KATHOLIEKE UNIVERSITEIT
LEUVEN

Faculty of Economics and
Applied Economics

Department of Economics

Nonparametric Tests of Optimizing Behavior in Public Service Provision: Methodology and an Application to Local Public Safety

by

Laurens CHERCHYE
Bruno DE BORGER
Tom VAN PUYENBROECK

Public Economics

Center for Economic Studies
Discussions Paper Series (DPS) 04.16
http://www.econ.kuleuven.be/ces/discussionpapers/default.htm

December 2004

DISCUSSION
PAPER

NONPARAMETRIC TESTS OF OPTIMIZING BEHAVIOR IN PUBLIC SERVICE PROVISION:
Methodology and an application to local public safety

Laurens Cherchye,
Bruno De Borger
Tom Van Puyenbroeck*

Abstract

We develop a positive non-parametric model of public sector production that allows us to test whether an implicit procedure of cost minimization at shadow prices can rationalize the outcomes of public sector activities. The basic model focuses on multiple C-outputs and does not imply any explicit or implicit assumption regarding the trade-offs between the different inputs (in terms of relative shadow prices) or outputs (in terms of relative valuation). The proposed methodology is applied to a cross-section sample of 546 Belgian municipal police forces. Drawing on detailed task-allocation data and controlling, among others, for the presence of state police forces, the cost minimization hypothesis is found to provide a good fit of the data. Imposing additional structure on output valuation, derived from available ordinal information, yields equally convincing goodness-of-fit results. By contrast, we find that aggregating the labor input over task specializations, a common practice in efficiency assessments of police departments, entails a significantly worse fit of the data.

Keywords: Public agencies, optimizing behavior, nonparametric production, local police departments
JEL codes: C14, C61, D21, D24

* Laurens Cherchye: Catholic University of Leuven (KUL) and Flemish Fund for Scientific Research (FWO-V); Bruno De Borger: University of Antwerp (UA); Tom Van Puyenbroeck: European University College Brussels (Ehsal). We are grateful to the Ministry of the Interior for financial support and for providing most of the data. The work in this paper does not necessarily reflect the official views of the Ministry of the Interior. All errors are ours.

1. INTRODUCTION

Non-parametric analysis of optimizing behavior is rooted in the works of, among others, Afriat (1972) and Hanoch and Rothschild (1972). They recognized that optimization implies some straightforward restrictions on observed choices that can be tested without imposing particular functional forms to describe preferences or technology. For example, the claim that firms minimize costs implies that, at given input prices, the cost of any observed production activity cannot exceed the cost of another production activity that yields at least as much output. This observation led Varian (1984) to introduce the Weak Axiom of Cost Minimization (WACM). He further developed an appropriate non-parametric technique to test for consistency with the axiom on the basis of sample data on firms' input-output quantities and input prices. More recently, this methodology has been extended and applied to other types of optimizing behavior. We refer to, for example, Blundell et al. (2003) and the references therein, for recent advances within a utility maximization context. Moreover, applications of profit maximization tests in a production setting are discussed in, among others, Hailu and Veeman (2001). Finally, Snyder (2001) uses non-parametric techniques to evaluate observed data for their consistency with Pareto optimal provision of public goods.

In this paper, we extend the non-parametric methodology referred to above to incorporate essential features of public sector behavior. The model we develop aims at testing whether the observed behavior of public agencies is consistent with an appropriately defined optimization problem that captures a number of important characteristics of the environment in which such agencies typically operate. Once these particular features are taken into account, can the activities of individual decision units in the public sector somehow be rationalized by optimizing behavior? The proposed methodology is applied to the provision of local public safety in Belgium, using data from a large sample of municipal police departments.

The idea that input-output combinations in the public sector may be the result of optimizing behavior raises a number of issues. First, on the input side, there is considerable doubt that federal and local public agencies would actually pursue cost minimization for given input prices. A variety of reasons have been mentioned in the literature, including agency problems, managerial slack, regulatory restrictions, and the imperfect link between public sector wages and labor productivity (for discussion of these arguments, see, e.g., Bös (1986), Pestieau and Tulkens (1993), and Mueller (2000)). A number of empirical studies have provided some support for these ideas. For example, Atkinson and Halvorsen (1986) and De Borger (1993) produced evidence of deviations from cost minimizing behavior at observed input prices for both public and regulated firms in the US and Europe. Similarly, recent studies on the behavior of local governments found substantial evidence of suboptimal choices on the input side (see, e.g., Grosskopf et al. (1995), Hayes et al. (1998), and Grossman et al. (1999)). One interpretation of this literature is that, at best, the public sector may be guided by the implicit use of (possibly agency-specific) shadow input prices that reflect the phenomena mentioned above.

A second problem is related to the output side of public sector production activities. Even if such activities are guided by some underlying optimization process, then what are the appropriate outputs entering this process? Ever since the seminal paper of Bradford, Malt and Oates (1969), economists have worried about the distinction between D-outputs and C-outputs. The former are outputs directly produced (e.g., the number of tax files administered by the tax authorities, operations performed, the total hours patrolled by local police forces, etc.), the latter refer to what is the ultimate concern of citizens (e.g., public safety, health, etc.). Empirical studies of public sector performance have often focused on direct outputs, if only because D-outputs are typically easier to measure (see, among many others, Gyiamah-Brempong (1989), Van Tulder (1994) and Grosskopf et al. (1995)). However, this emphasis on direct outputs is questionable. If indeed C-outputs are citizens' ultimate concern, public sector officials will be held accountable for their performance relative to these outputs. As a consequence, C-outputs may be quite relevant in guiding public agencies' decisions. For example, suppose that citizens evaluate local police performance in terms of changes in local safety as captured by, e.g., the reduction is criminal offenses of various types. Then it seems plausible that local officials take such safety objectives into account when making decisions, and that they do not limit their attention to intermediate outputs such as the number of police patrols. Unfortunately, introducing C-outputs in empirical analyses of public sector performance raises several other issues of concern. Given that prices for such outputs are not available, it is not clear how to model the relative valuation of the various outputs by decision makers. Moreover, it has forcefully been argued that the production of C-outputs may strongly depend on environmental characteristics that are exogenous to the decision maker (see, e.g., Ruggiero (1996a,b) and MacDonald (2002)).

With these considerations in mind, the purpose of this paper is twofold. First, we develop a positive non-parametric model of (local) public sector production that allows us to test whether the outcomes of public sector activities can be rationalized by an implicit procedure of cost minimization at shadow prices (i.e., consistency with the WACM, but at shadow rather than market prices). Obviously, our focus is on necessary, not sufficient, conditions for optimizing behavior. The basic model we develop focuses on multiple C-outputs and does not impose any explicit or implicit assumption regarding the trade-offs between the different inputs (in terms of relative shadow prices) or outputs (in terms of relative valuation). It also takes into account the presence of characteristics of the production process that are exogenous to the decision maker. The methodology we propose to test for deviations of optimizing behavior builds on Varian's (1990) suggestion to interpret standard non-parametric efficiency measures as goodness-of-fit indicators. He convincingly argued that such measures provide information on the closeness of observed outcomes with respect to hypothesized optimizing behavior; this insight underlies the tests developed below.

