximately 0.3. Moreover, the estimated coefficients of labour and capital are also in line with previous work, with $\widehat{\alpha_K} = 0.179$ and $\widehat{\alpha_L} = 0.741$, and a resulting elasticity of scale equal to 0.92.

We then turn to the estimation of the non-additive specification using the IV-W-H method, which is the main interest of this paper. The estimated average effect of innovation (0.272) is very close to that estimated using the additive model IV-W (0.285). Note that the results presented in Table 2 are all obtained using heteroskedasticity-robust standard errors. As stressed in subsection 2.3, when adopting the IV-W-H approach, bootstrapping the standard errors can be a viable solution to account for the fact that mean-centred variables are used for the interaction terms. We also apply a bootstrap with 1000 replications, which does not affect the results. The second important result that emerges is that the estimated parameters associated with the interaction terms, $\widehat{\alpha_{KI}}$ and $\widehat{\alpha_{LI}}$, appear to be statistically significant. This is a crucial result from this paper indicating that innovation does not have a neutral effect on production output. In fact, the production technology of innovative firms differs significantly from that of non-innovative firms, and this difference is produced by two elements: 1) a significant estimated parameter allowing for a shift, $\widehat{\alpha_I}$; 2) a significant change in the slope parameters, which is measured by $\widehat{\alpha_{KI}}$ and $\widehat{\alpha_{LI}}$, thus also affecting the shape of the production function.

Moreover, the additive specification can be tested against the non-additive one using a modified F test (Wooldridge, 1995). Unlike the standard F test, this statistic uses the sum of squared residuals from the second stage of the IV-2SLS regressions in the second step of the IV-W and IV-W-H approaches. The test rejects the additive specification in favour of the non-additive specification at the 5% significance level. In other words, the test rejects the condition (12) under which EHN is true. Further, we examine the condition under which HN and IHN hold in the non-additive specification using a Wald-type test; the test rejects, at the 5% level, the null hypothesis that (11) holds. Therefore, all definitions of Hicks neutrality are firmly rejected.

By examining the estimated parameters, we can obtain further insights into how technology differs between innovative and non-innovative firms. The estimated elasticity of labour for innovative firms equals 0.94 and is much higher than that for non-innovative firms, which is estimated at 0.43. The opposite holds for capital, with elasticity values equal to 0.33 for non-innovative and to 0.08 for innovative firms. As explained below, we find evidence of a capital-saving innovation. This supports the idea that innovation has an heterogeneous effect on the production process, which depends substantially on the production input with which it interacts. The consequences of the different interaction effects are visible in the shape of the estimated production function (fig.1).

The estimated $ATE(\mathbf{X})$ is equal to $0.272+0.510\left(\ln L-\overline{\ln L}\right)-0.248(\ln K-\overline{\ln K})$, where $\widehat{\alpha_I}=0.272$ is the estimated ATE, and we can focus our attention on the estimation of the underlying density function using the kernel approach. We use a second-order Gaussian kernel and cross-validation to choose the smoothing parameter. The estimated density is plotted in fig.2. As noted above, in our case, the ATE corresponds to the innovation coefficient. The estimated ATE returns an increase in value added of approximately 27% if innovation is introduced. Note that the ATE is the mean value of the $ATE(\mathbf{X})$. The estimated $ATE(\mathbf{X})$ is positive for most of the domain of the inputs: it ranges from -.677 to 1.453 and is positive for approximately 82% of the observations. Innovative firms are less productive than non-innovative firms only for very low values of labour associated with relatively high values of capital. On the contrary, the highest $ATE(\mathbf{X})$ appears for high values of labour associated with low values of capital. These results are directly related to the notion of locally progressive TC, which according to Chambers (1988), is fairly common in practice and has relevant implications. To the best of our knowledge, this is the first paper providing empirical evidence of such a situation.

Next, we discuss the elasticity of scale (see Basu and Fernald, 1997, for a through discussion on estimated returns to scale). While non-innovative firms are characterized by decreasing returns to scale, with an estimated elasticity of scale $\widehat{\alpha_K} + \widehat{\alpha_L} = 0.75$, innovative firms exhibit slightly increasing returns to scale, with an estimated elasticity of scale $\widehat{\alpha_K}+\widehat{\alpha_{KI}}+\widehat{\alpha_L}+\widehat{\alpha_{LI}} = 1.02$. Using Wald tests, the hypothesis of constant returns to scale is not rejected at the 10% significance level for innovative firms, while it is rejected at the 0.1% level for non-innovative firms. Instead, when we estimated a model assuming a common technology (the additive one), the elasticity of scale equals 0.92. In this case, the hypothesis of constant returns to scale is rejected at the 0.1% level. This indicates that when estimating a production function with added innovation, we face a kind of heterogeneity bias since the value 0.92 is obtained mingling two heterogeneous technologies: that of innovative firms characterized by constant returns to scale and that of non-innovative firms with decreasing returns to scale.

A final object that is of great relevance is the MRTS, the analysis of which provides us with information on the nature of technological progress. To obtain insights into such an object, we first use contour plots (in figure 3) to represent the estimated production function. Such isolines have a direct economic interpretation as the estimated isoquants. Given the shapes of the isoquants in figure 3, we find that the MRTS – the slope of the isoquants – is higher for innovative than for non-innovative firms. The substitution opportunities are reduced for innovative firms. The values of the relative MRTS tell us that to compensate for a 1% change in labour, an innovative firm should change capital by approximately 12%, while a non-innovative firm needs a change in capital of only approximately 1.3%.

To obtain complementary information, we also calculate the MRTS using (10) and then focus on the estimation of its density function (figure 4). We specifically estimate the conditional density of the MRTS, conditional to innovation. With innovation being a discrete variable, we adopt the approach of (Hall et al., 2004), which uses generalized product kernels to deal with mixed data and cross-validation to choose the smoothing parameters. Interestingly, the smoothing parameter associated with innovation goes to zero. This not only suggests that innovation is relevant – meaning that the two densities are not the same –but also indicates that the generalized estimator collapses to the standard frequency estimator. This result goes further in the direction of fully heterogeneous technology.

