PRODUCTION WITH STORABLE AND DURABLE INPUTS: NONPARAMETRIC ANALYSIS OF INTERTEMPORAL EFFICIENCY

LAURENS CHERCHYE, BRAM DE ROCK, AND PIETER JAN KERSTENS

Abstract. We propose a nonparametric methodology for intertemporal production analysis that accounts for durable as well as storable inputs. Durable inputs contribute to the production outputs in multiple consecutive periods. Storable inputs are non-durable and can be stored in inventories for use in future periods. We explicitly model the possibility that firms use several vintages of the durable inputs, i.e. they invest in new durables and scrap older durables over time. Furthermore, we allow for production delays of durable inputs and storage costs for storable inputs. We characterize production behavior that is dynamically cost efficient, which allows us to evaluate the efficiency of observed production decisions. For cost inefficient behavior, we propose a measure to quantify the degree of inefficiency. An attractive feature of this measure is that it can be decomposed in period-specific cost inefficiencies. We demonstrate the usefulness of our methodology through an application to the Italian manufacturing sector of fabricated metal products over the period 1995-2007 drawn from the AMADEUS dataset.

Keywords: data envelopment analysis, cost minimization, storable and durable inputs, production delay and storage costs, dynamic efficiency

JEL Classification: D21, D24, D92

(L. Cherchye) Department of Economics, University of Leuven (KU Leuven). E. Sabbelaan 53, B-8500 Kortrijk, Belgium. E-mail: laurens.cherchye@kuleuven.be. Laurens Cherchye gratefully acknowledges the European Research Council (ERC) for his Consolidator Grant 614221 and the Research Fund KU Leuven for the grant STRT1/08/004.
(B. De Rock) ECARES, Université Libre de Bruxelles and Department of Economics, University of Leuven (KU Leuven). Avenue F. D. Roosevelt 50, CP 114, B-1050 Brussels, Belgium. E-mail: bderock@ulb.ac.be. Bram De Rock gratefully acknowledges FWO and FNRS for their support.
(P. J. Kerstens) Department of Food and Resource Economics, University of Copenhagen. Rolighedsvej 25, DK-1958 Frederiksberg C, Denmark. E-mail: pjk@ifro.ku.dk.
Date: February 13, 2018.
The authors gratefully acknowledge three anonymous referees for their valuable comments and suggestions. The authors also thank Jaap Bos, Knox Lovell and Tom Van Puyenbroeck for helpful comments and suggestions, and Marijn Verschelde for generously providing the data used in our empirical application. Any remaining errors are the responsibility of the authors.

1. Introduction

Many production decisions have long term consequences for production and are capital-intensive: should a firm merely buy new machines to replace older machines that have reached (physical) end of life? Or should the firm invest in new machines to expand its production capacity? These capital-intensive investments are "durable" by nature, because they have a long term impact on production. Furthermore, firms often buy far larger quantities of inputs than they currently need. This can be economically rational for a number of reasons: there are discounts on bulk purchases of inputs or firms expect input prices to rise in the near future. These "storable" inputs can be stored in inventories and are used over several time periods. These durable and storable inputs used in production limit the flexibility of a firm in adjusting its input mix.

In this paper, we introduce a novel methodology for economic (cost) efficiency analysis that explicitly takes these intertemporal aspects of firms' production behavior into account. This obtains a more realistic modeling of intertemporal relations in production situations where storability and durability of inputs are relevant, which is often the case in real-life settings.

The existing literature has devoted much attention to the analysis of dynamically efficient production behavior from a technical perspective (see Fallah-Fini et al. (2014) for a recent review). Such technical efficiency analysis then focuses, for example, on the modeling of production delays, inventories, capital (quasi-fixed factors in general), adjustment costs and learning. By contrast, far less work has tackled the issue from an economic perspective.[1] Importantly, however, the distinction between economic and technical efficiency analysis becomes particularly relevant in dynamic decision settings.

For durable inputs, it has long been known that firms do not immediately scrap old (durable) capital equipment when new equipment becomes available. The process of replacing capital equipment is rather gradual. Firms deciding on new capital equipment face different substitution possibilities between inputs before (ex ante) and after (ex post) the purchase: once capital equipment is installed, it remains in use until the end of its predetermined lifetime (Forsund and Hjalmarsson, 1974; Johansen, 1959). Thus, while firms might seem inefficient from a technical perspective, they may actually be efficient from an economic perspective.[2]

Similarly, when deciding upon storable inputs, firms typically plan their production in advance for a certain time horizon. They form expectations on prices and demand and then decide on the amount of necessary inputs to acquire. Clearly, if prices of storable

[1]Notable exceptions include Nemoto and Goto (1999, 2003); Ouellette and Yan (2008); Silva and Stefanou (2003). We discuss the relation between our framework and this existing work in Section 2.
[2]Wibe (2008) coined the term "rational inefficiency" to mark this difference.

inputs vary over time, this can again generate significant discrepancies between technical and economic efficiency analysis.

In this paper, we present a unifying framework to analyze intertemporal cost minimizing behavior with both durable and storable inputs. For durable inputs, our framework explicitly models the possibility that firms use several vintages: they invest in new durables and scrap older durables over time. Furthermore, we allow for production delays of durable inputs and storage costs for storable inputs. We also show how our framework can incorporate alternative hypotheses such as degressive write-off of durables over time. Thus, our framework incorporates 3 out of 5 factors identified by Fallah-Fini et al. (2014) to which intertemporal dependence of production decisions can be attributed: (i) production delays, (ii) inventories and (iii) capital or generally quasi-fixed inputs, but not (iv) adjustment costs and (v) incremental improvement and learning models (i.e., "disembodied" technical change).

A main distinguishing feature of our methodology is that it is intrinsically nonparametric (in the spirit of Afriat (1972), Varian (1984) and Banker and Maindiratta (1988)): it can analyze production behavior without imposing any (usually non-verifiable) functional structure on the production technology. We characterize production behavior that is intertemporal cost efficient, which allows us to evaluate the efficiency of observed production decisions. For cost inefficient behavior, we propose a measure that quantifies the degree of inefficiency. This intertemporal inefficiency measure has the attractive property that it can be decomposed in period-specific cost inefficiencies.

The remainder of this paper unfolds as follows. In Section 2, we discuss the connection between our work and the closely related literature on both intertemporal production models and efficiency analysis. Sections 3 to 5 formally introduce our methodology. After introducing our general set-up in Section 3, we first consider the case where one has full information on allocations of storables and write-offs of durables in Section 4, to subsequently present the case where limited or no such information is known in Section 5. Section 6 presents some extensions to the basic framework. Section 7 contains the empirical application of our methodology to the Italian manufacturing sector of fabricated metal products over the period 1995-2007 drawn from the AMADEUS dataset. Specifically, this analysis will demonstrate the relevance of accounting for the intertemporal nature of both investment and material costs. Finally, Section 8 concludes and points out a number of interesting extensions.

2. Related literature

Our framework for intertemporal production analysis bears close connections with a number of existing studies on the analysis of efficient production behavior. Most of this earlier work appeared under the label Data Envelopment Analysis (DEA), which is often used to refer to the nonparametric analysis of production efficiency. In what follows, we discuss the relation with earlier literature on network DEA, efficiency analysis with quasi-fixed inputs, and DEA with lagged input effects. In turn, this will allow us to articulate the specificities of our own contribution.

First, our work is closely related to the literature on network DEA (Färe and Grosskopf, 2000) and dynamic DEA (Färe and Grosskopf, 1996). In an early contribution to this literature, Färe (1986) showed how to measure output efficiency by allowing for inputs that are allocatable over time, which are similar in nature to what we call storable inputs. He makes a distinction between inputs for which the allocation over time is known and inputs for which (only) the total amount is known but not how this amount is allocated over time. Importantly, however, he does not consider the intermediate case with new inputs in every period that are to be allocated over multiple time periods. In a similar fashion, Färe et al. (1997) model fixed but allocatable inputs over outputs and develop an output efficiency measure that locates potential efficiency gains due to the reallocation of inputs over the outputs. Färe et al. (2010) consider the problem of resource allocation over time by distinguishing between the decision of when to start the allocation and over how many periods to allocate the resources. These authors also consider this problem under specific returns-to-scale assumptions, capacity constraints and technical change. Again, they do not consider the problem of allocating new inputs after the start period. In the current setting we consider the time frame fixed, but allow for new inputs in every time period that need to be allocated in subsequent time periods. Finally, inventories are also explicitly modeled in Hackman and Leachman (1989)'s general framework of production.

Similarly to our use of durable inputs, Färe et al. (2007) construct a network DEA model with durable and instantaneous inputs to model technology adoption, where one of the technologies is vintage. Durable inputs are vintage-specific, and the adoption of a new technology is accomplished by diverting instantaneous inputs away from the vintage technology to the new technology. Kao (2013) models a dynamic DEA model where the intertemporal dependence among production processes is modeled by quasi-fixed inputs or intermediate products. The overall system efficiency measure can be decomposed as a weighted sum of per-period efficiency measures. These per-period efficiency measures are not necessarily unique and hence not comparable among different DMUs. In a two-stage DEA model Kao and Hwang (2008) maximize the efficiency of stage 1 while maintaining the overall system efficiency. In this way stage 1 efficiency is maximal for all DMUs and can be compared across DMUs.

All these network DEA models have in common that they measure technical efficiency (without price information) and not economic efficiency (with price information). Such technical efficiency analysis requires specific assumptions regarding the nature of the production technology.[3,4]

Furthermore, our concept of durable inputs is also related to the notion of quasi-fixed inputs. Nemoto and Goto (1999, 2003) model adjustment costs due to quasi-fixed inputs and develop an efficiency measure. They treat quasi-fixed inputs as intermediate outputs which are used as inputs in subsequent periods. Their model was extended by Ouellette and Yan (2008) by weakening the restrictions on capital investment. Similarly, Silva and Stefanou (2003) develop nonparametric tests for investment in quasi-fixed inputs with internal adjustment costs in the spirit of Varian (1984).

Next, Chen and van Dalen (2010) incorporate lagged effects of inputs on outputs in DEA efficiency measurement. The relation between output and delayed inputs is fixed parametrically. Thus, they assume that these productive effects are known a priori and estimate these by a fixed effect panel vector autoregressive model in their empirical application. This makes their efficiency measure highly dependent on their parametric specification of the productive effects.

Basically, our contribution is that we present a unifying framework to nonparametrically analyze economic (cost) efficiency in intertemporal production with both storable and durable inputs. We explicitly model the fact that these two types of inputs are used over several time periods: storable inputs are allocated over multiple periods, and durable "vintage" inputs are not immediately replaced by newer durable inputs (thus following Johansen (1959) and Forsund and Hjalmarsson (1974)). In addition, we also allow for production delays of durable inputs over time and storage costs for storable inputs. Next, we propose a cost inefficiency measure that can be decomposed in per-period inefficiencies.[5] Finally, as compared to the literature on quasi-fixed inputs, we do not focus on the issue of adjustment costs, but rather consider the replacement of vintages of durables over time from a cost perspective (see also the introduction of Section 3).

[3] In our concluding Section 8 we will indicate the possibility to conduct a technical efficiency analysis in the intertemporal framework (for economic efficiency analysis) that we develop in the following sections. These technical efficiency formulations could subsequently establish a formal link with the existing network DEA models.

[4] See the early contribution of Shephard and Färe (1980) for a formal treatment in terms of set representation and distance functions.

[5] We note that Kao (2013) proposed a similar decomposition of efficiency scores in per-period efficiencies for a DEA model with quasi-fixed inputs.

3. Set-up

We assume an unbalanced panel setting with K firms in which each firm k is observed over T_k periods. For each firm k and time period t, we observe the S-dimensional output $\mathbf{y}_{k,t} \in \mathbb{R}^S_+$, the N-dimensional storable input $\mathbf{q}_{k,t} \in \mathbb{R}^N_+$, the M-dimensional durable input $\mathbf{Q}_{k,t} \in \mathbb{R}^M_+$ and the corresponding discounted input prices $\mathbf{p}_{k,t} \in \mathbb{R}^N_{++}$ and $\mathbf{P}_{k,t} \in \mathbb{R}^M_{++}$, respectively.[6] For every $k = 1, \ldots, K$, this defines the dataset

$$\mathcal{S}_k = \left\{ (\mathbf{p}_{k,t}, \mathbf{q}_{k,t}, \mathbf{P}_{k,t}, \mathbf{Q}_{k,t}, \mathbf{y}_{k,t}) | t = 1, \ldots, T_k \right\}.$$

To keep our exposition simple, we assume that firms have perfect foresight, i.e., they exactly anticipate the future prices. In fact, it is fairly easy to extend our method to account for predicted prices that deviate from the prices that are realized ex post.[7] But this would only complicate our reasoning without really adding new insights. Also, the fact that we evaluate a firm's cost efficiency in terms of realized prices makes that we may interpret measured inefficiencies as (ex post) prediction errors.