A second purpose of the paper is to provide, as far as we know, the first attempt to test for deviations from shadow cost minimizing behavior by individual public agencies[1]. We apply our methodology to a cross-section sample covering 546 Belgian municipal police forces. The empirical analysis draws on detailed labor task-allocation data, and it explicitly controls for the local presence of state police forces as an alternative provider of local public safety. Several versions of the model are considered. We first apply the basic model described above. We then compare the results with those of two variants in which more structure is added: in one case we aggregate labor inputs, as is common in the literature; the other variant uses a priori information on the decision makers' relative valuation of the different outputs. Among other findings, the results suggest that the basic model does indeed provide a good fit to the data. However, aggregating labor inputs, a typical characteristic of

[1] In an interesting recent paper Grosskopf et al. (1995) do test for cost minimizing behavior in the production of public safety. However, unlike the current paper, they test for cost minimization at observed input prices, they use parametric techniques (distance functions), and they do not focus on direct outputs.

previous police efficiency studies, is found to yield a substantial reduction in the explanatory power of the model.

Note from the previous discussion that there are obvious technical similarities between the methodology suggested to assess non-optimizing behavior and the efficiency measurement literature. Empirical efficiency studies are scattered throughout the literature and include several evaluations of the performance of local police forces (see, e.g., Van Tulder (1994), Thanassoulis (1995), and Drake and Simper (2003)). The non-parametric efficiency literature typically imposes some behavioral and technological assumptions (that are often non-verifiable, such as convexity), and interprets deviations of observations from the estimated frontier as inefficiencies. Unfortunately, for public sector activities little is known about both the production technology (for example, can convexity be assumed?) and the trade-offs that implicitly guide public decision-makers, so that this interpretation may be unwarranted[2]. In the current paper, we therefore take a totally different perspective. We do not impose non-verifiable assumptions such as convexity, and do not restrict the allowed trade-offs among inputs and between different outputs. We then ask the question whether the data are consistent with a particular behavioral model, viz. cost minimizing behavior at unobservable shadow prices.

The structure of the paper is as follows. In Section 2 we take up the methodological aspects to adapt the nonparametric methodology to public sector behavior, incorporating shadow prices and output valuation functions. Application of the methodology to local public safety provision by police departments starts in Section 3 with a discussion of our input, output, and environmental data. Section 4 reports our empirical results on the validity of the optimizing assumptions for three alternative specifications of the shadow cost minimization model. Finally, section 5 contains a summary and some concluding remarks.

[2] Imposing potentially unwarranted assumptions implies that many observations may inappropriately be labeled as inefficient. For example, if the true underlying technology is not convex then an analysis that interprets deviations from a convexified frontier as inefficiencies will obviously incorrectly label a number of observations as inefficient.

2. METHODOLOGY

In this section we first review Varian's (1984, 1990) WACM condition to test for consistency with cost minimizing behavior, and his proposal to interpret non-parametric efficiency values as goodness-of-fit indicators with respect to this assumption (subsection 2.1). We then extend the methodology to cover public sector behavior in a multiple output, shadow price framework (subsection 2.2); the basic model imposes no restrictions on relative shadow input prices and on the relative valuation of outputs. The procedure to introduce environmental variables, i.e., factors that are outside the control of the decision maker but do influence the production process, is presented in subsection 2.3. Finally, some extensions to the use of a priori information on the relative valuation of outputs is presented in subsection 2.4.

2.1 THE WACM AND GOODNESS-OF-FIT

Varian's (1984) original WACM condition pertains to a setting where an input vector $x \in \Re^l_+$ is used for the production of a single output $y \in \Re_+$. All technologically feasible input-output combinations are contained within the production possibility set $T \equiv \{(x,y) \in \Re^{l+1}_+ \mid x \text{ can produce } y\}$, which is equivalently expressed in terms of input requirement sets $TI(y) \equiv \{x \in \Re^l_+ \mid (x,y) \in T)\}$. In order to verify whether each observed production unit provides its total output at minimal cost, it is necessary to value the physical inputs in cost terms. Given T (or $TI(y)$) and an input price vector $p \in \Re^l_+$, the minimal cost associated with a particular output can be denoted as $C^T(y;p) \equiv \{px \mid x \in TI(y)\}$.

Under complete information, consistency with cost minimizing behavior of observed productive activities is readily tested. Consider a set $S \subseteq \Re^{l+1}_+$ of observed input-output vectors that are subject to the production possibility set T. The set S fully complies with the WACM if

$$\forall (x,y) \in S: \quad px = C^T(y;p) \tag{1}$$

Using $px > 0$, the proposal is to directly measure the degree of consistency with the WACM at the level of each individual production unit by means of the "efficiency measure" :

$$\varphi\ (x,y;p) \equiv \frac{C\ (y;p)}{px};\ 1 \geq \varphi^T(x,y;p) \geq 0. \tag{2}$$

Interpretation is obvious: the measure $\varphi^T(x,y;p)$ reveals to what extent each observation $(x,y) \in S$ contributes to consistency of S with the WACM. It tells us how close production behavior is to optimizing behavior, where closeness has a direct economic meaning. For example, suppose that $\varphi^T(x,y;p) < 1$. Then the measure reveals by how much actual cost should be reduced for a given output vector to be cost minimizing.

Of course, the possibility set T is typically unknown and, hence, the WACM cannot be tested directly. The non-parametric approach suggests to approximate the theoretical set T by the observed set S, and to evaluate each individual observation relative to this set. This yields the empirical WACM condition

$$\forall (x,y) \in S: \quad px = C^S(y;p). \tag{3}$$

Under the assumption of free output disposal (any output reduction remains producible with no change of inputs, see Färe et al. (1985) for a general definition), this empirical condition is checked by means of the cost efficiency measure:

$$\rho^S(x,y;p) \equiv \frac{1}{px}\left(\min_{x'}\left\{px' \middle| (x',y') \in S \wedge y' \geq y\right\}\right), \tag{4}$$

where the reference (minimal) cost level is computed over the input vectors x' that are observed in combination with an output y' that equals at least y. By construction, $1 \geq \rho^S(x,y;p) \geq 0$. In addition, $\rho^S(x,y;p) \geq \varphi^T(x,y;p)$ if $S \subseteq T$ and outputs are freely disposable; i.e., $\rho^S(x,y;p) = 1$ is a necessary condition for $\varphi^T(x,y;p) = 1$.

The interpretation of $\rho^S(x,y;p)$ is analogous to that of its notional counterpart $\varphi^T(x,y;p)$. As stressed by Varian (1990) and Färe and Grosskopf (1995), it can be considered as an *empirical goodness-of-fit* measure for the cost minimization hypothesis. Note that goodness-of-fit values below unity have been given (combinations of) at least two different interpretations which, unfortunately, cannot be separately identified in empirical work:

- *Imperfect programming*: First, using Afriat's (1973) terminology, it can indicate that the production unit choosing $(x, y) \in T$ is inefficient in the sense of imperfect programming: even though the agency pursues cost minimization for the given output value, some inefficiency results. In a public sector setting, imperfect programming may be the result of having incomplete insight into the production possibilities or it may reflect uncertainty in the decision making process. Alternatively, $\rho^S(x, y; p) < 1$ may be due to monitoring problems in principal-agent relations.
- *Data problems*: Second, $\rho^S(x, y; p) < 1$ may reflect data problems (assuming perfect production programming). For example, measurement problems may cause the implicit actual cost to be overestimated and/or the reference (minimal) cost to be underestimated (i.e., $S \not\subset T$), while production behavior is effectively cost minimizing for the given output value (i.e., $\varphi^T(x, y; p) = 1$). In addition, $\rho^S(x, y; p) < 1$ may be due to the omission of relevant input and/or output variables in the empirical analysis (even though for the true but unobserved technology it is the case that $\varphi^{T,V}(x, y; p) = 1$).[3]

Summarizing, the measure in (4) gives an idea about the extent to which observed behavior can be considered as cost minimizing behavior. It allows one to test the null hypothesis of consistency of production behavior with the WACM by means of a nonparametric gauge with a straightforward economic interpretation.