4 Conclusion

In standard econometric models, innovation additively enters the production function, imposing a strict condition of Hicks neutrality. We depart from this restrictive framework by considering a production function that allows a heterogeneous effect of innovation on the production process and relaxes the Hicks neutrality assumption. We derive conditions, for common parametric specifications, under which neutrality holds and that are easily testable through common Wald-type tests. Further, taking into consideration the endogenous character of innovation, we estimate the model by adopting an instrumental variables approach that addresses the problem of many instruments when estimating a model with interactions. The econometric analysis rejects Hicks neutrality and indicates that innovation produces a non-neutral effect on the production process, which is obtained because of the joint presence of a shift in the production technology and a change in the slope of the isoquants. The latter indicates that innovative firms are *capital saving* compared with non-innovative firms. Moreover, as a consequence of the joint effect described above, a *locally progressive* technical change is also observed, because, while for most of the domain of the inputs, innovation has a positive effect, for a small part of it, innovative firms are less productive than non-innovative firms. Overall, our results indicate fully heterogeneous production technologies when comparing innovative to non-innovative firms. These findings have interesting policy implications, as they highlight the complex fashion in which innovation affects the production process. To the best of our knowledge, this is the first study supporting such evidence.

Further studies may consider extending the analysis to other countries or to specific sectors. Methodological extensions may be achieved by considering a panel data framework to account for the time dimension or by adopting nonparametric methods to highlight the potential presence of localized technical change.

References

Acemoglu, D. (2015). Localised and biased technologies: Atkinson and Stiglitz's New view, induced innovations, and directed technological change. *Economic Journal*, 125(583):443–463.

Angrist, J. (1991). Instrumental variables estimation of average treatment effects in econometrics and epidemiology. Technical report, National Bureau of Economic Research, Inc.

Angrist, J. D. and Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *The Journal of Economic Perspectives*, 15(4):69–85.

Antle, J. M. and Capalbo, S. M. (1988). An introduction to recent developments in production theory and productivity measurement. *Agricultural productivity: Measurement and explanation*, pages 17–95.

Atkinson, A. and Stiglitz, J. (1969). A new view of technological change. *The Economic Journal*, 79(315):573–578.

Basu, S. and Fernald, J. G. (1997). Returns to Scale in U.S. Production: Estimates and Implications. *Journal of Political Economy*, 105(2):249–283.

Blackorby, C., Knox Lovell, C. A., and Thursby, M. C. (1976). Extended Hicks Neutral Technical Change. *The Economic Journal*, 86(344):845–852.

Bound, J., Jaeger, D., and Baker, R. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, 90(430):443–450.

Cerulli, G. (2014). Ivtreatreg: A command for fitting binary treatment models with heterogeneous response to treatment and unobservable selection. *Stata Journal*, 14(3):453–480.

Chambers, R. G. (1988). *Applied production analysis: a dual approach*. Cambridge University Press.

Crépon, B., Duguet, E., and Mairessec, J. (1998). Research, Innovation And Productivi[Ty: An Econometric Analysis At The Firm Level. *Economics of Innovation and New Technology*, 7(2):115–158.

Griliches, Z. (1998). Patent Statistics as Economic Indicators: A Survey. In Griliches, Z., editor, *R&D and Productivity: The Econometric Evidence*, pages 287–343. University of Chicago Press, Chicago.

Hall, B. H. (2011). Innovation and productivity. *Nordic Economic Policy Review*, 2:167–203.

Hall, B. H., Lotti, F., and Mairesse, J. (2008). Employment, innovation, and productivity: evidence from italian microdata. *Industrial and Corporate Change*, 17(4):813–839.

Hall, B. H., Lotti, F., and Mairesse, J. (2009). Innovation and productivity in smes: empirical evidence for italy. *Small Business Economics*, 33(1):13–33.

Hall, P., Racine, J., and Qi, L. (2004). Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association*, 99(468):1015–1026.

Hansen, C., Hausman, J., and Newey, W. (2008). Estimation with many instrumental variables. *Journal of Business & Economic Statistics*, 26(4):398–422.

Hicks, J. (1963). *The theory of wages*. Springer.

Jones, R. W. (1965). The structure of simple general equilibrium models. *Journal of Political Economy*, 73(6):557–572.

Kelejian, H. H. (1971). Two-stage least squares and econometric systems linear in parameters but nonlinear in the endogenous variables. *Journal of the American Statistical Association*, 66(334):373–374.

Kennedy, C. and Thirlwall, A. (1977). Extended hicks neutral technical change-a comment. *Economic Journal*, 87(348).

Kennedy, C. and Thirlwall, A. P. (1972). Surveys in applied economics: technical progress. *The Economic Journal*, 82(325):11–72.

Lööf, H., Mairesse, J., and Mohnen, P. (2016). CDM 20 Years After CDM 20 Years After.

Mairesse, J. and Mohnen, P. (2010). Using innovation surveys for econometric analysis. In *Handbook of the Economics of Innovation*, volume 2, pages 1129–1155. Elsevier.

Mohnen, P. and Hall, B. H. (2013). Innovation and Productivity: An Update. *Eurasian Business Review*, 3(1):47–65.

Montiel Olea, J. L. and Pflueger, C. (2013). A robust test for weak instruments. *Journal of Business & Economic Statistics*, 31(3):358–369.

Morimoto, Y. (1974). Neutral technical progress and the separability of the production function. *The Economic Studies Quarterly (Tokyo. 1950)*, 25(3):66–69.

Musolesi, A. and Huiban, J.-P. (2010). Innovation and productivity in knowledge intensive business services. *Journal of Productivity Analysis*, 34(1):63–81.

Pakes, P. and Griliches, Z. (1984). Patents and R&D at the Firm Level: A First Look. In Griliches, Z., editor, *R&D, Patents, and Productivity*, pages 55–71. Chicago University Press, Chicago.

