

Accepted Manuscript

This is an Accepted Manuscript of the following article:

Rødseth K L. Productivity growth in urban freight transport: An index number approach.

Transport Policy. 56 (May), 2017, 86-95.
0967-070X

The article has been published in final form by Elsevier at

<http://dx.doi.org/10.1016/j.tranpol.2017.02.009>

© 2017. This manuscript version is made available under the CC-BY-NC-ND 4.0 license

<http://creativecommons.org/licenses/by-nc-nd/4.0/>

It is recommended to use the published version for citation.

The final publication is available in: *Transport Policy*. 2017, 56, 86-95. <https://doi.org/10.1016/j.tranpol.2017.02.009>

Productivity Growth in Urban Freight Transport: An Index Number Approach

Abstract: Improvement of operational efficiency is a common goal of most governmental freight transport policies. While productivity and efficiency analysis consequently provides a sound knowledge base, applications to freight transport are scarce. This paper illustrates how axiomatic production theory can be applied to model road freight transport, and proposes a logistics efficiency measure as the function representation. Based thereon, a logistics productivity index that decomposes into technical, cargo mix, vehicle capacity, and efficiency changes is established to determine the rate and drivers of growth. Emphasizing urban logistics, the paper discusses the limited access to reliable data at the micro level and illustrates how local or regional freight transport can be evaluated applying pseudo panel techniques to national freight surveys. Correspondingly, the theoretical productivity index is implemented on a pseudo panel covering the 24 largest cities in Norway between 2008 and 2012, when 12 of them entered a collaboration agreement to promote efficient transport. The results indicate a modest 0.6 percent average productivity growth. Efficiency change is the key driver of growth, countered by technical stagnation and regress. Negative productivity growth is expected if this trend continues. Moreover, the results do not reveal productivity gains from urban agglomeration or membership of the collaboration agreement, suggesting that prevailing transport and land use policies have so far been unable to foster productivity growth in urban freight transport.

Keywords: Urban freight transport; Logistics efficiency measure; Productivity change decompositions; Data Envelopment Analysis

JEL-codes: D24; C43; C61; R40

1. Introduction

On the one hand, urban freight transport is a necessary condition for sustaining urban settlements and for maintaining the urban way of life. On the other, it produces a wide range of external costs such as noise, air pollution, accidents, and congestion. Because of high population densities in urban areas, these external costs are also typically very high¹. With increasing urbanization and transportation, urban freight transport has therefore become an important issue on the political agenda worldwide; see e.g. European Commission (2011).

There are several ways to tackle the negative impacts of urban freight transport, including establishing eco-zones, delivery time restrictions, and vehicle weight restrictions². One of the most promising measures to reduce the negative impacts is to minimize the number of trips required for freight movements (Eidhammer & Jean-Hansen, 2008), i.e., to foster productivity growth and efficiency improvements in urban logistics. This approach to improving the sustainability of urban freight transport, by decoupling the movement of goods from transport activities, is the focus of my paper. More precisely, it develops and decomposes a Logistics Productivity Index (LPI) to identify the rate and drivers of productivity growth, and illustrates empirical implementation of the index in the context of urban freight transport. Thus, it establishes a management tool for freight transport policies in general, but pays special attention to urban freight transport. The availability of data for implementing the index at the city level is discussed.

¹ See Quack (2008) for more on the pros and cons of urban freight transport.

² See Browne et al. (2007) for an overview and discussion of possible political measures to deal with the negative impacts of freight transportation.

Caplice and Sheffi (1994) distinguish between two types of logistics performance measures; productivity measures (i.e., the ratio of outputs to inputs; e.g., goods lifted per truck or trip) and utilization measures (i.e., the capacity used to the total capacity available). As noted by McKinnon (2015a), while ton kilometers per truck per annum have risen steeply in most countries as trucks have increased in size, weight, and power rating, this does not necessarily mean that trucks are on average running fuller than before. Consequently, he advocates the need for a separate set of utilization metrics in addition to productivity measures when assessing the operational efficiency of freight transport. A key objective of this paper is to illustrate that logistics performances need not be deduced from a set of indicators. Instead, an index that comprises productivity and utilization can be established and decomposed to identify their relative importance to intertemporal changes in logistics performances.

This paper illustrates how axiomatic production theory can be applied to model road freight transport, when the number of trips (or vehicle kilometers) and vehicle carrying capacities are modeled as inputs and the tons lifted of various cargo types as outputs. The modeling approach has several merits. First, as noted by McKinnon (2015a), measuring the degree of utilization is a challenging task; for dense commodities, the vehicle weight limit is critical, while for low-density products with high “stackability” the main constraint is cubic capacity. The model proposed in this paper deals with the problem by modeling freight transport as a multi-output production process, in which different cargo types have different input requirements. Second, the model framework allows measuring logistics productivity and efficiency given vehicle capacities, and to disentangle the impact of changing vehicle capacities on productivity. Third, the model framework is adopted from the productivity and efficiency analysis literature, and is thus ideal for benchmarking road freight transport. The

proposed approach identifies best practices from identified practices, as opposed to comparing current practices to theoretical – and perhaps unattainable – maxima (cf., the lading factor)³. Fourth, production analysis is equipped to control for contextual variables that may influence logistics productivity, e.g., urban form (Allen et al., 2012).

Based on the model framework, I propose a LPI that allows assessing intertemporal changes in logistics productivity. This index is preferred to traditional productivity indices such as the Malmquist (1953) index because the LPI is easy accessible to stakeholders in transport by reporting intertemporal changes in goods lifted per trip (or per vehicle kilometer). The index decomposes into frontier shifts and efficiency improvements, where the frontier shift component can be further decomposed into input-output mix and technical changes, and the efficiency component can be decomposed into pure and scale efficiency changes. The LPI thus allows pinpointing the sources of intertemporal changes in logistics productivity, and is consequentially highly useful for evaluating the outcomes of policies aimed to improve urban logistics performances. Frontier-based techniques to measure performances are particularly helpful to competition-based policies that distribute financial support among cities based on their previous efforts to and successes in promoting sustainable freight transportation.

Several previous studies apply index number theory to analyse freight transport. Some examples include Kveiborg and Fosgerau (2007) who use a Divisia index to decompose the relative contributions of economic activities, the composition of commodities, the weight to value ratios, the handling of commodities, and the average load and trip length to the

³ As noted by McKinnon (2015b), under-utilization of vehicle capacity may not be an indicator of inefficiency, but rather an indicator that logistics companies make rational trade-off among transport efficiency and other objectives.

development of road freight traffic and transport in Denmark, Sorrell et al. (2009) who decompose the contributions of eleven key factors including GDP to intertemporal changes in road freight energy use based on the log-mean Divisia index approach, and Alises et al. (2014) who conduct a decomposition analysis to identify the drivers of the evolution of the road freight transport intensities of the United Kingdom and Spain. The approach introduced in this paper differs from these studies by being in the Malmquist (1953) index tradition⁴, using frontier analysis to disentangle technical and efficiency changes. While the author is unaware of previous attempts to evaluate the operational efficiency of freight transport using frontier-methods, they have been employed to assess the productivity and efficiency of transport companies. Crujssen et al. (2010) use Data Envelopment Analysis (DEA) to assess the economic efficiency of 82 Belgium road transportation companies. Heng et al. (2012) account for air pollution emissions when assessing the efficiency and productivity growth of trucking in U.S. states between 2002 and 2005. Zhang et al. (2015) propose a Malmquist CO₂ emission performance index that is used for assessing the dynamic performance of the Chinese regional transportation industry.

While the reviewed literature on index theory focuses on the development of freight transport at the national or sector level, this paper emphasizes urban freight transport. Betanzo-Quezada and Romero (2010) present an urban freight transport index, focusing on the attention of authorities in dealing with freight transport issues within cities. Their index ranks cities against a theoretical benchmark, while the index presented in this paper identifies

⁴ Grifell-Tatjé and Lovell (1996) argue that the main advantages of the Malmquist index over the Törnqvist and Fischer productivity indices are i) its weaker behavioural assumptions, ii) that its computation does not require price data, and iii) that it allows decomposing productivity changes into technical and efficiency changes.

best practices from observed practices. I illustrate the usefulness of the LPI by analyzing the intertemporal development in logistics performances of the 24 largest cities in Norway in a period when 13 of them entered a collaboration agreement with the central government to reduce greenhouse gas emissions and to make the cities a better place in which to live. The agreement, also known as the Cities of the Future agreement, became binding in 2008 and expired in 2014. Land use and transport is naturally one of the most important areas of the collaboration agreement.

