UNITED NATIONS
INDUSTRIAL DEVELOPMENT ORGANIZATION

Inclusive and Sustainable Industrial Development Working Paper Series
WP 7 | 2016

THE DYNAMICS OF VEHICLE ENERGY EFFICIENCY: EVIDENCE FROM THE MASSACHUSETTS VEHICLE CENSUS

DEPARTMENT OF POLICY, RESEARCH AND STATISTICS

WORKING PAPER 7/2016

The dynamics of vehicle energy efficiency: Evidence from the Massachusetts Vehicle Census

Sheng Zhong
UNU-MERIT

UNITED NATIONS INDUSTRIAL DEVELOPMENT ORGANIZATION

Vienna, 2016

The designations employed, descriptions and classifications of countries, and the presentation of the material in this report do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations Industrial Development Organization (UNIDO) concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries, or its economic system or degree of development. The views expressed in this paper do not necessarily reflect the views of the Secretariat of the UNIDO. The responsibility for opinions expressed rests solely with the authors, and publication does not constitute an endorsement by UNIDO. Although great care has been taken to maintain the accuracy of information herein, neither UNIDO nor its member States assume any responsibility for consequences which may arise from the use of the material. Terms such as “developed”, “industrialized” and “developing” are intended for statistical convenience and do not necessarily express a judgment. Any indication of, or reference to, a country, institution or other legal entity does not constitute an endorsement. Information contained herein may be freely quoted or reprinted but acknowledgement is requested. This report has been produced without formal United Nations editing.

The dynamics of vehicle energy efficiency:

Evidence from the Massachusetts Vehicle Census

Sheng Zhong

UNU-MERIT

Abstract: Using a rich quarterly panel dataset containing about 3.9 million vehicles in Massachusetts over the period 2008q1 - 2011q4, this paper is attempts to improve the micro level empirical basis of the study of population-level vehicle energy efficiency, and to provide some evidence that supports policy making related to sustainable development with regard to road vehicles. It (1) presents an aggregate vehicle energy efficiency indicator (state and municipality level) by taking into account vehicle heterogeneity, (2) investigates the contribution of changes in the structure of the vehicle population that affect aggregate vehicle energy efficiency and its growth, paying particular attention to vehicles' entry and exit, (3) explores the convergence and the Ergodic distribution of aggregate vehicle energy efficiency between municipalities, and (4) checks the socio-economic factors affecting the distribution of vehicles over locations, and vehicle exit event. The results confirm the importance of structural chance in the vehicle population, convergence of aggregate vehicle energy efficiency between municipalities and the crucial role of socio-economic factors in shaping vehicles distribution.

Keywords: vehicle level micro data, vehicle energy efficiency dynamics, decomposition, reallocation, convergence, socio-economic factors

1. Introduction

Sustainable development is a process with many dimensions (e.g., efficient use of resources and raw materials, increasing energy efficiency, reducing CO2 emissions, etc.). Therefore, it also requires a wide range of policies, aimed at all those dimensions, e.g., the EU 2020 Strategy (European Commission, 2010). This, in turn, requires careful study of the detailed processes that are going on in in the policy-relevant areas. One of those areas is energy efficiency of road vehicles. Energy use of road vehicles (broadly speaking, cars and trucks) is substantial. Due to technological change, new vehicle models that are more energy efficient are entering the market. But the diffusion of new vehicles is a long term process, which involves not only the acquisition of new and efficient vehicles, but also vehicle substitution (new vehicles displacing existing vehicles) and vehicle migration (inflow or outflow of vehicles from a particular location).[1] The displaced vehicles, however, do not necessarily exit the vehicle population immediately, they may be traded through the second-hand market, and possibly migrate to other places. If this is the case, the increase in energy efficiency at the level of the entire vehicle

[1] Upsizing of vehicle population would be observed if there is a net inflow of vehicles.

population becomes a function of which vehicles are replaced, and what is the energy efficiency of new vehicles. The analysis in this paper is aimed at contribute to a better understanding of the nature and timing of vehicle energy efficiency dynamics at the vehicle population level, and how this is affected by the life cycle of vehicles, including the introduction of new vehicles and the replacement of old ones. This will provide policymakers as well as entrepreneurs with better insight into how energy efficiency of road vehicles will change over time.

The analysis will describe the dynamics of vehicle energy efficiency in the state of Massachusetts in the US, over the period from 2008 to 2011. By linking vehicle level micro data to macro level municipal units, the analysis tries to answer the following questions with respect to the evolution of aggregate vehicle energy efficiency (municipality level and state level): (i) Does aggregate vehicle energy efficiency in Massachusetts improve as time progresses? (ii) What are the driving forces that cause the dynamics of vehicle energy efficiency? (iii) Is there a tendency of convergence (or divergence) in vehicle energy efficiency for municipalities in Massachusetts? What are the socio-economic factors that affect these dynamics?

These questions reflect several fundamental issues that are relevant for aggregate energy efficiency of the road vehicles population. First, there is a lack of suitable energy performance measurement indicators that are specific for a field in which policy can usefully be designed. For example, aggregate energy use per unit of GDP gives a general impression of how serious the sustainability pressure is, but it does not provide concrete guidance for policies. Indicators specifically designed for (for example) vehicles fill this gap. Second, as a durable consumer product, there is an extremely large number present in the market (commercial use as well as in households), and changes in the composition of this vehicle population can expected to be slow. Moreover, the major fuel used in the current global vehicle population is still fossil fuels like gasoline and diesel, which bring lots of environmental pressures. Third, and related to the previous point, there are important differentials in the levels of energy efficiency across vehicles related to vehicle characteristics (e.g., model year, vehicle make, type of vehicle, fuel used by the vehicle, etc.) and the way the vehicle is used and maintained. When new vehicles diffuse through a population, these differences will determine aggregate energy efficiency in a complicated way. In particular, when different vehicle populations interact, for example through the existence of second-hand markets, introduction in new vehicles in one population will also have (indirect) effects on the vehicle energy efficiency in other populations. These dynamics can be analyzed by looking at a set of such interacting populations (in our case municipalities) within a larger context (in our case, the state), and analyze how the overall dynamics of energy efficiency evolves (convergence or divergence). Fourth, the dynamics within and between subpopulations of vehicles will be influences by socio-economic (and other) conditions (e.g., income, population, unemployment, etc.), and these factors provide an important avenue for policy.

The data used in the analysis come from the Massachusetts Vehicle Census, a rich vehicle level panel dataset developed by the Metropolitan Area Planning Council of Massachusetts, which makes this paper depart from branch of literature on energy efficiency or productivity growth using aggregate data on sectoral or country level.[2] The advantages of using micro data are that (1) it allows to track the location and status of almost every vehicle over time in Massachusetts, (2) some interesting questions related to heterogeneity can be addressed effectively with micro data only (e.g., Bartelsman & Doms, 2000), and (3) micro data based methods can produce more plausible estimates (Olley & Pakes, 1996).

Figure 1 shows the total number of vehicles in Massachusetts over time. The total number of vehicles increased by around 8.18% between 2008q4 and 2011q4, rising from about 3.6 million to 3.9 million.[3] However, the rapid growth of total vehicles between 2008q1 and 2008q4 may be the result of the completion of the census dataset, and hence may not reflect a real trend. Nevertheless, it is clear that Massachusetts underwent significant entry and exit of vehicles. From 2008q4 to 2010q4, there was only a 0.58% increase in the total number of vehicles, but as illustrated in Table 1, the entry and exit of vehicles account for 13.79% and 13.29% of total vehicles in 2010q4 respectively. This implies that restructuring and reallocation of the vehicles population probably plays an important role.

Figure 1: Total number of vehicles in Massachusetts; 2008q1 – 2011q4

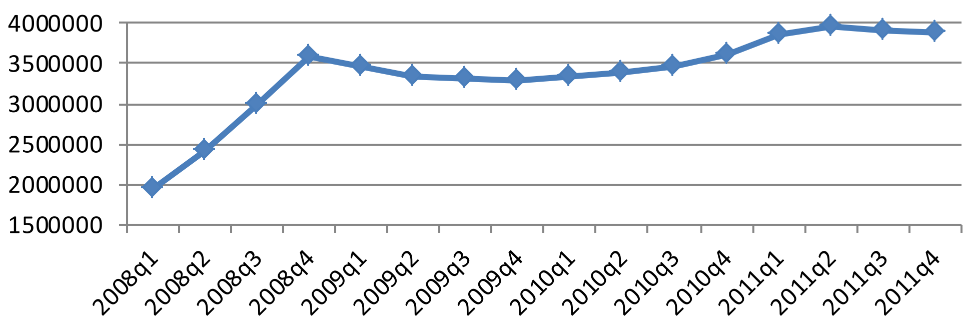


Table 1: Summary statistics of vehicles in Massachusetts between 2008q4 and 2010q4

Category	Number of vehicles	Share in the total vehicles
Entering vehicles: newly brought vehicles or vehicles coming from other states	574987	13.79%
Exiting vehicles: out-of-service vehicles or vehicles moving to other states or other countries	554117	13.29%
Continuing vehicles: vehicles staying in Massachusetts	3041663	72.93%

Source: own elaboration.

As will be shown below, the diversity among vehicles and municipalities is impressive. There are high energy efficiency entrants and low energy efficiency exiting vehicles (causing net efficiency gains), but also low energy efficiency entrants and high energy efficiency exiting vehicles (causing net efficiency

[2] The data in this paper are probably subject to the issues of misreporting of true vehicle efficiency, the so-called Volkswagen scandal. There is currently no way to avoid this.

[3] The last quarter of both years are selected. According to Figure 1, there was a rapid growth of total vehicles between 2008q1 and 2008q4, which may be explained by the process of completing the census dataset.

losses), municipalities with rapid growth of vehicle energy efficiency and municipalities that have small vehicle energy efficiency growth or even undergo deterioration in vehicle energy efficiency.

In order to deal with the diversity of issues and data trends, this paper applies a number of empirical methods. First, it applies methods from the study of productivity dynamics in firm level micro data to the field of energy efficiency dynamics.[4] This will look at Miles per Gallon (MPG) at the population level, and how its evolution can be explained by changes in the structure of vehicle (sub)population(s). The purpose of this part of the analysis is to find an appropriate model that can link micro data to aggregate energy efficiency (municipality level and state level), paying particular attention to the heterogeneity of vehicles. Two different decomposition approaches are introduced. The first decomposes the state level vehicle energy efficiency in each quarter from 2008q1 to 2011q4 into an unweighted average of vehicle level energy efficiency and a covariance between vehicle energy efficiency and the share of miles per day. The second looks at a between effect, within effect, entry effect and exit effect that contribute to the growth of vehicle energy efficiency on municipal level between 2008q4 and 2010q4.