Parisi, M. L., Schiantarelli, F., and Sembenelli, A. (2006). Productivity, innovation and R&D: Micro evidence for Italy. *European Economic Review*, 50(8):2037–2061.

Steedman, I. (1985). On the 'Impossibility' of Hicks-Neutral Technical Change. *The Economic Journal*, 95(379):746–758.

UNICREDIT (2008). Decima indagine sulle imprese manifatturiere italiane. Technical report, UNICREDIT.

Uzawa, H. and Watanabe, T. (1961). A note on the classification of technical inventions. *The Economic Studies Quarterly (Tokyo. 1950)*, 12(1):68–72.

Verbeek, M. (2008). *A guide to modern econometrics.* John Wiley & Sons.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, pages 1–25.

Wooldridge, J. M. (1995). Score diagnostics for linear models estimated by two stage least squares. *Advances in econometrics and quantitative economics: Essays in honor of Professor CR Rao*, pages 66–87.

Wooldridge, J. M. (2003). Further results on instrumental variables estimation of average treatment effects in the correlated random coefficient model. *Economics letters*, 79(2):185–191.

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

Table 1: Descriptive statistics

	lnVA			lnK			lnL		
	All firms	I=0	I=1	All firms	I=0	I=1	All firms	I=0	I=1
minimum	2.52	2.52	2.75	3.65	4.11	3.65	2.30	2.30	2.30
median	7.43	7.25	7.53	7.71	7.50	7.81	3.50	3.30	3.58
mean	7.46	7.32	7.53	7.60	7.46	7.68	3.52	3.40	3.59
maximum	9.32	9.29	9.33	9.78	9.77	9.78	5.50	5.48	5.50
st. deviation	0.81	0.82	0.80	1.11	1.13	1.09	0.73	0.73	0.72

Table 2 - Main results: lnVA as dependent variable

	OLS	*IV-W*	*IV-W-H*
lnL	.750***	.741***	.428***
	(.0166)	(.0218)	(.1044)
lnK	.185***	.179***	.327***
	(.0108)	(.0135)	(.0639)
Innovation	.0189	0.285***	0.272***
	(.0159)	(.0774)	(.0806)
IlnL			0.510***
			(.1545)
IlnK			-0.248**
			(.0970)
intercept	3.228***	2.870***	3.120***
	(.0578)	(.1203)	(.2618)
N	2924	2239	2239
R^2	.758	.727	.704
adj. R^2	.755	.724	.701
Endogeneity test		12.177***	23.341***
Montiel-Pflueger F		168.204	
[critical value]		[37.418]	
ATE	.019	.285	.272
Elasticity of labour	.75	.74	
Elasticity of labour ($I = 0$)			.43
Elasticity of labour ($I = 1$)			.94
Elasticity of capital	.19	.18	
Elasticity of capital ($I = 0$)			.33
Elasticity of capital ($I = 1$)			.08
Elasticity of scale	.94	.92	
Elasticity of scale ($I = 0$)			.75
Elasticity of scale ($I = 1$)			1.02
Relative MRTS	-4.12	-4.14	
Relative MRTS ($I = 0$)			-1.31
Relative MRTS ($I = 1$)			-11.78

Montiel-Pflueger test for weak instruments, null hypothesis that instruments are weak. $\tau = 5\%$, confidence level $\alpha = 5\%$, test not applicable in regressions with one endogenous variable.
Sectors not presented in the table.
Wooldridge's (1995) robust score test for overidentifying restrictions not applicable.
Endogeneity test according to Wooldridge's (1995) score test.
Robust standard errors, Huber/White/sandwich estimator
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, robust standard errors in parentheses

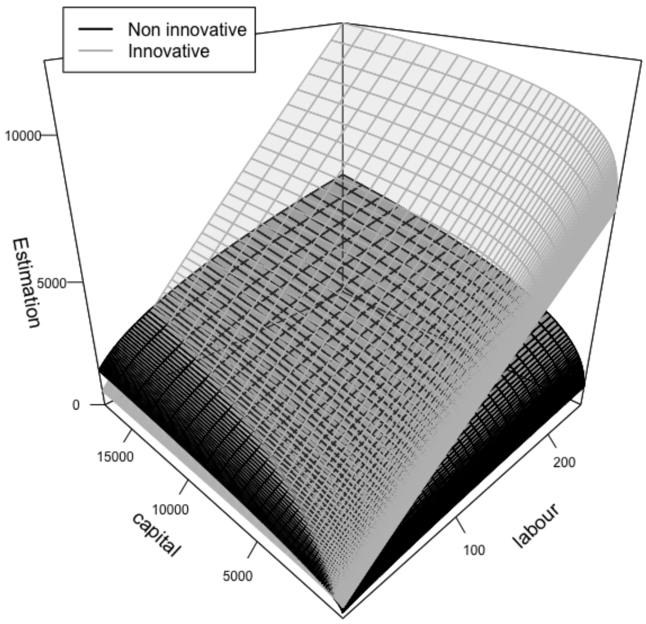

Figure 1: Estimated production function (IV-W-H method)