A major obstacle to monitoring logistics performances at the micro level is the limited accessibility to reliable data. This paper analyzes how local or regional freight transport can be evaluated using pseudo panel techniques based on the raw-data from national freight surveys. To that end, it utilizes DEA to empirically implement the LPI on a pseudo panel covering urban road freight transport in the 24 largest cities in Norway between 2008 and 2012.

This paper is structured as follows. The next section describes the theoretical foundations of the productivity index. Section 3 presents the dataset and the results, while section 4 concludes.

2. Methods

Consider freight transport as a production process in which inputs (i.e., the number of trips and vehicle capacities) are used to produce outputs (i.e. the quantity or weight of the cargo throughput). Denote inputs by $\mathbf{x} \in \mathfrak{R}_+^2$ and outputs by $\mathbf{y} \in \mathfrak{R}_+^M$. Assume that the production process is observed in $s = 1, \dots, S$ time periods. The technological possibilities for freight

transport in period s may then formally be summarized by a technology set. In this paper I consider two contemporaneous reference technologies; the variable returns to scale (VRS) technology

$$T_{VRS}^s = \left\{ (\mathbf{x}^s, \mathbf{y}^s) : \mathbf{x}^s \text{ can produce } \mathbf{y}^s \right\}, \quad s = 1, \dots, S \quad [1]$$

and the constant returns to scale (CRS) technology

$$T_{CRS}^s = \left\{ (\mathbf{x}^s, \mathbf{y}^s) : \mathbf{x}^s \text{ can produce } \mathbf{y}^s \right\}, \quad T_{CRS}^s = \delta T_{CRS}^s \text{ for all } \delta > 0, \quad s = 1, \dots, S \quad [2]$$

Following the usual convention, I assume that the technology sets satisfy the standard neo-classical axioms. That is, T is a compact and convex set satisfying the no free lunch and inactivity axioms, and free disposability of inputs and outputs. See Färe and Primont (1995) for more details on these axioms.

While the set theoretical representation of the technology is useful from an analytical perspective, it is insufficient for empirical analysis. Instead, function representations of the technology that can be estimated from data must be considered. In the case with only one output, the production function is a useful function representation that defines the maximal producible output for any given input vector. Distance functions are generally preferred function representations in cases with multiple inputs and outputs. These functions measure how far a given decision making unit is from the best-practice frontier by means of contracting inputs and/or expanding outputs, and are thereby useful measures of (in)efficiency. See Färe and Primont (1995) for more details on distance functions.

The production- and distance functions do not take into account that the objective of a decision making unit may be to maximize a ratio rather than to minimize its input use and/or

maximizing production. Environmental standards are for example often defined by the maximal allowable amount of pollution per unit of good output produced. It may be more useful for governments and public agencies to evaluate the gross national product per capita (the ratio of net present value to the overall budget) rather than the gross national product (the net present value). The amount of goods transported per trip or per vehicle kilometer are also easily understandable measures of logistics performance, and they will therefore be considered in this paper.

Hampf and Rødseth (2015) recently proposed using the ratio of a good to a bad output as a function representation of the technology, and introduced a new efficiency measure – the Ratio Efficiency Measure (REM) – based on this function representation. The current paper builds on and extends their contribution.

2.1. Ratio function representations and efficiency measures

Following Hampf and Rødseth (2015), a partial ratio measure (PRM) for the CRS technology – in our case maximizing the ratio of the tons lifted of a specific cargo type (y_j) to the number of trips (x_i) – can be defined:

$$PRM_{CRS}^s = \sup_{x_i^s, y_j^s} \left\{ \frac{y_j^s}{x_i^s} : (\mathbf{x}^s, \mathbf{y}^s) \in T_{CRS}^s \right\}, \quad s = 1, \dots, S \quad [3]$$

Eq. 3 maximizes the ratio of one output to one input (e.g., wet bulk per trip), keeping vehicle capacities and other outputs (e.g., dry bulk; general cargo; containerized cargo) constant. However, considering that the throughput of different cargo types is measured in a weight

unit (e.g., tons), it is also useful to define a general measure that characterizes the maximal load in tons per trip. A general ratio measure (GRM) is defined by:

$$GRM_{CRS}^s = \sup_{x_i^s, y^s} \left\{ \frac{\mathbf{1}'y^s}{x_i^s} : (\mathbf{x}^s, \mathbf{y}^s) \in T_{CRS}^s \right\}, \quad s = 1, \dots, S \quad [4]$$

where $\mathbf{1}$ is the unit vector⁵.

Eqs. 3 and 4 both assume that inputs and outputs could simultaneously be adjusted in order to maximize tons lifted per trip. This may not be an appropriate assumption in several cases, including the one presented in this paper. It is more likely that logistics companies choose the mix of inputs and outputs to maximize *profits* rather than to maximize *capacity utilization per trip*; cf. Eq. 4. There may also be cases where the operators cannot freely choose their input or output mix. Of importance to this study, which later considers freight transport within the largest cities in Norway, is that the cities' pattern of production and consumption will determine the volumes and types of cargo to be carried.

Whenever the logistics operators' objective function is inconsistent with the GRM, it may be a less relevant tool for evaluating logistics performances. In such cases, it would be more convenient to consider input-oriented (i.e., minimizing trips for given outputs) or output-oriented (i.e., maximizing freight deliveries for given trips) ratio measures. Assuming that an input-orientation is suitable for evaluating logistics performances (i.e., treating cargo-flows as exogenously given to logistics companies or cities), I propose an input-oriented GRM (IGRM) that defines the minimal number of trips necessary for a given freight delivery:

⁵ Note one may convert equation 3 into a more traditional economic measure by replacing the unit vector by the corresponding freight price vector. The GRM would then maximize freight revenues per trip.

$$\begin{aligned}
IGRM_{CRS}^s &= \sup_{x_i^s} \left\{ \frac{\mathbf{1}'\mathbf{y}^s}{x_i^s} : (\mathbf{x}^s, \mathbf{y}^s) \in T_{CRS}^s \right\} & [5] \\
&= \frac{\mathbf{1}'\mathbf{y}^s}{\inf_{x_i^s} \left\{ x_i^s : (\mathbf{x}^s, \mathbf{y}^s) \in T_{CRS}^s \right\}} = \frac{\mathbf{1}'\mathbf{y}^s}{x_{i,CRS}^s(x_{-i}^s, \mathbf{y}^s)}, \quad s = 1, \dots, S
\end{aligned}$$

where $x_{i,CRS}^s(x_{-i}^s, \mathbf{y}^s)$ is the minimal amount of trips needed to transport goods \mathbf{y}^s , given vehicle capacities, x_{-i}^s , in time-period s . Note that the superscript s on the outside of the brackets of $x_{i,CRS}^s(x_{-i}^s, \mathbf{y}^s)$ refers to the time period for the contemporaneous reference technology while the superscripts inside the brackets refer to the time period when the data on inputs and outputs were observed. For example, $x_{i,CRS}^t(x_{-i}^t, \mathbf{y}^t)$ corresponds to the minimal feasible number of trips for the period- t CRS technology using the data on vehicle capacities and outputs observed in period $t \in s$ (and thus, $x_{i,CRS}^{t+1}(x_{-i}^t, \mathbf{y}^t)$ denotes the minimal amount of trips needed to transport the goods, \mathbf{y}^t , given the vehicle capacities, x_{-i}^t – both observed in period t – for the $t+1$ CRS reference technology).

Following Hampf and Rødseth (2015), the input-oriented ratio efficiency measure (I-REM) is defined:

$$I-REM^s = \frac{\frac{\mathbf{1}'\mathbf{y}^s}{x_i^s}}{\frac{\mathbf{1}'\mathbf{y}^s}{x_{i,CRS}^s(x_{-i}^s, \mathbf{y}^s)}} = \frac{x_{i,CRS}^s(x_{-i}^s, \mathbf{y}^s)}{x_i^s}, \quad s = 1, \dots, S \quad [6]$$

The input-oriented ratio measure is defined by the ratio of the actual amount of cargo transported per trip to the maximal amount of cargo transported per trip. Since the observed ratio always is smaller or equal to the optimal ratio, the I-REM is always smaller or equal to

one. When the measure takes a value of one it indicates that the decision making unit under evaluation is efficient.

Figure 1 provides a graphical illustration of the GRM, the IGRM, and the I-REM in the case with one input and output. Note that I assume a piece-wise linear reference VRS technology for illustrative purposes.