2.2. Testing for optimising behaviour in the public sector: the basic model

Let us now focus on the extension of the above ideas to test for optimizing behavior in the public sector. As argued in the introduction, the behavior of public sector managers may be implicitly guided by unobservable shadow prices rather than (possibly observed) market prices. Moreover, we focus on direct C-output production which is typically multidimensional in nature.

First, consider the fact that behavior may be based on unobservable and observation-specific shadow input prices. This implies that the relevant price vector p in (4) is not observed. Given the complete absence of information on shadow input prices, we propose to use "most favorable" input prices to test for possible optimizing

[3] Equivalently, $\rho^S(x, y; p) < 1$ can reveal that different input-output combinations are subject to different technological constraints; i.e., the possibility set T differs over observations within the set S for the given input-output selection, which essentially boils down to omitted input-output dimensions.

behavior in the public sector. For each observation it is checked whether there exists at least one set of shadow input prices that rationalizes observed behavior as the outcome of a cost minimization process at these prices. If no such price vector can be identified, the deviation from cost minimizing behavior is evaluated using the most favorable set of prices, i.e., the shadow input price vector that brings the unit closest to cost minimizing behavior. Note that the suggested procedure is consistent with a search for necessary conditions for cost minimization; it leaves observations the benefit of the doubt by choosing the price vector most favorable with respect to this hypothesis.

To implement the described methodology, first note that (4) is equivalently expressed as:

$$\rho^S(x, y; p) \equiv \min_{(x', y') \in S} \left\{ \frac{px'}{px} \middle| y' \geq y \right\}$$

Evaluation on the basis of most favorable shadow prices is then obtained by formulating the efficiency measure:

$$\sigma^S(x, y) \equiv \max_{p \in \Re_+^l : px > 0} \min_{(x', y') \in S} \left\{ \frac{px'}{px} \middle| y' \geq y \right\}, \tag{5}$$

where the selection of most favorable prices is reflected in the max operator. It is easily verified that $1 \geq \sigma^S(x, y) \geq \rho^S(x, y; p)$.[4]

Second, consider the problem of multidimensional C-outputs and their valuation by decision makers. In other words, unlike in (5), y is a multidimensional vector; i.e., $y \in \Re_+^m$ with $m > 1$; moreover, the relative importance of changes in the various outputs to the decision maker may differ substantially. For example, those responsible for local police operations may be interested in reducing both property crime and accidents, but their relative valuation of marginal reductions in these outputs may be quite different. In general, the decision maker's trade-offs between the various outputs could be captured by an *output valuation function* $V : \Re_+^m \to \Re_+$, which associates an overall output value *V(y)* with the *m*-dimensional vector *y*. The

[4] For efficient production units, the shadow price vector reveals the implicit monetary trade-off between inputs that makes production efficient relative to the reference production set as approximated by *S*. For such observations there may in fact be multiple optimal solutions (i.e. ranges of shadow prices), but this aspect is not important for the analysis to follow.

corresponding measure to test public sector consistency with cost minimizing behavior (WACM) is then defined as[5]:

$$\sigma^S(x,y) \equiv \max_{p\in\Re^l_+ : px>0} \min_{(x',y')\in S} \left\{ \frac{px'}{px} \middle| V(y') \geq V(y) \right\}. \tag{6}$$

Unfortunately, of course, the valuation function V is not known, and to presuppose a particular functional form for V is obviously inconsistent with the nonparametric approach. We therefore proceed axiomatically by imposing some minimal structure on V.[6] Although additional structure could be imposed (see subsection 2.4 below), here we only assume that $V(y)$ is monotonically increasing in outputs; i.e., $y' \geq y$ implies $V(y') \geq V(y)$. The assumption "more is never valued worse" seems quite plausible in the case of public services such as safety, educational achievements, recreational facilities, etc. Formally, letting M denote the set of such monotone functions, we assume $V \in \mathrm{M}$; this gives the following goodness-of-fit measure, which is just the multidimensional counterpart of (5):

$$\theta^S(x,y) \equiv \max_{p\in\Re^l_+ : px>0} \min_{(x',y')\in S} \left\{ \frac{px'}{px} \middle| y' \geq y \right\}, \tag{7}$$

with $\theta^S(x,y) \geq \sigma^S(x,y)$ for $V \in \mathrm{M}$. Expression (7) then provides an appropriate measure to test for consistency with the WACM.

In the computations of our empirical application, we will use the linear programming version of (7), which can be expressed as (using $px > 0$)

$$\theta^S(x,y) = \max_{p\in\Re^l_+} \left\{ u \middle| px = 1;\ u \leq px' \quad \forall (x',y') \in S : y' \geq y \right\}, \tag{8}$$

where u denotes the minimal cost level for given endogenously selected shadow prices. The normalization $px = 1$ evidently does not affect the value of $\theta^S(x,y)$. In practice, one first identifies the set $D^S(y;z) \equiv \{(x',y') \in S \mid y' \geq y\}$ (involving a straightforward vector dominance check) after which (8) can be applied.

---

[5] Note the subtle difference between the conditions $y' \geq y$ in (5) and $V(y') \geq V(y)$ in (6). The former is a purely technological condition (viz. free output disposability), the latter captures the implicit valuation of outputs by decision makers.

[6] In principle, an alternative would have been to make the valuation function operational by, again, sets of unobservable shadow (output) prices. Importantly, this is complicated by the fact that there is no reason why overall output value should, in general, be linear in outputs. In other words, the shadow prices may themselves depend on outputs.

2.3. INCORPORATING EXOGENOUS ENVIRONMENTAL FACTORS

In this subsection, we briefly turn to the issue that C-output production may depend on local environmental characteristics for which public decision makers cannot be held accountable. More specifically, we use the procedure set out by Ruggiero (1996a,b) to integrate such dependencies into our model. The principle is as follows. Let *z* denote a vector of community characteristics which, together with direct inputs *x*, shape the final C-outputs. Public production can then be characterized by implicit production possibilities sets, which depend on the value of *z*, i.e. $T(z) \equiv \{(x,y) \in \Re_+^{l+m} \mid x \text{ can produce } \; y \text{ for given } z\}$. In principle, one could then immediately test the WACM for a given, identical *z* (i.e., holding z constant).

It is obvious, however, that in many applications few observations face exactly the same environment, which makes it problematic to implement the above procedure in practice. Ruggiero's (1996a,b) proposal is therefore to focus on the case where $z \leq z'$ can be meaningfully interpreted as *z'* representing a relatively more favorable environment to produce *y*. In that case, it is reasonable to argue that $T(z) \subseteq T(z')$. This in turn implies that the performance of an observation can be assessed by reference to other observations facing an environment that is "at least as harsh". Using this assumption then leads to the following environment-corrected version of the above WACM measure (7):

$$\theta^{S*}(x,y;z) \equiv \max_{p \in \Re_+^l : px > 0} \min_{(x',y',z') \in S*} \left\{ \frac{px'}{px} \middle| y' \geq y; z' \leq z \right\}, \tag{9}$$

where we use the observed set $S* = \{(x',y',z') \mid (x',y') \in S \text{ for given } z'\}$ as the environment-adjusted version of the original set *S*. Measure (9) thus computes cost efficiency with respect to the conditional reference set $\{(x',y') \in S \mid x' \text{ produces } y' \text{ for given } z' \leq z\}$ rather than the unconditional set S (which is employed in (8)). The linear programming version of (9) is then expressed as

$$\theta^{S*}(x,y;z) = \max_{p \in \Re_+^l} \{ u \mid px = 1; u \leq px' \quad \forall (x',y',z') \in S* : y' \geq y; z' \leq z \}, \tag{10}$$

which has a parallel interpretation as (8).