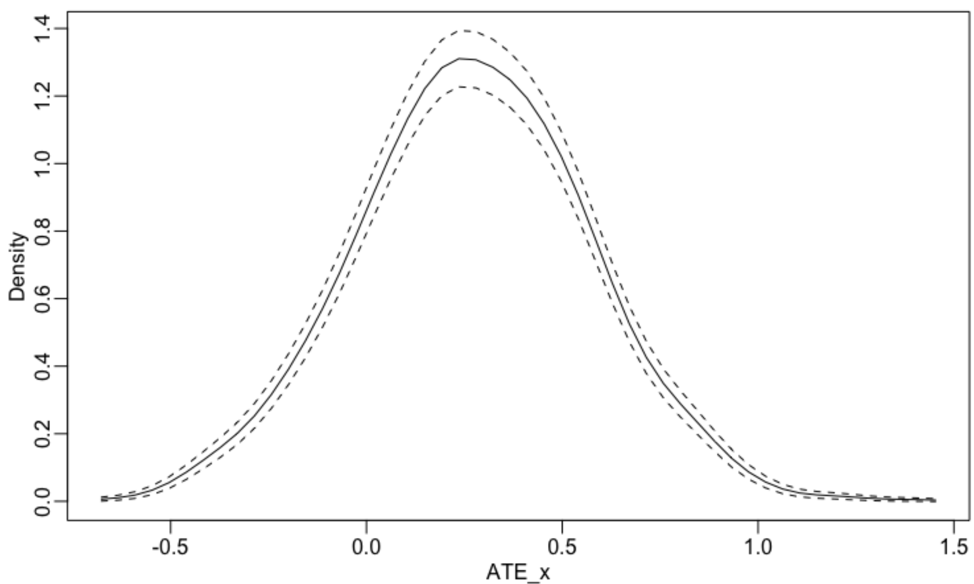

Figure 2: Distribution of the ATE($\mathbf{X}$), estimated using gaussian kernels. The dashed lines represent 95% bootstrapped confidence bands. The bandwidth (0.07) is selected by least squares cross validation. The mean value of ATE ($\mathbf{X}$) corresponds to ATE and is equal to 0.272.

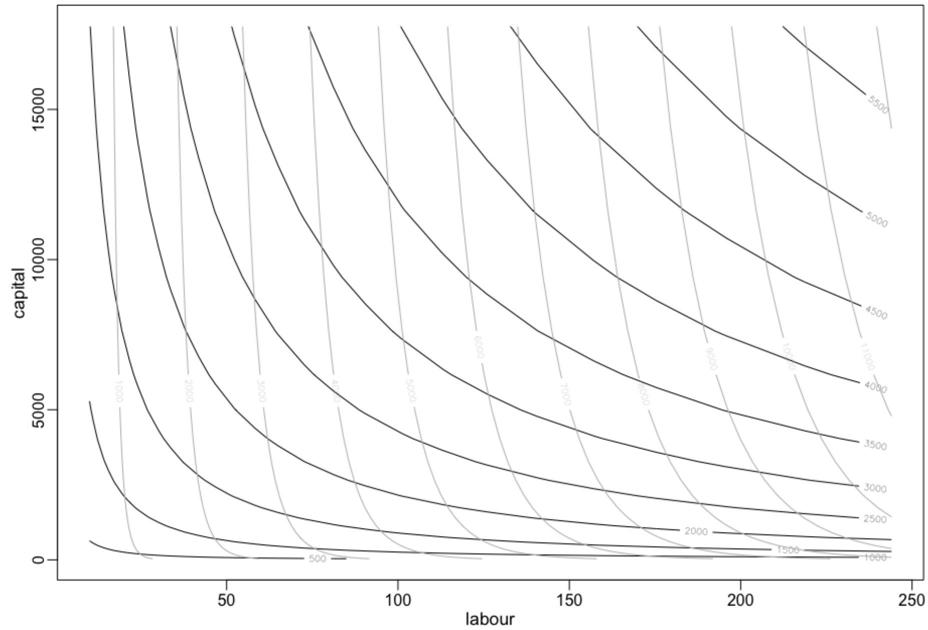

Figure 3: Contours of the production function estimated with IV-W-H. The grey lines correspond to the contours of the innovative firms, while the black lines correspond to those of non-innovative firms.

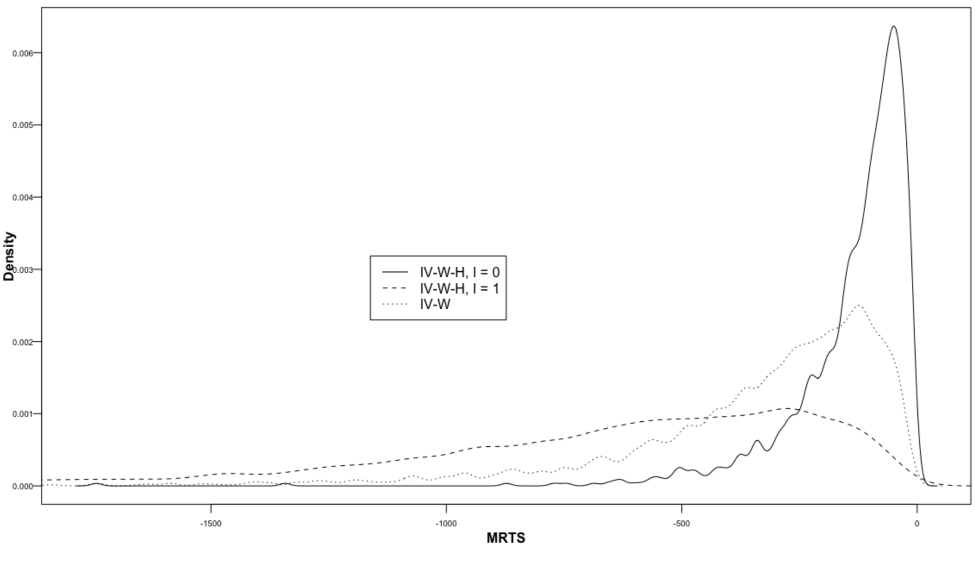

Figure 4: Density distribution of the MRTS. The dotted line corresponds to the MRTS estimated by the IV-W approach. The solid line and the dashed line correspond to the MRTS of non-innovative and innovative firms in the IV-W-H case.

Supplementary Material

Appendix A: Relationships among HN, IHN and EHN

The definitions of HN, IHN and EHN are described in subsection 2.1. In this appendix we provide additional insights into the work of Blackorby et al. (1976, sec.4). Specifically, we provide an illustrative, simplified presentation of the conditions under which the above definitions are equivalent, and we highlight their relationships using the following figure.