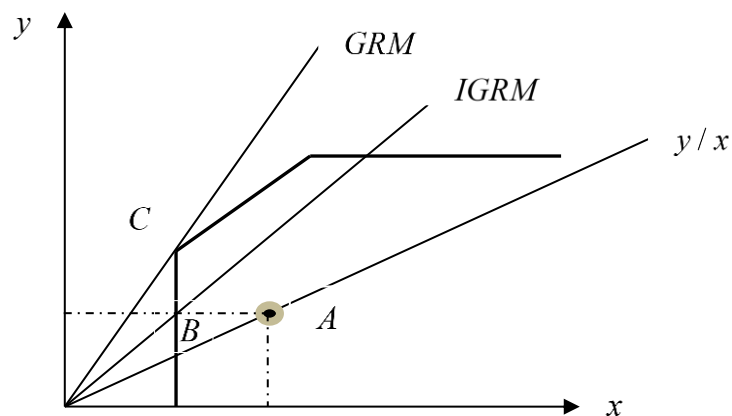


Figure 1: Ratio measures

The reference technology is represented by the three bold lines. A given decision making unit (DMU) is allocated at the point A, indicated by a circle. The DMU's current input consumption and production are indicated by the two dotted lines. The ratio of the two (i.e., DMU A's productivity) is given by the slope of the ray which passes through point A.

It can easily be seen from the figure that the DMU in question is inefficient, i.e., it operates in the interior of the technology set. Assume now that the DMU can move to the technology frontier in a way that maximizes the ratio of its output to its input consumption. If the output vector is assumed to be constant, the optimal allocation is at point B. The corresponding

optimal output-to-input ratio (i.e., productivity) is defined by the ray through B, which is equivalent to the IGRM. The input-oriented ratio efficiency measure can be calculated by taking the ratio of the y/x -ray to the IGRM ray. If, on the other hand, the DMU could also reallocate the output, it could obtain maximal productivity at point C. The slope of the ray through C is exactly the GRM.

Eq. 6 shows that the I-REM can be written as the ratio of the minimal number of trips (given the outputs and vehicle capacities) to the actual number of trips. It follows readily that the measure also has a traditional efficiency interpretation. Consider the short-run Farrell (1957) technical efficiency measure that minimizes the number of trips given the outputs and vehicle capacities:

$$TE_{CRS}^s(\mathbf{x}^s, \mathbf{y}^s) = \inf_{\theta_{CRS}^s} \left\{ \theta_{CRS}^s : (\theta_{CRS}^s x_i^s, x_{-i}^s, \mathbf{y}^s) \in T_{CRS}^s \right\}, \quad s = 1, \dots, S \quad [7]$$

The technical efficiency measure shrinks the point x_i^s to the period- s technology frontier. Since only one input is minimized (i.e., the number of trips), the solution to Eq. 7 is consequentially:

$$\theta_{CRS}^s x_i^s = x_{i,CRS}^s(x_{-i}^s, \mathbf{y}^s) \Leftrightarrow \theta_{CRS}^s = \frac{x_{i,CRS}^s(x_{-i}^s, \mathbf{y}^s)}{x_i^s} \quad [8]$$

Comparing Eq. 8 to Eq. 6, it is clear that the I-REM is equivalent to the short-run Farrell technical efficiency measure. Pictorially, it means that the I-REM is equivalently represented by the distance from point B to point A in figure 1.

It is well-known that the Farrell (1957) technical efficiency measure is the inverse of Shephard's (1953) input distance function. Thus, the I-REM is the inverse of the corresponding short-run input distance function.

2.2. Productivity index

So far, I have treated logistics efficiency measurement for a given time period s . I will now simultaneously consider two adjacent time-periods $(t, t+1) \in s$, in order to develop a framework for evaluating intertemporal changes in logistics productivity. Eq. 9 defines the logistics productivity index:

$$\begin{aligned}
 LPI^{t,t+1} &= \frac{\mathbf{1}'\mathbf{y}^{t+1} / x_i^{t+1}}{\mathbf{1}'\mathbf{y}^t / x_i^t} & [9] \\
 &= \frac{\mathbf{1}'\mathbf{y}^{t+1} / x_{i,CRS}^{t+1} (x_{-i}^{t+1}, \mathbf{y}^{t+1})}{\mathbf{1}'\mathbf{y}^t / x_{i,CRS}^t (x_{-i}^t, \mathbf{y}^t)} \times \frac{\frac{\mathbf{1}'\mathbf{y}^{t+1} / x_i^{t+1}}{\mathbf{1}'\mathbf{y}^{t+1} / x_{i,CRS}^{t+1} (x_{-i}^{t+1}, \mathbf{y}^{t+1})}}{\frac{\mathbf{1}'\mathbf{y}^t / x_i^t}{\mathbf{1}'\mathbf{y}^t / x_{i,CRS}^t (x_{-i}^t, \mathbf{y}^t)}} \\
 &= \underbrace{\frac{\mathbf{1}'\mathbf{y}^{t+1} / x_{i,CRS}^{t+1} (x_{-i}^{t+1}, \mathbf{y}^{t+1})}{\mathbf{1}'\mathbf{y}^t / x_{i,CRS}^t (x_{-i}^t, \mathbf{y}^t)}}_{\text{Frontier change}} \times \underbrace{\frac{\theta_{CRS}^{t+1}}{\theta_{CRS}^t}}_{\text{Efficiency change}}
 \end{aligned}$$

where the second equality follows by algebraic manipulation and the third equality follows from Eqs. 6 and 8.

Eq. 9 shows that the intertemporal change in logistics productivity can be decomposed into two components that describe the contributions of intertemporal i) *frontier changes* and ii) *efficiency changes* to the overall capacity utilization changes. Both components take values greater than 1 when they contribute to better logistics performances in period $t+1$ than in

period t , values equal to 1 when they contribute equally to logistics performances in both periods, and values less than 1 when they contribute to intertemporal regress in logistics productivity.

Recall that the Farrell technical efficiency measure is the inverse of Shephard's input distance function. The efficiency change component in Eq. 9 is thereby the inverse of the efficiency change component of the well-known Malmquist (1953) productivity index (assuming a short-run specification of the input distance function in which only trips are minimized).

By exploiting the relationship to the Malmquist index, the efficiency change measure can further be decomposed into pure efficiency changes and scale efficiency changes using the approach of Färe, Grosskopf and Margaritis (2008)⁶:

$$\frac{\theta_{CRS}^{t+1}}{\theta_{CRS}^t} = \underbrace{\frac{\theta_{VRS}^{t+1}}{\theta_{VRS}^t}}_{\text{Pure efficiency change}} \times \underbrace{\frac{\theta_{CRS}^{t+1} / \theta_{VRS}^{t+1}}{\theta_{CRS}^t / \theta_{VRS}^t}}_{\text{Scale efficiency change}} \quad [10]$$

where $TE_{VRS}^s(\mathbf{x}^s, \mathbf{y}^s) = \inf_{\theta_{VRS}^s} \{ \theta_{VRS}^s : (\theta_{VRS}^s x_i^s, x_{-i}^s, \mathbf{y}^s) \in T_{VRS}^s \}$ and $\theta_{VRS}^s \geq \theta_{CRS}^s, s = 1, \dots, S$. The efficiency change sub-components take values larger than 1 when technical and/or scale efficiencies have improved between periods t and $t+1$ in s .

Having decomposed the efficiency change component, I now turn to the frontier change component in Eq. 9. I aim to decompose it further, to evaluate the contributions of i) intertemporal changes *along* the technology frontier (due to intertemporal changes in the

⁶ Note that although the decomposition in Eq. 10 has been criticized by among other Ray and Desli (1997) and Lovell (2003), it is nevertheless a popular approach. See e.g. Walden et al. (2010) for a recent empirical application.

cargo mix and/or the vehicles' carrying capacities) and ii) changes *in* the technology frontier (i.e., technical changes) to the observed changes in logistics productivity.

In the productivity index literature, technical changes are often evaluated using “mixed period” estimates, e.g. by comparing the observed inputs and outputs in period $t+1$ to the contemporaneous frontier constructed from observations in period t . However, this specification of the productivity index violates the index property known as circularity. To resolve the problem, I follow Pastor and Lovell (2005) and consider a global reference technology that is constructed based on data from all time-periods under evaluation:

$$T_{CRS}^G = \text{conv}\{T_{CRS}^1 \cup \dots \cup T_{CRS}^S\} \quad [11]$$

The minimum feasible number of trips given x_{-i}^t and \mathbf{y}^t is in this case defined by $x_{i,CRS}^G(x_{-i}^t, \mathbf{y}^t)$

. Since $T_{CRS}^G \supseteq T_{CRS}^s$, $s = 1, \dots, S$, by Eq. 11, it follows readily that $x_{i,CRS}^s(x_{-i}^t, \mathbf{y}^t) \geq x_{i,CRS}^G(x_{-i}^t, \mathbf{y}^t)$.