Storable inputs are divisible and we assume they are used over J periods: a fraction is used in each period, while the remaining part is stored for the next periods. We assume the firm incurs no cost for storing storable inputs to keep the exposition simple. While this might be an unrealistic assumption, we show later in Section 6 how our methodology can account for storage costs. Storable inputs can only be used once and are nondurable. Durable inputs are indivisible and usable in multiple periods before reaching end of life status. This is where they differ from storable inputs. Durable inputs are related to quasi-fixed inputs in that they have an effect over multiple periods, but differ from quasi-fixed inputs because they may also be adjusted instantaneously (e.g., one can stop using a laptop or company car immediately). In that sense, we can see quasi-fixed inputs as a subset of durable inputs. In general, durable inputs are seen as investments: a firm intends to use the durable input for a number of periods and writes off the cost of investment over these periods. Examples of durable inputs include machines, equipment, company cars, etc. To keep the exposition simple, we also assume they are used over J periods. We show in Section 6 how this assumption can be relaxed.

---

[6]Our theoretical set-up accounts for the possible use of discounted prices. In such a case, the undiscounted prices $\tilde{\mathbf{p}}_{k,t} \in \mathbb{R}^N_{++}$ and $\tilde{\mathbf{P}}_{k,t} \in \mathbb{R}^M_{++}$ are turned into discounted prices as $\mathbf{p}_{k,t} = \tilde{\mathbf{p}}_{k,t}/(1+\rho)^t$ and $\mathbf{P}_{k,t} = \tilde{\mathbf{P}}_{k,t}/(1+\rho)^t$, where $\rho \geq 0$ is the common discount factor at time t. We remark that the specification of the discount rule does not interfere with the applicability of our methodology. In our empirical application in Section 7, our main goal will be to illustrate discrepancies between a static ($J = 1$) model and a dynamic ($J = 2$) model stemming from intertemporal dependencies of inputs. Therefore, we will not discount these input prices (i.e. we set $\rho = 0$), because this could further enlarge these discrepancies.

[7]For example, a simple solution consists of verifying the cost efficiency conditions that we define below for alternative specifications of (anticipated) prices, as a robustness analysis.

Our behavioral hypothesis is that firms are intertemporally cost minimizing. To formalize this assumption, we represent firm technologies in terms of input requirement sets $\mathcal{I}_t(\mathbf{y}_{k,t})$ for the output of firm k produced at time period t. These sets are defined in the usual way, i.e.,

$$\mathcal{I}_t(\mathbf{y}_{k,t}) = \left\{(\mathbf{q}, \mathbf{Q}) \in \mathbb{R}_+^{N+M} | (\mathbf{q}, \mathbf{Q}) \text{ can produce } \mathbf{y}_{k,t}\right\}.$$

Next, we make use of quantity allocations $(\mathfrak{q}_t^1, \ldots, \mathfrak{q}_t^J)_{t=1}^{T_k}$ of storable inputs and price write-offs $(\mathfrak{P}_t^1, \ldots, \mathfrak{P}_t^J)_{t=1}^{T_k}$ of durable inputs. These allocations and write-offs will be used to distribute firm k's input costs over the J relevant time periods, and are subject to the adding-up restrictions $\mathbf{q}_{k,t} = \sum_{j=1}^J \mathfrak{q}_{k,t}^j$ and $\mathbf{P}_{k,t} = \sum_{j=1}^J \mathfrak{P}_{k,t}^j$.[8] Then, we say that firm k minimizes its total production costs over the time horizon $[J, \ldots, T_k]$ if it chooses the allocation $(\mathfrak{q}_t^1, \ldots, \mathfrak{q}_t^J)_{t=1}^{T_k}$ and write-off $(\mathfrak{P}_t^1, \ldots, \mathfrak{P}_t^J)_{t=1}^{T_k}$ that solves

$$\min_{\substack{(\mathfrak{q}_t^1, \ldots, \mathfrak{q}_t^J)_{t=1}^{T_k} \\ (\mathfrak{P}_t^1, \ldots, \mathfrak{P}_t^J)_{t=1}^{T_k}}} \sum_{t=J}^{T_k} \sum_{j=t-J+1}^{t} \left(\mathbf{p}_{k,j} \mathfrak{q}_j^{t-j+1} + \mathfrak{P}_j^{t-j+1} \mathbf{Q}_{k,j}\right) \tag{1a}$$

$$\text{s.t.} \quad \left(\sum_{j=1}^J \mathfrak{q}_{t-j+1}^j, \sum_{j=1}^J \mathbf{Q}_{k,t-j+1}\right) \in \mathcal{I}_t(\mathbf{y}_{k,t}) \qquad \forall t = J, \ldots, T_k, \tag{1b}$$

where the last feasibility constraint states that the allocation of storables and durables effectively admits the production of the output $\mathbf{y}_{k,t}$ for the given technology. The fact that the storable and durable input quantities are summed over J present and past periods reveals the intertemporal dependency of firm k's production decisions.

Table 1 sharpens the intuition of the above concepts through a simple example that shows a firm's observed costs and production costs over time for $J = 2$. The table illustrates two crucial points. First, observed costs and production costs generally differ. Thus, any efficiency comparison using observed costs instead of production costs is potentially overly pessimistic. Second, the lack of information on allocations and write-offs beyond the observed time frame limits any test of (1) to the time period $[2, \ldots, T_k]$ when $J = 2$ and $[J, \ldots, T_k]$ in general. This explains why we only consider the period $[J, \ldots, T_k]$ in (1) instead of $[1, \ldots, T_k]$ in our minimization program. Note that in practice we often encounter situations where the lifetime J exceeds the length of the panel data (i.e., $J > T_k$). In these situations our methodology, as any other method dealing with intertemporal dependencies, cannot be applied.

[8] In Appendix A, we explain the economic intuition of the write-offs $(\mathfrak{P}_t^1, \ldots, \mathfrak{P}_t^J)_{t=1}^{T_k}$ as representing (in monetary terms) marginal productivities of the durable inputs.

t	observed cost	production cost	
		storable inputs	durable inputs
1	$\mathbf{p}_1\mathfrak{q}_1$	$\mathbf{p}_1\mathfrak{q}_1^1 + ?$	$\mathfrak{P}_1^1\mathbf{Q}_1 + ?$
2	$\mathbf{p}_2\mathfrak{q}_2$	$\mathbf{p}_2\mathfrak{q}_2^1 + \mathbf{p}_1\mathfrak{q}_1^2$	$\mathfrak{P}_2^1\mathbf{Q}_2 + \mathfrak{P}_1^2\mathbf{Q}_1$
3	$\mathbf{p}_3\mathfrak{q}_3$	$\mathbf{p}_3\mathfrak{q}_3^1 + \mathbf{p}_2\mathfrak{q}_2^2$	$\mathfrak{P}_3^1\mathbf{Q}_3 + \mathfrak{P}_2^2\mathbf{Q}_2$
4	$\mathbf{p}_4\mathfrak{q}_4$	$\mathbf{p}_4\mathfrak{q}_4^1 + \mathbf{p}_3\mathfrak{q}_3^2$	$\mathfrak{P}_4^1\mathbf{Q}_4 + \mathfrak{P}_3^2\mathbf{Q}_3$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
t	$\mathbf{p}_t\mathfrak{q}_t$	$\mathbf{p}_t\mathfrak{q}_t^1 + \mathbf{p}_{t-1}\mathfrak{q}_{t-1}^2$	$\mathfrak{P}_t^1\mathbf{Q}_t + \mathfrak{P}_{t-1}^2\mathbf{Q}_{t-1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$T_k - 1$	$\mathbf{p}_{T_k-1}\mathfrak{q}_{T_k-1}$	$\mathbf{p}_{T_k-1}\mathfrak{q}_{T_k-1}^1 + \mathbf{p}_{T_k-2}\mathfrak{q}_{T_k-2}^2$	$\mathfrak{P}_{T_k-1}^1\mathbf{Q}_{T_k-1} + \mathfrak{P}_{T_k-2}^2\mathbf{Q}_{T_k-2}$
T_k	$\mathbf{p}_{T_k}\mathfrak{q}_{T_k}$	$\mathbf{p}_{T_k}\mathfrak{q}_{T_k}^1 + \mathbf{p}_{T_k-1}\mathfrak{q}_{T_k-1}^2$	$\mathfrak{P}_{T_k}^1\mathbf{Q}_{T_k} + \mathfrak{P}_{T_k-1}^2\mathbf{Q}_{T_k-1}$
$T_k + 1$	$?$	$? + \mathbf{p}_{T_k}\mathfrak{q}_{T_k}^2$	$? + \mathfrak{P}_{T_k}^2\mathbf{Q}_{T_k}$

Table 1. Overview of production costs with storable and durable inputs for $J = 2$.

4. Complete information

We next turn to deriving operational conditions for cost minimizing behavior as defined in (1). As indicated in the Introduction, we derive nonparametric conditions in the spirit of Afriat (1972); Varian (1984) and Banker and Maindiratta (1988), which make minimal assumptions regarding the production technology. To set the stage, we first consider the limiting case that is characterized by full information on the quantity allocations of the storable inputs and the price write-offs of the durable inputs: i.e., for each firm k the empirical analyst observes the allocations $(\mathfrak{q}_{k,t}^1, \mathfrak{q}_{k,t}^2, \ldots, \mathfrak{q}_{k,t}^J)$ and the write-offs $(\mathfrak{P}_{k,t}^1, \mathfrak{P}_{k,t}^2, \ldots, \mathfrak{P}_{k,t}^J)$ at every time period $t = 1, ..., T_k$.

Such complete information greatly simplifies matters. From (1), it is easy to verify that, for a given specification of storable allocations and durable write-offs, the production costs for any time period t are defined independently of the production costs for other time periods. As an implication, firm k behaves consistently with (1) if and only if it solves, for every $t = J, ..., T_k$,

$$\min_{\substack{(\mathfrak{q}_j^1,\ldots,\mathfrak{q}_j^J)_{j=t-J+1}^t \\ (\mathfrak{P}_j^1,\ldots,\mathfrak{P}_j^J)_{j=t-J+1}^t}} \sum_{j=t-J+1}^{t} \left(\mathbf{p}_{k,j}\mathfrak{q}_j^{t-j+1} + \mathfrak{P}_j^{t-j+1}\mathbf{Q}_{k,j}\right) \tag{2a}$$

$$\text{s.t.} \quad \left(\sum_{j=1}^{J} \mathfrak{q}_{t-j+1}^j, \sum_{j=1}^{J} \mathbf{Q}_{k,t-j+1}\right) \in \mathcal{I}_t(\mathbf{y}_{k,t}) \tag{2b}$$

Putting it differently: dynamically cost minimizing behavior under complete information can be represented as statically cost minimizing behavior for every period t. Varian (1984) developed the nonparametric characterization of such static cost minimization.[9] Thus, we can obtain our empirical condition for dynamic cost efficiency by translating Varian's reasoning to our particular setting.

Throughout, we will adopt the next two axioms regarding the production technology (given by $\mathcal{I}_t(\mathbf{y}_{k,t})$):

Axiom 1 (observability means feasibility)**.** *For all* $k = 1, \ldots, K$ *and* $t = 1, \ldots, T_k$: $(\mathbf{p}_{k,t}, \mathfrak{q}^1_{k,t}, \ldots, \mathfrak{q}^J_{k,t}, \mathfrak{P}^1_{k,t}, \ldots, \mathfrak{P}^J_{k,t}, \mathbf{Q}_{k,t}, \mathbf{y}_{k,t}) \in \mathcal{S}_k \Rightarrow \left(\sum_{j=1}^J \mathfrak{q}^j_{k,t-j+1}, \sum_{j=1}^J \mathbf{Q}_{k,t-j+1}\right) \in \mathcal{I}_t(\mathbf{y}_{k,t})$.

Axiom 2 (nested input sets)**.** *For all* $k, s = 1, \ldots, K$ *and* $t = 1, \ldots, T_k : \mathbf{y}_{s,t} \geq \mathbf{y}_{k,t} \Rightarrow \mathcal{I}_t(\mathbf{y}_{s,t}) \subseteq \mathcal{I}_t(\mathbf{y}_{k,t})$.[10]

In words, Axiom 1 says that there are no significant measurement errors in the data.[11] Axiom 2 says that, for a given time period t, input requirement sets are nested: if firm s produces at least the same output as firm k (i.e., $\mathbf{y}_{s,t} \geq \mathbf{y}_{k,t}$), then the input set for s must be contained in the set for k (i.e., $\mathcal{I}_t(\mathbf{y}_{s,t}) \subseteq \mathcal{I}_t(\mathbf{y}_{k,t})$).[12] Intuitively, this means that outputs are freely disposable. These are the only two production axioms that we will assume in the sequel of this paper.

Then, we define

$$c_{k,t} = \min_{s \in D^t_k} \left\{ \sum_{j=t-J+1}^{t} \left(\mathbf{p}_{k,j} \mathfrak{q}^{t-j+1}_{s,j} + \mathfrak{P}^{t-j+1}_{k,j} \mathbf{Q}_{s,j} \right) \right\}. \tag{3}$$

for

$$D^t_k = \left\{ s | \mathbf{y}_{s,t} \geq \mathbf{y}_{k,t} \right\}, \tag{4}$$

[9] Varian (1984) characterized cost minimizing production behavior in terms of the so-called Weak Axiom of Cost Minimization (WACM). Basically, Proposition 1 will state this WACM criterion for our intertemporal setting.