A second group of methods used below explores the tendency of movement of below the aggregate vehicle energy efficiency, i.e., at the municipality level. Here we look at the relationship between initial relative vehicle energy efficiency on municipal level and its growth rate by performing OLS, nonparametric regression and bootstrapping quantile regression, and at a Markov-transition approach that maps the evolution of aggregate vehicle energy efficiency distribution. Lastly the socio-economic factors affecting the distribution of vehicles and the initially existing vehicle's staying or exiting decision are investigated.

2. Literature review

One of the challenges this paper faces is that there is little literature on the dynamics of vehicle energy efficiency in economics or in related areas. In fact, there seems to be no suitable indicator that can properly reflect the vehicle energy efficiency at population level, while such an aggregate indicator is greatly useful for policymaking regarding vehicles, given the extremely large number of vehicles in modern industrial societies and their large amount of fossil fuels consumption. A commonly used vehicle level technical indicator of vehicle's energy efficiency is miles per gallon (MPG). It is also a commonly used indicator for producer's marketing activities. But it turns out to be extremely difficult to obtain a suitable aggregate indicator of vehicle energy efficiency at the population level. The reasons are mainly twofold.

First, the data availability issue. the classical definition of energy efficiency proposed by Patterson (1996) is simply a ratio:

$$\frac{Useful\ output\ of\ a\ process}{Energy\ input\ into\ a\ process}$$

[4] Typical examples include: (Aw, Chen, & Roberts, 2001; Baily, Bartelsman, & Haltiwanger, 2001; Baily, Hulten, & Campbell, 1992; E. J. Bartelsman & Doms, 2000; Lucia Foster, Haltiwanger, & Krizan, 1998; L. Foster, Haltiwanger, & Krizan, 2006; Olley & Pakes, 1996).

As a non-production unit, a vehicle does not generate any useful physical output, but provides useful services measured by the miles it travels. Generating an aggregate indicator of vehicle energy efficiency for an administrative unit, say, a municipality, requires full information about miles and vehicle's energy consumption in that municipality. Such aggregate information is often not available.

Second, the heterogeneity issue. The aggregation approach mentioned above might not be a suitable one even that aggregate information is available, since it does not take into account the heterogeneity of vehicles. Following insightful views by Bartelsman and Doms (2000) and Basu and Fernald (1997) in the firm productivity studies, the complexity of the interactions between the factors correlated with aggregate energy efficiency growth, such as consumer behaviors regarding the usage of vehicles, vehicle level MPG, and vehicle movement (staying, entering or exiting), calls for a better aggregation method and a simple aggregation over all the heterogeneous units works. The preferred aggregation method taking into account heterogeneity is a weighted average of vehicle level MPG, where the weights are related to the importance of the vehicle, as suggested by several crucial studies in the area of productivity (Baily et al., 1992; Bartelsman & Doms, 2000; Olley & Pakes, 1996).

However, applying such an aggregation approach has an even higher requirement for the data: it needs big data that cover every vehicle over time. For a long time this has not been feasible: given the extremely large number of vehicles, the data collection would be greatly expensive and time-consuming and the processing of the data would be restricted by the computing ability of the computer. Fortunately, the Massachusetts Vehicle Census, one of the few public used databases providing vehicle level micro data, has been available recently. Benefiting from this dataset, this paper aims to fill the existing research gap on energy efficiency dynamics based on micro data.

The methods used here heavily rely on the studies on productivity that use longitudinal micro data (LMD), following large numbers of firms over time. As summarized by Foster et al. (1998), lots of studies in this field pay great attention to the reallocation and restructuring of inputs and outputs, especially the reallocation due to entry and exit of firms, for example, in the US retail trade sector (Foster et al., 2006), in the U.S. manufacturing sector (Baily, Bartelsman, & Haltiwanger, 1996; Baily et al., 2001; Baily et al., 1992; Lee & Mukoyama, 2015; Olley & Pakes, 1996), in the Taiwanese manufacturing sector (Aw et al., 2001) and in the Israeli industry (Griliches & Regev, 1995). Moreover, using plant level micro data on manufacturing establishments in China, India and the U.S., Hsieh and Klenow (2009) estimate the loss of aggregate productivity due to the misallocation of resources.

To some extent energy efficiency can be viewed as a special productivity measure. But the nature of vehicles is less complex as compared to the nature of a firm. For example, as a production unit, a firm's productivity growth can be a consequence of complex interactions between factors like worker skills, technological usage, management techniques and quality of output (Bartelsman & Doms, 2000). In this respect, the most important distinction between a vehicle and a firm is that plant level productivity can be estimated following a typical production function incorporating all the key factors and using plant level micro data, while this is not the case for a vehicle. The simple non-production nature of vehicle restricts the application of many well-established theoretical or empirical approaches from productivity studies in this paper. In fact, this paper mostly synthesizes and extends the empirical methods in productivity studies regarding the decomposition of aggregate productivity and its growth. The methodological details will be discussed in the next section.

In addition, there are several crucial studies on the product substitution in the vehicle market. Brownstone and Train (1999) generalize the basic choice model to mixed logits model that can approximate any product substitution pattern among gas, electric, methanol and CNG vehicles with various attributes. Berry, Levinsohn, and Pakes (2004) utilize rich micro data on consumer's choice of vehicles to precisely estimate the demand for differentiated vehicles. Both studies show that micro data are very useful in understanding the process of product substitution. These studies require micro data on consumers but the data available in Massachusetts Vehicle Census can only provide information on vehicles, not the owner of the vehicle. This implies that the product substitution cannot be directly observed and assessed in this dataset[5], although it actually happens.

3. Data and methodology

3.1 Data description

The main data source is the Massachusetts Vehicle Census, constructed by the Metropolitan Area Planning Council of Massachusetts. The census covers the period 2008q1 – 2011q4 and has a large and statistically representative sample of the heterogeneous vehicle population. For the computation of aggregate vehicle energy efficiency at both state and municipality levels and its decomposition, the entire dataset is used. For the decomposition of the growth of aggregate vehicle energy efficiency, the data for the period 2008q4 through 2010q4 are used. The reason is the data quality: in the first three quarters in 2008 there is a sharp increase in the vehicle population in the dataset (see Figure 1), probably due to a catch-up in data collection, while in 2011 lots of vehicles' data on miles per day are missing in the dataset.

Vehicle energy efficiency is measured by miles per gallon (MPG), a commonly used vehicle level technical indicator measuring how many miles a vehicle can travel by consuming one gallon of fuel. In Massachusetts Vehicle Census, two versions of MPG indicators are reported: the first is MPG2008 that is based on 2008 EPA ratings, and the second one is adjusted MPG that is adjusted to account for deterioration of energy efficiency over time and due to usage. In this study adjusted MPG is used as the only indicator for vehicle energy efficiency[6], allowing us to track the performance of individual vehicles in energy efficiency over time. In the census the odometer of every individual vehicle during the quarter under investigation is recorded twice, so that the average miles per day that the vehicle travels can be obtained. Certainly there are some measurement errors in the dataset. For example, some vehicles' miles per day are more than 1 million. Another issue is related to the number of days covered by two inspections. For some vehicles the number of days between by two inspections of odometer is small but during this short period of time the vehicle is frequently used, which leads to a very high average miles

[5] For example, when a new vehicle enters the state (it appears in the dataset), it is unknown whether the new vehicle is used to substitute the old one or not. It might be caused by migration.

[6] The adjusted MPG is estimated by employing a formula that includes vehicle age and mileage. See Daly and Ó Gallachóir (2011).

per day, say, more than 1000. The high miles per day of some vehicles based on a short measurement period may not represent the usage of the vehicles in the whole quarter. To reduce the impact of these kinds of problems, the study employs a data cleaning strategy in which in each quarter the vehicles are excluded in the analysis if their miles per day are above the 99% quantile of the state.

Additionally, the dataset contains information about the municipality where every individual vehicle is located in each quarter, which permits us to generate aggregate vehicle energy efficiency at municipality level. In addition, some socio-economic data at municipality level are used in this study, so that the results related to vehicle energy efficiency can be linked to socio-economic factors. The mean household income and median household income at municipality level in Massachusetts over the period 2008 to 2011 are extracted from American Community Survey (ACS). Municipality level data for population, labor force and unemployment are available from Department of Revenue of Massachusetts.

3.2 Aggregate vehicle energy efficiency and its decomposition over the period 2008q1 – 2011q4

On the basis of previous studies (Aw et al., 2001; Baily et al., 1992; Olley & Pakes, 1996), the relationship between micro level vehicle energy efficiency and aggregate vehicle energy efficiency can be described as follows (take municipality level aggregate vehicle energy efficiency as example):

$$\ln MPG_t = \sum_i \theta_{i,t} \ln MPG_{i,t} \qquad (1)$$

Where MPG_t is a certain municipality's aggregate vehicle energy efficiency in time t; $MPG_{i,t}$ is the energy efficiency of vehicle *i* at time t, measured by the vehicle's adjusted MPG; $\theta_{i,t}$ is vehicle *i*'s share of miles per day in the municipality at time t, defined as $\theta_{i,t} = \frac{m_{i,t}}{\sum_i m_{i,t}}$; $m_{i,t}$ is the vehicle *i*'s miles per day at time t.

In this approach, the aggregate vehicle energy efficiency is viewed as a relative share-weighted average of individual vehicle energy efficiency. Shifts in the shares of miles per day from low energy efficiency vehicles to high energy efficiency vehicles will contribute positively to the growth of aggregate vehicle energy efficiency. The basic idea behind this formulation is to describe two aspects revolving around cross-sectional distribution of vehicle energy efficiency: innovation generates new vehicle models and market competition determines the diffusion and adoption of these new vehicle models. On the other hand, the market share in terms of sales of a certain vehicle model does not provide useful information about its usage, while usage of a the vehicle is an indicator as important as its technical energy efficiency in shaping the dynamics of aggregate vehicle energy efficiency.

Following the methods by Olley and Pakes (1996) and Aw et al. (2001), the deviation between an individual vehicle and the unweighted mean over all the vehicles is derived from Equation (1), as shown in the equation below:

$$\ln MPG_t = \sum_{i=1}^{N_t} \theta_{i,t} \ln MPG_{i,t} = \sum_{i=1}^{N_t} (\overline{\theta_t} + \Delta\theta_{i,t})(\overline{\ln MPG_t} + \Delta \ln MPG_{i,t}) \qquad (2)$$

where $\Delta\theta_{i,t} = \theta_{i,t} - \overline{\theta_t}$, $\Delta \ln MPG_{i,t} = \ln MPG_{i,t} - \overline{\ln MPG_t}$, and a bar over the variable indicates the unweighted mean of the variable at time t.