Having established the global technology, the frontier change component decomposes as:

$$\begin{aligned} \frac{\mathbf{1}'\mathbf{y}^{t+1} / x_{i,CRS}^{t+1}(x_{-i}^{t+1}, \mathbf{y}^{t+1})}{\mathbf{1}'\mathbf{y}^t / x_{i,CRS}^t(x_{-i}^t, \mathbf{y}^t)} &= \frac{\mathbf{1}'\mathbf{y}^{t+1} / x_{i,CRS}^{t+1}(x_{-i}^{t+1}, \mathbf{y}^{t+1})}{\mathbf{1}'\mathbf{y}^{t+1} / x_{i,CRS}^G(x_{-i}^{t+1}, \mathbf{y}^{t+1})} \frac{\mathbf{1}'\mathbf{y}^{t+1} / x_{i,CRS}^G(x_{-i}^{t+1}, \mathbf{y}^{t+1})}{\mathbf{1}'\mathbf{y}^t / x_{i,CRS}^G(x_{-i}^t, \mathbf{y}^t)} \frac{\mathbf{1}'\mathbf{y}^t / x_{i,CRS}^G(x_{-i}^t, \mathbf{y}^t)}{\mathbf{1}'\mathbf{y}^t / x_{i,CRS}^t(x_{-i}^t, \mathbf{y}^t)} \quad [12] \\ &= \underbrace{\frac{\mathbf{1}'\mathbf{y}^{t+1} / x_{i,CRS}^G(x_{-i}^{t+1}, \mathbf{y}^{t+1})}{\mathbf{1}'\mathbf{y}^t / x_{i,CRS}^G(x_{-i}^t, \mathbf{y}^t)}}_{\text{Mix change}} \underbrace{\frac{x_{i,CRS}^G(x_{-i}^{t+1}, \mathbf{y}^{t+1}) / x_{i,CRS}^{t+1}(x_{-i}^{t+1}, \mathbf{y}^{t+1})}{x_{i,CRS}^G(x_{-i}^t, \mathbf{y}^t) / x_{i,CRS}^t(x_{-i}^t, \mathbf{y}^t)}}_{\text{Best practice gap}} \end{aligned}$$

The best practice gap measures the intertemporal change in the gap between optimal ratios evaluated at the global and contemporaneous frontiers (equality 2 in Eq. 12) or equivalently, the intertemporal change in the difference between the minimal input requirements for the global and contemporaneous frontiers (equality 3 in Eq. 12), measured along rays

(x_{-i}^s, \mathbf{y}^s) , $s = t, t+1$. In other words, it indicates whether the contemporaneous technologies are moving towards or away from the global frontier over time. The best practice gap component takes a value larger than 1 if the period $t+1$ frontier is closer to the global frontier than the period t frontier (indicating technical progress) and takes a value less than 1 if the frontiers are further apart in period $t+1$ than in period t (indicating technical regress).

The mix change component indicates the contributions of changes in the cargo mix and vehicle carrying capacities to the overall frontier change. That is, it summarizes all frontier changes that relate to movements “along” the (global) frontier, and not to technical changes. The mix change component takes a value greater than 1 if the optimal ratio is higher in period $t+1$ than in period t .

It is useful to further decompose the mix change component, to deduce the contributions of i) changes in the vehicles’ carrying capacities and ii) changes in the cargo mix to the overall mix change. The mix change component can be rewritten as:

$$\frac{\mathbf{1}'\mathbf{y}^{t+1} / x_{i,CRS}^G(x_{-i}^{t+1}, \mathbf{y}^{t+1})}{\mathbf{1}'\mathbf{y}^t / x_{i,CRS}^G(x_{-i}^t, \mathbf{y}^t)} = \frac{\left(\frac{\mathbf{1}'\mathbf{y}^{t+1}}{\mathbf{1}'\mathbf{y}^t}\right)^{0.5}}{\underbrace{\left(\frac{x_{i,CRS}^G(x_{-i}^{t+1}, \mathbf{y}^{t+1})}{x_{i,CRS}^G(x_{-i}^t, \mathbf{y}^{t+1})} \frac{x_{i,CRS}^G(x_{-i}^{t+1}, \mathbf{y}^t)}{x_{i,CRS}^G(x_{-i}^t, \mathbf{y}^t)}\right)^{0.5}}_{\text{Capacity induced mix change}}} \times \frac{\left(\frac{\mathbf{1}'\mathbf{y}^{t+1}}{\mathbf{1}'\mathbf{y}^t}\right)^{0.5}}{\underbrace{\left(\frac{x_{i,CRS}^G(x_{-i}^{t+1}, \mathbf{y}^{t+1})}{x_{i,CRS}^G(x_{-i}^{t+1}, \mathbf{y}^t)} \frac{x_{i,CRS}^G(x_{-i}^t, \mathbf{y}^{t+1})}{x_{i,CRS}^G(x_{-i}^t, \mathbf{y}^t)}\right)^{0.5}}_{\text{Output induced mix change}}} \quad [13]$$

Summing up, I have showed that the productivity index can be decomposed as follows:

$$LPI^{t,t+1} = \underbrace{\text{Capacity mix change} \times \text{Output mix change} \times \text{Best practice gap}}_{\text{Frontier change}} \times \underbrace{\text{Pure efficiency change} \times \text{Scale efficiency change}}_{\text{Efficiency change}} \quad [14]$$

2.3. Nonparametric estimation

The relevant function representations of the technology can be estimated from data using parametric or nonparametric techniques. I prefer the latter because they do not require me to choose a functional form a priori. Assume there are $l=(1,..,L)$ urban areas in the dataset, each using inputs $\mathbf{x}^{l,s} = (x_i^{l,s}, x_{-i}^{l,s}) \in \mathfrak{R}_+^2$ to transport goods $\mathbf{y}^{l,s} = (y_1^{l,s}, \dots, y_M^{l,s}) \in \mathfrak{R}_+^M$, $s = 1, \dots, S$. Let $\lambda^{l,s}$, $l=(1,..,L)$, define intensity variables. The minimal feasible amount of trips in urban area l' in period s , given this urban area's observed deliveries of goods and its aggregate vehicle capacity in period s , is for the contemporaneous DEA VRS-technology then defined as:

$$x_{i,VRS}^s(x_{-i}^s, \mathbf{y}^s) = \min_{x_i, \lambda} \left\{ x_i : \begin{aligned} \sum_{l=1}^L \lambda^{l,s} y_m^{l,s} &\geq y_m^{l',s}, & m = 1, \dots, M \\ \sum_{l=1}^L \lambda^{l,s} x_i^{l,s} &\leq x_i \\ \sum_{l=1}^L \lambda^{l,s} x_{-i}^{l,s} &\leq x_{-i}^{l',s} \\ \lambda^{l,s} &\geq 0, & l = 1, \dots, L \\ \sum_{l=1}^L \lambda^{l,s} &= 1 \end{aligned} \right\} \quad [15]$$

Note that the left hand sides of the inequalities (the sum of the products of intensity variables and the observed period- s data) represent the frontier of the contemporaneous technology

set, while the right hand sides define the data which is compared to this frontier (more specifically, the period s data on outputs and vehicle capacities for area l'). The minimal trip requirement for the CRS technology is obtained by omitting the summing-up condition

$$\sum_{l=1}^L \lambda^{l,s} = 1 \text{ from Eq. 15.}$$

The minimal trip requirement for the global technology is calculated in a similar fashion, by including observations from all periods under consideration in the estimation of the boundary of the technology set (i.e., the left-hand sides of the inequalities):

$$x_{i,CRS}^G(x_{-i}^s, \mathbf{y}^s) = \min_{x_i, \lambda} \left\{ x_i : \begin{aligned} & \sum_{s=1}^S \sum_{l=1}^L \lambda^{l,s} y_m^{l,s} \geq y_m^{l',s} \quad m = 1, \dots, M \\ & \sum_{s=1}^S \sum_{l=1}^L \lambda^{l,s} x_i^{l,s} \leq x_i \\ & \sum_{s=1}^S \sum_{l=1}^L \lambda^{l,s} x_{-i}^{l,s} \leq x_{-i}^{l',s} \\ & \lambda^{l,s} \geq 0, \quad s = 1, \dots, S, l = 1, \dots, L \end{aligned} \right\} \quad [16]$$

3. Dataset and results

3.1. Compiling the dataset

The dataset contains information about urban freight transport in 24 Norwegian municipalities – in which Norway's largest urban areas are situated. I limit the dataset to include deliveries that are internal to the urban areas (i.e., the origin and destination of any given delivery is one and the same). There is a fundamental technological difference between urban freight distribution and long haul, both in terms of the vehicle types used and in

efficiencies and capacity utilization (Hovi and Andersen, 2010). Thus, this step is taken to ensure that the units under comparison are homogenous.