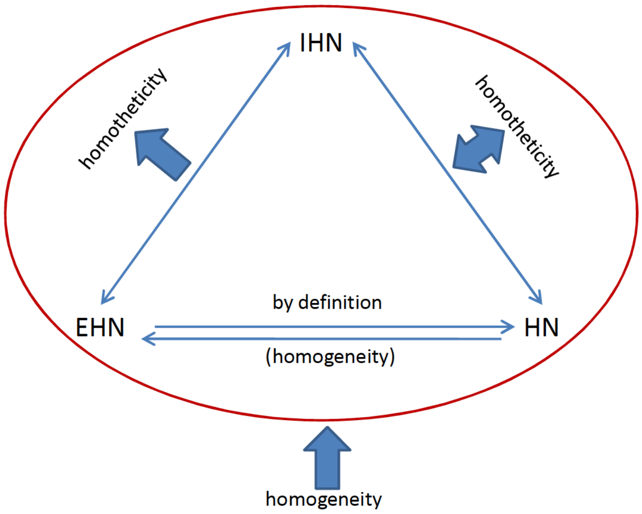

Figure 5: Illustration of the relationships among the different definitions of Hicks neutrality; double arrows symbol implies equivalence.

The relationship between EHN and HN

Consider the production function $Y = f(\mathbf{X}, I)$, where $\mathbf{X}$ is a vector of inputs and I is a variable that measures innovation. If I is EHN, then I is strongly separable from $\mathbf{X}$. Therefore, it is also weakly separable from $\mathbf{X}$. For definitions of weak and strong separability, we refer to Chambers (1988, p. 42-46). Morimoto (1974) proves that weak separability of I from $\mathbf{X}$ is equivalent to I being HN. Therefore, if I is EHN, it is also HN. In general, the converse does not hold, that is, I being HN does not necessarily imply that it is also EHN. Uzawa and Watanabe (1961) prove that in the case of an input-homogeneous production function f, I is HN if and only if f is multiplicatively decomposable into a function of I and a function of $\mathbf{X}$. Therefore, if the production function is input homogeneous, then HN implies EHN.

The relationship between HN and IHN

Moreover, HN and IHN are generally not equivalent terms. Blackorby et al. (1976, fig.I) describe technical progress that is HN but not IHN, and vice versa. They prove that HN and IHN are equivalent if and only if the production function is input homothetic. An obvious consequence of this theorem is that in a homothetic production function, if I is not HN (IHN), then it is also not IHN (HN). Moreover, if I is HN (IHN) but not IHN (HN), then f is not input homothetic.

The relationship between IHN and EHN

In general, IHN and EHN are also not equivalent. Blackorby et al. (1976) prove that if I is IHN and EHN, then the production function is input homothetic. By negation, it can be shown that if f is homothetic and I is EHN (IHN), then I should also be IHN (EHN). For example, in (14), we describe an empirical specification of a TL function with an added I. In this case, I is EHN but, generally, not IHN, unless certain homotheticity conditions hold.

Finally, note that according to the above, if the production function is input homogeneous, then HN, IHN and EHN are equivalent. Indeed, baring that homogeneity implies homotheticity, if f is homogeneous and I follows one of these definitions of neutrality, then I should also follow the other two (see also fig.5).

Appendix B: Hicks neutrality and the translog specification

In this study, we have also estimated by IV-2SLS, IV-W and IV-W-H the TL specifications that are presented in (14) and (17) below. Nevertheless, the results of the F-type tests (see Wooldridge, 1995) that compare the CD and the TL models do not reject the CD specification and the estimations of the TL model provide poor results in terms of magnitude and significance level for most of the coefficients.[2] For these reasons, the main interest in this paper is on the CD specification. However, as the TL specification is the most popular Diewert-flexible form and a widely used direct generalization of the CD specification, we extend the analysis in section 2.2 to account for the TL case. Particularly, in this appendix, we provide insights into HN within a TL framework and provide easily testable conditions for the presence of HN in TL specifications. Therefore, the following analysis may be useful in empirical studies that adopt a TL framework while relaxing HN.

Under the assumption that innovation enters additively, the econometric specifica-

[2]The results of the estimations of the TL specifications are available upon request.

tion of a TL production function with conventional inputs K and L is:

$$\begin{aligned}\ln Y_i &= \ln f^{\mathtt{TL}}\left(K_i, L_i, I_i, \epsilon_i\right) = \\ &= \alpha + \alpha_I I_i + \alpha_K \ln K_i + \alpha_L \ln L_i + \alpha_{KL} \ln K_i \ln L_i + \alpha_{K2}\frac{1}{2}\ln K_i^2 + \alpha_{L2}\frac{1}{2}\ln L_i^2 + \epsilon_i,\end{aligned} \tag{14}$$

and the MRTS between L and K is described by:

$$MRTS_i^{\mathtt{TL}} = -\frac{\alpha_L + \alpha_{KL}\ln K_i + \alpha_{L2}\ln L_i}{\alpha_K + \alpha_{KL}\ln L_i + \alpha_{K2}\ln K_i}\frac{K_i}{L_i}, \tag{15}$$

according to which I is HN but not IHN, because the MRTS is not constant along a ray from the origin, unless $\alpha_{KL} + \alpha_{L2} = 0$ and $\alpha_{KL} + \alpha_{K2} = 0$. Moreover, EHN holds because $f^{\mathtt{TL}}$ is multiplicatively decomposable into a function of I and a function of the inputs. Alternatively, it can be shown that:

$$\frac{\partial}{\partial K_i} E\left[\ln Y_i | K_i, L_i, I_i\right] = \left(\alpha_K + \alpha_{KL}\ln L_i + \alpha_{K2}\ln K_i\right)\frac{1}{K_i} \tag{16a}$$

$$\frac{\partial}{\partial L_i} E\left[\ln Y_i | K_i, L_i, I_i\right] = \left(\alpha_L + \alpha_{KL}\ln K_i + \alpha_{L2}\ln L_i\right)\frac{1}{L_i} \tag{16b}$$

Note that while in the case of a CD production function with added innovation, all the above definitions of neutrality are imposed, the TL specification with added innovation imposes HN and EHN but not IHN.