Then aggregate adjusted MPG (in log) can be written as a sum of an unweighted average MPG (in log) and a simple covariance term between individual vehicles' MPG and its share of miles per day. The following equation can be obtained[7]:

$$\ln MPG_t = \overline{\ln MPG_t} + \sum_{i=1}^{N_t} \Delta\theta_{i,t} \Delta \ln MPG_{i,t} \qquad (3)$$

This equation is very useful, because it provides insight into the effects of structural change in the vehicle population by separating out micro level indicators from an aggregate one in the time series. The larger the covariance term, the larger the share of miles per day that is contributed by higher energy efficiency vehicles and the more energy efficient is the vehicle. By employing Equation (3), state level aggregate vehicle energy efficiency over the period 2008q1 and 2011q4 is decomposed, in order to gain a general picture of the restructuring of vehicles and learn whether the cross-sectional reallocation is energy efficiency enhancing over time or not.

3.3 Decomposition of growth of aggregate vehicle energy efficiency between 2008q4 and 2010q4

One limitation of Equation (3) is that entry and exit of vehicles across municipalities during a period of time is not reflected. According to Table 1, however, there is a large entry and exit between 2008q4 and 2010q4, which motivates an attempt to quantify the important role of the reallocation dynamics in aggregate vehicle energy efficiency dynamics.

In the group of entrants to a certain municipality during a period of time, one may roughly find two different sources depending on the status of the vehicle's initial location. One is the vehicles that initially exist in other municipalities within the same state, including immigration of vehicles from other municipalities (the owner moves and takes the vehicle), and vehicles coming from other municipalities through the second-hand market. The other source is the vehicles that come from outside Massachusetts, either through migration or trade (including the buying of brand-new vehicles and second-hand vehicles). Similarly, exiting vehicles from a certain municipality can be grouped into two categories, depending on their final locations. If the vehicle in end period still stays in Massachusetts, it might migrate or is sold to another municipality in Massachusetts. Otherwise the vehicle leaves Massachusetts through migration, trade in the second-hand market, or goes out of service. Table 2 below summarizes all the scenarios related to vehicle's entering and exiting in a certain municipality.

[7] This is because $\ln MPG_t$

$$= \sum_{i=1}^{N_t} \overline{\theta_t}\, \overline{\ln MPG_t} + \sum_{i=1}^{N_t} \overline{\theta_t} \left(\ln MPG_{i,t} - \overline{\ln MPG_t}\right) + \sum_{i=1}^{N_t} \overline{\ln MPG_t} \left(\theta_{i,t} - \overline{\theta_t}\right) + \sum_{i=1}^{N_t} \Delta\theta_{i,t} \Delta \ln MPG_{i,t}$$

$$= N_t \overline{\theta_t}\, \overline{\ln MPG_t} + \overline{\theta_t} \left(\sum_{i=1}^{N_t} \ln MPG_{i,t} - N_t \overline{\ln MPG_t}\right) + \overline{\ln MPG_t} \left(\sum_{i=1}^{N_t} \theta_{i,t} - N_t \overline{\theta_t}\right) + \sum_{i=1}^{N_t} \Delta\theta_{i,t} \Delta \ln MPG_{i,t}$$

$$= N_t \overline{\theta_t}\, \overline{\ln MPG_t} + \sum_{i=1}^{N_t} \Delta\theta_{i,t} \Delta \ln MPG_{i,t} = \overline{\ln MPG_t} + \sum_{i=1}^{N_t} \Delta\theta_{i,t} \Delta \ln MPG_{i,t}$$

Table 2: Entry and exit of vehicles in a certain municipality

Movement of vehicle	Initially existing in Massachusetts	Source
Entry	Yes	Immigration from other municipalities and vehicles brought from other municipalities through second-hand market
	No	Immigration from other states and vehicles brought from other states (including brand-new and second-hand vehicles)

Movement of vehicle	Final location in Massachusetts	Direction
Exit	Yes	Migration to other municipalities and vehicles sold to other municipality through second-hand market
	No	Out of service, migration to other states and vehicles sold to other states through second-hand market

Source: own elaboration

The growth of aggregate vehicle energy efficiency between two quarters is approximated by using the difference term of a logarithm function.[8] Taking the first order difference on both sides of Equation (1) yields the growth of aggregate vehicle energy efficiency between time t and $t-1$:

$$\Delta \ln MPG_t = \sum_i \theta_{i,t} \ln MPG_{i,t} - \sum_i \theta_{i,t-1} \ln MPG_{i,t-1} \qquad (4)$$

Taking into account all the possible scenarios related to vehicle's entering and exiting illustrated in Table 2, the following sets of vehicles are introduced:

$Stayer$: vehicles that stay in the municipality from start to end period (quarter);

$Entry$: vehicles that move to the municipality between the start and end quarter. This set consists of two subsets: $E1$, the subset of vehicles that move to the municipality from other municipalities in the state, and $E2$, the subset of vehicles that directly move to the municipality from outside the state;

$Exit$: a set of vehicles that leave the municipality between the start and end quarter; this set can be disaggregated into two subsets: $X1$, the subsets of vehicles that move to other municipalities in the state, and $X2$, the subset of vehicles that go out of service or move to other states (or countries).

The basic decomposition equation that is used was proposed by Baily et al. (1992), as shown below:

$$\Delta \ln MPG_t = \sum_{i \in Stayer} (\theta_{i,t} \ln MPG_{i,t} - \theta_{i,t-1} \ln MPG_{i,t-1}) + \sum_{i \in Entry} \theta_{i,t} \ln MPG_{i,t} - \sum_{i \in Exit} \theta_{i,t-1} \ln MPG_{i,t-1}$$

[8] According to the definition of derivative, $f(x)' = \lim_{\Delta x \to 0} \frac{\Delta f(x)}{\Delta x}$. When Δx is very small, approximately, $f(x)' \approx \frac{\Delta f(x)}{\Delta x}$. Let $f(x) = \ln x$. Then $f(x)' = (\ln x)' = \frac{1}{x} \approx \frac{\Delta f(x)}{\Delta x}$, that is, $\Delta(\ln x) \approx \frac{\Delta x}{x}$. The growth of x can be approximated by the difference of its logarithm function.

$$= \sum_{i \in Stayer} (\theta_{i,t} \ln MPG_{i,t} - \theta_{i,t-1} \ln MPG_{i,t} + \theta_{i,t-1} \ln MPG_{i,t} - \theta_{i,t-1} \ln MPG_{i,t-1}) + \sum_{i \in Entry} \theta_{i,t} \ln MPG_{i,t} - \sum_{i \in Exit} \theta_{i,t-1} \ln MPG_{i,t-1}$$

$$= \sum_{i \in Stayer} (\theta_{i,t} - \theta_{i,t-1}) \ln MPG_{i,t} + \sum_{i \in Stayer} \theta_{i,t-1} (\ln MPG_{i,t} - \ln MPG_{i,t-1}) + \sum_{i \in Entry} \theta_{i,t} \ln MPG_{i,t} - \sum_{i \in Exit} \theta_{i,t-1} \ln MPG_{i,t-1}$$

The growth of aggregate vehicle energy efficiency in a certain municipality can be decomposed into the contributions of the stayers, entrants and the exiters. The sets $Entry$ and $Exit$ can be further broken down into subsets $E1\ and\ E2$, $X1\ and\ X2$, respectively, that is,

$$\Delta \ln MPG_t = \sum_{i \in Stayer} (\theta_{i,t} - \theta_{i,t-1}) \ln MPG_{i,t} + \sum_{i \in Stayer} \theta_{i,t-1} (\ln MPG_{i,t} - \ln MPG_{i,t-1}) + \sum_{i \in E1} \theta_{i,t} \ln MPG_{i,t} + \sum_{i \in E2} \theta_{i,t} \ln MPG_{i,t} - \sum_{i \in X1} \theta_{i,t-1} \ln MPG_{i,t-1} - \sum_{i \in X2} \theta_{i,t-1} \ln MPG_{i,t-1}$$

The contribution of the stayers consists of two elements: the effect due to changes in individual vehicles' shares of miles per day, holding vehicle level MPG constant (or the "between effect", the first term) and the effect due to changes in MPG in each vehicle separately (or the "within effect", the second term). Unlike firm productivity, however, the deterioration of a vehicle's energy efficiency is irreversible, which implies that the within effect in the basic decomposition equation is negative. The third term and the forth term summarize the contribution of two kinds of entrants. The last two terms represent the effects of exiting vehicles. The net entry effect is then defined as the sum of the last four terms, all the effects of entering and exiting vehicles, reflecting any differences in the shares of miles per day between entering and exiting vehicles and any differences in levels of individual vehicle's MPG.

One disadvantage of this basic decomposition method is that the weights used in the between effect and within effect represent different time periods: the between effect uses ending-level vehicle-level MPG as the weight while the within effect uses starting-level share of vehicle's miles per day. As addressed by Foster et al. (1998), the between effect used in the basic decomposition method actually captures a covariance term as well, which might lead to a bias towards a positive between term and a negative net entry term in the analysis of firm productivity. They develop a modified version compared to Baily et al. (1992) and Baily et al. (1996), by separating out the covariance term from the between effect and using the deviations from initial aggregate MPG in the decomposition. This method is extended to the field of vehicle energy efficiency dynamics and labeled as Method 2, as shown in Table 3.

In Method 2, the growth of aggregate vehicle energy efficiency is decomposed into seven parts. The first three add up to the contribution of stayers to the growth of aggregate vehicle energy efficiency.

Compared to the basic decomposition method, the components of Method 2 are distinguished as follows. The within term does not change: it is weighted by initial shares of miles per day. The between term is weighted by the deviation of initial vehicle level MPG (in log points) and initial aggregate MPG (in log points). The newly introduced covariance term is measured as a sum of vehicle level MPG growth and vehicle's share change. The entry and exit effects are weighted by end period shares and initial shares, respectively, and revised by using the deviation terms between initial vehicle level MPG (in log points) and initial aggregate MPG (in log points).