[10] Throughout $\mathbf{y}_{s,t} \geq \mathbf{y}_{k,t}$ should be interpreted as vector inequalities, implying that the inequality needs to hold for all components.

[11] Clearly, this axiom may often be problematic in practical situations. In such instances, we can use alternative techniques to explicitly account for errors. For example, one may adjust our methodology by integrating it with the probabilistic method which Cazals, Florens, and Simar (2002) and Daraio and Simar (2005, 2007) originally proposed in a DEA context. To focus our discussion, we do not consider this extension here.

[12] We remark that this assumes that different firms s and k face the same technology in period t. Obviously, we can also use other hypotheses regarding technological homogeneity/heterogeneity across firms and time periods. For example, we may assume homogeneous technologies (only) for subsets of firms (e.g., defined on the basis of observable firm characteristics), or firm-specific technologies that are constant over time. For compactness, we will again not explicitly implement this.

i.e., the set of observed firms s that produce at least the same output as firm k in period t (i.e., $\mathbf{y}_{s,t} \geq \mathbf{y}_{k,t}$). By construction, we have $k \in D_k^t$, so that $D_k^t \neq \emptyset$. In words, $c_{k,t}$ represents the minimal cost over this set D_k^t. Obviously, we can compute $c_{k,t}$ by simply enumerating over all $s \in D_k^t$ if the allocations $(\mathfrak{q}_{k,t}^1, \mathfrak{q}_{k,t}^2, \dots, \mathfrak{q}_{k,t}^J)$ and write-offs $(\mathfrak{P}_{k,t}^1, \mathfrak{P}_{k,t}^2, \dots, \mathfrak{P}_{k,t}^J)$ are given.

We can now state the following result.

Proposition 1. *Firm k solves (1) for a production technology that satisfies Axioms (1) and (2) if and only if, for all $t = J, ..., T_k$:*

$$\sum_{j=t-J+1}^{t} \left(\mathbf{p}_{k,j} \mathfrak{q}_{k,j}^{t-j+1} + \mathfrak{P}_{k,j}^{t-j+1} \mathbf{Q}_{k,j} \right) = c_{k,t}. \tag{5}$$

Proof. We use the equivalence between (1) and (2). Then, the result follows from Theorem 1 (statements (1) and (2)) of Varian (1984). □

This results directly suggests the next measure of cost inefficiency for every period t:

$$CE_k^t \equiv \sum_{j=t-J+1}^{t} \left(\mathbf{p}_{k,j} \mathfrak{q}_{k,j}^{t-j+1} + \mathfrak{P}_{k,j}^{t-j+1} \mathbf{Q}_{k,j} \right) - c_{k,t}, \tag{6}$$

Obviously, firm k meets the empirical cost minimization criterion (5) in Proposition 1 if and only if $CE_k^t = 0$. More generally, we have $CE_k^t \geq 0$, and the value of CE_k^t indicates how much firm k deviates from cost minimizing behavior at time t.

When aggregating over all $t = J, \dots, T_k$, we can similarly define an overall cost inefficiency measure as

$$CE_k \equiv \sum_{t=J}^{T_k} \sum_{j=t-J+1}^{t} \left(\mathbf{p}_{k,j} \mathfrak{q}_{k,j}^{t-j+1} + \mathfrak{P}_{k,j}^{t-j+1} \mathbf{Q}_{k,j} \right) - \sum_{t=J}^{T_k} c_{k,t}. \tag{7}$$

By construction, we have

$$CE_k = \sum_{t=J}^{T_k} \left(\sum_{j=t-J+1}^{t} \left(\mathbf{p}_{k,j} \mathfrak{q}_{k,j}^{t-j+1} + \mathfrak{P}_{k,j}^{t-j+1} \mathbf{Q}_{k,j} \right) - c_{k,t} \right) = \sum_{t=J}^{T_k} CE_k^t, \tag{8}$$

which yields the next result.

Proposition 2. $CE_k = 0 \Leftrightarrow CE_k^t = 0 \; \forall t = J, \dots, T_k$.

Proof. The result follows from (7) and the definitional fact that $CE_k^t \geq 0$. □

In words, firm k minimizes its total production costs over the full period $[J, \dots, T_k]$ if and only if its production costs are minimal in every single period t. Essentially, this result shows that our overall cost inefficiency measure CE_k satisfies the aggregate indication axiom of Blackorby and Russell (1999).

5. Incomplete information

The previous section assumed an ideal scenario in which the empirical analyst had full knowledge of the allocations $(\mathfrak{q}^1_{k,t}, \mathfrak{q}^2_{k,t}, \ldots, \mathfrak{q}^J_{k,t})$ and the write-offs $(\mathfrak{P}^1_{k,t}, \mathfrak{P}^2_{k,t}, \ldots, \mathfrak{P}^J_{k,t})$. In practice, however, only very limited information on allocations and write-offs is often available. It may even happen that such information is completely absent. This section shows how to proceed in such (more realistic) instances.

Formally, we will assume that the available information is captured by the polyhedron

$$\Theta(\mathbf{A}, \mathbf{b}) \equiv \left\{ \boldsymbol{\rho} \in \mathbb{R}_+^{T_k J(N+M)} : \mathbf{A}\boldsymbol{\rho} \geq \mathbf{b} \right\}, \tag{9}$$

which represents L restrictions on the allocations of storable inputs and on the write-offs of durable inputs. Specifically, $\mathbf{A}$ is a $L \times T_k J(N+M)$ matrix and $\mathbf{b}$ a $L \times 1$ vector, and $\boldsymbol{\rho}$ represents all vectors that satisfy the constraints imposed by $\mathbf{A}$ and $\mathbf{b}$.

To structure our discussion, we will first consider the limiting case in which we cannot use any information on firms' allocations and write-offs, which corresponds to $\Theta = \mathbb{R}_+^{T_k J(N+M)}$. Subsequently, we will discuss the intermediate scenario where some information is available: i.e., $\Theta \subset \mathbb{R}_+^{T_k J(N+M)}$.

5.1. No information on allocations and write-offs.

5.1. No information on allocations and write-offs. In the absence of full information on storable allocations and durable write-offs, we can no longer check the condition (2) independently for every single time period t. In this case, we verify if there exists at least one possible specification of $(\mathfrak{q}^1_{k,t}, \mathfrak{q}^2_{k,t}, \ldots, \mathfrak{q}^J_{k,t})$ and $(\mathfrak{P}^1_{k,t}, \mathfrak{P}^2_{k,t}, \ldots, \mathfrak{P}^J_{k,t})$ that makes firm k's behavior consistent with the overall cost minimization condition (1). More specifically, we define feasible allocations and write-offs that present firm k as efficient as possible. This evaluates firm k in the most favorable light and, thus, gives this firm the benefit-of-the-doubt in the absence of full information.[13]

[13] This benefit-of-the-doubt idea is intrinsic to DEA efficiency evaluations. See, for example, Cherchye et al. (2007) for a detailed discussion of the benefit-of-the-doubt interpretation of DEA models in the specific context of composite indicator construction.

The following linear program operationalizes this idea:

(10a)
$$\min_{\substack{c_{k,t}\geq 0,\\ (\mathfrak{q}^1_{s,t},\ldots,\mathfrak{q}^J_{s,t})_{t=1}^{T_k}\geq 0,\\ (\mathfrak{P}^1_{k,t},\ldots,\mathfrak{P}^J_{k,t})_{t=1}^{T_k}\geq 0}} \sum_{t=J}^{T_k}\left(\sum_{j=t-J+1}^{t}\left(\mathbf{p}_{k,j}\mathfrak{q}^{t-j+1}_{k,j}+\mathfrak{P}^{t-j+1}_{k,j}\mathbf{Q}_{k,j}\right)-c_{k,t}\right)$$

(10b)
$$\text{s.t. } c_{k,t}\leq \sum_{j=t-J+1}^{t}\mathbf{p}_{k,j}\mathfrak{q}^{t-j+1}_{s,j}+\mathfrak{P}^{t-j+1}_{k,j}\mathbf{Q}_{s,j} \qquad \forall s\in D^t_k,\ \forall t=J,\ldots,T_k,$$

(10c)
$$\sum_{j=1}^{J}\mathfrak{q}^j_{s,t}=\mathbf{q}_{s,t} \qquad \forall s\in D^t_k,\ \forall t=1,\ldots,T_k,$$

(10d)
$$\sum_{j=1}^{J}\mathfrak{P}^j_{k,t}=\mathbf{P}_{k,t} \qquad \forall t=1,\ldots,T_k,$$

In this program, the objective minimizes firm k's cost inefficiency (as defined in (7)) in terms of the chosen allocation and write-off schemes. The first constraint imposes that $c_{k,t}$ effectively represents the minimal cost to produce the output $\mathbf{y}_{k,t}$ (over the set D^t_k). The second and third constraints impose the adding-up restrictions that apply to feasible specifications of $(\mathfrak{q}^1_{k,t},\mathfrak{q}^2_{k,t},\ldots,\mathfrak{q}^J_{k,t})$ and $(\mathfrak{P}^1_{k,t},\mathfrak{P}^2_{k,t},\ldots,\mathfrak{P}^J_{k,t})$. Intuitively, cost inefficiency occurs as soon as some other firm s is characterized by a lower production cost than firm k no matter what allocations and write-offs are used.[14]

The allocations and write-off schemes of inputs acquired in periods $[T_k-J+2,\ldots,T_k]$ deserve some discussion at this point. These inputs are used beyond the time horizon T_k (until T_k+J-1 for inputs acquired in period T_k) and for them we can predict the optimal allocation of the LP solver. Since shifted cost allocations to periods $[T_k+1,\ldots,T_k+J-1]$ do not enter the objective of (10), optimal choices for firm k's allocations and write-offs of $[T_k-J+2,\ldots,T_k]$ consist of distributing these costs entirely over the periods beyond T_k. Thus, firm k is efficient by default for the periods $[T_k-J+2,\ldots,T_k]$. This is an inherent feature of the benefit-of-the-doubt idea. Furthermore, this prediction becomes invalid when partial information on allocations and write-offs is added.

[14]Similarly as in conventional multiplier formulations of DEA, the allocations and write-offs that solve (10) are not necessarily unique. As a result, the efficiency decomposition in (8) is also not necessarily unique. However, one can use a similar procedure as in Kao and Hwang (2008) to find a unique decomposition by maximizing per-period efficiency while maintaining overall efficiency.

Recall from Section 3 that we cannot evaluate dynamic cost efficiency for periods $[1, \ldots, J-1]$ due to data limitations. Together with our discussion in the previous paragraph, this implies that we can only effectively discriminate among firms within the time window $[J, \ldots, T_k - J + 1]$. Note the impact of both J and T_k on the size of this time window: the larger J (T_k), the smaller (larger) this time window and the more (fewer) period observations are efficient by default. This reveals a clear trade-off between J and T_k. While T_k is often determined by the data at hand, one can vary J to check sensitivity of the results.

5.2. **Partial information on allocations and write-offs.** In many practical situations, it is possible to put some additional restrictions on the feasible allocation and write-off schemes. Such partial information can be incorporated by suitably specifying $\Theta(\mathbf{A}, \mathbf{b})$. Correspondingly, we can append the restriction

$$(\mathfrak{q}^1_{s,t}, \ldots, \mathfrak{q}^J_{s,t}, \mathfrak{P}^1_{k,t}, \ldots, \mathfrak{P}^J_{k,t})^{T_k}_{t=1} \in \Theta(\mathbf{A}, \mathbf{b}) \tag{11}$$

to program (10), and solve the resulting (linear) problem. Clearly, by using restriction (11) we constrain the solution space, which will generally result in higher values of the computed cost inefficiencies.

To take a specific instance, let $(\mathfrak{q}^{u,A}_{k,v})_{u \in U \subseteq [1,\ldots,J]}$ represent lower bounds on the quantity allocations of the storable inputs for firm k and time period(s) $v \in V \subseteq [J, \ldots, T_k]$. Similarly, let $(\mathfrak{P}^{w,A}_{k,z})_{w \in W \subseteq [1,\ldots,J]}$ be known lower bounds on the price write-offs of the durable inputs for time period(s) $z \in Z \subseteq [J, \ldots, T_k]$. We then define

$$\Theta = \Big\{ \mathfrak{q}^u_{k,v} \geq \mathfrak{q}^{u,A}_{k,v}, \ \forall u \in U, \ \forall v \in V \\ \mathfrak{P}^w_{k,z} \geq \mathfrak{P}^{w,A}_{k,z}, \ \forall w \in W, \ \forall z \in Z \Big\}.$$

As a limiting case, instantaneous input consumption complies with $\mathfrak{q}^{u,A}_{k,v} = (\mathbf{q}_{k,v}, 0, \ldots, 0)$ or, equivalently, $\mathfrak{P}^{w,A}_{k,z} = (\mathbf{P}_{k,z}, 0, \ldots, 0)$.