2.4. Including a priori information: potential refinements

The methodology presented so far imposes very minimal prior information regarding the trade-offs between the different inputs and outputs. In some applications, however, additional information regarding these trade-offs may be available, allowing one to impose more structure on the relative shadow prices or output valuations. On the input side, for example, some studies have restricted the shadow price vector $p \in P \subset \Re^{l}_{+}$ by a priori imposing the condition that jobs with greater responsibility are associated with higher shadow wages, or to reflect the condition that the same jobs should everywhere be awarded (at least approximately) the same salaries[7]. Similarly, a priori constraints on the relative valuation between outputs may be imposed if there are good reasons to do so. It has recently been recommended implementing such information in nonparametric assessments whenever relevant (see, e.g., Thanassoulis et al. (2004)).

In our application, substantial differences between police departments in the unobservable shadow prices can be reasonably expected. Even observable salaries differ in a non-negligible way between police personnel in larger cities and their colleagues employed in less densely populated municipalities; moreover, other shadow price variations are likely in view of differences in the composition of the police force, population structure, etc. The basic empirical model therefore does not restrict the shadow input prices at all. In one alternative specification of the model, however, we did test the appropriateness of labor input aggregation procedures typically used in the literature. This boils down to restricting relative shadow input prices for various labor categories (see section 4).

As explained in more detail below, our data set does include ordinal information that can be used to impose some additional structure on the trade-offs of

[7] See, e.g., Kuosmanen and Post (2001) and Kuosmanen et al. (2004). The former study contains an in-depth discussion regarding specifications of the price set P that preserve the linear programming structure of (8). Essentially, they make explicit the shadow price interpretation of 'weight restrictions' used in the context of nonparametric efficiency evaluation (often under the label 'Data Envelopment Analysis' (DEA)).

local policy makers between different outputs[8]. To see how such information can be incorporated, suppose that an output A is valued higher than output B in the sense that $\partial V / \partial y^A > \partial V / \partial y^B$ holds everywhere, with $y = (y^1, ..., y^m)$ and $A, B \in \{1,..,m\}$. In that case, one unit more of output A may always compensate for one unit less of output B, but not *vice versa*. To implement such an ordering in the design of our goodness of fit tests, we redefine the output vector y such that the output y^B is replaced by the sum $y^A + y^B$ (while maintaining the original output y^A). In other words, we impose that a sufficient condition for achieving a higher output valuation level consists in generating at least the same value for $y^A + y^B$ *and* at least the same value for y^A. Note that this suggested procedure implicitly assigns a higher marginal valuation to the output A in a way which is consistent with a focus on necessary efficiency conditions[9]. We will illustrate the use of this type of ordinal information in our empirical application.

3. Application to local public safety: inputs, outputs and environmental characteristics

In the following sections we apply the methodology to assess the consistency of observable police activities in Belgium with cost minimizing behavior at unobservable shadow prices, using non-parametric efficiency measures as goodness of fit indicators. In this section, we first briefly relate the analysis of this paper to the recent literature on performance measurement of police work, and then proceed to a discussion of the data used in the empirical analysis.

As argued in the introduction, a substantial literature exists on evaluating the efficiency of police departments (see, among many others, Thanassoulis (1995),

[8] Specifically, it turns out that policy makers value reductions in violent crime consistently different from reductions in non-violent crime. See Section 4 below.

[9] Consider by way of illustration the following numerical example. Suppose we want to evaluate a particular public agency E that produces 5 units of output A and 5 units of output B, i.e. (y^A, y^B) = $(5,5)$. Next, let there be two possible reference partners $R1$ and $R2$; for $R1$ the combination (y^A, y^B) = $(6,4)$ while for $R2$ we have (y^A, y^B) = $(4,6)$. The procedure we described above implies that $R1$ yields a higher output valuation than E under our assumptions. On the contrary, $R2$ does not because it produces less of output A. Output A is implicitly valued more highly than B. Of course, note that several extensions of this ordinal output weighting procedure are conceivable. For example, an analogous procedure may be followed for introducing a multi-layer ordinal structure (e.g., A is valued more than B and B is more important than C).

Drake and Simper (2003), Van Tulder (2000)). These studies typically calculate inefficiency scores assuming a particular behavioral model (e.g., cost minimization, output maximization, etc.). Moreover, the aggregation of labor inputs typically employed implies strong restrictions regarding the relative prices of different types of labor. Finally, in the few cases where multiple C-outputs have been used, quite stringent assumptions have been imposed on the relative valuation of these outputs[10]. Not surprisingly, therefore, despite some useful insights that have been obtained from these studies, this efficiency approach to evaluate police performance has been subject to some critique (see, in particular, Stone's (2002) comments on Spottiswoode's (2000) report on police force efficiency in the UK). One argument is that efficiency measurement boils down to a normative application of optimizing models, for which the underlying assumptions (e.g., assuming cost minimizing behavior) may not be appropriate. Furthermore, the input aggregation as commonly applied may be misleading, and the restrictions imposed on relative output valuations are difficult to justify.

In this paper, we therefore take a positive view towards non-parametric analysis of police production. Rather than imposing a particular behavioral model, we test the consistency of observed police activities with an appropriately defined cost minimization model. Specifically, our basic model tests the goodness-of-fit of the assumption of cost minimization at unobservable shadow input prices, given a monotone output valuation function. Unlike the efficiency literature, we impose only very weak restrictions on the trade-offs between inputs (relative shadow prices) and between marginal output valuations.

Our police data are mostly taken from a rich collection of statistics gathered by the Belgian Interior Ministry for the year 2000. Our cross-section sample consists of 546 observations. Since there are 589 Belgian municipalities, less then 8% had to be excluded from the original sample due to lack of data. Note that the data refer to the period just prior to a huge consolidation operation (2002-2003) in which the Belgian

---

[10] For example, Van Tulder (1994) has used imputed output prices on the basis of the average sentences imposed by judicial authorities for certain crime types. Other studies based output prices on the value of stolen goods (Darrough and Heineke, 1979). Recently, several authors advocate a subjective method where police output "prices" are derived from public or expert opinion (see, e.g., Carr-Hill (2000) and Stone (2002)). We are, however, not aware of published empirical applications that effectively use this 'subjective' approach.

police structure was profoundly reshaped: police operations were reorganised into 195 local police zones that replaced the previous 'municipal' organisation[11].

We use four variables to describe the C-outputs. Treating crime indicators and accidents as 'bads', the inverse of the following variables is used:

(i) local traffic accidents[12],

(ii) non-violent property crimes and extortion,

(iii) violent crimes, and

(iv) all other reported crimes.