Some other studies (Aw et al., 2001; Baily et al., 2001; Bartelsman & Wolf, 2014; Griliches & Regev, 1995) employ a more elegant decomposition approach that contains deviations from average aggregate MPG between start and end periods, instead of deviations from initial aggregate MPG. This method is labeled as Method 3, and illustrated in Table 3. Method 3 is the preferred decomposition approach used in this paper. The reasons are the following. First, conceptually Method 2 is more reasonable since it separates out the covariance term from between effect, but this method focuses on measurement errors (Baily et al., 2001). In fact, as shown later in the state level growth decomposition results in this paper, Method 2 generates results that are very similar to those from Method 3. Second, using mean values between start and end period in Method 3, there is no need to specify initial or ending level weights.

In Method 3, the growth of aggregate vehicle energy efficiency is decomposed into six components, distinguished as follows. The first term is a within effect, that is, the changes in stayer's MPG weighted by the average share of miles per day between start and end periods. The second term is a between effect reflecting changes in stayer's share of miles per day, weighted by the deviation of vehicle level average MPG (in log points) from average aggregate MPG (in log points). This implies that an increase in the share of miles per day can only have positive impact on the growth of aggregate MPG if the vehicle's average MPG is higher than the average aggregate MPG. In other words, increasing usage of high energy efficiency vehicles can positively contribute to the growth of aggregate vehicle energy efficiency. Similarly, the entry and exit effects are broken down into their subsets of entrants and exiting vehicles, weighted by end period and initial period shares respectively. But both entry and exit effects are revised by using deviations between vehicle level MPG and average aggregate MPG. The basic idea behind such a formulation is that the exit of low energy efficiency vehicles is equivalent to the entry of high energy efficiency vehicles in improving aggregate MPG.

Method 3 is applied to the decomposition of aggregate vehicle energy efficiency growth at municipality level between 2008q4 and 2010q4. For the growth decomposition at the whole state level, results are obtained by employing all the three methods, as shown in the result section.

Table 3: Two alternative methods of decomposition of aggregate vehicle energy efficiency growth

	Method 2: equation containing deviation from initial aggregate MPG	**Method 3**: equation containing deviation from average aggregate MPG between two quarters
Main references	(Baily et al., 1996; E. J. Bartelsman & Doms, 2000; Lucia Foster et al., 1998; L. Foster et al., 2006)	(Aw et al., 2001; Baily et al., 2001; E. J. Bartelsman & Wolf, 2014; Christensen, Cummings, & Jorgenson, 1981; Griliches & Regev, 1995; Olley & Pakes, 1996)
Growth of vehicle energy efficiency	$\Delta \ln MPG_t$	$\Delta \ln MPG_t$
Factor 1: within effect	$\sum_{i \in stayer} \theta_{i,t-1}(\ln MPG_{i,t} - \ln MPG_{i,t-1})$	$\sum_{i \in stayer} (\frac{\theta_{i,t} + \theta_{i,t-1}}{2})(\ln MPG_{i,t} - \ln MPG_{i,t-1})$
Factor 2: between effect	$\sum_{i \in stayer} (\theta_{i,t} - \theta_{i,t-1})(\ln MPG_{i,t-1} - \ln MPG_{t-1})$	$\sum_{i \in stayer} (\theta_{i,t} - \theta_{i,t-1})(\frac{\ln MPG_{i,t} + \ln MPG_{i,t-1}}{2} - \frac{\ln MPG_t + \ln MPG_{t-1}}{2})$
Factor 3: covariance term	$\sum_{i \in stayer} (\theta_{i,t} - \theta_{i,t-1})(\ln MPG_{i,t} - \ln MPG_{i,t-1})$	
Factor 4: entry effect vehicles coming from other municipalities	$\sum_{i \in E1} \theta_{i,t}(\ln MPG_{i,t} - \ln MPG_{t-1})$	$\sum_{i \in E1} \theta_{i,t}(\ln MPG_{i,t} - \frac{\ln MPG_t + \ln MPG_{t-1}}{2})$
Factor 5: entry effect due to newly bought vehicles or vehicles coming from other states	$\sum_{i \in E2} \theta_{i,t}(\ln MPG_{i,t} - \ln MPG_{t-1})$	$\sum_{i \in E2} \theta_{i,t}(\ln MPG_{i,t} - \frac{\ln MPG_t + \ln MPG_{t-1}}{2})$
Factor 6: exit effect due to vehicles moving to other municipalities	$-\sum_{i \in X1} \theta_{i,t-1}(\ln MPG_{i,t-1} - \ln MPG_{t-1})$	$-\sum_{i \in X1} \theta_{i,t-1}(\ln MPG_{i,t-1} - \frac{\ln MPG_t + \ln MPG_{t-1}}{2})$
Factor 7: exit effect due to vehicles moving to other states or out of service	$-\sum_{i \in X2} \theta_{i,t-1}(\ln MPG_{i,t-1} - \ln MPG_{t-1})$	$-\sum_{i \in X2} \theta_{i,t-1}(\ln MPG_{i,t-1} - \frac{\ln MPG_t + \ln MPG_{t-1}}{2})$

Source: Own elaboration

3.4 Convergence and transition of aggregate vehicle energy efficiency

Decompositions break down an effect observed in the entire population, in this case the state vehicle population, into different parts, as explained above. But they do not provide insight into what happens at the subpopulation level, i.e., within municipalities, or what are the underlying factors leading to the observed effects. In order to investigate these issues, we apply two other methods. For example, we would like to know whether the observed trends in energy efficiency, e.g., a rise or decline at the state level, are shared among all subpopulations (municipalities), or not, and why. This is what the methods introduced below will address.

3.4.1 Econometric approach

The primary focus of the analysis in this section is on the dynamic evolution of aggregate vehicle energy efficiency, investigating important features of convergence or divergence in particular, by extending well-established empirical approaches used in growth theory, to energy efficiency dynamics. There has been a debate about the advantages and disadvantages of various empirical methods and recent research effort on growth and convergence has been paid more towards the laws that generate growth distribution rather than the realizations of growth convergence (Maasoumi, Racine, & Stengos, 2007). The purpose of the analysis in this section is to shed light on the features or stylized facts about the distribution of aggregate vehicle energy efficiency, instead of testing any specific theory. Traditional empirical methods can serve this purpose properly.

The first empirical specification used in this section is motivated by a series works on examining the relationship between per capita growth rates of countries and starting level income per capita (Barro, 1991; Barro & Salaimartin, 1991, 1992; Sala-i-Martin, 1996). Using municipality level aggregate vehicle energy efficiency indicators based on vehicle level micro data, a similar empirical strategy is employed: investigating the relationship between the growth of relative aggregate vehicle energy efficiency and initial relative aggregate vehicle energy efficiency[9] in a cross section of municipalities in Massachusetts. Specifically, all the relevant variables are defined as follows:

$$Relative\ aggregate\ MPG = \frac{Municipality\ level\ aggregate\ MPG}{State\ level\ aggregate\ state\ MPG}$$

$$Growth\ of\ relative\ aggregate\ MPG = \frac{Difference\ of\ relative\ aggregate\ MPG\ between\ 2011q4\ and\ 2008q1}{Relative\ aggregate\ MPG\ in\ 2008q1}$$

The question behind this analysis is whether municipalities with low aggregate vehicle energy efficiency tend to improve faster than municipalities with high aggregate vehicle energy efficiency, i.e., whether a convergence tendency exists. The relationship between the growth of relative aggregate vehicle energy efficiency and initial relative aggregate vehicle energy efficiency is estimated by using three methods in a cross section of municipalities in Massachusetts. The first is the standard ordinary least squares regression (OLS) that estimate the average effect of the independent variable. However, the average effect estimated by OLS might not be robust or even misleading, because not all the municipalities' effects of initial relative aggregate vehicle energy efficiency are on the average level and there are some outliers. To obtain a more comprehensive picture of the relationship between the two variables of interest, the quantile regression technique for cross-sectional data developed by Koenker and Bassett (1978) is employed.[10] As addressed by Bartelsman et al. (2014), quantile regression can provide two advantages: first, the method is robust to outliers and heavy-tailed distributions (see the example of the US wage structure provided by Buchinsky, 1994), and second, quantile regression relaxes Gaussian hypotheses on the error terms, compared to classical methods based on least squares residuals (Koenker & Bassett, 1982). Specifically, this paper follows the basic idea of birth weight model example provided by Koenker and Hallock (2001): the quantile regression is applied at various quantiles of the growth of relative aggregate vehicle energy efficiency (ranging from 0 to 1), conditional on initial relative aggregate vehicle energy efficiency. Standard errors are estimated by using a bootstrapping technique.

[9] Relative to the state level aggregate vehicle energy efficiency, measured as MPG.

[10] The analysis in this section only uses cross-sectional data. The quantile regression has been extended to panel data analysis. See Bartelsman, Dobbelaere, and Peters (2014) for a discussion.

In contrast to the first two estimation methods, as another robustness check, the empirical relationship between the two variables of interest is estimated by using a nonparametric approach. The main advantage of nonparametric approach compared to parametric methods is that there is no need to assume any a priori parametric specification (misspecification of parametric models can lead to misleading results). Recent development of nonparametric estimation has extended the analysis to multivariate data (Li & Racine, 2007) and recent research on growth has applied nonparametric estimation to convergence analysis (Maasoumi et al., 2007). Specifically, this paper chooses kernel regression (using Epanechnikov kernel as the weight function) to investigate the relationship between the two variables of interest for a cross section of municipalities.[11]

3.4.2 Distribution approach

Standard econometric approaches as outlined above collapse the dynamics of aggregate vehicle energy efficiency into growth rates over the period in a cross-sectional structure. In other words, information on the dynamic characteristics is lost. To make full use of the whole distribution of the variable of interest and its dynamic behavior over time, transition matrix has been proven to be a very useful tool and has been used in the area of productivity distributions (Baily et al., 1992; Bartelsman & Dhrymes, 1998). Specifically, this paper extends the well-established distribution approach to aggregate vehicle energy efficiency distributions based on important works about income distribution literature (Quah, 1993, 1996, 1997), as shown below.

The basic data are relative aggregate MPG over the period 2008q1 – 2010q4. All the municipalities in Massachusetts are divided into ten groups on the basis of their relative aggregate MPG. The classification bounds are based on the decile of initial relative aggregate MPG, so that all the municipalities in the initial period can be divided into ten equal-sized groups (ranging from the lowest aggregate vehicle energy efficiency to the highest aggregate vehicle energy efficiency). Let F_t be the distribution of relative aggregate MPG across municipalities at time t. Then the distribution at time $t+1$ can be viewed as a standard first-order autoregression and can be described by the equation as follows:

$$F_{t+1} = M * F_t$$

Where M is a 10*10 Markov chain transition matrix mapping the distribution at time t to the distribution at time $t+1$. The value in cell (i, j) of matrix M represents the probability that a municipality in Group i moves to Group j after one quarter, $1 \leq i, j \leq 10, i, j \in N^*$.