Further, a TL function with interaction terms between innovation and the explanatory variables is described by:

$$\begin{aligned}\ln Y_i = {} & \alpha + \alpha_I I_i + \alpha_K \ln K_i + \alpha_{KI} I_i \ln K_i + \alpha_L \ln L_i + \alpha_{LI} I_i \ln L_i + \alpha_{KL} \ln K_i \ln L_i + \\ & + \alpha_{KLI} I_i \ln K_i \ln L_i + \alpha_{K2}\frac{1}{2}\ln K_i^2 + \alpha_{K2I}\frac{1}{2} I_i \ln K_i^2 + \alpha_{L2}\frac{1}{2}\ln L_i^2 + \alpha_{L2I}\frac{1}{2} I_i \ln L_i^2 + \epsilon_i.\end{aligned} \tag{17}$$

In this case, the MRTS between L and K is given by the following equation:

$$MRTS_i^{\mathtt{TLh}} = -\frac{\alpha_L + \alpha_{LI} I_i + (\alpha_{KL} + \alpha_{KLI} I_i)\ln K_i + (\alpha_{L2} + \alpha_{L2I} I_i)\ln L_i}{\alpha_K + \alpha_{KI} I_i + (\alpha_{KL} + \alpha_{KLI} I_i)\ln L_i + (\alpha_{K2} + \alpha_{K2I} I_i)\ln K_i}\frac{K_i}{L_i}. \tag{18}$$

Generally, the definitions of HN and IHN are not satisfied because the MRTS in (18) is dependent on I. By contradiction, it is proven that the definition of EHN does not hold either, unless $\alpha_{LI} = \alpha_{KI} = \alpha_{KLI} = \alpha_{K2I} = \alpha_{L2I} = 0$. Finally, assuming non-zero coefficients, HN holds only if at least one of the following conditions is satisfied:

$$\frac{\alpha_K}{\alpha_L} = \frac{\alpha_{KL}}{\alpha_{L2}} = \frac{\alpha_{K2}}{\alpha_{KL}} = \frac{\alpha_{KI}}{\alpha_{LI}} = \frac{\alpha_{K2I}}{\alpha_{KLI}} = \frac{\alpha_{KLI}}{\alpha_{L2I}}, \text{ or} \tag{19a}$$

$$\frac{\alpha_{LI}}{\alpha_L} = \frac{\alpha_{KI}}{\alpha_K} = \frac{\alpha_{KLI}}{\alpha_{KL}} = \frac{\alpha_{K2I}}{\alpha_{K2}} = \frac{\alpha_{L2I}}{\alpha_{L2}} \tag{19b}$$

Under condition (19a), I in (17) is also IHN because (8) and (18) coincide. Under condition (19b), (18) is identical to (15). In this case, I is IHN if, in addition to (19b), it also holds that $\alpha_{KL} + \alpha_{L2} = 0$ and $\alpha_{KL} + \alpha_{K2} = 0$.

In summary, in the case of a TL production function with interactions, the definitions of HN, IHN and EHN are not satisfied, unless particular conditions are met. These conditions are summarized in Table 3.

Table 3 - Conditions for neutrality

Type	TL additive	TL non-additive
HN	–	$\frac{\alpha_K}{\alpha_L} = \frac{\alpha_{KL}}{\alpha_{L2}} = \frac{\alpha_{K2}}{\alpha_{KL}} = \frac{\alpha_{KI}}{\alpha_{LI}} = \frac{\alpha_{K2I}}{\alpha_{KLI}} = \frac{\alpha_{KLI}}{\alpha_{L2I}}$ or $\frac{\alpha_{LI}}{\alpha_L} = \frac{\alpha_{KI}}{\alpha_K} = \frac{\alpha_{KLI}}{\alpha_{KL}} = \frac{\alpha_{K2I}}{\alpha_{K2}} = \frac{\alpha_{L2I}}{\alpha_{L2}}$
IHN	$\alpha_{KL} + \alpha_{L2} = 0$ and $\alpha_{KL} + \alpha_{K2} = 0$	$\frac{\alpha_K}{\alpha_L} = \frac{\alpha_{KL}}{\alpha_{L2}} = \frac{\alpha_{K2}}{\alpha_{KL}} = \frac{\alpha_{KI}}{\alpha_{LI}} = \frac{\alpha_{K2I}}{\alpha_{KLI}} = \frac{\alpha_{KLI}}{\alpha_{L2I}}$ or $\frac{\alpha_{LI}}{\alpha_L} = \frac{\alpha_{KI}}{\alpha_K} = \frac{\alpha_{KLI}}{\alpha_{KL}} = \frac{\alpha_{K2I}}{\alpha_{K2}} = \frac{\alpha_{L2I}}{\alpha_{L2}}$, $\alpha_{KL} + \alpha_{L2} = 0$, $\alpha_{KL} + \alpha_{K2} = 0$
EHN	–	$\alpha_{LI} = \alpha_{KI} = \alpha_{KLI} = \alpha_{K2I} = \alpha_{L2I} = 0$

Appendix C: Robustness checks

In this appendix, we provide robustness checks of the main results presented in subsection 3.2. A natural approach is first to estimate the additive model by IV-2SLS and focus attention on the choice of the IVs. We apply the IV selection method described below and arrive at different choices of sets **Z** of strong and valid instruments. Then, we apply Wooldridge's approach to perform IV-W and IV-W-H estimations using the **Z** sets selected previously. As presented below, the results are very robust in all estimations.

We first select an initial set of potential IVs according to the literature on KPF (see, e.g., Hall et al., 2009; Musolesi and Huiban, 2010). This set is given in the table below.