5.3. **Write-off hypotheses.** By using this approach, we can actually include (and check) alternative hypotheses regarding the allocation of the durable costs to individual time periods (i.e., specific write-off schemes). On the one hand, write-off schemes are often dictated by standard accounting practices for specific durable inputs, which makes write-off schemes public information. Imposing these write-off schemes in the LP then allows to check whether these write-off schemes are cost efficient for a firm. On the other hand, durable inputs are often an aggregate of multiple durable inputs (e.g., "investment" in our own empirical application), so that it is unclear a priori what the exact write-off scheme is. Imposing alternative write-off hypotheses in the LP can then help to clarify this. For example, it might often be reasonable to assume that the firm's valuation of a durable input diminishes over time. In our framework, this corresponds to

$$\mathfrak{P}_{k,t}^{1} \geq \mathfrak{P}_{k,t}^{2} \geq \ldots \geq \mathfrak{P}_{k,t}^{J}, \tag{12}$$

which complies with a degressive write-off of investment costs. From our above explanation, it follows that this is also consistent with the assumption of technological improvement, where older machines are scrapped and replaced by newer – technologically improved – ones over time.

Alternatively, a linear write-off of investment corresponds to

$$\mathfrak{P}_{k,t}^{1} = \mathfrak{P}_{k,t}^{2} = \ldots = \mathfrak{P}_{k,t}^{J}, \tag{13}$$

implying that $\mathfrak{P}_{k,t}^{j} = \mathbf{P}_{k,t}/J\ \forall j = 1, \ldots, J$. Both hypotheses can be tested by adding (12) or (13) for $t = J, \ldots, T_k$ to $\Theta(\mathbf{A}, \mathbf{b})$.

6. Extensions

We next focus on a number of extensions of our basic framework set out in the previous section. These extensions highlight the versatility of our framework and, of course, are not exhaustive. First, we show how to convert our (difference) cost inefficiency measures (6) and (7) in ratio form. Then, we discuss the extension of our framework to allow for heterogeneous input lifetime, production delays and storage costs. Finally, we indicate how to proceed in the absence of input price information by applying shadow pricing. Here, we will also explain the decomposition of cost inefficiency as defined above in terms of technical and allocative inefficiency. We will illustrate the different extensions in our empirical application in Section 7.

6.1. **Ratio measures of inefficiency.** A downside of our cost inefficiency measure in (7) is that it is not invariant to rescaling of prices and inputs. However, one can turn this difference measure into a ratio measure of cost inefficiency by an appropriate normalization. In principle, a multitude of normalizations are possible. A natural choice is to divide by the actual cost: i.e.,

$$RCE_k \equiv \frac{CE_k}{\sum_{t=J}^{T_k} \sum_{j=t-J+1}^{t} \left(\mathbf{p}_{k,j} \mathfrak{q}_{k,j}^{t-j+1} + \mathfrak{P}_{k,j}^{t-j+1} \mathbf{Q}_{k,j} \right)} \tag{14}$$

This relative measure is situated between 0 and 1 and expresses the proportion of total production costs that can be saved by minimizing total production costs over the periods $[J, \ldots, T_k]$.[15]

[15] This normalization mirrors the one used by Chambers et al. (1998) for profit efficiency.

Analogously to (8), we can decompose this overall ratio measure in terms of per-period measures.[16] In this case, we have that RCE_k equals a weighted sum of per-period cost inefficiencies in ratio form RCE_k^t. Specifically, it uses the period-specific weights

$$w_k^t \equiv \frac{\sum_{j=t-J+1}^{t}\left(\mathbf{p}_{k,j}\mathfrak{q}_{k,j}^{t-j+1}+\mathfrak{P}_{k,j}^{t-j+1}\mathbf{Q}_{k,j}\right)}{\sum_{t=J}^{T_k}\sum_{j=t-J+1}^{t}\left(\mathbf{p}_{k,j}\mathfrak{q}_{k,j}^{t-j+1}+\mathfrak{P}_{k,j}^{t-j+1}\mathbf{Q}_{k,j}\right)}, \tag{15}$$

which represent the proportions of total production costs allocated to every period t. This obtains

$$\begin{aligned}
RCE_k &= \sum_{t=J}^{T_k} w_k^t \frac{CE_k^t}{\sum_{j=t-J+1}^{t}\left(\mathbf{p}_{k,j}\mathfrak{q}_{k,j}^{t-j+1}+\mathfrak{P}_{k,j}^{t-j+1}\mathbf{Q}_{k,j}\right)}\\
&= \sum_{t=J}^{T_k} w_k^t \left(1-\frac{c_{k,t}}{\sum_{j=t-J+1}^{t}\left(\mathbf{p}_{k,j}\mathfrak{q}_{k,j}^{t-j+1}+\mathfrak{P}_{k,j}^{t-j+1}\mathbf{Q}_{k,j}\right)}\right)\\
&= \sum_{t=J}^{T_k} w_k^t RCE_k^t, \qquad (16)
\end{aligned}$$

As a final note, we indicate that $1-RCE_k$ and $1-RCE_k^t$ give the conventional cost efficiency measures, i.e., minimal cost divided by actual cost.

6.2. **Heterogeneous input lifetime, production delays and storage costs.** Until now, we have assumed that all durable inputs have the same lifetime J. Admittedly, this may sometimes be a too strong assumption. In addition, our current specification does not allow for production delays or storage costs. As we show next, we can solve these issues by making use of the concept of delay matrices.

Specifically, let $\mathbf{D}^D = (\mathbf{d}_1^D, \ldots, \mathbf{d}_J^D) \in \{0,1\}^{M\times J}$ denote a binary delay matrix, where each row represents a durable input. For example, for the durable input m we may use one of the following specifications:

- $(1, \underbrace{0, \ldots, 0}_{J-1})$ if input m is an instantaneous input;
- $(1, \ldots, 1)$ if input m is a durable input with lifetime J;
- $(\underbrace{0, \ldots, 0}_{U}, \underbrace{1, \ldots, 1}_{J-U})$ if input m is a durable input usable after a delay of $U < J$ periods with a lifetime of $J-U$ periods;
- $(\underbrace{1, \ldots, 1}_{U}, \underbrace{0, \ldots, 0}_{J-U})$ if input m is a durable input usable over $U < J$ periods.

[16]The following decomposition parallels Färe and Zelenyuk (2003)'s decomposition of industry revenue efficiency as a weighted sum of firms' revenue efficiency.

Similarly, let $\mathbf{D}^S = (\mathbf{d}_1^S, \ldots, \mathbf{d}_J^S) \in \mathbb{R}_+^{N \times J}$ represent a delay matrix for the storable inputs. In addition to allowing for heterogeneous input lifetimes and production delays as before, this matrix can be used, for example, to incorporate storage costs by using $\mathbf{d}_1^S \leq \ldots \leq \mathbf{d}_J^S$ with $\mathbf{d}_j^S \geq 1$ for all j: this captures that storage costs increasingly penalize carry-overs of storable inputs to later periods of their lifetime. Then, we can formulate the next modified optimization problem of firm k:

$$\min_{\substack{(\mathfrak{q}_t^1, \ldots, \mathfrak{q}_t^J)_{t=1}^{T_k} \\ (\mathfrak{P}_t^1, \ldots, \mathfrak{P}_t^J)_{t=1}^{T_k}}} \sum_{t=J}^{T_k} \sum_{j=t-J+1}^{t} \mathbf{p}_{k,j} \mathbf{d}_{t-j+1}^S \mathfrak{q}_j^{t-j+1} + \mathfrak{P}_j^{t-j+1} \mathbf{d}_{t-j+1}^D \mathbf{Q}_{k,j} \tag{17a}$$

$$\text{s.t.} \quad \left(\sum_{j=1}^{J} \mathbf{d}_j^S \mathfrak{q}_{t-j+1}^j, \sum_{j=1}^{J} \mathbf{d}_j^D \mathbf{Q}_{k,t-j+1} \right) \in \mathcal{I}_t(\mathbf{y}_{k,t}) \qquad \forall t = J, \ldots, T_k \tag{17b}$$

Closer inspection reveals that $\mathbf{d}_j^S \mathfrak{q}_{t-j+1}^j$ and $\mathbf{d}_j^D \mathbf{Q}_{k,t-j+1}$ only selects those inputs that are usable in period j. In this case, J stands for the maximum lifetime over all durable and storable inputs. We further note that we can include a resale value for all durable inputs simply by setting J to the maximum lifetime + 1. The final write-off of the durable inputs is then the resale value of the durable inputs, which can be subtracted from the per-period cost to reflect resale of the durable inputs.

The associated linear program is

(18a)
$$\min_{\substack{c_{k,t}\geq 0,\\ (\mathfrak{q}^1_{s,t},\ldots,\mathfrak{q}^J_{s,t})_{t=1}^{T_k}\geq 0,\\ (\mathfrak{P}^1_{k,t},\ldots,\mathfrak{P}^J_{k,t})_{t=1}^{T_k}\geq 0}} \sum_{t=J}^{T_k}\left(\sum_{j=t-J+1}^{t}\left(\mathbf{p}_{k,j}\mathbf{d}^S_{t-j+1}\mathfrak{q}^{t-j+1}_{k,j}+\mathfrak{P}^{t-j+1}_{k,j}\mathbf{d}^D_{t-j+1}\mathbf{Q}_{k,j}\right)-c_{k,t}\right)$$

(18b)
$$\text{s.t. } c_{k,t}\leq \sum_{j=t-J+1}^{t}\mathbf{p}_{k,j}\mathfrak{q}^{t-j+1}_{s,j}\mathbf{d}^S_{t-j+1}+\mathfrak{P}^{t-j+1}_{k,j}\mathbf{d}^D_{t-j+1}\mathbf{Q}_{s,j} \qquad \forall s\in D^t_k,\ \forall t=J,\ldots,T_k,$$

(18c)
$$\sum_{j=1}^{J}\mathfrak{q}^j_{s,t}\mathbf{d}^S_j=\mathfrak{q}_{s,t} \qquad \forall s\in D^t_k,\ \forall t=1,\ldots,T_k,$$

(18d)
$$\sum_{j=1}^{J}\mathfrak{P}^j_{k,t}\mathbf{d}^D_j=\mathbf{P}_{k,t} \qquad \forall t=1,\ldots,T_k,$$

(18e)
$$(\mathfrak{q}^1_{s,t},\ldots,\mathfrak{q}^J_{s,t},\mathfrak{P}^1_{k,t},\ldots,\mathfrak{P}^J_{k,t})_{t=1}^{T_k}\in\Theta(\mathbf{A},\mathbf{b}).$$

It is easy to verify that this program reduces to (10) for $\mathbf{D}^S=\mathbb{1}_{N\times J}$ and $\mathbf{D}^D=\mathbb{1}_{M\times J}$. Furthermore, any zero values in $\mathbf{D}^S$ and $\mathbf{D}^D$ immediately imply zero values for the corresponding allocations and write-offs. In other words, the use of delay matrices allows us to impose a priori restrictions on the storable allocations and durable write-offs. We also remark that, in principle, we can specify firm-specific delay matrices if this seems desirable.

6.3. **Shadow prices and technical inefficiency.** So far, we have focused on economic (cost) efficiency, which requires price information for the relevant inputs. By contrast, technical efficiency analysis does not require such price information and, thus, can be used if limited price information is available.

Generally, technical efficiency criteria/measures can be characterized as economic efficiency criteria/measures evaluated at so-called "shadow prices".[17] Thus, by establishing

[17] In DEA terminology, this shadow price characterization of technical efficiency corresponds to the "multiplier" formulation of DEA models. Practical applications often make use of DEA models in "envelopment" form, which is dual to this multiplier formulation. In our set-up, the envelopment formulation can be obtained as the dual of the linear program that we define below (to compute TE_k). We refer to Färe and Primont (1995) for a more detailed discussion.

the shadow price representation of our dynamic efficiency concepts, we can define technical efficiency notions that explicitly account for the dynamic (storable and durable) nature of the inputs.

It is easy to use shadow pricing if the exact allocation of storable inputs over time periods (i.e., $(\mathfrak{q}_{s,t}^1, \ldots, \mathfrak{q}_{s,t}^J)_{t=1}^{T_k}$) is known to the empirical analyst. In that case, it suffices to solve (18) with the input prices $(\mathbf{P}_{k,t})_{t=1}^{T_k}$ and $(\mathbf{p}_{k,t})_{t=1}^{T_k}$ as additional free variables that are subject to a non-negativity constraint and the normalization

$$\sum_{t=J}^{T_k} \left(\sum_{j=t-J+1}^{t} \mathbf{p}_{k,j} \mathfrak{q}_{k,j}^{t-j+1} \mathbf{d}_{t-j+1}^{S} + \mathfrak{P}_{k,j}^{t-j+1} \mathbf{d}_{t-j+1}^{D} \mathbf{Q}_{k,j} \right) = 1.$$

We remark that this price normalization implies $CE_k = RCE_k$, but their respective per-period inefficiencies CE_k^t and RCE_k^t do not necessarily coincide. The resulting shadow cost inefficiency measure (or technical inefficiency measure TE_k) measures cost inefficiency using shadow prices that give the benefit-of-the-doubt to the firm. Thus, these shadow prices are chosen as to maximize the shadow cost efficiency of the firm.