Note that this choice of outputs is neither very novel nor uncontested. Three types of critique on using this kind of output measures have been reported in the literature. First, it has been argued that the type of output indicators we use should preferably be regarded as exogenous outputs in police production (see, e.g., Van Tulder (1994, 2000)), or even as inputs (see Thanassoulis (1995)). Second, it has been claimed that these indicators do not capture all offences that are effectively committed (i.e., the so-called dark-number problem, as recently discussed by MacDonald (2002)). Third, and most importantly, several authors have pointed at the possibility that the recorded crime-rate goes up following an increase in police inputs, creating a perverse link between resources and the outputs used to assess public safety. For example, Cameron (1988) extensively documents the empirical lack of support for a negative relationship between the size of police forces and crime rates. Similarly, Schwab and Zampelli (1987) mention this perverse relationship as one of the prime reasons for not using such output measures when estimating the characteristics of the production technology or income and price effects of the demand for public safety. More recent research, carefully controlling for the problem of endogenous police inputs (i.e., the latter may be high just because there is a high crime rate) did find evidence for a negative relation between police inputs and the crime rate; see, e.g., Levitt (1997). It is clear, however, that the issue remains unsettled.

Although the above arguments have quite some merit, the key motivation for using the above listed outputs for our empirical tests is threefold. First, as argued in

[11] To the extent that they are available, more recent data may be less suited for the purposes of the current paper in view of possible transition effects associated with the move to a completely new institutional environment.

[12] 'Local traffic accidents' refers to accidents with personal injuries on local (non-highway) roads.

the introduction, C-outputs rather than direct outputs do form the major basis for public accountability with respect to public safety. Indeed, crime and accident statistics are among the most important indicators underlying intensive public debate, as witnessed by the media attention and political debate which surrounds their publication. Second, in view of the positive and explanatory nature of our tests, it seems hardly sustainable that police officials aiming at crime-reduction in effect base allocative decisions on *unreported* incidents.[13] In a similar vein, it seems ill-advised to make use of *cleared* crimes. In Belgium as well as in most other countries, this particular measure heavily relies on the behavior of actors further down the judicial chain, and it is therefore not well-suited to evaluate local police work as such. Third, by focusing on possible deviations from cost minimizing behavior, the emphasis in this paper is clearly on allocative decisions of local police managers on the input side, conditional on output levels. To the extent that policy makers do base their input allocation decisions on unreported crimes, one expects this to result in a poor goodness-of-fit relative to the shadow cost minimization hypothesis.

Turning to the input side, an interesting feature of our data set is that it allows us to distinguish between different labor allocation categories. For each local police department, the Belgian Interior Ministry collected detailed statistics on personnel allocation over different tasks. We were thus able to distinguish labor allocated to:

(i) community policing,

(ii) intervention squads,

(iii) victim aid,

(iv) criminal investigation, and

(v) administrative/managerial services.

In each of these tasks, we only retained policemen or civilians that were assigned to this specific task on at least a 0.8 full-time basis. The remainder of ("non-specialized") labor inputs was grouped into:

(vi) a residual labor allocation category.

[13] As Carr-Hill (2000) observes, even countries that emphasize zero tolerance still base their policy on reported incidents. He correspondingly argues that survey information is more appropriate than reported crime figures only if public confidence in the police, rather than crime reduction, is the relevant objective of the analysis. .

Finally, these labor data were complemented with information on:

(vii) the total hours per week that the local police unit could be contacted.

Although deviating from most earlier empirical studies on police activities, disaggregating the labor input along these functional lines is extremely valuable to grasp realistic allocative decisions of police managers. What they have to decide upon is not so much the optimal use of capital versus labor, but rather how to allocate personnel over the different functional categories so as to contribute as much as possible to the safety objectives they have in mind. This yields a somewhat richer approach than typically in the empirical literature. If not lumped together, "uniformed" and "civilian" personnel are the standard two labor categories often considered (e.g., Grosskopf et al. (1995), Van Tulder (2000), Drake and Simper (2002)). In contrast, note that we do not consider capital inputs in the analysis below, as there is an outspoken proportional relationship in local police units between labor and capital equipment indicators in our sample.[14]

Finally, we included several variables that are important in shaping the production possibilities of local police services but that are outside the control of local decision makers; these are the z-variables discussed in section 2.3. First, an interesting and novel feature of our data set is that information is included on the inputs provided by the state police organism (the so-called *Rijkswacht/Gendarmerie*) as a complimentary public provider of local safety[15]. To the best of our knowledge, the fact that in most countries several police forces co-exist and influence public safety simultaneously has not yet been accounted for in an empirical analysis. To some extent this can be attributed to an evident non-overlapping jurisdiction problem, which implies that state or federal police resource data are typically not itemized on the municipal level. But in our particular sample –a cross-section snapshot on the eve of a merger of Belgian municipal police forces with decentralized state police

---

[14] For example, the sample correlation of total police personnel with the total number of vehicles in a local police force, an often used capital measure, amounts to 0.96. For another commonly used capital measure, available PC's, these figures are very similar. It is a well-known feature of the nonparametric goodness-of-fit measures used in this paper that their value is hardly affected when introducing an additional input that correlates almost perfectly with another input that was already taken up in the evaluation model.

[15] These have been integrated with local police forces in the recent reorganization. In 2000, however, they co-existed with local police; although their activities extended beyond providing local safety services, they did provide important inputs.

brigades– we do have the appropriate data. Specifically, we use for each municipality the state police personnel figures that the Belgian Ministry of Interior calculated for the year 2000. Our use of the input by state police forces as an exogenous input in the analysis implies that any municipal force is only compared, ceteris paribus, with other observations that face an input environment which is "at least as harsh", i.e., with observations that have at most the same level of state police personnel. Importantly, consistent with the idea that these inputs are exogenous, the figures are not additionally used to calculate the goodness-of fit value of an observation.

Second, several other variables have been considered for possible inclusion as exogenous z-variables. In view of the procedure to incorporate exogenous factors, as explained in section 2.3, note the importance of a monotone relation of potential z-variables with cost efficiency. Recently, Daraio and Simar (2003) developed a probabilistic approach to test for the presence of such monotone relationships. Application of this methodology provided a strong corroboration of the hypothesis of a monotone relation between population size and cost efficiency[16]. However, the Daraio-Simar procedure did not additionally reveal similarly obvious patterns for other frequently considered environmental factors such as median income, population density or municipality area size. Therefore, only population was included in the empirical analysis[17].

In Table 1 we provide an overview of the variables selected for the analysis, together with some summary statistics describing sample characteristics. To facilitate the interpretation, we report the original crime and accident figures; as argued before, we use their inverse as output measures in the analysis. Moreover, as the distributions of the input and output data are quite skewed for almost all variables, we also report averages for a number of subgroups in our sample such as the large cities, regional cities and small rural municipalities. Clearly, the major cities are outliers in the sample. Further observe the low average values for some of the labor input categories, such as victim aid and criminal investigation. It is not unusual for smaller

[16] For the sake of brevity, we do not include these results in this paper, but they are available from the authors upon request.

[17] One could argue whether this particular dimension is an exogenous output –it has been used as a proxy for heterogeneous police outputs in older studies by Hirsch (1959) or Popp and Seebold (1972)–, or rather an environmental variable that captures the relative ease by which a higher (absolute) safety level can be achieved in a municipality. While we do prefer the second interpretation, we note that both are similarly reconcilable with Ruggiero's (1996a,b) methodological framework.

municipalities not to assign close to full time personnel to these categories; moreover, at the time the data were collected by the Ministry of the Interior, aid to victims was a relatively new full-time and distinct activity. Finally, note that the relation between the presence of state police and the size of municipalities is not monotonically declining: state police input per 1000 inhabitants is higher in both the large cities and small rural municipalities as compared to regional cities. One plausible explanation for this phenomenon is that state police was used to guarantee a minimum acceptable scale in rural (large surface area) municipalities.