Table 4 - Initial set of IVs

Category	Variables	Description	Type
Firm Characteristics	NW, NE, C, S, Age, Group, Consortium	Firm's characteristics including geographical location; age and group or consortium membership	binary (0,1)
Human capital	RD_Personnnel	Percentage of employees in Research and Development activities	numeric
Objectives of investment	BetterProd, MoreProd, NewProd, Env, CostRed, Advert, SellNet, SellAss	Objectives including ameliorating the product, produce more, introduce a new one, reduce the environmental impact, reduce costs, to increase the selling network or to ameliorate it, respectively	binary (0,1)
Market penetration	MarkPenEU15, MarkPenEU2004, MarkPenRussia, MarkPenOtherEU, MarkPenAfrica, MarkPenAsia, MarkPenCina, MarkPenUSMex, MarkPenSouthAm, MarkPenOce	Market penetration in different world regions, including EU member states, Africa, Asia, China, U.S., Canada, Mexico, South America and Oceania	binary (0,1)
Commercial agreements	CommAgrEU15, CommAgrEU2004, CommAgrRussia, CommAgrOtherEU, CommAgrAfrica, CommAgrAsia, CommAgrCina, CommAgrUSMex, CommAgrSouthAm, CommAgrOce	Commercial agreements in world regions, as mentioned above	binary (0,1)
Patent acquisition	PatBuyEU15, PatBuyEU2004, PatBuyRussia, PatBuyOtherEU, PatBuyAfrica, PatBuyAsia, PatBuyCina, PatBuyUSMex, PatBuySouthAm, PatBuyOce	Location of the aforementioned world regions where the firm acquired patents	binary (0,1)
Production overseas	ProdAbroadEU15, ProdAbroadEU2004, ProdAbroadRussia, ProdAbroadOtherEU, ProdAbroadAfrica, ProdAbroadAsia, ProdAbroadCina, ProdAbroadUSMex, ProdAbroadSouthAm, ProdAbroadOce	Production located in the aforementioned world regions	binary (0,1)
Competitiveness	LowCompet, HighCompet, SmallProdScale	Perceived level of competitiveness and scale of production compared to competitors	binary (0,1)
Financial specs	ListedComp, FinanIncent	Listed company or receiving financial incentives	binary (0,1)

As highlighted in Bound et al. (1995), a major pitfall that results in inconsistency and large finite sample bias exists when selecting instruments that are weakly correlated with the endogenous variable. To avoid the presence of weak instruments and ensure the validity of the IVs, we follow a two-step procedure.

First, we regress innovation on $\mathbf{X}$ and the above set and adopt a backward selection algorithm to choose an initial set of potential IVs that could be strongly correlated with innovation. The sets corresponding to a 10% and a 5% threshold are presented in the table below.

Table 5 - IV sets

set	instruments
10% set	FinanIncent, BetterProd, MarkPenEU15, C, EUCompet, NewProd, MarkPenEU2004, ProdAbroadEU15, Age, CommAgrAfrica, RD_Personnnel, MoreProd, HighCompet
5% set	FinanIncent, BetterProd, MarkPenEU15, C, EUCompet, NewProd, HighCompet, Age
IV1	BetterProd, MarkPenEU15, EUCompet
IV2	FinanIncent, BetterProd, MarkPenEU15, C, EUCompet
IV3	FinanIncent, BetterProd, MarkPenEU15, EUCompet, NewProd

Generally, the bias of the IV-2SLS estimator increases as the correlation between the IVs and the endogenous variable decreases and as the number of instruments increases. For this reason, in a second step, we estimate by IV-2SLS the additive specification using all the possible combinations of IVs from the 10% set. Since heteroskedasticity-robust standard errors are considered, for each specification, we apply the robust score tests by Wooldridge (1995) to test endogeneity and overidentifying restrictions. We also use the Montiel-Pflueger test to detect the presence of weak instruments (Montiel Olea and Pflueger, 2013). Unlike traditional tests, the Montiel-Pflueger test also accounts for heteroskedastic, serially correlated errors.

The above post-estimation tests indicate IV sets of strong and valid instruments. According to the Montiel-Pflueger test, 42 combinations provide strong IVs. We also find that for 54 sets, the robust score test cannot reject the validity of the IVs. We finally select 13 sets of valid and strong instruments for which the exogeneity test is not rejected. These combinations provide very robust results; both the estimated coefficients and the significance levels are stable across choices of sets. For reasons of brevity, in Table 6, we present the IV-2SLS, IV-W and IV-W-H estimations using these three sets[3], while Table 7 shows the averages of the IV-2SLS, IV-W and IV-W-H estimations for the CD specifications using the above 13 sets.

The first three columns of Table 6 present the IV-2SLS results of the additive specification. The estimated effect of innovation on productivity is between 0.24 and 0.31 and is significant at 0.01 level. In columns 4 to 6, we present the results of the IV-W estimations. The estimated effects of labour, capital and innovation are similar to those from the IV-2SLS estimations, in terms of both the magnitude of the coefficients and the confidence levels. The estimated innovation parameter is between 0.23 and 0.28 and significant at the 0.01 level. Finally, in the last three columns, we also present the results of the IV-W-H estimation. The estimated innovation parameter is significant at the 0.05 level, ranges between 0.22 and 0.31 and is similar to the IV-2SLS and IV-W estimates. An exhaustive presentation of the IV-W-H results and the comparison to the other approaches is given in subsection 3.2.

[3]The remaining sets and the results they provide are available upon request.