Let TE_k represent the "technical inefficiency" measure that is obtained as the solution of the resulting linear program. Further, let RCE_k represent the relative cost inefficiency measure found by solving (18) as before and applying (14). By construction, we have $TE_k \leq RCE_k$. The difference between TE_k and RCE_k gives us a measure AE_k of "allocative inefficiency": i.e.,

$$TE_k + AE_K = RCE_k \Leftrightarrow AE_k = RCE_k - TE_k. \tag{19}$$

Allocative inefficiency is the difference between cost inefficiency using market prices and shadow cost inefficiency using optimal shadow prices. This shows that reducing cost inefficiency can be achieved by a combination of two effects: (i) reducing technical inefficiency by minimizing the level of inputs conditional on outputs and (ii) selecting the cost minimizing input mix out of all technically efficient inputs. Allocative inefficiency can occur when the realized a posteriori prices deviate from the expected a priori input prices. In our context of durable and storable inputs, intertemporal dependencies can exacerbate this allocative inefficiency: buying storable and/or durable inputs right before a persistent price drop can cause a persistent per-period inefficiency over subsequent periods, because the cost of these inputs is distributed over J periods. Thus, one slip up can resonate for many subsequent time periods.

Interestingly, using (16) we can also decompose AE_k in relative period-specific allocative inefficiencies as follows:

$$AE_k \equiv \sum_{t=J}^{T_k} \left(w_k^t RCE_k^t - \tilde{w}_k^t TE_k^t \right) = \sum_{t=J}^{T_k} \hat{w}_k^t AE_k^t, \tag{20}$$

where $\tilde{w}_k^t$ ($\hat{w}_k^t$) are the corresponding weights in the per-period decomposition of TE_k (AE_k). Since the per-period weights $\tilde{w}_k^t$ are computed using shadow prices, the weights $\hat{w}_k^t$ of per-period allocative inefficiency AE_k^t represent the shift in the per-period cost allocation resulting from the use of observed versus optimal (shadow) prices for the firm.

Finally, matters are more complicated when the allocation $(\mathbf{q}_{s,t}^1, \ldots, \mathbf{q}_{s,t}^J)_{t=1}^{T_k}$ is unobserved. In that case, the analogue of the programming problem (18) becomes nonlinear in unknown prices and quantities. We can restore linearity by making specific assumptions regarding the storable input allocation. For example, if we are willing to assume that all DMUs allocate their storable inputs in the same way over time, then we can use a similar procedure as outlined in Cook et al. (2000) and Cherchye et al. (2013).

7. Empirical application

We use firm-level data from AMADEUS as provided by Bureau van Dijk. AMADEUS covers firm-level data of 19 million companies in 43 European countries and is based on income statements and balance sheets in a common format as gathered from national providers. We follow the same data cleaning procedure as described in Verschelde et al. (2016) and, in addition, refer to Merlevede et al. (2015) for more specific details on the data.[18]

In order to avoid extreme effects due to outliers and noise in the data, we removed observations with one or more improbable inputs (i.e., employment costs, deflated tangible fixed assets or deflated material costs less than 1000 EUR) or outputs (i.e., deflated turnover less than 1000 EUR) and removed observations per sector-year whose growth rate was below the 1 or exceeded the 99 percentile. Finally, we removed firms with interruptions in their yearly observations because for these firms we cannot evaluate dynamic cost inefficiency over the entire period of observation. In the following, a firm s can only serve as a benchmark for firm k (i.e. $s \in D_k^t$) if, in addition to dominating k in outputs, we observe it over the same subperiod as firm k. This ensures that we always benchmark firm k with respect to the firms s over the same time period.

We focus on the "Manufacture of fabricated metal products, except machinery and equipment" sector (NACE rev. 1.1 code 28) in Italy over the period 1995-2007. This sector is frequently considered a typical manufacturing sector and is by far the largest sector in our sample in terms of observations. The sector consists of micro (staff in FTE from 10 to 20; 2560 observations), small (staff in FTE from 20 to 50; 2406 observations), medium (staff in FTE from 50 to 250; 1488 observations) and large (staff in FTE larger

[18]The subsequent analysis is carried out in MATLAB using an LP solver provided by the freely available OPTI Toolbox. The program codes are available as supplementary material.

or equal to 250; 141 observations) firms. Firms of different size are quite well represented in our sample.

The AMADEUS dataset contains information on employment costs, material costs, tangible fixed assets, turnover, as well as the total number of employees (in FTE). While we do not observe investments directly, we derive them using the perpetual inventory method (for capital K_t and investment I_t in year t):

$$K_t = (1-\delta)K_{t-1} + I_{t-1}$$
$$\Leftrightarrow I_{t-1} = K_t - (1-\delta)K_{t-1},$$

with $\delta = 0.08$ a constant yearly depreciation rate. We select only firms with a net positive investment $I_t \geq 0$. This leaves us with an unbalanced panel of 6595 observations in 13 years. Table 2 shows that the number of firms in every year ranges between 171 and 773, with an average of 507.31 firms.

Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Number of firms	171	338	395	436	430	478	546	753	613	455	475	773	732

TABLE 2. Number of firms per year

We can find an average wage by dividing employment costs by the total number of employees, and we use industry-level price indexes obtained from EU KLEMS as prices for material costs, investments and turnover. Since our main goal is to illustrate discrepancies between a static ($J = 1$) model and a dynamic ($J = 2$) model stemming from intertemporal dependencies of inputs, we do not discount these input prices (i.e. $\rho = 0$) because this will only further enlarge these discrepancies.

7.1. **Output and input specification.** We use number of employees, deflated material costs and deflated investments as inputs and deflated turnover as output. Table 3 shows summary statistics for our inputs and output. The firms in our sample are very heterogeneous in size as, for example, can be seen from their staff numbers, which range from 10 to 1045. On average, material costs make up the largest share of the total costs, followed by employment costs and investment.

We consider labor as an input that is instantaneously consumed (i.e., not storable or durable). Next, as motivated above, deflated investment forms a prime example of a durable input. Finally, although deflated material expenditures typically do not represent instantaneous consumption, it is not directly clear whether they should be treated as storable or durable. Therefore, we will treat them as storable now but, in a following step we will also treat them as durable and compare the associated efficiency results. For the general model specification (with deflated investment and deflated material costs

	Mean	Std.	Min	Max
Staff numbers (in FTE)	46.015	65.653	10	1045
Deflated materials	3287199.564	8526417.361	1938.991	1.462×10^8
Deflated investments	460386.479	1675022.028	0	6.479×10^7
Wage	29843.722	7424.776	10061.6	83100
Deflator of material costs	1.171	0.122	1	1.387
Deflator of tangible fixed assets	1.161	0.076	1	1.261
Costs of employees (EUR)	1431268.885	2215141.355	100616	2.351×10^7
Material costs (EUR)	3833511.637	1.008×10^7	2410	2.027×10^8
Investment costs (EUR)	542729.86	2048850.239	0	8.172×10^7
Costs of employees as share of total costs	0.378	0.202	0.016	0.998
Material costs as share of total costs	0.528	0.218	0.002	0.973
Investment costs as share of total costs	0.094	0.122	0	0.833
Deflated turnover	7053533.366	1.439×10^7	191129.085	1.968×10^8
Deflator of turnover	1.114	0.113	0.995	1.315

TABLE 3. Summary statistics (6595 observations).

usable in J years), this obtains the $2 \times J$ delay matrix

$$\mathbf{D}^D = \begin{pmatrix} 1 & 0 \ldots 0 \\ 1 & \underbrace{1 \ldots 1}_{J-1} \end{pmatrix},$$

with labor and deflated investment corresponding to row 1 and 2, respectively, and the $1 \times J$ delay matrix

$$\mathbf{D}^S = \begin{pmatrix} 1 & \underbrace{1 \ldots 1}_{J-1} \end{pmatrix},$$

for deflated material costs.

Figure 1 shows the 25 percentile, median and 75 percentile of material costs and employment costs in every year. Employment costs gradually drop over the period 1995-2002, rise from 2002 until 2005, and drop again between 2005 and 2007. The evolution in the levels of material costs throughout the considered period is similar to that of employment costs. Analogously, Figure 2 reports the 25 percentile, median and 75 percentile of tangible fixed assets and investments in every year. The trend for the tangible fixed assets is declining between 1995 and 2002. Over the same subperiod investments fluctuate around 1.244×10^5, with some larger investment peaks in 1995, 1997 and 1999. This seems to indicate that, over this subperiod, the tangible fixed assets stock is gradually decreased by only replacing some of the decommissioned equipment. Following this downward trend, investment picks up in 2002-2005 and increases the available tangible fixed assets stock in 2002-2005. Finally, between 2005 and 2007 the tangible fixed assets stock decreases again and investment only serves to replace decommissioned equipment.

It follows from our discussion in the previous sections that treating investment as a durable input and material costs as a storable input requires us to specify the lifetime

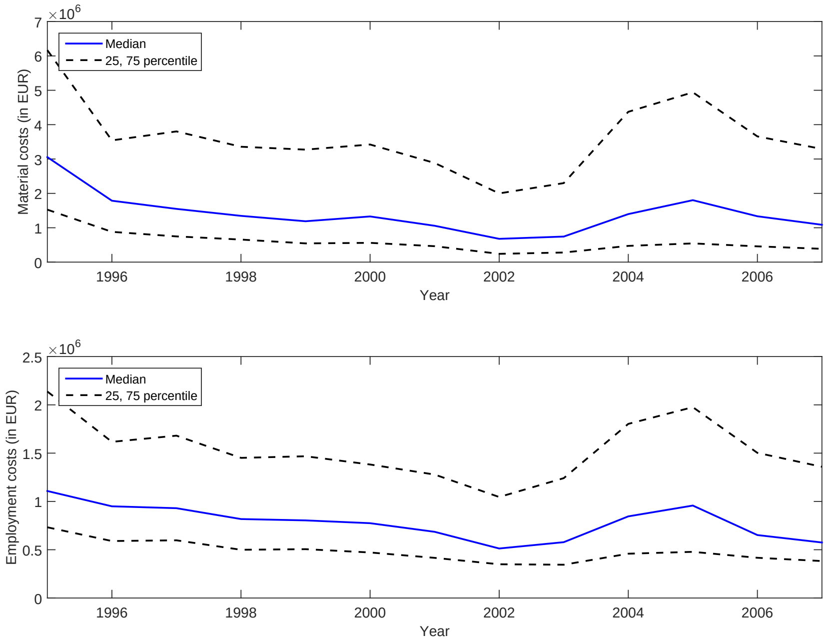


FIGURE 1. Material costs and employment costs over time

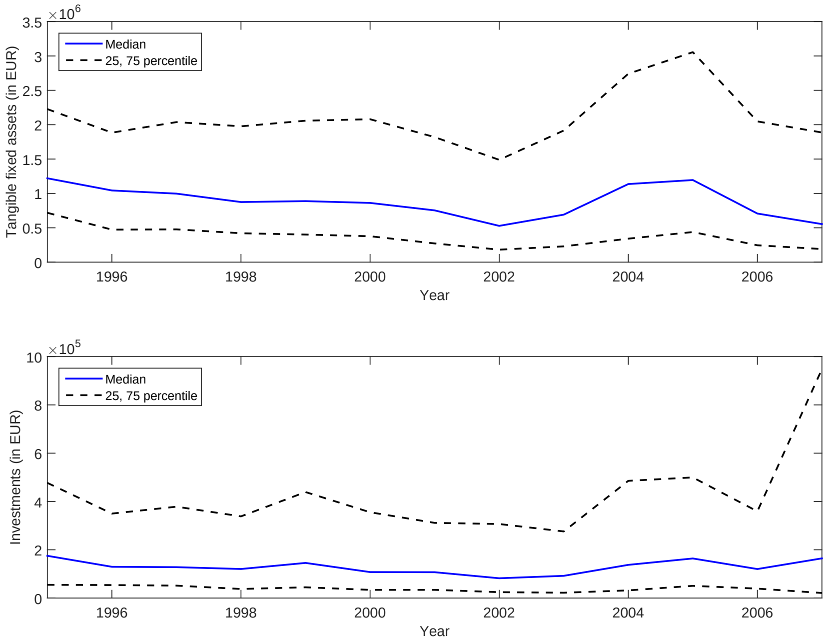


FIGURE 2. Tangible fixed assets and investments over time

(i.e., J). Both inputs are essentially aggregates whose underlying equipment and materials can have very different lifetimes. This makes it hard to specify the exact lifetime of both. For this reason, and to clearly demonstrate the potential impact of intertemporal dependencies between inputs, we will mainly focus on a minimalistic scenario with $J = 2$. In a following analysis, we will also consider alternative values for J to check robustness of our main conclusions.

7.2. **Cost inefficiency.** In summary, the dynamic nature of our empirical analysis relates to one durable input (investment) and one storable input (material costs). Moreover, these inputs represents a fairly large fraction of the total costs relative to the employment costs (see our discussion of Table 3). In what follows, we will show that ignoring the intertemporal aspect can substantially affect the efficiency analysis. Obviously, these distortions will generally be more pronounced in production settings where durable and/or storable inputs form an even more important fraction of total costs.