	Sample Average	*Coeff. Of var.*	*75th Percentile*	*Average For Big Cities*	*Average for Regional Cities*	*Average for Small Rural Municipalities*
	(n=546)	*(n=546)*	*(n=546)*	*(n=5)*	*(n=17)*	*(n=165)*
Accident and Crime rates						
Traffic accidents	0.078	0.665	0.083	0.736	0.441	0.025
Non-violent property	0.761	0.305	0.567	23.356	3.734	0.125
Violent	0.272	0.460	0.264	5.422	1.407	0.073
All other crimes	0.472	0.446	0.462	9.793	2.177	0.134
Controllable inputs						
Labor Inputs per category						
Community policing	0.142	0.538	0.212	1.514	0.338	0.088
Intervention squads	0.231	0.590	0.430	1.060	0.919	0.021
Victim aid	0.008	0.303	0	0.028	0.018	0.008
Criminal investigation	0.038	0.288	0	0.313	0.124	0.003
Administrative/Staff	0.134	1.070	0.184	0.530	0.186	0.106
Non-specialized	0.821	1.604	1.018	2.777	1.219	0.722
Other						
Contact hours (*)	52	0.63	50	718	115	35
Exogenous inputs						
State police personnel	0.823	1.948	0.981	1.022	0.766	1.102
Other						
Population (**)	18 126	0.642	19 400	237 886	71 152	6 226

Note: all average and 75-th percentile figures per 1000 inhabitants, except (*: hours per week) and (**). See footnote 11 for the definition of traffic accident data. The population data are those reported by the National Institute of Statistics. All other data were provided by the Belgian Interior Ministry.

4. EMPIRICAL RESULTS

In this section we present empirical results based on the methodology discussed in section 2. In the remainder of this section, we discuss the goodness of fit results for three alternative specifications of the model. We also compare the statistical significance of the differences in goodness of fit measures obtained for the various model specifications. Indeed, Varian (1990) has convincingly argued that the full distribution of the efficiency values should be studied in practical applications, claiming that “the pattern of violations [of the efficiency conditions] may tell us a lot about what is going on in the data” (p. 131). The three alternative specifications of the model differ from one another in terms of the restrictions imposed on the shadow input prices and the relative valuations of the outputs:

(a) The basic model (denoted *Cost Efficiency (CE)*) determines the cost efficiency measures calculated on the basis of all inputs and outputs discussed in section 3; it also includes the presence of state police forces and population as exogenous variables, following the Ruggiero (1996a,b) procedure discussed in section 2.3.

(b) A second version of the model imposes restrictions on the relative shadow input prices. As mentioned in section 3, it is common practice in the existing literature to use an aggregate labor input for analyzing the productive efficiency of police departments. In terms of our model, such a procedure implies the assumption that the shadow price of personnel is independent of its task specialization. To check the sensitivity of the goodness-of-fit results with respect to this additional assumption, we consider a model (denoted *Inputs Weighted* (*IW*)) that considers only two inputs: the first one is the aggregate of the original labor inputs (i)-(vi), the second one is the original input (vii); the latter cannot be taken up in the aggregation procedure as it is expressed in a different measurement unit (i.e., opening hours instead of full-time equivalents).

(c) A third model specification introduces some available (ordinal) a priori information regarding the relative marginal valuation of outputs by local policy makers. This model is denoted *OW (Outputs Weighted*); it follows the methodology explained in section 2.4. The a priori information on output valuation is taken from a nationwide survey conducted in 2000 by the Belgian Ministry of the Interior. A consistent finding in this survey is that citizens rank traffic accidents and burglary (rather common offenses, so that the probability of citizens to become a victim sooner or later is high) substantially higher on their priority list than violent crime (which is, fortunately, still rather uncommon). Of course, this does not imply that people are less upset by violent crime, when it occurs, than by traffic accidents. It only means that their overall day to day safety feeling is less affected by types of offenses with which they are not, or very infrequently, confronted[18]. To see to what extent the use of this a priori information, which is available to local policy makers so that they can respond to it, affects the goodness of fit results, we impose the constraint on the model that "traffic accidents" and "property crimes" (outputs (i) and (ii)) are more problematic than "violent crimes" (output (iii)). The procedure to incorporate this type of ordinal information was explained in section 2.4.

We now turn to the empirical results. Table 2 provides summary statistics for the three distributions (*CE*, *OW* and *IW*) obtained for the alternative model specifications. The table first gives information on the goodness-of-fit (efficiency) values calculated for the various models, and it reports summary information on the number of comparison partners. This refers to the number of police departments that dominate an evaluated department, in the sense of generating more outputs with fewer inputs in an environment that is at least equally harsh (see the discussion of (8) and (10) in section 2). Next, the lower part of Table 2 gives for each of the three models the number of observations (on a total of 546) that pass certain threshold efficiency values.

[18] The survey ('Veiligheidsmonitor') asked people to rank various types of crime and offenses on a scale from 'considered not very problematic at all' to 'considered very problematic'. Although there were of course differences in citizens' ranking of different safety threats over various regions and types of municipalities, the lesser importance of violent crimes relative to traffic offences and burglary was a consistent finding. Despite this observation, the *OW*-model should be considered as illustrative at best, if only because the crime categories used in the safety survey (traffic offenses, burglary, etc.), although obviously closely related in spirit, are not exactly the same as the ones we discern in our empirical analysis (traffic accidents, property crimes, etc.).

Table 2: efficiency distributions

	Cost Efficiency (CE)		Outputs Weighted (OW)		Inputs Weighted (IW)	
	Efficiency Scores	*Comparison partners*	*Efficiency scores*	*Comparison partners.*	*Efficiency scores*	*Comparison Partners.*
Average	95.90%	10.60	94.87%	7.14	77.07%	10.60
standard deviation	11.17%	24.79	12.05%	18.30	31.14%	24.79
maximum	100.00%	223	100.00%	197	100.00%	223
Minimum	30.41%	0	30.41%	0	7.14%	0
Threshold efficiency values…	*percentage*	*number*	*percentage*	*Number*	*percentage*	*number*
55%	97.99%	535	97.62%	533	94.51%	516
60%	97.07%	530	96.70%	528	92.31%	504
65%	95.79%	523	95.24%	520	89.19%	487
70%	94.87%	518	93.41%	510	86.45%	472
75%	94.14%	514	92.49%	505	82.23%	449
80%	92.12%	503	89.93%	491	78.39%	428
85%	88.46%	483	85.53%	467	73.99%	404
90%	86.45%	472	82.23%	449	69.60%	380
95%	84.07%	459	79.67%	435	65.38%	357
100%	81.68%	446	76.37%	417	63.55%	347

Note: "percentage" stands for the percentage of observations that have an efficiency value that equals at least the value in the left column and "number" stands for the corresponding absolute number of observations.