Table 6 - Cobb Douglas estimations

	IV-2SLS			*IV-W*			*IV-W-H*		
	IV1	IV2	IV3	IV1	IV2	IV3	IV1	IV2	IV3
lnL	.735***	.737***	.737***	.737***	.738***	.737***	.457***	.471***	.482***
	(.0197)	(.0200)	(.0200)	(.0196)	(.0200)	(.0200)	(.1002)	(.0939)	(.0900)
lnK	.179***	.181***	.181***	.179***	.181***	.181***	.294***	.294***	.300***
	(.0125)	(.0126)	(.0126)	(.0123)	(.0125)	(.0125)	(.0646)	(.0601)	(.0595)
$Innovation$	.312***	.247***	.242***	.283**	.233***	.229**	.308***	.221**	.219**
	(.1097)	(.0862)	(.0890)	(.1102)	(.0862)	(.0903)	(.117)	(.0903)	(.0935)
$IlnL$							.451***	.438***	.420***
							(.1536)	(.1430)	(.1372)
$IlnK$							-.193*	-.189**	-.198**
							(.1020)	(.0933)	(.0923)
intercept	2.850***	2.943***	2.949***	2.885***	2.960***	2.964***	3.256***	3.270***	3.189***
	(.1481)	(.1223)	(.1246)	(.1489)	(.1230)	(.1265)	(.2969)	(.2535)	(.2586)
N	2635	2474	2474	2635	2474	2474	2635	2474	2474
R^2	.725	.736	.737	.731	.738	.738	.710	0.723	0.724
adj. R^2	.723	.733	.734	.728	.735	.736	.707	.720	.721
Robust score test	4.521	5.573	6.492						
Endogeneity test	7.685***	6.878***	6.004**	6.094**	5.974**	5.182**	16.044***	16.187***	15.904***
Montiel-Pflueger F	24.814	21.055	20.787	74.160	109.329	104.643			
[critical value]	[15.760]	[20.981]	[20.751]	[37.418]	[37.418]	[37.418]			
	First stage regression			Probit					
FinanIncent		.110***	.109***		.333***	.331***			
		(.0224)	(.0224)		(.0698)	(.0699)			
MarkPenEU15	.135***	.126***	.123***	.431***	.422***	.414***			
	(.0268)	(.0275)	(.0275)	(.0943)	(.0994)	(.0994)			
BetterProd	.120***	.106***	.115***	.348***	.310***	.333***			
	(.0196)	(.0203)	(.0207)	(.0585)	(.0606)	(.0612)			
EUCompet	.058***	.064***	.062**	.176**	.191***	.182**			
	(.0233)	(.0239)	(.0240)	(.0699)	(.0729)	(.0727)			
C		.075***			.206***				
		(.0259)			(.0739)				
NewProd			.099**			.280**			
			(.0405)			(.1224)			
1st-stage partial R^2	.0241	.0369	.0358	.0248	.0382	.0362			
1st-stage robust F	25.47***	22.73***	22.61***	74.16***	109.33***	104.64***			
ATE	.312	.247	.242	.283	.233	.229	.308	.221	.219
Elasticity of labour	.74	.74	.74	.74	.74	.74			
Elasticity of labour $(I=0)$							.46	.47	.48
Elasticity of labour $(I=1)$							.91	.91	.90
Elasticity of capital	.18	.18	.18	.18	.18	.18			
Elasticity of capital $(I=0)$							.29	.29	.30
Elasticity of capital $(I=1)$							.10	.11	.10
Elasticity of scale	.91	.92	.92	.92	.92	.92			
Elasticity of scale $(I=0)$							.75	.77	.78
Elasticity of scale $(I=1)$							1.00	1.01	1.00
Relative MRTS	-4.11	-4.07	-4.07	-4.12	-4.08	-4.07			
Relative MRTS $(I=0)$							-1.55	-1.60	-1.61
Relative MRTS $(I=1)$							-8.99	-8.66	-8.84

Montiel-Pflueger test for weak instruments, null hypothesis that the instruments are weak, $\tau = 5\%$, confidence level $\alpha = 5\%$
Sectors not presented in the table. Only the instruments are presented in the first-stage regression of the IV-2SLS and in the probit of the IV-W.
Results of the probit regression in the IV-W-H estimation are same as in the IV-W case.
Wooldridge's (1995) robust score test for overidentifying restrictions. Endogeneity test according to Wooldridge's (1995) score test.
Robust standard errors, Huber/White/sandwich estimator
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, robust standard errors in parentheses

Regarding the IV-W and IV-W-H estimations, the results presented in Table 6 are similar, in both values and significance levels, to the respective results in Table 2. The robustness of the results is clearly shown in Table 7, where we present the averages of the estimated parameters that are obtained by the estimations that use the 13 sets mentioned above. The standard deviations of the parameters are also presented in parentheses. The results show that the estimated coefficients are stable across the different IV sets, for all estimation approaches.

In summary, the above sensitivity analysis shows that the results on the estimated ATE are very stable across the different estimation methods: IV-2SLS, IV-W and IV-W-H. Moreover, the results are very robust across different sets of **Z** and, further, strongly support the results presented in the main text.

Table 7 - Cobb Douglas estimation averages

	IV-2SLS	*IV-W*	*IV-W-H*
lnL	.737	.738	.440
	(.002)	(.002)	(.054)
lnK	.179	.180	.310
	(.002)	(.002)	(.027)
Innovation	.285	.266	.256
	(.056)	(.055)	(.053)
IlnL			.489
			(.090)
IlnK			-.218
			(.048)
intercept	3.176	3.180	3.233
	(.021)	(.020)	(.073)
ATE	.285	.266	.256
Elasticity of labour	.74	.74	
Elasticity of labour ($I = 0$)			.44
Elasticity of labour ($I = 1$)			.93
Elasticity of capital	.18	.18	
Elasticity of capital ($I = 0$)			.31
Elasticity of capital ($I = 1$)			.09
Elasticity of scale	.92	.92	
Elasticity of scale ($I = 0$)			.75
Elasticity of scale ($I = 1$)			1.02
Relative MRTS	-4.12	-4.11	
Relative MRTS ($I = 0$)			-1.44
Relative MRTS ($I = 1$)			-10.85

Average values of the estimated parameters, ATEs, elasticities and rel. MRTS. Standard deviations of the estimated parameters in parentheses.
Use of 13 IV sets of strong and valid instruments, for which the exogeneity test is rejected at 0.05.
Sectors not presented in the table.