Then the future distribution at time $t+s$ can be obtained through iteration:

$$F_{t+s} = M^s F_t$$

Letting the time difference $s \to +\infty$ yields a likely long run distribution of relative aggregate MPG across municipalities, or the Ergodic distribution. The gap in aggregate MPG between municipalities is

[11] The kernel regression is estimated by using "kernreg1" package in Stata, based on a series works by (Salgado-Ugarte, Shimizu, & Taniuchi, 1994, 1996a, 1996b).

narrowing if lots of municipalities tend to move to the group that contains the state level relative aggregate MPG, that is, 1. Polarization would occur if lots of municipalities move to distant groups in the Ergodic distribution, for example, municipalities distributed in groups with very low and very high relative aggregate MPG.

4. Results

This section presents results for the decomposition of aggregate vehicle energy efficiency at state level over the period 2008q1 to 2011q4, the decomposition of the growth of aggregate vehicle energy efficiency between 2008q4 and 2010q4, and the empirical exploration of distribution dynamics and mobility of aggregate vehicle energy efficiency. In addition, the relationship between a municipality's vehicle population and socio-economic factors, and the socio-economic factors that affect initially existing vehicles staying or exiting decision are investigated.

4.1 Decomposition of aggregate vehicle energy efficiency over the period 2008q1 – 2011q4

On the basis of Equation (3) the state level aggregate vehicle energy efficiency (in log points, measured as aggregate MPG) is decomposed into a sum of an unweighted average state level aggregate MPG and a covariance term for the period 2008q1 – 2011q4, as illustrated in Table 4 and Figure 2 below.

Overall, the shapes of state level aggregate MPG and unweighted average aggregate MPG are somehow similar: both of them have a sharp increase in 2008 and a significant drop in 2011. One possible explanation is that because of data collection many existing vehicles were not included in the dataset in early 2008 as shown by the sharp increase in vehicle population in 2008 in Figure 1, and the data quality in 2011 is not very good; lots of vehicles' miles per day data are missing. Accordingly there might be some measurement errors in early 2008 and in 2011.

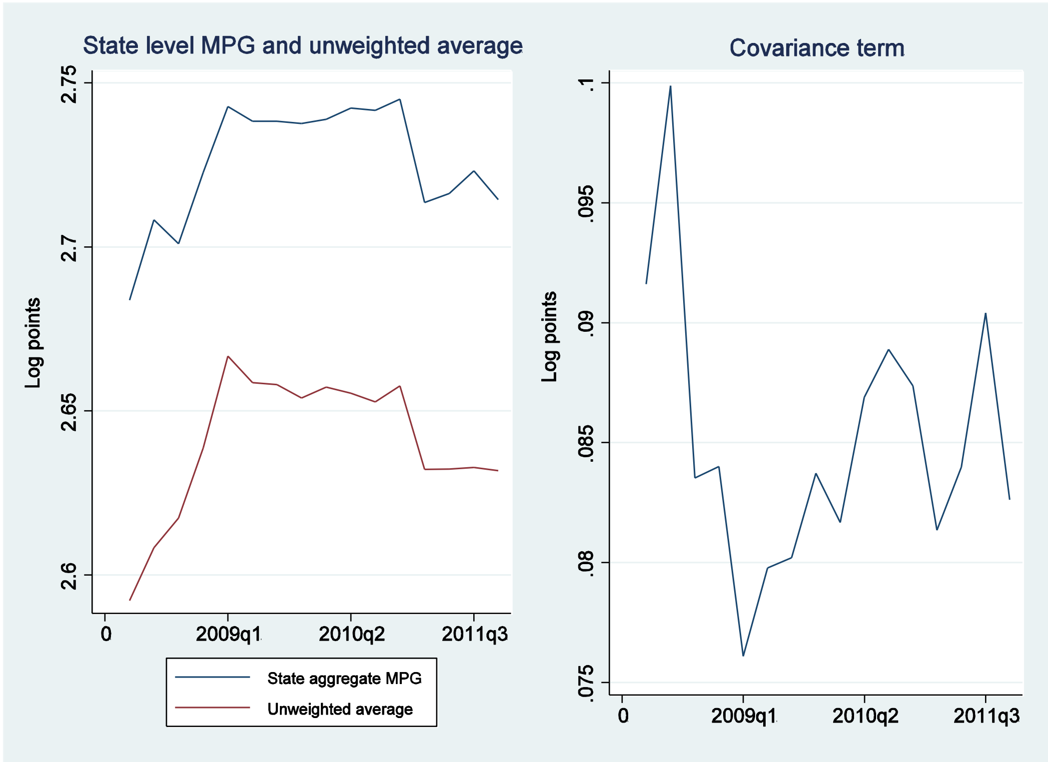
For the period 2008q4 – 2010q4, with more reliable data, the state level aggregate MPG and the unweighted average aggregate MPG show slightly different dynamic patterns. Compared to 2008q4, there is a small increase in the state level aggregate MPG in 2010q4, growing from 2.7227 to a peak of 2.7450, while a decline can be observed in between the two quarters. This indicates that the state is becoming more efficient in terms of vehicle energy efficiency during 2008q4 – 2010q4, but only to a small extent. On the other hand, the unweighted average aggregate MPG generally shows a downward trend: reducing to 2.6577 in 2010q4 after reaching its peak in 2009q1. This difference is certainly caused by the covariance term.

In general, the covariance term exhibits a cyclical dynamics and it is positive in all the quarters. The covariance term consists of two important components: one is the deviation of individual vehicle's MPG from the unweighted average aggregate MPG and the other is the deviation of individual vehicle's miles per day from the state level average. The positive covariance term clearly indicates the reallocation of high energy efficiency vehicles: the larger the covariance term, the larger share of miles per day contributed by high energy efficiency vehicles and the more energy efficient is the vehicle. The covariance term reaches its peak (0.0999) in 2008q2 and sharply decreases to its minimum level (0.0761) in 2009q1. The explanation for this significant decline might be the 2008 financial crisis that might lower purchase and usage of high energy efficiency vehicles. Between 2008q4 and 2010q4, the unweighted average aggregate MPG shows a downward trend, which generally implies the ageing of vehicle population in Massachusetts: vehicle energy efficiency is decreasing over time. The state level aggregate MPG increases by about 2.23% during this period due to the reallocation of high energy efficiency vehicles. The reallocation effect can be caused by increasing usage of existing or entering high energy efficiency vehicles, and equivalently by decreasing usage of existing or exiting low energy efficiency vehicles. However, the decomposition of state level aggregate MPG cannot provide more detailed information about the reallocation effect. Several crucial factors contributed to the growth of aggregate MPG are explored in the next section.

Table 4: Decomposition of state level aggregate MPG over the period 2008q1 - 2011q4

Quarter	$\ln MPG_t$: Log of state level aggregate MPG	$\overline{\ln MPG_t}$: Log of unweighted average state level aggregate MPG	$\sum_{i=1}^{N_t} \Delta\theta_{i,t} \Delta \ln MPG_{i,t}$: covariance term
2008q1	2.6838	2.5921	0.0916
2008q2	2.7082	2.6083	0.0999
2008q3	2.7009	2.6174	0.0835
2008q4	2.7227	2.6387	0.0840
2009q1	2.7428	2.6667	0.0761
2009q2	2.7384	2.6586	0.0798
2009q3	2.7383	2.6581	0.0802
2009q4	2.7376	2.6539	0.0837
2010q1	2.7389	2.6573	0.0817
2010q2	2.7423	2.6554	0.0869
2010q3	2.7416	2.6527	0.0889
2010q4	2.7450	2.6577	0.0874
2011q1	2.7136	2.6322	0.0814
2011q2	2.7163	2.6323	0.0840
2011q3	2.7232	2.6328	0.0904
2011q4	2.7144	2.6318	0.0826

Figure 2: Decomposition of state level aggregate MPG over the period 2008q1 - 2011q4



Source: Own elaboration.

4.2 Decomposition of growth of aggregate vehicle energy efficiency between 2008q4 and 2010q4

In this section the results for decomposition of growth of aggregate MPG between 2008q4 and 2010q4 are presented. Table 5 below shows the decomposition results at state level. Method 3 is the preferred decomposition method, but the state level results based on the other two methods are included in this table for comparison purpose.

At the state level the subsets $E1$ and $E2$ are combined in calculating an aggregate entry effect and the subsets $X1$ and $X2$ are combined in calculating an aggregate exit effect. As shown in Table 5, the growth rates of aggregate MPG using three different methods are the same, and they are very close to the actual growth, 2.26%.[12] Method 2 and Method 3 generate quite similar results, but they greatly differ from the results calculated using Method 1, except for the within effect. According to Method 1's results, the state level aggregate MPG is lowered by 8.41% due to decreasing energy efficiency of all vehicles, by 3.19% due to the ageing of the vehicle population and by 32.14% due to efficiency drain by the exiting of vehicles. But the entry of vehicles contributes a much higher efficiency gain, 45.97%, so that overall the aggregate MPG increases by 2.24% in the end quarter. The reason of such a big difference in the magnitude between factors is that weights for different quarters are used and Method 1 does not explicitly distinguish effects between high energy efficiency vehicles and low energy efficiency vehicles. Logically, lower usage or the exit of low energy efficiency vehicles should improve the aggregate MPG,

[12] There is a very small difference between the growth of aggregate MPG calculated from decomposition and the actual growth, since the difference of log points is used as an approximation of the actual growth.

but in Method 1 they are viewed as equivalent as detrimental effects caused by less usage or the exiting of high energy efficiency vehicles.

Accordingly, Method 2 and Method 3 provide more reasonable revised results that take into account different effects due to high and low energy efficiency vehicles. The high energy efficiency vehicle in Method 2 is the one whose MPG is higher than the initial level aggregate MPG, while in Method 3 the comparison threshold is the average of aggregate MPG between initial and ending levels. The conclusion changes between Method 2 and Method 3. Method 2 has an additional positive covariance term (0.0020), which means that the reallocation effect among staying vehicles contributes only 0.02% to the growth of aggregate MPG. The between effect becomes positive: 1.5% in Method 2 and 1.64% in Method 3, which refers to the increasing usage of high energy efficiency vehicles or decreasing usage of low energy efficiency vehicles among the staying vehicles. The within effect reflecting the ageing of the vehicle population is still negative, 3.19% in Method 2 and 3.09% in Method 3. The positive entry effect, 2.51% in Method 2 and 2.33% in Method 3, indicates that on average the entering vehicles have high energy efficiency, leading to an efficiency gain. Interestingly, compared to Method 1, the exit effect becomes positive, 1.22% in Method 2 and 1.35% in Method 3. This finding shows that the exiting vehicles have low energy efficiency.