We begin by comparing the overall cost inefficiency scores relative to total deflated turnover (i.e., $CE_k/\sum_{t=J}^{T_k} Turnover_{k,t}$) of the dynamic ($J = 2$) model and the static ($J = 1$) model. Summary statistics and the total number of efficient firms are presented in Table 4. There are substantially more firms efficient in the dynamic model than in the static model, i.e. 1879 firms (84.37%) vs. 123 firms (5.52%). The estimated potential cost savings for the dynamic model are also much smaller than for the static model. Interestingly, the coefficient of variation shows that the estimated potential cost savings for the dynamic model ($0.127/0.026 \approx 4.885$) are much more dispersed than for the static model ($0.546/0.378 \approx 1.444$).

	#Efficient	Mean	Std.	Min	Max
Dynamic ($J = 2$)	1879 (84.37%)	0.026	0.127	0	3.655
Static ($J = 1$)	123 (5.52%)	0.378	0.546	0	9.828

TABLE 4. Summary statistics of $CE_k/\sum_{t=J}^{T_k} Turnover_{k,t}$ for dynamic ($J = 2$) vs. static ($J = 1$) model.

Next, we consider the per-period inefficiency scores (CE_k^t) of the dynamic ($J = 2$) model and the static ($J = 1$) model. Recall that these per-period inefficiency scores directly indicate potential cost savings in every period in monetary terms. Figure 3 compares the difference in per-period inefficiency relative to the deflated yearly turnover of the $1, 25, 50, 75$ and 99 percentile on a yearly basis. This difference then measures over- or underestimation of potential cost savings by the static model where a negative difference reflects overestimation. From the percentiles we can deduce that for an overwhelming majority of firms and years the detected overestimation is large. It amounts to no less than 0.739 times yearly turnover for the average firm-year observation. For most years (except 1996-1998 and 2007) there are some firms at the extreme end of the distribution for which the dynamic model estimates larger potential cost savings (i.e., between 0.183 and 1.635 times yearly turnover at the 99 percentile) than the static model. This illustrates that there may well be instances where the dynamic model detects more inefficiency than the static model, even though one may generally expect the dynamic model to detect less inefficiency than the static model (because the dynamic model allows for more flexibility to allocate storable and durable inputs over time periods).

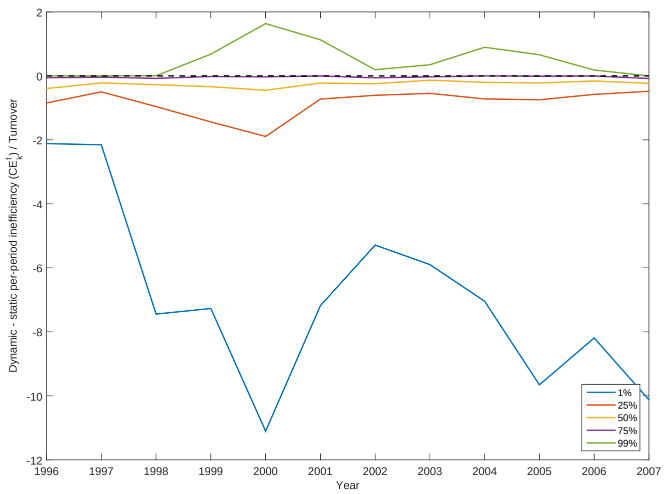
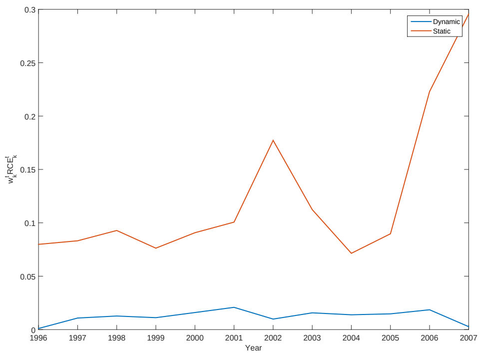


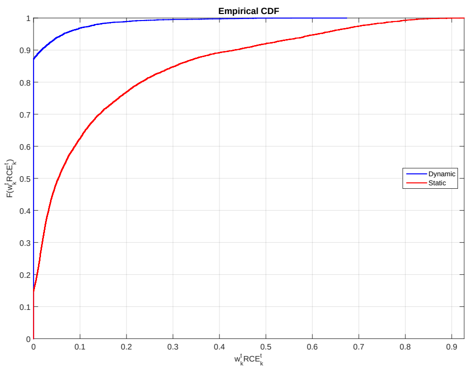
FIGURE 3. 1,25,50,75 and 99 percentiles on a yearly basis of (dynamic CE_k^t - static CE_k^t) / Turnover.

Figure 4a shows that, at the aggregate sample level, the average per-period relative inefficiency is markedly higher for the static model. While the relative inefficiency fluctuates from year to year, average static inefficiency increased by a factor 3.7 over the period as a whole (from 0.08 in 1996 to 0.296 in 2007). The trajectory of average dynamic inefficiency is much smoother over the whole period: throughout the entire period it fluctuates around the average 0.012 with a peak in 2001 at 0.021.

Furthermore, we observe from Figure 4b that in about 87% of cases the dynamic model finds no per-period inefficiency. In contrast, the static model detects no per-period inefficiency in about 15% of cases, and generally reveals much higher levels of inefficiency (i.e., the maximum relative inefficiency level is 0.927 for the static model vs. 0.674 for the dynamic model). When using the nonparametric smoothing test of Li et al. (2009) and adopting a significance level of 1%, we can reject the hypothesis of equality of the $w_k^t RCE_k^t$-distributions for the dynamic and static model for all years $1996 - 2007$ (all respective p-values were effectively or numerically very close to zero).[19]



(A) Mean $w_k^t RCE_k^t$ in every year.



(B) Cumulative density of $w_k^t RCE_k^t$ for dynamic ($J = 2$) vs. static ($J = 1$) model.

FIGURE 4. Dynamic ($J = 2$) vs. static ($J = 1$) model.

Figure 5 reveals some interesting patterns at the firm size level in terms of average dynamic and static per-period inefficiency. The bottom panel in the figure clearly shows that average static per-period inefficiency decreases with increasing firm size. The average micro and small firms experience increased static inefficiency over the entire period, with a drop between 2002 and 2005. The average medium and large firms experience a decline in static inefficiency until 2001, after which inefficiency steadily increases again. Across all firm sizes the average firm ends up with a higher static inefficiency in 2007 than in 1996. Note that these increases in inefficiency for the average large and medium

[19]The test of Li et al. (2009) applies kernel density estimation and considers differences in the entire distribution (unlike, e.g., a Kolmogorov-Smirnov test).

firm occur over the subperiod (i.e., 2002-2005) where investment picks up as well (cfr., Figure 2).

The top panel of Figure 5 shows a rather different picture for average dynamic per-period inefficiency. Starting from very low levels of dynamic per-period inefficiency in 1996, the per-period inefficiency level of the average large firm stays roughly constant until 2000. After that, it enters a subperiod of complete efficiency in 2001-2005. Finally, per-period inefficiency rises and falls over the next two years. At the end of the sample period, the average large firm is roughly as inefficient as in 1996. The per-period inefficiency of the average medium firm initially rises from 0 in 1996 and then fluctuates around 0.01 throughout the entire considered period. The average small firm experiences the largest fluctuations in per-period inefficiency: starting from almost complete efficiency in 1996 the trajectory follows a sawtooth pattern before ending in 2007 at the highest level of per-period inefficiency of all. Finally, the average micro firm experiences a steadily increasing level of per-period inefficiency until 2001 followed by a drop in the subperiod 2001-2003. After that per-period inefficiency level rises to its maximum in 2004 and then steadily decreases again. The average micro firm ends up with the lowest level of inefficiency in 2007.

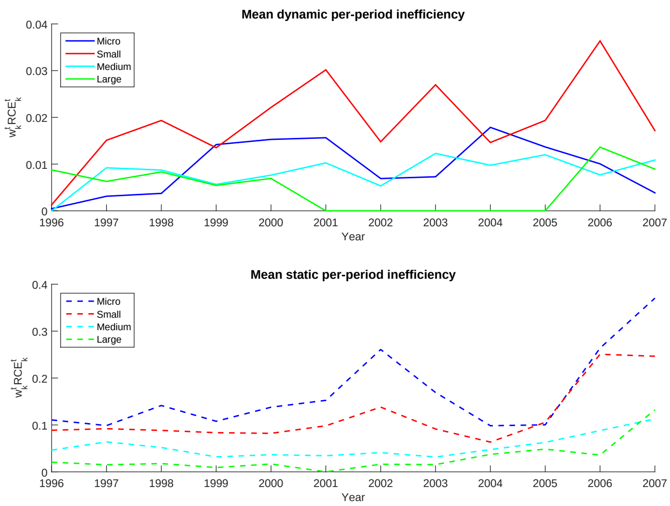
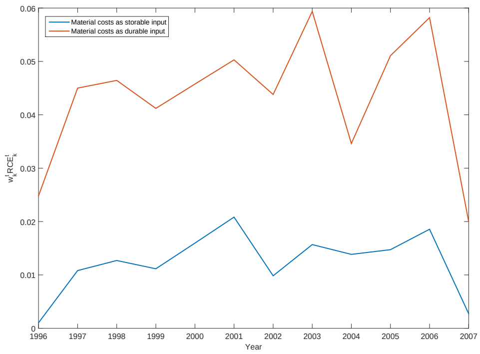
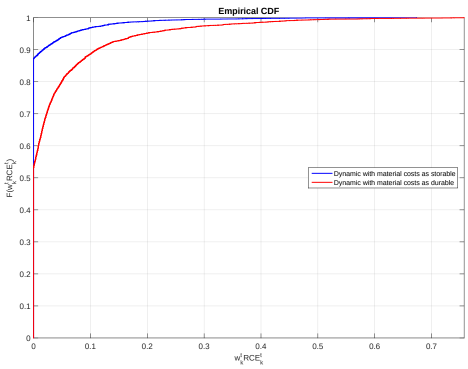


FIGURE 5. Mean dynamic (top) and static (bottom) $w_k^t RCE_k^t$ in every year by firm size.

Finally, in light of our earlier discussion on how to treat material costs, we repeat our dynamic efficiency analysis, but now using both material costs and investment as durable inputs, and we compare the associated efficiency results with our previous results. Figure 6a again shows mean relative per-period inefficiency per year, and Figure 6b shows the cumulative density of the $w_k^t RCE_k^t$-distributions, for the following two specifications: (i) material costs as a storable and investments as a durable input, and (ii) material costs and investments as durable inputs. It should be clear from both figures that the choice between storability and durability makes a difference for the efficiency results. For all years, the nonparametric smoothing test of Li et al. (2009) strongly rejects the hypothesis of equality of the $w_k^t RCE_k^t$-distributions for the two specifications at the 1% significance level (all p-values were effectively or numerically very close to zero). When treating material costs as a durable input the mean relative per-period inefficiency fluctuates much more and is substantially larger than when treating these costs as a storable input.



(A) Mean $w_k^t RCE_k^t$ in every year.



(B) Cumulative density of $w_k^t RCE_k^t$.

FIGURE 6. Material costs as a storable input vs. material costs as a durable input.

7.3. **Technical and allocative inefficiency.** As a following exercise, we redo our dynamic efficiency analysis by using shadow prices. As explained above, this effectively computes the technical inefficiency measure TE_k, which we can further use to calculate the aggregate allocative inefficiency measure AE_k (in (19)) as well as per-period allocative inefficiencies AE_k^t (in (20)). Table 5 shows summary statistics of the TE_k and AE_k results for all firms. We find that technical inefficiency is rather small on average (0.049) for the firms under study, but it varies between firms. The maximal TE_k-value amounts to 0.629. In contrast, the AE_k-values are quite high on average (0.131) and vary about the same as the TE_k-values (standard deviations amount to 0.142 and 0.110 respectively). The worst performing firm has an allocative inefficiency of 0.734, which is higher than the maximal technical inefficiency of 0.629. Figure 7 also shows that there is substantial variation in the average AE_k^t-values over time. One possible explanation for the large discrepancy between the TE_k- and RCE_k-results is that the assumption of perfect price foresight is too strong for the application setting under study.

	mean	std.	min.	max.
RCE_k	0.180	0.168	0.00	0.757
TE_k	0.049	0.110	0.00	0.629
AE_k	0.131	0.142	0.00	0.734

TABLE 5. Summary statistics on RCE_k with material costs and investments as durable inputs, TE_k and AE_k.

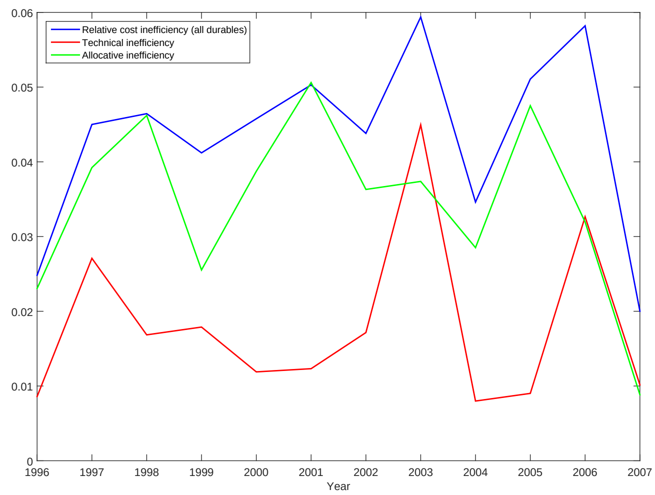


FIGURE 7. Per-period mean dynamic cost $w_k^t RCE_k^t$ with material costs and investments as durables, technical $\tilde{w}_k^t TE_k^t$ and allocative inefficiencies $\hat{w}_k^t AE_k^t$.