The dataset is extracted from Statistics Norway's database on road freight transportation. Statistics Norway reports the data on a quarterly basis, and the municipality level is the lowest level of subdivision. The data is based on survey responses from freight transport companies. New surveys are issued every week, and in total about 1800 surveys – each related to a specific truck – are issued every quarter. The overall population is classified into four strata before randomly drawing the survey participants. Moreover, the register of vehicles is used to append information about the vehicles (e.g., on their carrying capacities) to the dataset.

After collecting the survey responses, Statistics Norway extrapolates the results to the strata level and, thus, to the overall population of trucks. While this approach produces a useful overview of freight transport at the *country level*, it does not result in representative statistics at the *municipality level* – which is my primary concern. Consequently, I find it more appropriate to base my analysis on the survey responses (i.e., on the micro or raw data). Preliminary studies of municipalities' logistics productivity based on the extrapolated data support this claim, as the growth rates are fluctuating and “inappropriately” high or low for several observations. On the one hand, Statistics Norway extrapolate the survey results to the strata level in five steps that include adjusting for attrition biases and underreporting of the cargo weight. Using the survey data as is, these adjustments are not considered and the results may therefore be vulnerable to the measurement biases. On the other, the index decomposition approach laid out in this paper is based on benchmarking, comparing each city to a benchmark constructed based on their best-practice peers. Unless the measurement biases are expected to vary systematically across urban areas and over time, the results are

likely to provide useful information about the rate and direction of changes in logistics performances.

The number of survey responses obtained for each of the 24 municipalities under consideration do – of course – only cover a small share of the total number of trips taking place within these municipalities (i.e., the cities) each year. The question is, however, if the sample is representative for the overall annual freight transportation taking place within the selected cities. In total, my dataset comprises 25,830 trips taking place in the 24 municipalities between 2008 and 2012. This implies an average of 215 trips per municipality per year. The minimum number of annual trips recorded at the municipality level is 12, while the maximum is 1809 trips. There is a clear relationship between the number of annual survey responses per city and the city size (proxied by the population size); the Pearson correlation coefficient is 0.92 and the Spearman correlation coefficient is 0.76. Both are statistically significant.

In the case where proper panel data do not exist but a series of independent cross sections are available, Deaton (1985) proposes the construction of a pseudo panel by grouping individual observations with similar characteristics into several homogenous cohorts. In our case, this corresponds to averaging survey responses in like areas, i.e., for each of the 24 municipalities under consideration. The analysis is thus at city (or municipality) level, which is suitable for our purpose of comparing the 24 cities in terms of their logistics performances. Verbeek and Nijman (1992) show that the pseudo panel approach is valid if the cohort sizes are sufficiently large, i.e., in the range of 100 to 200 individual observations. Aiming to meet this criterion, I group the data in two adjacent years into one period to boost the number of observations per cohort per period. The four periods under considerations are thus 2008-2009, 2009-2010, 2010-2011, and 2011-2012. This grouping of data is generally referred to as

Windows Analysis (Charnes et al, 1985) in the DEA literature, and is often used as a remedy when the number of decision making units are few but there are many relevant input and output variables. Windows analysis is based on moving averages, and is consequently useful for identifying performance trends. I prefer the windows approach because with the five years of data, the alternative will be to construct only two periods that do not overlap (e.g., 2008-2009 and 2011-2012) for examining productivity changes. This approach masks the intertemporal performance development in the period under consideration⁷.

The pseudo panel data approach has very interesting implications in the DEA setting⁸, which to my knowledge has not been addressed in the literature. Let $K^{l',s}$ denote the cohort size in city l' in period s , $\mathbf{x}^{k',l',s} \in \mathfrak{R}_+^2$ denote the input of observation k' in city l' in period s , and let $\mathbf{y}^{k',l',s} \in \mathfrak{R}_+^M$ denote the cargo throughput of observation k' in city l' in period s . The DEA model in Eq. 15 can then be restated as:

⁷ I am indebted to an anonymous referee for pointing this out. A potential drawback of the windows analysis is that adjacent periods, while being treated as independent, clearly are not because of the way the windows are constructed. To determine how this influences the statistical properties of the pseudo panel approach that is developed based on independent cross sections is beyond the scope of this paper. In general, I advise avoiding overlapping periods in future applications whenever possible.

⁸ Application of DEA to a pseudo panel data can be found in e.g. Paul et al. (2004).

$$\begin{aligned}
x_{i,VRS}^s(x_{-i}^s, \mathbf{y}^s) = \min_{\tilde{x}_i, \lambda} \left\{ \frac{\tilde{x}_i}{K^{l',s}} : \right. & \sum_{l=1}^L \lambda^{l,s} \left(\frac{K^{l',s}}{K^{l,s}} \sum_{k^l=1}^{K^l} y_m^{k^l,l,s} \right) \geq \sum_{k^l=1}^{K^l} y_m^{k^l,l',s}, m=1, \dots, M \\
& \sum_{l=1}^L \lambda^{l,s} \left(\frac{K^{l',s}}{K^{l,s}} \sum_{k^l=1}^{K^l} x_i^{k^l,l,s} \right) \leq \tilde{x}_i \\
& \sum_{l=1}^L \lambda^{l,s} \left(\frac{K^{l',s}}{K^{l,s}} \sum_{k^l=1}^{K^l} x_{-i}^{k^l,l,s} \right) \leq \sum_{k^l=1}^{K^l} x_{-i}^{k^l,l',s} \\
& \lambda^{l,s} \geq 0, \quad l=1, \dots, L \\
& \left. \sum_{l=1}^L \lambda^{l,s} = 1 \right\}
\end{aligned} \tag{17}$$

where I have harmlessly used the equality $x_i = \tilde{x}_i / K^{l',s}$ for the endogenous variable. Eq. 17 shows that the pseudo panel data approach is equivalent to scaling the data prior to the estimation to adjust for differences in cohort sizes. More precisely, the approach scales the data such that the size of each cohort corresponds to the size of the cohort under consideration (i.e., unit l'). This means, for example, that the number of trips in each cohort are assumed to be equivalent to the number of trips observed taking place in city l' . The drawback of this approach is that since all units are assumed to undertake the same number of trips in a given time-period, the variable returns to scale model by definition assumes that the minimum number of trips needed to process the cargo throughput of city l' in time-period s , given its aggregate vehicle capacities, is equivalent to the city's actual number of trips. Hence, the approach does not allow detecting technical inefficiency, but attributes inefficiencies solely to differences in scale efficiencies.

I consider operationalizing logistics productivity as the freight load *per trip* or *per kilometer* of transport. There are pros and cons to considering the number of trips or distances as inputs. Particularly, contextual variables that are outside of the logistics operators' control are likely to influence vehicle kilometers. For example, the population density and the city size in terms

of land area are likely to play a direct role in determining the average trip length. Hence, using vehicle kilometers as input, city characteristics are easily mistaken for differences in logistics performances. This issue might be avoided by considering the number of trips instead. The model then compares the cities' output per trip, which is an intuitive measure of logistics productivity. On the other hand, using the number of trips as the input variable, cities which have taken steps to reduce the vehicle miles travelled within their city limits are not explicitly rewarded for their actions by the logistics performance assessment.

Since reducing the vehicle miles travelled by trucks is not an explicit goal of the Cities of the Future agreement, and since differences in the characteristics of urban areas in Norway may largely affect the miles travelled, I utilize the number of trips as a transport input in the subsequent empirical analysis.

The raw-data classifies the cargo into eight aggregate categories: food and beverages, consumer goods, industrial goods, consolidated goods, chemical products, building materials, petroleum, and bulk and waste products. They are not homogenous with respect to their characteristics, e.g., heaviness. It is essential to control for weight characteristics of the cargo since the outputs (i.e., goods deliveries) are measured in a weight unit (i.e., kilograms) rather than in volume (Hovi and Andersen, 2010). This means that the capacity utilization for goods which are spacious, but which also are very light, will be regarded low when vehicle capacities are defined in terms of tonnage. I therefore prefer a multi-output approach, which allows me to control for differences in various cargo types input requirements.

There is a trade-off involved in selecting the number of outputs. If I include 8 outputs (i.e., cargo types) in the analysis, the "degrees of freedom" in the DEA model are low and it becomes difficult to discriminate efficient from inefficient units. However, if I aggregate up to

one output I neglect weight differences between the goods categories, which in turn may result in biased efficiency rankings. I therefore prefer to aggregate the 8 goods categories up to 4 outputs that are used in the subsequent empirical analysis. These categories are also common aggregates in the logistics literature. Table 1 presents the four goods categories and their subcategories:

Table 1: Aggregation of cargo types

Aggregate (outputs)	Specific goods types
Dry bulk	• Bulk and waste products
Wet bulk	• Petroleum • Chemical products
General cargo	• Food and beverages • Consumer goods • Industrial goods • Building materials
Consolidated goods	• Consolidated goods

Summary statistics of the dataset is provided by table 2.