Another finding based on Method 2 and Method 3 is that the within effect is much stronger than the between effect (Method 3) or the between effect and covariance term combined (Method 2). This seems somehow pessimistic, because it means that among the large number of staying vehicles the deterioration of aggregate vehicle efficiency caused by the ageing effect is more influential than the improvement due to increase of the usage of high energy efficiency vehicles, or reducing the usage of low energy efficiency vehicles. Because of efficiency gains provided by both entry effect and exit effects, however, the deterioration of aggregate vehicle energy efficiency has been offset. In other words, the net entry effect (the sum of entry effect and exit effect) is the main driving force of the growth of aggregate MPG. This finding has important policy implication related to innovation and technological change.

Table 5: Decomposition of aggregate vehicle energy efficiency at state level between 2008q4 and 2010q4

	Growth of aggregate MPG	Between effect: change in the share of miles per day	Within effect: change in the vehicles' adjusted MPG	Covariance term	Entry effect due to vehicles coming from other states or newly brought vehicles	Exit effect due to out of service or vehicles moving to other states
Method 1	0.0224	-0.0841	-0.0319		0.4597	-0.3214
Method 2	0.0224	0.0150	-0.0319	0.0020	0.0251	0.0122
Method 3	0.0224	0.0164	-0.0309		0.0233	0.0135

Source: Own elaboration.

Note that Method 2 has a between term weighted by deviation from initial aggregate MPG and a covariance term. Method 3 has a between term weighted by deviation from average over all the vehicles' MPG between two quarters.

To have a deeper and more comprehensive understanding of the vehicle energy efficiency dynamics, Method 3 is employed to decompose the growth of aggregate MPG between 2008q4 and 2010q4 for 350 municipalities in Massachusetts.[13] A summary of the decomposition results is shown in Table 6 and relevant histograms are shown in Figure 3.

Generally the findings at municipality level are consistent with the decomposition at state level. The average growth of aggregate MPG over all the 350 municipalities is about 2.38% and most municipalities have improved their aggregate MPG during the period under research. But there are certainly some outliers in the growth of aggregate MPG: the best performing municipality has a surprisingly high growth of aggregate MPG (37.65%), but the poorest performing municipality has a 15.63% decline. The standard deviation of the growth is 5.01%, much larger than its mean, 2.38%, which means there is a very big difference in the growth across municipalities. This might be indicative of a large regional imbalance of vehicle energy efficiency. The same conclusion can be found in the between effect: most municipalities have a positive between effect but a large gap exists between the best performing and the poorest performing municipalities (29.37% and -11.09%, respectively). The ageing of the vehicle population leads to a significant decline in aggregate MPG in all the municipalities. For most municipalities this decline caused by ageing of vehicle population is between -3% and -2%, and stronger than the growth contributed by the between effect. On the other hand, most municipalities also have a larger net entry effect that can compensate the energy efficiency reduction, which means the net entry effect is the main contributor of the growth of aggregate MPG. Specifically, it is evident that most municipalities actually benefit from the efficiency gain brought by vehicles entering due to purchase of new vehicles or migration of vehicles (from other municipalities or other states). Interestingly, for most municipalities the strongest factor causing efficiency drain is the exit effect due to vehicle's moving to other municipalities, probably through trade or migration. This implies that those vehicles belonging to this category are of high energy efficiency. If the vehicle is no longer in Massachusetts (for example, the vehicle moves to other states through trade or migration, or the vehicle goes out of service), the exit effect then exhibits an efficiency gain, which implies the exiting of low energy efficiency vehicles.

[13] There are 351 official municipalities in Massachusetts. But for the municipality Gosnold there are no data for calculating the growth of aggregate MPG during the period under research.

Table 6: Summary of decomposition of growth of aggregate vehicle energy efficiency on municipal level using Method 3, 2008q4 – 2010q4

	25% quantile	50% quantile	75% quantile	Mean	Std.dev.	Min.	Max.
Growth of aggregate MPG	-0.0013	0.0175	0.0427	0.0238	0.0501	-0.1563	0.3765
Between effect	-0.0003	0.0104	0.0266	0.0145	0.0317	-0.1109	0.2937
Within effect	-0.0269	-0.0260	-0.0253	-0.0258	0.0021	-0.0321	0.0000
Net entry effect	0.0200	0.0331	0.0460	0.0352	0.0332	-0.1063	0.3648
Entry effect: from other municipalities	0.0069	0.0119	0.0173	0.0125	0.0115	-0.0422	0.0747
Entry effect: from other states or newly brought vehicles	0.0172	0.0248	0.0300	0.0235	0.0179	-0.1077	0.0901
Exit effect: to other municipalities	-0.0211	-0.0149	-0.0104	-0.0162	0.0150	-0.0801	0.1116
Exit effect: to other states or out of service	0.0067	0.0123	0.0201	0.0153	0.0225	-0.0334	0.3333

Source: Own elaboration

The regional distribution of the growth of aggregate MPG at municipality level between 2008q4 and 2010q4 is visualized in the map in Figure 4. The darker (lighter) the color, the higher (lower) is the growth of aggregate MPG. Best performing municipalities seem to be distributed across the whole state and the urban areas (for example, the greater Boston area and municipalities close to Boston) seem to have an average level or below average level growth of aggregate MPG. Thus, Figure 4 gives an impression of a regional imbalance of vehicle energy efficiency, as previously shown in Table 6. It is additionally evident that poor performing municipalities (areas with light color) are distributed in most parts of the state and there is a big discrepancy in the growth among neighboring municipalities. For example, the best performing and the poorest performing municipalities are close to each other in the northwestern part of the state.

Figure 3: Histogram graphs of decomposition of growth of aggregate vehicle energy efficiency at municipality level using Method 3, 2008q4 – 2010q4

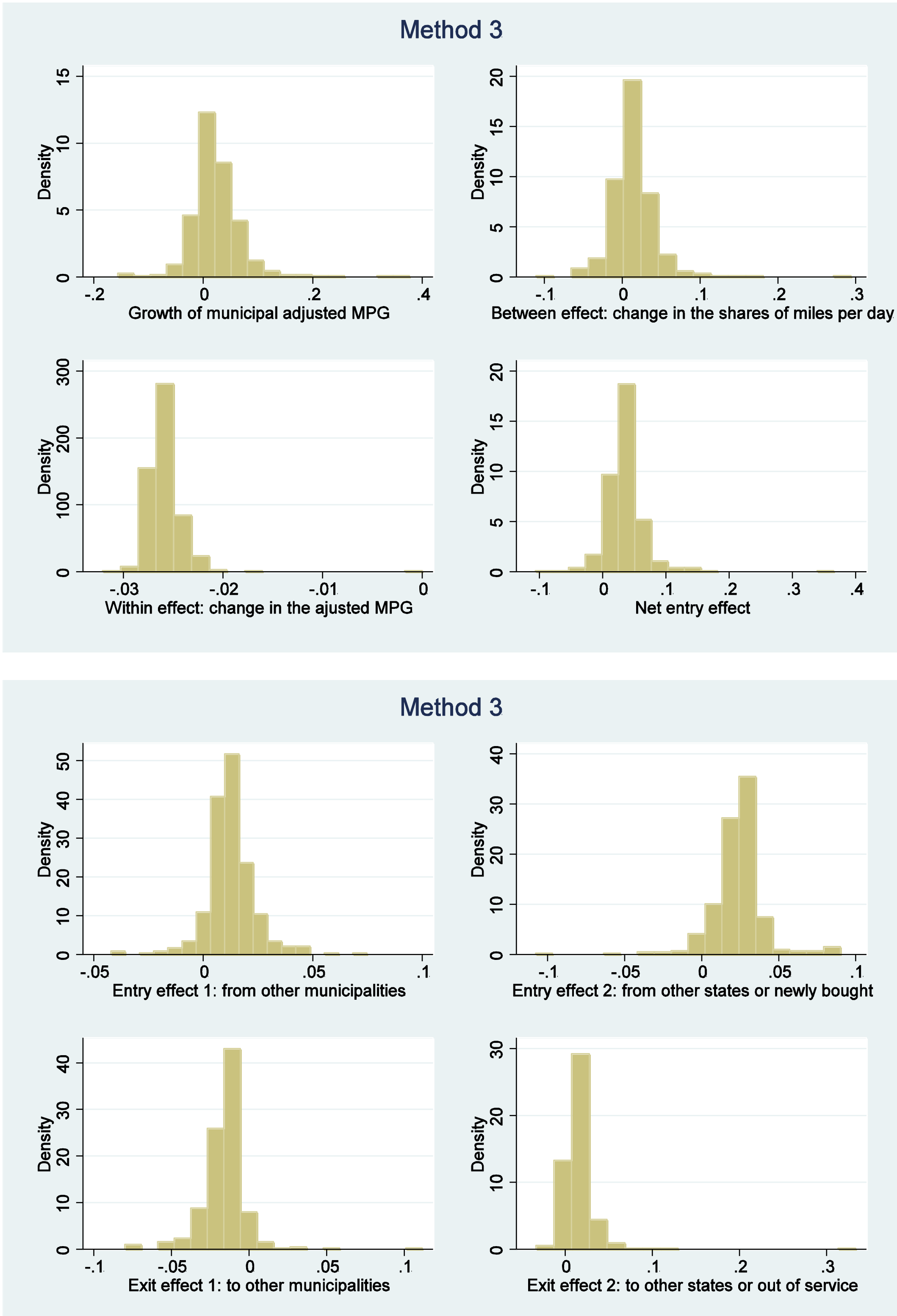
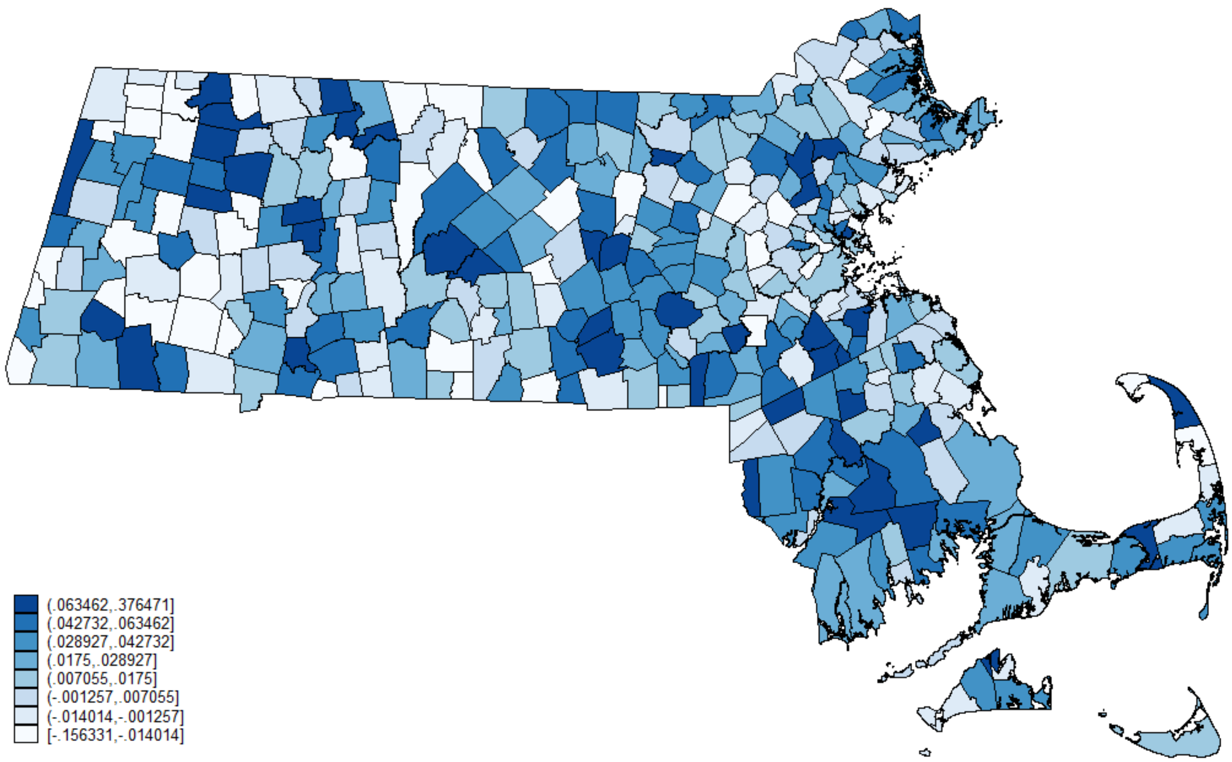