A first observation is that the basic *CE* model generally provides a good fit of the data. On average, the results imply small deviations from cost minimizing behavior; average scores amounts to almost 96%, and the standard deviation is fairly small. Moreover, for more than 80% of all observations a set of shadow input prices existed that induced consistency with cost minimizing behavior: indeed, some 81.68% of the observations attain a calculated efficiency value of 100%. Reassuringly, it is found that the efficiency scores have been computed with reference to on average some 11 (viz. 10.6) comparison partners. This allows us to be reasonably confident in the favorable goodness-of-fit results, as they can hardly be attributed to the systematic presence of a small number of comparison partners; the latter would signal low power of the WACM tests. A first conclusion is, therefore, that cost minimization at unobservable shadow prices is a behavioral model that has good explanatory power for the observed task allocation within Belgian police departments.

Second, however, a fair number of local police services shows quite substantial deviations from cost minimizing behavior: for example, we find that the minimum efficiency value amounts to no more than some 30%. Table 2 also reveals that 16 out of the 546 observations have an efficiency value below 60 %. Since we focus on the overall validity of the public sector variant of the WACM model, disentangling the relative importance of alternative interpretations (pure inefficiency versus data measurement problems; see section 2.1) and actually explaining these low scores for individual observations is not pursued in the current study. Still, given the very weak conditions imposed on the model, for example the use of observation specific and most favorable shadow input prices, the large deviations from cost minimizing behavior for some police departments are a remarkable finding.

In a next step we compare the results for the *CE* model with those for the alternative *OW* and *IW* models. In general, Table 2 indicates that both these alternative model formulations yield lower goodness-of-fit values than the basic *CE* model. In a sense, of course, this is not that surprising, because the *OW* and *IW* models impose additional structure on the behavioral model. Moreover, it is well known that efficiency measurement is not totally insensitive to the number of dimensions, and the *IW* model does reduce the number of dimensions by aggregating inputs.

Still, the differences between the models are relevant from the goodness-of-fit perspective we take in this paper. Importantly, the results for the *IW* model suggest that aggregating labor inputs does strongly affect the goodness-of-fit with respect to the cost minimization hypothesis. For example, average efficiency amounts to only 77%, and the *IW* model finds only some 63% of all observations in line with the WACM hypothesis. The model does not generally provide a good fit to the data. We previously argued that, from a managerial perspective, there are good reasons for a functional decomposition of the labor inputs, because this is more in line with realistic allocation decisions by local police departments than the aggregate inputs typically employed in empirical studies. The differences in results between the basic *CE* model, based on functional disaggregation, and the *IW* model, reflecting aggregation of labor, suggest that the latter performs much worse as an explanatory model of the behavior of police departments.

In contrast, the figures reported for the *OW* model indicate that the inclusion of the additional assumption that police officers are responsive to the (ordinal) information regarding output valuation of citizens turns out to be quite harmless. Average test scores for the *OW* model are only marginally affected, amounting to almost 95%; the same holds for the number of observations deviating substantially from cost minimization. Moreover, the *OW* model still identifies some 76% of the sample as 100% in accordance with the associated WACM conditions.

To conclude, we evaluate the statistical significance of the differences in the efficiency values caused by the alternative model specifications. To do so, we use non-parametric (Kernel-based) tests. We abstract from an in-depth discussion of the test procedure, but refer to Kumar and Russell (2002) for an application in a similar context as the one of this paper[19]. To account for possible sensitivity of our test results with respect to the Kernel bandwidth specification, we have carried out the tests for three alternative bandwidth selections, viz.. 0.01, 0.05 and 0.1.

The formal test results, which are tabulated in Table 3, confirm our earlier impression: while the *CE* distribution does not systematically differ from the *OW* distribution for any reasonable significance level (and for all bandwidth selections), there is a significant difference with the *IW* distribution. In other words, we may indeed expect the usual input aggregation practice (underlying the *IW* model) to distort the efficiency analysis, so possibly leading to ill-justified conclusions. Our results suggest that aggregation of functional labor categories is potentially distorting. As a consequence, we plead for a labor input disaggregation, which results in a much better overall fit of the data. In addition, putting additional structure on the output valuation function, which is based on the limited external information that is available, enhances the statistical power of the nonparametric WACM tests and does an equally convincing job as the basic *CE* model in terms of goodness-of-fit[20].

---

[19] See also Pagan and Ullah (1999, p. 68-69) for a formal introduction to the test procedure. Given our specific setting, we effectively exploit Li's (1996) argument that these Kernel-based tests remain valid for comparing the distributions of dependent variables. The latter is indeed the case here, as the different goodness-of-fit measures are computed with respect to the same sample of observations.

[20] An alternative strategy may consist of imposing additional structure on the production possibilities. For example, many efficiency studies of police departments employ production technology assumptions such as convexity or constant returns-to-scale (see e.g. Färe et al. (1994) for a methodological introduction into the nonparametric modeling of such assumptions). We performed similar tests as those reported in Table 3 under such assumptions. It was found that there was a significant deterioration of the goodness-of-fit results when imposing convexity or constant returns-to-scale. As argued in the introduction, given that such technology properties are usually non-verifiable, we favor the (trade-off-based) strategies discussed in Section 2.

Table 3: test results

		Cost Efficiency	Outputs Weighted	Inputs Weighted
normalized statistics	***CE***	0.000	0.674	10.553*
(bandwidth: 0.01)	**OW**	0.674	0.000	5.781*
	IW	10.553*	5.781*	0.000
normalized statistics	***CE***	0.000	0.703	11.649*
(bandwidth: 0.05)	***OW***	0.703	0.000	6.625*
	IW	11.649*	6.625*	0.000
normalized statistics	***CE***	0.000	0.589	10.571*
(bandwidth: 0.10)	***OW***	0.589	0.000	6.219*
	IW	10.571*	6.219*	0.000

Note: "normalized statistics" stands for the values of the test statistic that should follow a standard normal distribution under the null hypothesis of equal efficiency distributions (see Pagan and Ullah (1999, p. 68-69) for its construction); "*" indicates a significant difference between the efficiency distributions at the 0.01 level.

5. CONCLUSION

We have developed a positive nonparametric model of (local) public sector production that allows us to test whether the outcomes of public sector activities can be rationalized by optimizing behavior. Our extensions essentially accommodate existing tools to a number of characteristic features of managerial decision making in the public sector. First, we take into account that public sector decision makers may well optimize behavior with respect to unobservable shadow input prices rather than (possibly observed) market prices. Second, an output valuation function was introduced which accounts for the typically multidimensional nature of local public sector output, without imposing constraints on the relative marginal valuation of the different outputs by local decision makers. Moreover, the evaluation model focuses on C-output production, i.e., on outputs that are the ultimate concern to citizens. Based on these concepts, a goodness-of-fit measure was introduced that can be used to test for the consistency of observed behavior with cost minimization at unobservable shadow input prices. This can be interpreted as a public sector version of the WACM condition introduced by Varian (1984).

Our application to a cross-section sample of Belgian municipal police forces deviates from the existing literature by drawing on detailed task-allocation data and by controlling for the presence of state police forces. We test for consistency with respect to the cost minimization hypothesis for three different model specifications. The first model does not impose any additional structure on the trade-offs between inputs (on the input shadow prices) or outputs (on the relative marginal output valuations). We found that it provides a good description of the data: the input-output combinations observed for many local police departments are consistent with the appropriately defined cost minimization problem, and average deviations from cost minimizing behavior are small.