Figure 8 shows the mean dynamic per-period technical and allocative inefficiency per firm size. In general, the average inefficiency levels per firm size are small (i.e. at most 0.08), but the fluctuation pattern differs per firm size. We first discuss technical inefficiency (upper part of Figure 8). The average large firm is almost entirely technical efficient throughout the considered period: there is only a very small level of technical inefficiency in 2002 and 2006. The average medium firm experiences slightly higher levels of technical inefficiency throughout the period: it is only fully technical efficient in 1999 and 2005 with a peak of technical inefficiency in 2003 at 0.018. Nevertheless, these technical inefficiency levels are almost always lower (except in 2000) than those of the average small or micro firm. The per-period technical inefficiency scores of the average small and micro firm follow the same pattern throughout with some crossovers in the considered period. The largest peaks in technical inefficiency are in 1997, 2003 and 2006. All but the average large firm end up with a higher technical inefficiency in 2007 than in 1996.

Next, we turn to the average per-period allocative inefficiency per firm size. Here, the story is a bit different. While the average large firm still has a subperiod of allocative efficient levels in 2001-2004, there are (i) more periods where it is allocative inefficient (with notable upticks in 2000 and 2005) and (ii) at higher levels than its technical inefficiency. Although it starts out with some level of allocative inefficiency in 1996 (0.02), the average large firm does end up allocative efficient in 2007. While the average medium firm initially experiences a slow decrease in allocative inefficiency in 1997-1999, allocative inefficiency slowly increases over the entire period. The pattern for the average small firm follows a sawtooth pattern throughout the period: upticks in allocative inefficiency follow downturns in allocative inefficiency. Just as for technical inefficiency, the average small firm ends up with the highest allocative inefficient level of all firm sizes in 2007. Interestingly, the per-period allocative inefficiency of the average small firm seems to converge to that of the average medium firm in the final two years. Finally, the average micro firm sees a steady increase in allocative inefficiency until 2000 after which a decreasing trend sets in so that it ends up at a lower level of allocative inefficiency in 2007 (i.e. 0.018) than in 1996.

Overall when comparing Figure 5 and Figure 8, we find that technical and allocative per-period inefficiency contribute differently to intertemporal per-period cost inefficiency for different firm sizes. The average large firm is almost intertemporal cost efficient throughout the period and any inefficiency is primarily allocative. Thus, the average large firm seems to have access to the latest state-of-the-art technologies, but mispredicts future prices sometimes. The average medium firm has similar but slightly higher levels of intertemporal cost inefficiency which is again mostly driven by allocative rather than technical inefficiency. Note though that technical inefficiency is not absent and larger than for the average large firm. Finally, both technical and allocative inefficiency contribute more equally to intertemporal per-period cost inefficiency for both the average small and micro firm. Thus, this seems to indicate that both the average small and micro firm (i) have limited (and/or delayed) access to the latest state-of-the-art technologies and (ii) are less able to accurately predict future prices.

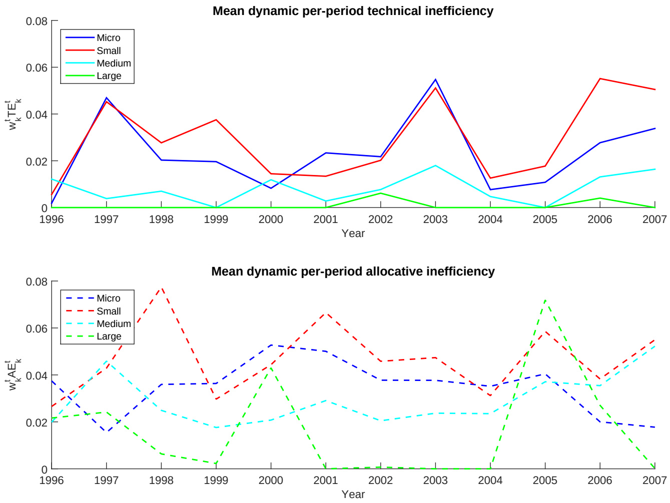


FIGURE 8. Per-period mean dynamic technical $\tilde{w}_k^t TE_k^t$ (top) and allocative (bottom) inefficiencies $\hat{w}_k^t AE_k^t$ by firm size.

7.4. **Robustness checks.** We conclude our empirical analysis by conducting a number of robustness checks. These additional exercises will further illustrate the versatility of our general framework, in terms of relaxing or imposing particular assumptions in the intertemporal efficiency assessment. First, we consider the effect of specifying J on our results. Second, we compute efficiency results when imposing particular (degressive and linear) structure on the write-off schemes used by the evaluated firms.

We begin by evaluating the sensitivity of the overall dynamic inefficiencies to alternative specifications of J (> 1). We find that the number of efficient firms (out of 2227) increases with J: 1867 (83.83%) efficient firms for $J = 2$; 2167 (97.31%) efficient firms for $J = 3$; 2213 (99.37%) efficient firms for $J = 4$ and 2225 (99.91%) efficient firms for $J = 5$. Thus, the results are sensitive to a proper choice of J. Generally, higher values of J weaken the efficiency criteria. Intuitively, higher J-values generally allow for more flexibility to allocate storable and durable inputs over time periods.

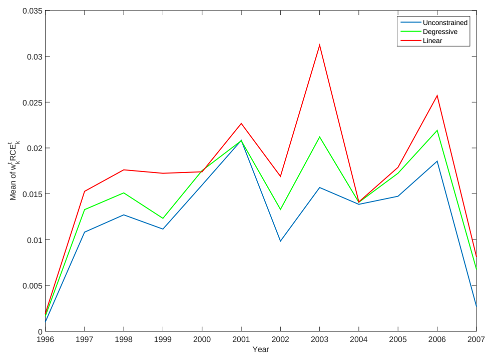
Finally, at the end of Section 5 we indicated that an interesting feature of our methodology is that it allows for imposing specific hypotheses regarding the allocation of the costs of durables to individual time periods (i.e., putting structure on the write-off schemes). As a last robustness check, we compute efficiency results for the degressive scheme in (12) and the linear scheme in (13).

Our results are summarized in Figure 9. Figure 9a shows the mean relative per-period inefficiency for the different write-off schemes. First of all, we find that mean inefficiency generally increases with the stringency of the write-off scheme constraint (i.e., linear $\geq$ degressive $\geq$ unconstrained). Actually, this could be expected a priori as the degressive and linear models put increasingly stringent structure on possible allocations of investments to successive time periods. Second, the trajectories of inefficiency for the different write-off schemes are generally similar over the entire period. The difference in per-period inefficiency across the different write-off schemes is clearly largest in the year 2003.

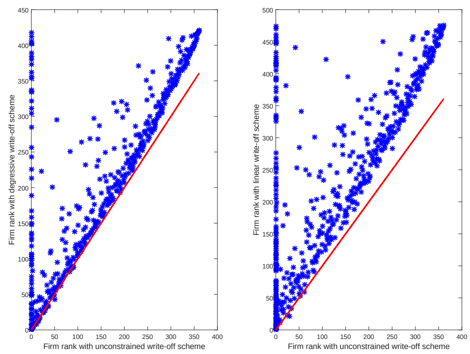
Next, Figure 9b compares the individual firm rankings for the degressive and linear write-off schemes with the firm rankings for the unconstrained write-off scheme. Firms situated below the 45 degree line attain a higher ranking (i.e., lower rank number) for the more restrictive scheme while, obviously, the opposite holds for firms above the 45 degree line. Note that all efficient firms receive rank 1 as we cannot further discriminate between efficient firms. All observations on the y-axis are efficient firms that become inefficient when imposing specific write-off schemes. The recurring pattern for both figures is that the lower the firm ranking when not imposing any specific write-off scheme, the larger the drop in the firm ranking after imposing a specific write-off scheme: production observations are generally situated further away from the 45 degree line as one progresses to the right on the x-axis. However, we also have quite many firm observations that are situated exactly on the y-axis. These are firms that can be diagnosed as dynamically efficient when not imposing a specific write-off scheme, but which experience an (often substantial) drop in ranking/efficiency when assuming a degressive or linear scheme. At the higher end of the ranking, we do observe some increases in firm ranking after imposing specific write-off scheme (i.e., some firm observations are below the 45 degree line). Generally, we can conclude that imposing particular write-off hypotheses can have a – sometimes significant – impact on the ranking of firms.

8. Conclusion

We have presented a methodology for a nonparametric analysis of cost efficient production behavior that can account for intertemporal considerations related to the use of storable and durable inputs. We do not require imposing (nonverifiable) functional



(A) Mean relative per-period inefficiency.



(B) Individual firm rank with different write-off schemes: unconstrained vs. degressive (left) and unconstrained vs. linear (right).

FIGURE 9. Comparison of different write-off schemes (unconstrained, degressive and linear).

structure on the production technology. The methodology is versatile in that it can account for production delays of durable inputs and storage costs for storable inputs. In addition, it allows for defining a cost inefficiency measure that can be decomposed in period-specific inefficiencies. These cost inefficiencies can be computed through simple linear programming.

Our application to the Italian manufacturing sector of fabricated metal products has shown the empirical usefulness of our methodology. Most notably, it showed that explicitly accounting for the dynamic nature of (in our case investment and materials) inputs can significantly impact the efficiency results. For a considerable number of firms, we found that per-period inefficiencies for our model, with capital investments as durable inputs and materials as storable inputs, differed substantially from the ones for the (static) model that ignores such intertemporal durability and storability. At a more general level, these empirical findings demonstrate the practical relevance of our methodology: erroneously disregarding intertemporal aspects of firms' production decisions may substantially distort the efficiency assessment and, therefore, also the managerial conclusions that are drawn from it.

We see multiple possible extensions. First, at the application level, it seems particularly interesting to use our methodology to assess efficiency in regulated industries. For such industries, it is vital that the regulator takes intertemporal dependencies of production into account in the regulation exercise. However, regulators generally do not incorporate these interdependencies in practice. This is sometimes motivated by a lack of panel data, which forces regulators to limit the analysis to cross-sectional data (Pollitt, 2005). However, also the used definition of capital costs (Shuttleworth, 2005) and capitalization policies across firms, countries or industries (Haney and Pollitt, 2013) can be contested.[20] Clearly, not taking these dependencies into account can lead to erroneous cost reduction targets. Shuttleworth (2005) reports a case where Ofgem, a UK electricity distribution regulator, imposed a too strong target for one distributor (Seeboard) while imposing a too loose target on another (Southern). This discrepancy was due to the fact that Ofgem only considered operational expenses, while disregarding capital expenses. And it happened that Southern was characterized by high capital expenses and low operational expenses, while the opposite applied to Seeboard. Balk et al. (2010) also discuss the importance of correctly considering capital expenses in the regulation analysis and note that there is no consensus on how to properly deal with capital expenses in the studies they considered.

The relation between regulatory regime and investment has received a lot of attention (see Guthrie (2006) for a discussion). Focusing on cross-sectional data can lead to penalization of firms that invest while rewarding those that delay investments. Nick and Wetzel (2015) conclude that firms have an incentive to cut investments when the regulator uses a static benchmarking model. Our empirical application has shown that the resulting dynamic efficiency conclusions may significantly differ from the ones that are based on a static efficiency analysis. In our opinion, this directly motivates the practical relevance of our methodology as these differences may substantially affect the regulatory policies that are based on the efficiency assessment. For our data, the intertemporal efficiency assessment generally reveals lower (but still frequently significant) potential cost savings than the static assessment. As the intertemporal efficiency results are based on a more realistic modeling of the evaluated production processes, they yield more credible

[20] There does exist general guidelines on capital measurement. We refer to the manual on capital measurement of the OECD (2009) for an example.

cost saving estimates than the static efficiency results. Generally, this makes that our novel efficiency assessment methodology can often provide a more convincing basis for designing (and enforcing) regulatory policies.

Next, at the theoretical level, we have been considering a single output setting in the current paper but ignored any output interdependencies, mainly to simplify our exposition. In practice, however, interdependencies among outputs often exist in the form of joint inputs. From this perspective, it seems particularly interesting to combine our methodology for dynamic production analysis with the (nonparametric) methodology for multi-output production analysis that was recently developed by Cherchye et al. (2013, 2014). This multi-output framework accounts for interdependencies between different output production processes through jointly used inputs, which are formally similar in nature to the durable inputs on which we focus in the current paper (i.e., they capture inter-period interdependencies between production decisions). Combining the two methodologies will further enhance the realistic modeling of production interdependencies (across outputs as well as time periods).