Figure 4: Massachusetts, growth of aggregate vehicle energy efficiency at municipality level between 2008q4 and 2010q4



Source: Own elaboration.

In order to get a better understanding of the variety of experiences at the municipality level, the 350 municipalities in Massachusetts are clustered into a much smaller number of groups on the basis of the results of the decomposition analysis, and a number of socio-economic characteristics of the municipalities. Municipalities that are similar to each other, in terms of these variables, are put in the same cluster. By analyzing the properties of the clusters, a better understanding of the extent and nature of variety between municipalities is obtained.

Five variables are taken into account in the clustering method: the net entry effect[14], the between effect, the within effect, the average unemployment rate between 2008 and 2010, and the average of mean household income between 2008 and 2010[15]. The introduction of the last two variables is intended to reflect job market characteristics and the living standard in the municipality, respectively. The average of

[14] The net entry effect is the sum of two entry effects and two exit effects.
[15] The municipality level unemployment data for 2008 and 2010 are available at the Massachusetts Department of Revenue. The municipality's mean household income data for 2008 and 2010 are taken from American Community Survey (ACS).

the two variables between the initial and ending years is used in the analysis. The commonly used Ward linkage is chosen to measure the dissimilarity between municipalities in clustering method. Then all the 350 municipalities are grouped into three clusters.[16]

The mean values of the standardized variables used in cluster analysis are presented by cluster in Table 7 below. The related Table 8 summarizes the mean values of the non-standardized version of all the five variables. Figure 5 is a visualization of the cluster analysis results, where municipalities that belong to the same cluster are labeled in the same color and all the three clusters are graphed onto a map of Massachusetts by using three colors (orange, green and purple).

There are several interesting findings from Table 8. The average within effect, or the ageing of vehicle population, does not significantly vary between the three clusters, and hence appears as a general tendency. Municipalities in Cluster 3 (in orange) have the highest average household income (132206.3 US dollars), the lowest average unemployment rate (5.47%), the smallest average within effect (-2.72%) as well as the lowest average net entry effect (2.72%). This is probably because well-off municipalities have fewer incentives to use energy efficient vehicles. Municipalities in Cluster 2 (in purple) have a medium level of average household income (83739.51 US dollars), but show a high level of average net entry effect (5.5%) and average between effect (4.38%), which means a large effect from restructuring of the vehicle population and high usage of energy efficient vehicles. Municipalities with the lowest average household income (Cluster 1, in green) have the lowest average between effect (only 0.12%), indicating the low usage of high energy efficiency vehicles, but the average net entry effect in Cluster 1 is not so different from that in Cluster 3 (2.85% for Cluster 1 and 2.72% for Cluster 3). Low income and high unemployment rate probably restrict the adoption and usage of (new) energy efficient vehicles.

In Figure 5 a clear geographical pattern is that the eastern part of Massachusetts is mainly dominated by municipalities belonging to Cluster 1 and Cluster 3. For example, the Boston area belongs to Cluster 1 and it is surrounded by rich municipalities belonging to Cluster 3. This may be explained by counter-urbanization. On the whole, municipalities in Cluster 3 with lower average net entry effect and average between effect are found more often in the eastern part of the state. As a result, the eastern part of the state generally has a lower growth rate of aggregate vehicle energy efficiency. This finding is consistent with Figure 4 where the eastern part of the map is lighter.

Table 7: Mean value of the five variables used in cluster analysis

Cluster	Standardized net entry effect	Standardized between effect	Standardized within effect	Standardized average unemployment rate between 2008 and 2010	Standardized average of the mean household income between 2008 and 2010
1	-0.2017	-0.4197	0.1701	0.1629	-0.4205
2	0.5974	0.9240	0.3114	0.3829	-0.3146
3	-0.2421	-0.1628	-0.6672	-0.7295	1.1628

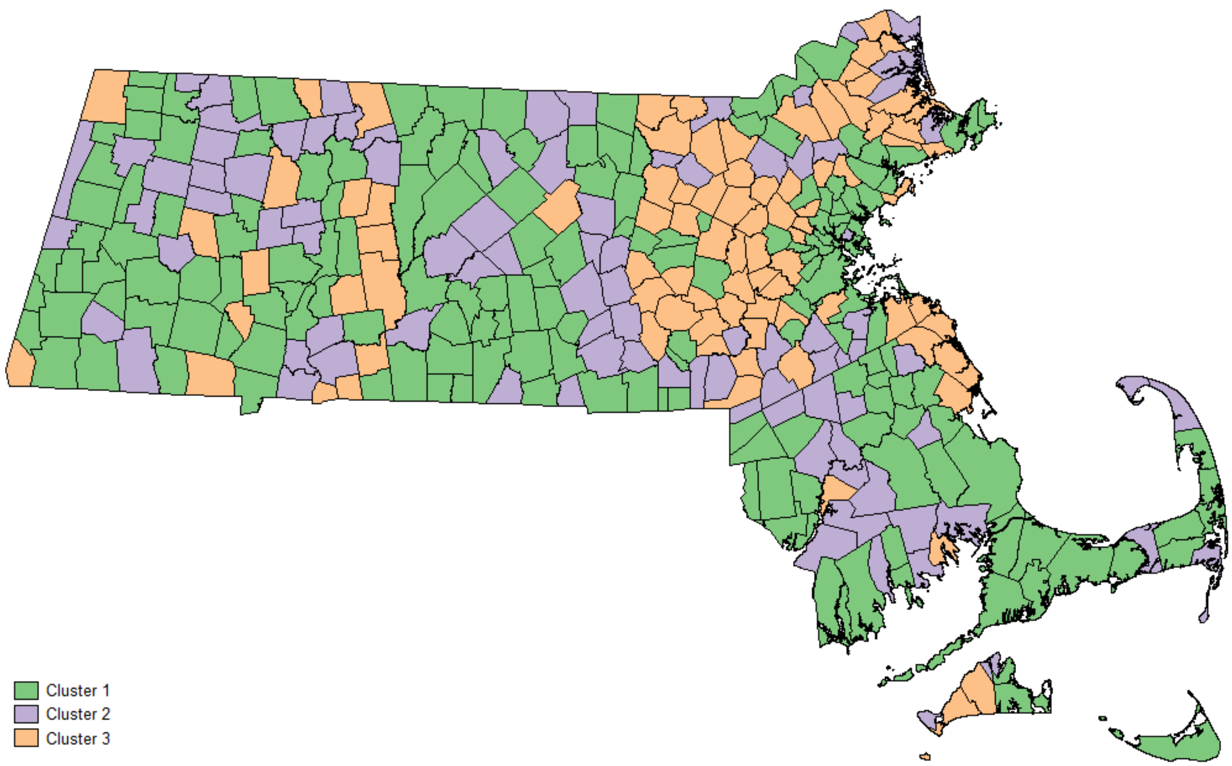
[16] The dendrogram of the hierarchical clustering analysis shows it is reasonable to group all the municipalities in to three clusters.

Table 8: Mean value of the five variables by cluster

Cluster	Average net entry effect (%)	Average between effect (%)	Average within effect (%)	Average unemployment rate between 2008 and 2010 (%)	Average of the mean household income between 2008 and 2010 (US$)
1	0.0285	0.0012	-0.0255	0.0742	80263.5500
2	0.0550	0.0438	-0.0252	0.0791	83739.5100
3	0.0272	0.0093	-0.0272	0.0547	132206.3000

Source: Own elaboration.

Figure 5: Massachusetts, municipality clusters based on five factors between 2008 and 2010



Source: Own elaboration.

Thus, the hierarchical clustering analysis in this section clearly shows that there are certain distributional patterns associated with municipal socio-economic characteristics hidden behind the growth decomposition results: reallocation effect and usage of energy efficient vehicles generally tend to be lower in both very rich and very poor municipalities. The growth decomposition results focused on the contributions of micro level vehicle related variables, but they cannot tell us whether and how these decomposition results differentiate across municipalities with various socio-economic characteristics. In fact, the socio-economic characteristics might affect people's behaviors regarding the movement and usage of vehicles, leading to significant differentials in decomposition factors across municipalities and then in the growth rates of aggregate MPG.