Table 2: Summary statistics

2008-2009 (average of 11,707 trips)					
	Capacity	Dry bulk	Wet bulk	General cargo	Consolidated
Mean	22346.7	2108.1	630.7	2514.1	885.9
St. dev	16312.2	5585.7	3830.3	6281.5	3567.5
Min	3545.0	0.0	0.0	0.0	0.0
Max	78040.0	35000.0	36000.0	40000.0	35000.0
2009-2010 (average of 10,810 trips)					
	Capacity	Dry bulk	Wet bulk	General cargo	Consolidated
Mean	23192.9	1966.1	493.6	2641.7	765.1
St. dev	16846.7	5354.7	3400.1	6519.5	3320.4
Min	3525.0	0.0	0.0	0.0	0.0
Max	73670.0	34338.0	36000.0	40000.0	34000.0

2010-2011 (average of 9,844 trips)					
	Capacity	Dry bulk	Wet bulk	General cargo	Consolidated
Mean	24704.9	1891.4	644.0	2756.7	917.4
St. dev	17691.8	5520.2	3894.6	6779.6	3702.5
Min	3525.0	0.0	0.0	0.0	0.0
Max	73670.0	32840.0	36000.0	37350.0	32000.0
2011-2012 (average of 8,862 trips)					
	Capacity	Dry bulk	Wet bulk	General cargo	Consolidated
Mean	25538.7	2199.3	813.9	2691.2	1007.8
St. dev	17973.3	6007.2	4385.5	6688.5	3915.7
Min	3505.0	0.0	0.0	0.0	0.0
Max	61335.0	33220.0	39990.0	36700.0	32950.0

- Vehicle capacities (including trailer capacities for all trips where trailers are used) and outputs are reported in kilograms

3.2. Results

Having established the dataset, I now turn to the empirical implementation of the theoretical index in section 2.2 using DEA. The LPI is transitive⁹, which means that the overall productivity change taking place in the entire period under consideration is defined by the product of the contemporaneous (i.e., adjacent period) productivity indices. Figure 2 presents the contemporaneous productivity indices (the bars) for the period from 2008 to 2012 and the overall change from 2008 to 2012 (the dots). Note that the contemporaneous indices are listed with respect to the first years of the aggregates used for windows analysis. For example, 2008-2009 refers to the productivity change between the two aggregate periods 2008-2009 and 2009-2010. The horizontal axis intersects the vertical axis at 1, which means that

⁹ The contemporaneous Malmquist index on the other hand is not transitive.

bars/dots above the horizontal axis indicate productivity growth, while bars/dots below the line indicate productivity decline. The productivities are sorted according to the cities' population size, ranging from Molde with its 22,676 inhabitants to Oslo with its 608,013 inhabitants. Population size is frequently perceived as a proxy for urban scale economies, and it is therefore useful to consider how productivity changes vary with this characteristic.

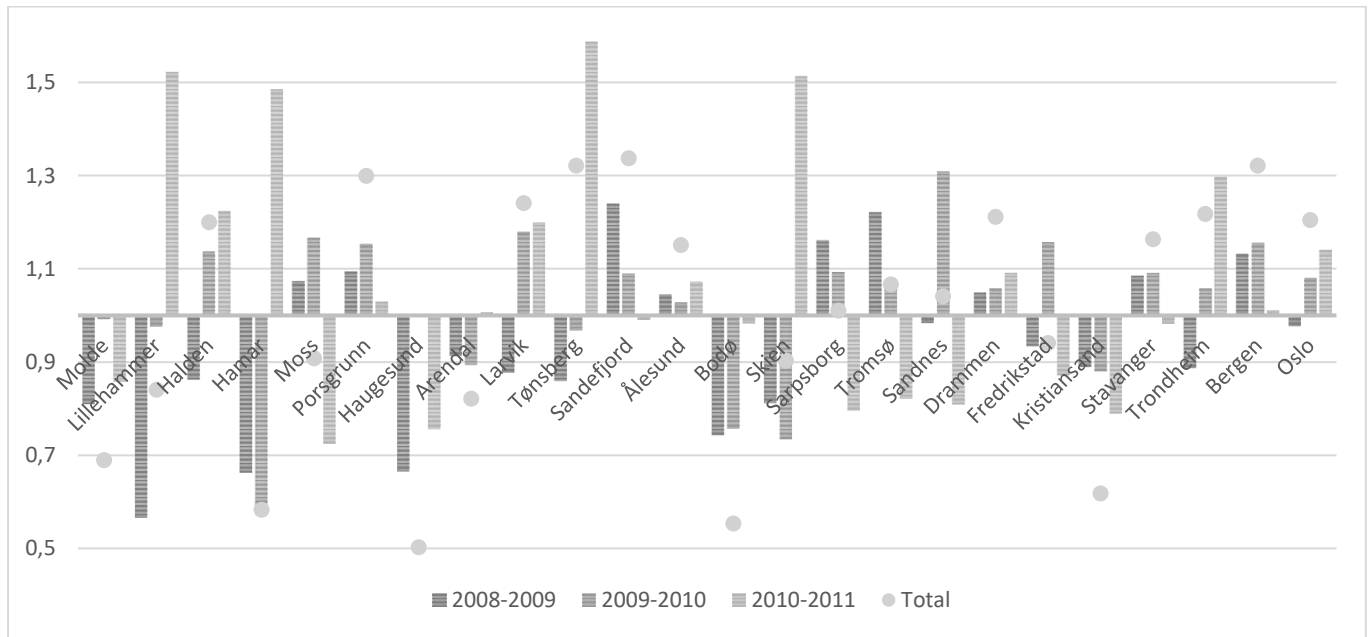


Figure 2: *The contemporaneous and overall productivity indices*

The figure shows that productivity changes fluctuate both across cities and within cities over the timespan under consideration. There is an immediate question as to whether fluctuation reflects the true intertemporal variation in productivities, or merely that our data is based on repeated cross-sectional surveys where different random samples are drawn from the population at different points in time. On the one hand, if the cross-sectional surveys are not generalizable to the entire population of trucks and the selection of trucks/trips change over time, fluctuations are expected. On the other, several steps have been taken to smooth

such effects. That is, the analysis is based on pseudo panel techniques and a windows analysis is used to ensure that the cohort sizes are sufficient large for the pseudo panel approach to be valid. This points to that the cities' logistics performances do in fact varying across time and place, which is an interesting finding.

The average LPI is 1.006 and only 14 of 24 cities show positive productivity growth for the entire period under consideration. 13 cities are found to have improved productivities in the last period (2010-2011) relative to the first period under consideration (2008-2009).

While the four largest cities experience productivity growth in the period between 2008 and 2012, there seems to be no clear-cut relationship among the population size and the logistics productivity change. There are several possible reasons for this. First, as previously suggested, vehicle kilometers travelled may be more sensitive to the urban form than the number of trips undertaken. Second, the city size may be more important for determining the relative performances of cities (i.e., logistics efficiencies) than for determining productivity changes. Third, a key factor is probably that the cities under consideration are quite small. In fact, 20 of the cities have less than 100,000 inhabitants. Hence, there may be too little variation with regards to city size, and the cities could be too small to realize economies of scale in logistics operations.

Having examined the rate of productivity growth, I now turn to its drivers. Figure 3 presents the average of the contemporaneous efficiency change (light grey bars), best practice gap (medium grey bars), and overall mix change (dark grey bars) components for each city. As in figure 2, the 24 cities are listed in ascending order based on their population sizes. The horizontal axis intersects the vertical axis at 1, which means that bars above the horizontal

axis indicate positive contributions to productivity, while bars below the line indicate negative contributions to productivity.