Third, our cost and technical inefficiency measures can be used to measure productivity by combining it with various productivity measures such as the cost Malmquist index of Maniadakis and Thanassoulis (2004) or the Malmquist index of Caves et al. (1982), Bjurek (1996)'s Hicks-Moorsteen index or the Luenberger indicator of Chambers et al. (1996), among others. These productivity measures have been proposed in the context of nonparametric (DEA) analysis of productive efficiency and, therefore, are easily combined with our novel methodology. This combination will lead to richer productivity analyses because it explicitly accounts for intertemporal production interdependencies through storable and durable inputs.

Fourth, the focus of this paper was on efficiency under the assumption of perfect price foresight: i.e., our efficiency measures are based on solutions of LPs that assume that the evaluated firm correctly predicts the prevailing prices. In practice, this assumption rarely holds so that inefficiency can, at least partially, be explained by the failure of this assumption. It would therefore be useful to also consider efficiency measures allowing for price uncertainty, where one considers efficiency under all possible (realistic) price situations. A good starting point in this respect may be the framework introduced in Kuosmanen and Post (2002). Similar extensions pertain to our assumptions of no adjustment costs and no incremental improvement or learning models. As indicated in the introductory section, Fallah-Fini et al. (2014) indicated that relaxing these assumptions can entail additional sources of intertemporal dependence of production processes.

Finally, Varian (1982) has developed a nonparametric approach to consumer demand analysis that is formally analogous to the nonparametric approach to production analysis to which we adhere here. Following this analogy, we may translate the insights developed in the previous sections towards a consumption setting to obtain a more realistic modeling of intertemporal aspects of consumer behavior.[21] Specifically, our concept of storable inputs corresponds to the notion of infrequent purchases in a consumption context, and durable inputs are similar in spirit to durable consumption goods (for example, cars, houses, etc.) in a demand setting.

References

Afriat, S. N. (1972): "Efficiency Estimation of Production Functions," *International Economic Review*, 13, 568–598.

Balk, B., J. Blank, J. Bos, and P. Koot (2010): "Onderzoek naar de generieke productiviteitsontwikkeling in de reguleringsmethode van Tennet TSO BV: Rapport voor de Energiekamer van de NMa," Tech. rep., Universiteit Delft, Instituut voor Public Sector Efficiency.

Banker, R. D. and A. Maindiratta (1988): "Nonparametric analysis of technical and allocative efficiencies in production," *Econometrica*, 56, 1315–1332.

Bjurek, H. (1996): "The Malmquist total factor productivity index," *Scandinavian Journal of Economics*, 303–313.

Blackorby, C. and R. R. Russell (1999): "Aggregation of efficiency indices," *Journal of Productivity Analysis*, 12, 5–20.

Caves, D. W., L. R. Christensen, and W. E. Diewert (1982): "The economic theory of index numbers and the measurement of input, output, and productivity," *Econometrica*, 1393–1414.

Cazals, C., J.-P. Florens, and L. Simar (2002): "Nonparametric frontier estimation: a robust approach," *Journal of Econometrics*, 106, 1 – 25.

Chambers, R. G., Y. Chung, and R. Färe (1998): "Profit, directional distance functions, and Nerlovian efficiency," *Journal of Optimization Theory and Applications*, 98, 351–364.

Chambers, R. G., R. Färe, and S. Grosskopf (1996): "Productivity growth in APEC countries," *Pacific Economic Review*, 1, 181–190.

Chen, C.-M. and J. van Dalen (2010): "Measuring dynamic efficiency: Theories and an integrated methodology," *European Journal of Operational Research*, 203, 749 – 760.

[21]See, for example Crawford (2010) and Crawford and Polisson (2014) for recent contributions to the nonparametric analysis of intertemporal consumer behavior.

Cherchye, L., B. De Rock, B. Dierynck, F. Roodhooft, and J. Sabbe (2013): "Opening the "Black Box" of Efficiency Measurement: Input Allocation in Multioutput Settings," *Operations Research*, 61, 1148–1165.

Cherchye, L., T. Demuynck, B. De Rock, and K. De Witte (2014): "Nonparametric Analysis of Multi-output Production with Joint Inputs," *The Economic Journal*, 124, 735–775.

Cherchye, L., W. Moesen, N. Rogge, and T. Van Puyenbroeck (2007): "An introduction to benefit of the doubt composite indicators," *Social Indicators Research*, 82, 111–145.

Cook, W. D., M. Hababou, and H. J. Tuenter (2000): "Multicomponent efficiency measurement and shared inputs in data envelopment analysis: an application to sales and service performance in bank branches," *Journal of Productivity Analysis*, 14, 209–224.

Crawford, I. (2010): "Habits Revealed," *Review of Economic Studies*, 77, 1382–1402.

Crawford, I. and M. Polisson (2014): "Testing for intertemporal nonseparability," *Journal of Mathematical Economics*, 52, 46–49.

Daraio, C. and L. Simar (2005): "Introducing environmental variables in nonparametric frontier models: a probabilistic approach," *Journal of Productivity Analysis*, 24, 93–121.

——— (2007): *Advanced Robust and Nonparametric Methods in Efficiency Analysis: Methodology and Applications*, Springer.

Fallah-Fini, S., K. Triantis, and A. L. Johnson (2014): "Reviewing the literature on non-parametric dynamic efficiency measurement: state-of-the-art," *Journal of Productivity Analysis*, 41, 51–67.

Färe, R. (1986): "A dynamic non-parametric measure of output efficiency," *Operations Research Letters*, 5, 83–85.

Färe, R., R. Grabowski, S. Grosskopf, and S. Kraft (1997): "Efficiency of a fixed but allocatable input: A non-parametric approach," *Economics Letters*, 56, 187–193.

Färe, R. and S. Grosskopf (1996): *Intertemporal production functions with dynamic DEA*, Kluwer Academic Publishers.

Färe, R. and S. Grosskopf (2000): "Network DEA," *Socio-Economic Planning Sciences*, 34, 35–49.

Färe, R., S. Grosskopf, and D. Margaritis (2010): "Time substitution with application to data envelopment analysis," *European Journal of Operational Research*, 206, 686–690.

Färe, R. and D. Primont (1995): *Multi-output production and duality: theory and applications*, Springer Science & Business Media.

Färe, R. and V. Zelenyuk (2003): "On aggregate Farrell efficiencies," *European Journal of Operational Research*, 146, 615–620.

Forsund, F. R. and L. Hjalmarsson (1974): "On the Measurement of Productive Efficiency," *The Scandinavian Journal of Economics*, 76, 141–154.

Färe, R., S. Grosskopf, and G. Whittaker (2007): "Network DEA," in *Modeling Data Irregularities and Structural Complexities in Data Envelopment Analysis*, ed. by J. Zhu and W. Cook, Springer US, 209–240.

Guthrie, G. (2006): "Regulating infrastructure: The impact on risk and investment," *Journal of Economic Literature*, 44, 925–972.

Hackman, S. T. and R. C. Leachman (1989): "A general framework for modeling production," *Management Science*, 35, 478–495.

Haney, A. B. and M. G. Pollitt (2013): "International benchmarking of electricity transmission by regulators: A contrast between theory and practice?" *Energy Policy*, 62, 267 – 281.

Johansen, L. (1959): "Substitution versus fixed production coefficients in the theory of economic growth: a synthesis," *Econometrica*, 27, 157–176.

Kao, C. (2013): "Dynamic data envelopment analysis: A relational analysis," *European Journal of Operational Research*, 227, 325–330.

Kao, C. and S.-N. Hwang (2008): "Efficiency decomposition in two-stage data envelopment analysis: An application to non-life insurance companies in Taiwan," *European Journal of Operational Research*, 185, 418–429.

Kuosmanen, T. and T. Post (2002): "Nonparametric efficiency analysis under price uncertainty: a first-order stochastic dominance approach," *Journal of Productivity Analysis*, 17, 183–200.

Li, Q., E. Maasoumi, and J. S. Racine (2009): "A nonparametric test for equality of distributions with mixed categorical and continuous data," *Journal of Econometrics*, 148, 186–200.

Maniadakis, N. and E. Thanassoulis (2004): "A cost Malmquist productivity index," *European Journal of Operational Research*, 154, 396–409.

Merlevede, B., M. De Zwaan, K. Lenaerts, and V. Purice (2015): "Multinational networks, domestic, and foreign firms in Europe," Tech. rep., Ghent University, Faculty of Economics and Business Administration.

Nemoto, J. and M. Goto (1999): "Dynamic data envelopment analysis: modeling intertemporal behavior of a firm in the presence of productive inefficiencies," *Economics Letters*, 64, 51 – 56.

——— (2003): "Measurement of Dynamic Efficiency in Production: An Application of Data Envelopment Analysis to Japanese Electric Utilities," *Journal of Productivity Analysis*, 19, 191–210.

NICK, S. AND H. WETZEL (2015): "The hidden cost of investment: the impact of adjustment costs on firm performance measurement and regulation," *Journal of Regulatory Economics*, 1–23.

OECD (2009): "Measuring Capital - OECD Manual 2009," Tech. rep., Organisation for Economic Co-operation and Development.

OUELLETTE, P. AND L. YAN (2008): "Investment and dynamic DEA," *Journal of Productivity Analysis*, 29, 235–247.

POLLITT, M. (2005): "The role of efficiency estimates in regulatory price reviews: Ofgem's approach to benchmarking electricity networks," *Utilities Policy*, 13, 279 – 288.

SHEPHARD, R. W. AND R. FÄRE (1980): *Dynamic theory of production correspondences*, vol. 50, Verlag Anton Hain.

SHUTTLEWORTH, G. (2005): "Benchmarking of electricity networks: Practical problems with its use for regulation," *Utilities Policy*, 13, 310–317.

SILVA, E. AND S. E. STEFANOU (2003): "Nonparametric dynamic production analysis and the theory of cost," *Journal of Productivity Analysis*, 19, 5–32.

VARIAN, H. (1982): "The nonparametric approach to demand analysis," *Econometrica*, 50, 945–973.

——— (1984): "The nonparametric approach to production analysis," *Econometrica*, 52, 579–597.

VERSCHELDE, M., M. DUMONT, G. RAYP, AND B. MERLEVEDE (2016): "Semiparametric stochastic metafrontier efficiency of European manufacturing firms," *Journal of Productivity Analysis*, 45, 53–69.

WIBE, S. (2008): "Efficiency: A dynamic approach," *International Journal of Production Economics*, 115, 86–91.

Appendix A. The economic meaning of write-off schemes

In this short section we clarify the economic intuition behind the write-off schemes for the durable inputs. For a moment, let us rewrite (1) by replacing input requirement sets with production functions:

$$\min_{\substack{(\mathfrak{q}_t^1,\dots,\mathfrak{q}_t^J)_{t=J}^{T_k} \\ (\mathfrak{P}_t^1,\dots,\mathfrak{P}_t^J)_{t=J}^{T_k}}} \sum_{t=J}^{T_k} \sum_{j=t-J+1}^{t} \mathbf{p}_{k,j}\mathfrak{q}_j^{t-j+1} + \mathfrak{P}_j^{t-j+1}\mathbf{Q}_{k,j} \tag{21a}$$

$$\text{s.t. } \mathbf{F}_t\left(\sum_{j=1}^{J} \mathfrak{q}_{t-j+1}^{j}, \sum_{j=1}^{J} \mathbf{Q}_{k,t-j+1}\right) \geq \mathbf{y}_{k,t} \qquad \forall t = J,\dots,T_k \tag{21b}$$

The first order conditions with respect to $\mathbf{Q}_{k,t}$, for all $t = J,\dots,T_k$, are

$$\sum_{j=1}^{J} \mathfrak{P}_{k,t}^{j} - \sum_{j=1}^{J} \lambda_{t+j-1} \frac{\partial \mathbf{F}_{t+j-1}}{\partial \mathbf{Q}_{k,t}} \geq 0 \Leftrightarrow \mathbf{P}_{k,t} - \sum_{j=1}^{J} \lambda_{t+j-1} \frac{\partial \mathbf{F}_{t+j-1}}{\partial \mathbf{Q}_{k,t}} \geq 0,$$

which holds with equality if $\mathbf{Q}_{k,t} > 0$. Rearranging shows that, when a durable input $\mathbf{Q}_{k,t}$ is purchased at time t, the discounted market prices reflect the expected marginal benefits to production of the durable inputs over their entire lifetime, i.e.,

$$\mathbf{P}_{k,t} = \sum_{j=1}^{J} \lambda_{t+j-1} \frac{\partial \mathbf{F}_{t+j-1}}{\partial \mathbf{Q}_{k,t}}, \tag{22}$$

which reveals a write-off scheme that defines the valuation of the firm for durable inputs in terms of their marginal effects on productivity in periods t to $t+J-1$. It is as if the firm invests in this input and writes off this investment for J periods. We capture this interpretation by (implicit) period-specific prices

$$\mathfrak{P}_{k,t}^{j} = \lambda_j \frac{\partial \mathbf{F}_j}{\partial \mathbf{Q}_{k,t}}. \tag{23}$$

Intuitively, these period-specific prices attribute part of the cost of the durable inputs to different periods t in accordance to the inputs' marginal productivities.