Next, we carried out a more refined WACM-test that incorporates more structure on the output valuation function; this additional structure was based on ordinal information provided by a survey of the Belgian Interior Ministry. The resulting evaluation model, which implies a more powerful analysis, obtains equally convincing goodness-of-fit results for the WACM. By contrast, we find that aggregating the personnel input for the different specialization tasks, which basically boils down to adding structure on the input trade-offs, entails a significantly worse fit of the data. This suggests that such an aggregation, which is common practice in efficiency studies of police departments, may *a priori* distort the analysis and, hence, lead to ill-justified conclusions.

References

Afriat, S. (1972), "Efficiency estimation of production functions", *International Economic Review 13*, 568-598.

Afriat, S. (1973), "On a System of Inequalities in Demand Analysis: An Extension of the Classical Method", *International Economic Review, 14*, 460-472.

Atkinson, S.E. and R. Halvorsen (1986), "The relative efficiency of public and private firms in a regulated environment", *Journal of Public Economics 29*, 281-284.

Blundell, R., Browning, M. and I. Crawford (2003), "Nonparametric Engel curves and revealed preferences, *Econometrica 71*, 205-240.

Bös, D. (1986), *Public enterprise economics*, North Holland, Amsterdam.

Bradford, D., R. Malt and W. Oates (1969), "The Rising Cost of Local Public Services: Some Evidence and Reflections", *National Tax Journal* 22, 185-202.

Cameron, S. (1988), "The Economics of Crime Deterrence: A Survey of Theory and Evidence", *Kyklos* 41, 301-323.

Carr-Hill, R. (2000), "Developing a Robust Resource Allocation Formulae for Police", *Policing and Society* 10, 235-2641.

Daraio, C. and Simar, L. (2003), "Introducing Environmental Variables in Nonparametric Frontier Models: a Probabilistic Approach", *Journal of Productivity Analysis*, to appear.

De Borger, B. (1993), "The economic environment and public enterprise behavior", *Economica* 60, 443-463.

Drake, L. and R. Simper (2002) "X-Efficiency and Scale Economies in Policing: A Comparative Study Using the Distribution Free Approach and DEA,", *Applied Economics*, 3,. 1859-1870.

Drake, L. and Simper, R. (2003), "The Measurement of English and Welsh Police Force Efficiency: A Comparison of Distance Function Models', *European Journal of Operational Research*, 147, pp. 165-186

Färe, R. and S. Grosskopf (1995), "Non-Parametric Tests of Regularity, Farrell Efficiency, and Goodness of Fit", *Journal of Econometrics*, 69, 415-425.

Färe, R., S. Grosskopf and C.A.K. Lovell (1994), *Production Frontiers*, Cambridge University Press, Cambridge.

Grosskopf, S., Hayes, K. and J. Hirschberg (1995), "Fiscal stress and the production of public safety: a distance function approach", *Journal of Public Economics 57*, 277-296.

Grossman, P., Mavros, P. and R. Wassmer (1999), "Public sector technical inefficiency in large US cities", *Journal of Urban Economics 46*, 278-299.

Gyiamah-Brempong, K. (1989), “Production of Public Safety: Are Socio-Economic Characteristics of Local Communities Important?”, *Journal of Applied Econometrics*, 4, 57-71.

Hailu, A. and T.S. Veeman (2001), “Non parametric productivity analysis with undesirable outputs: an application to the Canadian pulp and paper industry”, *American Journal of Agricultural Economics 83*, 605-616.

Hanoch, G. and M. Rothschild (1972), “Testing the assumptions of production theory: a nonparametric approach”, *Journal of Political Economy* 80, 265-275.

Hayes, K., Razzolini, L. and L. Ross (1998), “Bureaucratic Choice and nonoptimal provision of public goods: theory and evidence”, *Public Choice 94*, 1-20.

Hirsch, W. (1959), “Expenditure Implications of Metropolitan Growth and Consolidation”, *The Review of Economics and Statistics*, 41, 323-341.

Hughes, A. en S. Yaisawarng (2000), “Efficiency of Local Police Districts: A New South Wales Experience”, in J. Blank, *Public Provision and Performance*, North-Hollmand-Elsevier, Amsterdam.

Kumar, S. and R.R. Russell (2002), “Technical Change, Technological Catch-Up and Capital deepening: Relative Contributions to Growth and Convergence”, *American Economic Review*, 92, 527-548.

Kuosmanen, T., Cherchye, L. and Sipiläinen, T. (2004), ”The Law of One Price in Data Envelopment Analysis: Restricting Weight Flexibility across firms”, *European Journal of Operational Research*, to appear.

Kuosmanen, T., and G.T. Post (2001), “Measuring Economic Efficiency with Incomplete Price Information: With an Application to European Commercial Banks”, *European Journal of Operational Research*, 134, 43-58.

Levitt, S. (1997), “Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime”, *American Economic Review* 87, 270-290.

Li, Q. (1996), “Nonparametric Testing of Closeness between Two Unknown Distribution Functions”, *Econometric Reviews*, 15, pp. 261–274.

MacDonald, Z. (2002), “Official Crime Statistics: Their Use and Interpretation”, *The Economic Journal* 112, 85-106.

Mueller, D.C. (2000), *Public Choice II*, Cambridge University Press, Cambridge.

Pagan, A. and Ullah, A. (1999), *Nonparametric Econometrics*, Cambridge University Press, Cambridge.

Pestieau, P. and Tulkens, H. (1993), “Assessing and Explaining the Performance of Public Enterprises”, *FinanzArchiv*, 50, 293-323.

Popp, D. and Seebold, F. (1972), “Quasi-Returns to Scale in the Provision of Police Services”, *Public Finance* 27, 46-60.

Ruggiero, J. (1996a), “On the Measurement of Technical Efficiency in the Public Sector”, *European Journal of Operational Research* 90, 553-565.

Ruggiero, J. (1996b), “Efficiency of Educational Production: An Analysis of New York School Districts”, *The Review of Economics and Statistics* 78, 499-509.

Schwab, R.M. and Zampelli, E. (1987), “Disentangling the Demand Function from the Production Function for Local Public Services: The case of Public Safety”, *Journal of Public Economics*, 33, 245-260.

Snyder, S.K. (2001), “Testable Restrictions of Pareto Optimal Public Goods Provision”, *Journal of Public Economics*, 71, 97-119.

Spottiswoode, C. (2000), *Improving Police Performance: a new approach to measuring police efficiency*, Public Services Productivity Panel, Her Majesty’s Treasury, Londen.

Stone, M. (2002), “How Not to Measure the Efficiency of Public Services (And How One Might)”, *Journal of the Royal Statistical Society, Series A*, 165, 405-434.

Thanassoulis, E. (1995), “Assessing Police Forces in England and Wales using Data Envelopment Analysis”, *European Journal of Operational Research*, 641-657.

Thanassoulis, E., Portela, M.C. and R. Allen (2004), “Incorporating value judgments in DEA”, in W.W. Cooper, L. Seiford and J. Zhu, *Handbook on Data Envelopment Analysis*, Kluwer Academic Publishers, Dordrecht.

Van Tulder, F.P. (1994), “Law Enforcement, Production and the Decision-Making Process: An Analysis of Dutch Police-Departments”, *Public Finance* 49, 259-276.

Van Tulder, F.P. (2000), “The Revenue Approach to Dutch Police Departments” in J. Blank, *Public Provision and Performance*, North-Holland, Amsterdam, p. 247-275.

Varian (1984), “The non-parametric approach to production analysis”, *Econometrica* 52, 279-297.

Varian (1990), “Goodness-of-fit in optimizing models”, *Journal of Econometrics* 46, 125-140.