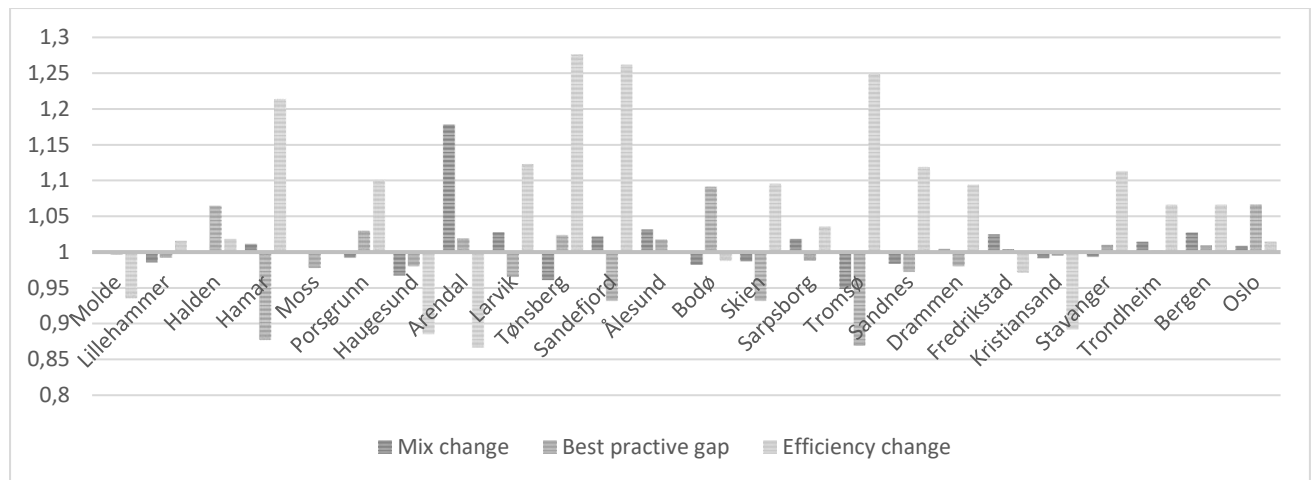


Figure 3: The LPI and its components

Except for Arendal, the average mix change is close to 1 for each city: On average, it is 1.006. Thus, changes in capacities and/or the output mix are not they key drivers of productivity changes. This is intuitively reasonable as for example commodity mixes depend on the cities' consumption and production patterns, which are likely to be approximately constant over the short time span under consideration. The average best practice gap is 0.991 while the average efficiency change is 1.059 over the period under consideration. Consequently, efficiency improvement appears to be the most important driver of productivity change. This result becomes even clearer when productivity changes are evaluated relative to the first period in the dataset (i.e., 2008-2009). In this case, the average mix change is 1.000, the average best practice gap is 0.883, while the average efficiency change is 1.137. Thus, there is evidence of technical stagnation and even regress in the period under consideration. Under technical

regress, the units will appear more efficient even if their logistics productivities have not improved.

While the overall mix change has little influence on the productivity growth, it is interesting to evaluate how capacity and output mix changes influence the overall mix change. Figure 4 reports the average of the contemporaneous capacity mix change (dark grey bars) and output mix change (light grey bars) components. As with the previous figures, the 24 cities are listed in ascending order based on their population sizes.

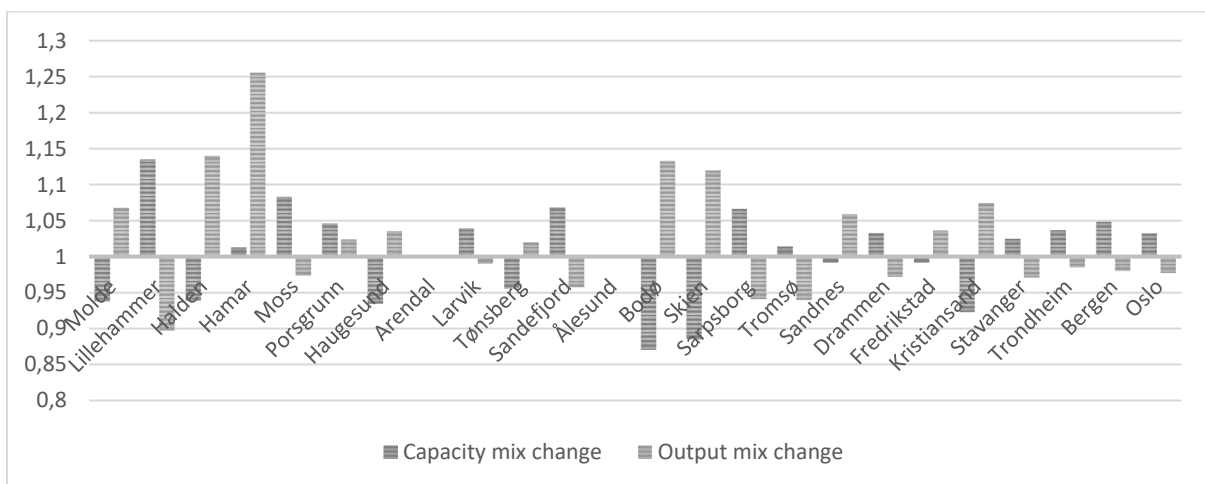


Figure 4: *The overall mix change and its components*

Figure 4 shows that the effects of changes in capacities and outputs on the overall mix change counteract each other for most cities. This is intuitively reasonable, as the cities are likely to

increase their cargo throughput when the aggregate vehicle capacities increase. The only exceptions from this rule are Hamar and Porsgrunn¹⁰.

Finally, I consider the implications of the Cities of the Future agreement on the LPI and its subcomponents. Several of the initiatives of the Cities of the Future agreement directly or indirectly affect urban freight transport. First, a major goal of the agreement is to reduce private car use in the cities, and a result-based financial compensation is granted annually to the cities that either can document success in reducing the use of private cars or which have implemented measures that are likely to result in declining car-use over time. By limiting the miles travelled by private cars in large cities, the accessibility of freight transport is likely to increase, thereby allowing freight companies' costs to decrease. This, on the other hand, *can* contribute to a reduction in logistics performances as lower costs are likely to reduce the lower limit of capacity utilization for which a trip is economically viable. Second, the Cities of the Future agreement also includes initiatives that directly target logistics performances. One initiative is for the city authorities to establish cooperation with companies and organizations to develop more efficient urban freight transport.

Since the program both targets reductions in the use of private cars in urban areas and increased freight transport performances, I hypothesize that the members of the program outperform the non-members in terms of productivity growth in the period between 2008 and 2012. I test this assumption using four non-parametric tests; the Kolomogorov-Smirnov,

¹⁰ No results are reported for Arendal and Ålesund because the solutions to the trip minimization problem for the hypothetical datapoints (x_{-i}^t, y^{t+1}) and (x_{-i}^{t+1}, y^t) are infeasible for all periods under consideration. The problem of infeasibility applies only to the sub-components of the overall mix change, and not to the efficiency, best practice gap, and overall mix change components.

ANOVA, Wilcoxon rank-sum, and Median tests. All tests examine the null hypothesis that the empirical results for members and non-members of the Cities of the future agreement are equal in terms of means and distributions. The test statistics and the corresponding P-values (in brackets) are reported in table 3.

Table 3: Non-parametric tests. Test scores (P-values).

Year	Test	LPI	Mix change	Best practice gap	Efficiency change
2008-2009	K-Smirnov	0.583 (0.034)	0.500 (0.100)	0.417 (0.249)	0.417 (0.249)
	ANOVA	5.770 (0.025)	0.320 (0.576)	0.580 (0.454)	0.350 (0.559)
	Wilcoxon	2.367 (0.018)	1.328 (0.184)	0.924 (0.356)	0.984 (0.325)
	Median	4.167 (0.041)	1.500 (0.221)	0.167 (0.683)	1.500 (0.221)
2009-2010	K-Smirnov	0.500 (0.100)	0.333 (0.518)	0.417 (0.249)	0.167 (0.996)
	ANOVA	1.850 (0.187)	2.180 (0.154)	1.430 (0.245)	0.030 (0.860)
	Wilcoxon	1.328 (0.184)	-1.097 (0.273)	1.328 (0.184)	-0.087 (0.931)
	Median	1.500 (0.221)	0.167 (0.683)	1.500 (0.221)	0.167 (0.683)
2010-2011	K-Smirnov	0.250 (0.847)	0.333 (0.518)	0.417 (0.249)	0.250 (0.847)
	ANOVA	0.970 (0.335)	0.120 (0.730)	2.260 (0.147)	0.020 (0.888)
	Wilcoxon	-0.751 (0.453)	0.000 (1.000)	-0.693 (0.488)	0.116 (0.908)
	Median	0.167 (0.683)	0.167 (0.683)	0.167 (0.683)	0.167 (0.683)

While the overall set of tests indicates differences in productivity changes in the period 2008-2010 among members and non-members of the agreement, the tests do not produce conclusive evidence of performance differences in the long term. I thus conclude that the agreement does not appear to have fostered productivity and efficiency changes that are widely different from the general development of other cities.

4. Summary and conclusions

Improvement of operational efficiency is a common goal of most governmental freight transport policies as inefficient freight operations impair business competitiveness and amplify the external costs of transport. Identifying potentials for improvements as well as the drivers of performance changes are thus of the essence. This paper has developed logistics efficiency and productivity measures and showed that they can be powerful tools for monitoring the development of urban logistics performances. The theoretical measures have been implemented on a dataset covering urban logistics in the 24 largest cities in Norway in 2008 and 2012, when 12 of the cities entered a collaboration agreement with the Norwegian government to reduce private car use and to improve freight transport efficiency. The results show a modest average productivity growth of 0.6 percent in urban freight transportation in Norway in this period. Efficiency changes are found to be the most important promotor of productivity growth, but appear to be countered by technical stagnation and regress. Because this study evaluates productivity changes over a relative short time span, it cannot be determined whether this development coincides with the general productivity development of urban logistics in Norway in recent years. However, if the trend continues, negative productivity growth is expected in the long run. Its drivers and how it may be reversed should consequently be addressed in future research.

The Cities of the Future agreement appears to have had limited impact on logistics productivity. This assessment illustrates the usefulness of the proposed method as a management tool for urban transport policies, allowing policy makers to monitor the development of urban freight transport and correspondingly to evaluate and adapt policies to achieve their objectives. However, one possible explanation for the agreement's current

lack of success is that its impacts may become visible only in the long run, while the timespan under consideration in this paper is relatively short. Lags do in general constitute an obstacle to allocating funds to cities based on their past performances, which is currently on the agenda in Norway.

Productivity changes are found to fluctuate across and within cities. This is likely to mirror fluctuations in the demand for freight transport services, which make it difficult for transport companies to match vehicle capacities and freight volumes on a day-to-day basis. Moreover, it may be reasonable for carriers to tailor vehicle capacities to expected peak loads rather than average loads, because operating smaller vehicles leads to additional driver costs. McKinnon (2015a) argue that there is a limited role for governments in correcting such failures, except for removing legal restrictions that prevent carriers from picking up backloads.

Taken at face value, the results do not indicate a relationship between the city size and productivity growth. Thus, logistics companies appear to be unable to reap agglomeration benefits, which indicates that the prevailing land use policies have not been successful in promoting efficient freight transport. However, as both the current study and the Norwegian cities are limited in scope, this research question deserves further attention.

This paper's approach to model road freight transport using set theoretical production analysis – in which the number of trips and vehicle capacities are treated as inputs and the throughput of cargo as outputs – is, to my knowledge, novel. While this paper demonstrates its usefulness for productivity and efficiency analysis of freight transport, it could also be useful for the empirical examination of a wider range of research questions: The production analysis apparatus allows examining substitution possibilities for trips and capacity (for given

freight deliveries), and the estimation of shadow prices for inputs and outputs; see Färe and Primont (1995) for details. It can also be extended to examine the impacts of contextual factors such as urban characteristics and policy changes on productivity and efficiency¹¹. This offers new tools for examining the economics of freight transport.

5. Acknowledgement

The author thanks Inger Beate Hovi and Benjamin Hampf, Jens Krüger, and the other participants of my research seminar at the Technical University of Darmstadt for helpful comments. The usual disclaimer applies.

6. References

- Alises, A., J.M. Vassala, & Guzmán Cherrett, A.F. (2014). Road freight transport decoupling: A comparative analysis between the United Kingdom and Spain. *Transport Policy*, 32, 186-193.
- Allen, J., M. Browne, & Cherrett, T. (2012). Investigating relationships between road freight transport, facility location, logistics management and urban form. *Journal of Transport Geography*, 24, 45-57.
- Betanzo-Quezada, E., & Romero, J. S. (2010). An urban freight transport index. *Procedia Social and Behavioral Sciences*, 2, 6312–6322.

¹¹ See for example Simar and Wilson (2007) for a discussion on the use of second-stage regressions to capture the impacts of contextual variables on DEA efficiency scores.

- Browne, M., J. Allen, T. Nemoto, J. Visser, & Wild, D. (2008). City access restrictions and the implications for goods deliveries. In E. Taniguchi and R. G. Thompson (eds.), *City Logistics V*, Crete:Institute for City Logistics
- Caplice, C., & Sheffi, Y. (1994). A review and evaluation of logistics metrics. *International Journal of Logistics Measurement*, 5, 11-28.
- Charnes, A., C.T. Clark, W.W. Cooper & Golany, B. (1985). A developmental study of Data Envelopment Analysis in measuring the efficiency of maintenance units in the U.S. air forces. *Annals of Operations Research*, 2, 95-112.
- Crujssen, F., W. Dullaert & Joro, T. (2010). Freight transportation efficiency through horizontal cooperation in Flanders. *International Journal of Logistics Research and Applications*, 13, 161-178.
- Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of Econometrics*, 30, 109-126.
- Eidhammer, O., & Jean-Hansen, V. (2008). Urban freight transport and logistics. In *TØI report 948/2008*. Oslo: Institute of Transport Economics.
- European Commission (2011) White paper: Roadmap to a single European transport area - Towards a competitive and resource efficient transport system. Brussels: European Commission
- Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society*, 120, Part 3, 253-281.
- Färe, R., Grosskopf, S., & Margaritis, D. (2008). *Efficiency and productivity: Malmquist and more*. In H.O. Fried, C.A.K Lovell, and S.S. Schmidt (eds.), *The Measurement of Productive Efficiency and Productivity Growth*. New York: Oxford University Press.
- Färe, R., & Primont, D. (1995). *Multi-output production and duality: theory and applications*. Boston: Kluwer Academic Publishers.
- Grifell-Tatjé, E., & Lovell, C. A. K. (1996). Deregulation and productivity decline: the case of Spanish savings banks. *European Economic Review*, 40, 1281-1303.

- Hampf, B., & Rødseth, K. L. (2015). Carbon dioxide emission standards for U.S. power plants: an efficiency analysis perspective. *Energy Economics*, 50, 140-153.
- Heng, Y., S.H. Lim & Chi, J. (2015). Toxic air pollutants and trucking productivity in the US. *Transportation Research Part D*, 17, 309-316.
- Hovi, I.B., & Andersen, J. (2010). Trends in the utilization and environmental performance of Norwegian freight vehicles. In *TØI report 1063/2010*. Oslo: Institute of Transport Economics.
- Kveiborg, O., & Fosgerau, M. (2007). Decomposing the decoupling of Danish road freight traffic growth and economic growth. *Transport Policy*, 14, 39-48.
- Lovell, C.A.K. (2003). The decomposition of Malmquist productivity indexes. *Journal of Productivity Analysis*, 20, 437-458.
- Malmquist, S. (1953). Index numbers and indifference surfaces. *Trabajos de Estadística*, 4, 209-242
- McKinnon, A. (2015a). Performance measurement in freight transport: In contribution to the design, implementation and monitoring of public policy. International Transport Forum: Prepared for the Roundtable on Logistics Development Strategies and their Performance Measurements (9-10-March 2015, Queretaro)
- McKinnon, A. (2015b). *Opportunities for improving vehicle utilization*. In: McKinnon, A., Browne, M., Piecyk, M. and Whitening, A. (eds.) *Green Logistics 3ed*. Kogan Page, London
- Pastor, J.T., & Lovell, C.A.K. (2005). A global Malmquist productivity index. *Economics Letters*, 88, 266-271.
- Paul, C., Nehring, R., Banker, D. & Somwaru, A. (2004). Scale economies and efficiency in U.S. agriculture: Are traditional farms history?. *Journal of Productivity Analysis*, 22, 185-205.
- Quack, H. (2008). *Sustainability of urban freight transport. Retail distribution and local regulations in cities.*, Erasmus University, Rotterdam; <http://repub.eur.nl/pub/11990/EPS2008124LIS9058921543Quak.pdf>.
- Ray, S.C., & Desli, E. (1997). Productivity growth, technical progress, and efficiency change in industrialized countries: comment. *American Economic Review*, 87, 1033-1039.

- Shephard, R.W. (1953). *Cost and production functions*. Princeton: Princeton University Press.
- Simar, L. & Wilson, P.W. (2007). Estimation and inferences in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, 136, 31-64.
- Sorrell, S., Lehtonen, M., Stapleton, L., Pujol, J. & Champion, T. (2009). Decomposing road freight energy use in the United Kingdom. *Energy Policy*, 37, 3115-3129.
- Verbeek, M. & Nijman, T. (1992). Can cohort data be treated as genuine panel data. *Empirical Economics*, 17, 9-23.
- Walden, J.B., Kirkley, J.E., Färe, R. & Logan, P. (2012). Productivity change under an individual transferable quota management system. *American Journal of Agricultural Economics*, 94, 913-928.
- Zhang, N., Zhou, P. & Kung, C. (2012). Total-factor carbon emission performance of the Chinese transportation industry: A bootstrapped non-radial Malmquist index analysis. *Renewable and Sustainable Energy Reviews*, 41, 584-593.