4.3 Convergence and transition of aggregate vehicle energy efficiency

The focus of the analysis now shifts to the mobility and distributional dynamics of municipality level aggregate vehicle energy efficiency, in order to obtain further insights into the variety of experiences in subpopulations of vehicles. Table 6 and Figure 4 show results for the period 2008q4 - 2010q4, and during this period an imbalanced development of aggregate vehicle energy efficiency is found. But the results shown so far do not illustrate how this trend is shared or differs between municipalities. In particular, do the effects related to vehicle mobility (between municipalities), as well as the other effects, lead to a convergence municipalities in terms of vehicle energy efficiency, or not?

First, we collapse the dynamic characteristics of MPG at the municipality level into a single summary statistic, that is, the growth of relative aggregate MPG between the initial and end quarter, and then regress this relative growth on the initial level of relative aggregate MPG. The basic idea underlying here is that a tendency for narrowing the gap of MPG between municipalities results if municipalities that have low aggregate MPG at the beginning, improve their vehicle energy efficiency faster than the municipalities that are initially energy efficient. Second, once the municipalities have been ranked by their aggregate MPG relative to the state level, in each quarter, they can be classified into quantile groups that have fixed bounds, and a transition matrix can be obtained that provides useful information about the long-run dynamics of the distribution. As emphasized by Quah (1993), however, such a steady-state distribution should not be interpreted as a precise forecast, since random shocks (for example, sudden change in government policy) are not taken into account.

Table 9 shows the results regarding the relationship between the initial level of relative aggregate MPG and the growth of relative aggregate MPG. The OLS regression in Column 1 is the basic estimation of this relationship. Since the average effect of the initial level of relative aggregate MPG estimated by OLS is not robust when there are outliers in the dataset, quantile regression is employed as a robustness check. Columns 2 – 4 in Table 9 show the results of quantile regression at 25%, 50% (median) and 75% quantiles of the dependent variable. Table 9 clearly gives some evidence that supports the convergence of relative aggregate MPG: the coefficients of the initial level relative aggregate MPG are negative and significant at 5% significance level in OLS and all the three quantile regressions, which implies that municipalities with lower initial level relative aggregate MPG tend to have a higher growth of relative aggregate MPG. Interestingly, when the quantile used in the quantile regression (25%, 50% and 75% quantile) increases, the coefficient of interest keeps decreasing (-0.3, -0.4653, and -0.6648, respectively).

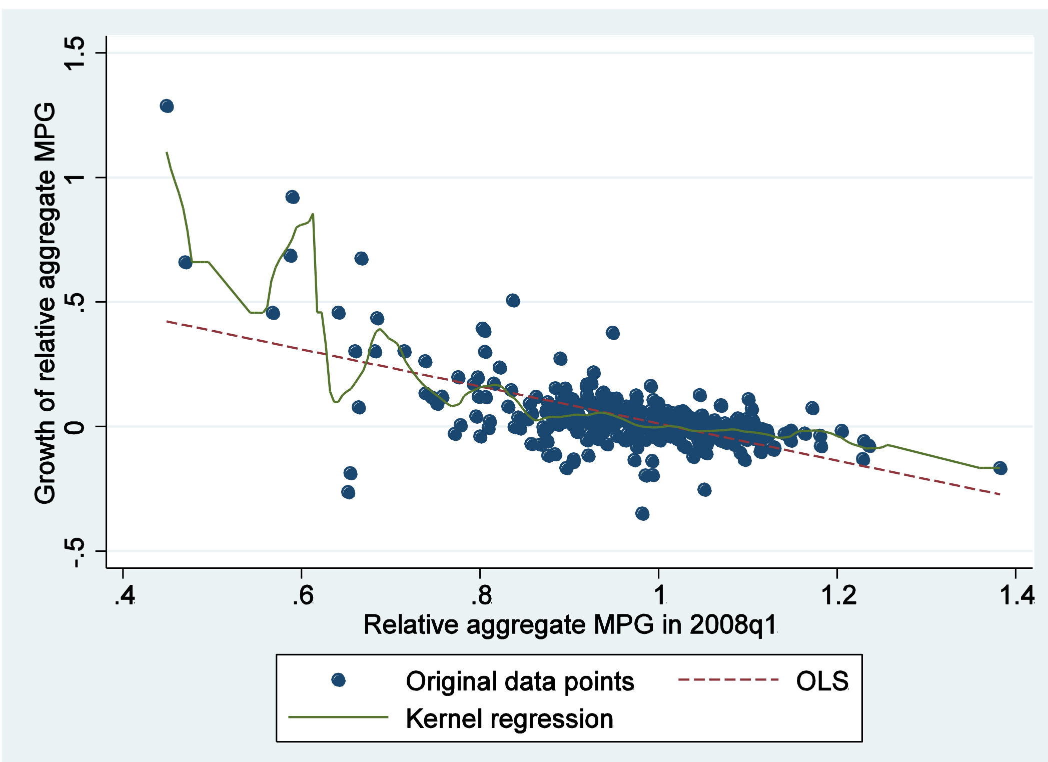
The graphical evidence obtained by using OLS and nonparametric approach for the convergence of relative aggregate MPG is illustrated in Figure 6. As in Table 9, the fitted line estimated by OLS shows a downward sloping trend. As another robustness check, the curve estimated by using nonparametric approach (kernel regression) is illustrated in Figure 6 as well. In general, the curve is downward sloping which indicates that municipalities with higher initial level of relative aggregate energy efficient vehicles have a lower growth rate. In the interval where the relative aggregate MPG is below 0.8, there is a big discrepancy in the growth of relative aggregate MPG across municipalities, so that the kernel regression curve significantly fluctuates. Among the municipalities whose relative aggregate MPG are higher than 0.8, the gap in the growth is much smaller and accordantly the kernel regression curve is smoothly downward sloping. To conclude, finding from nonparametric approach is consistent with the one from OLS, and convergence between municipalities in terms of the energy efficiency of their vehicle populations is observed.

Table 9: The relationship between relative aggregate MPG in 2008q1 and growth of relative aggregate MPG; OLS and quantile regression at 25%, 50% and 75% quantile

Variable	OLS	Quantile regression 25%	Quantile regression 50%	Quantile regression 75%
Relative aggregate MPG in 2008q1	-0.7442*** (0.0505)	-0.3000*** (0.0437)	-0.4653*** (0.0344)	-0.6648*** (0.0498)
Constant	0.7565*** (0.0494)	0.2773*** (0.0427)	0.4781*** (0.0337)	0.7151*** (0.0488)
Number of observations	350	350	350	350
Standard errors in parentheses; * p<.05, ** p<.01, *** p<.001				

Source: Own elaboration.

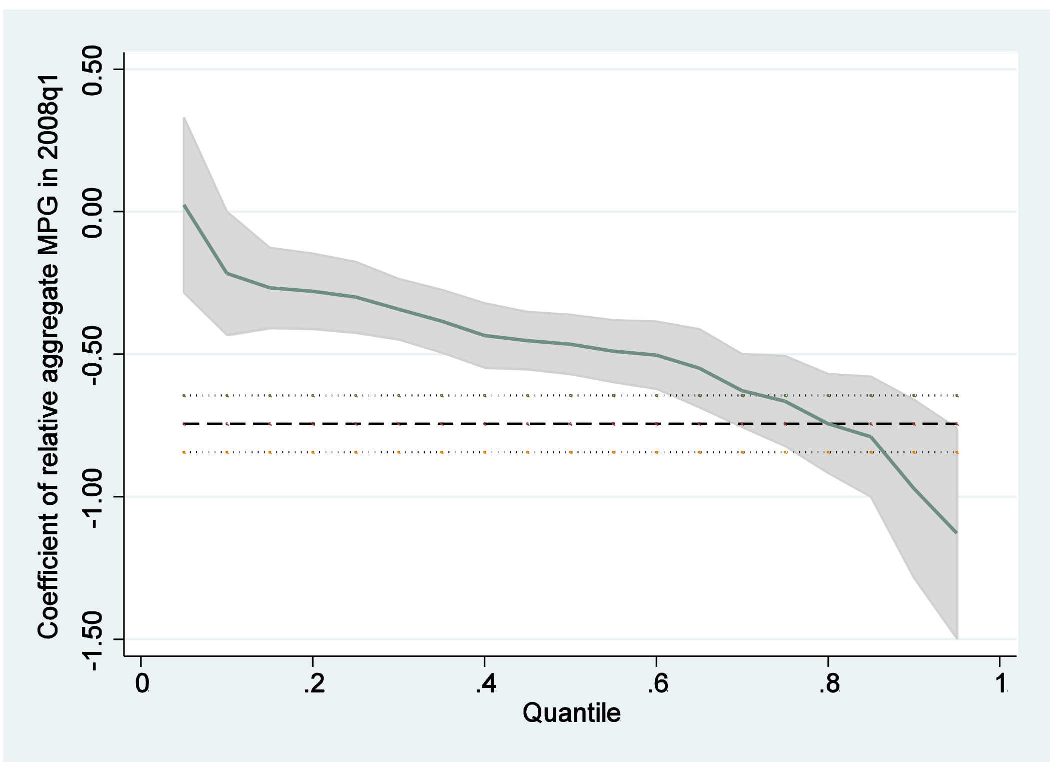
Figure 6: The relationship between relative aggregate MPG in 2008q1 and growth of relative aggregate MPG; OLS and nonparametric approach



Source: Own elaboration.

Table 9 provides the quantile regression results at three representative quantiles only. A more comprehensive and robust examination of the empirical relationship is provided by applying bootstrapping quantile regression at various quantile from 0 to 1. Figure 7 below illustrates the graphic results of bootstrapping quantile regression and the OLS as a comparison. The horizontal axis represents the quantile of the dependent variable and the vertical axis stands for the coefficient of the initial level relative aggregate MPG estimated by quantile regression at a given quantile. It is evident that the coefficient estimated by quantile regression at any quantile is negative and with the increase in quantile used in quantile regression, the coefficient of interest is constantly decreasing (i.e., the convergence effect is getting stronger). These findings are consistent with Table 9 as well. When regressing at low and high quantiles (below 0.2 and above 0.8, respectively), there is a sharp decline in the coefficient. The coefficient estimated by OLS is roughly similar to the one obtained by quantile regression at 80% quantile. The bootstrapping quantile regression results are very robust and provide strong support to the convergence hypothesis. Municipalities with low initial aggregate MPG actually improve much faster, while those initially energy efficient municipalities have a smaller growth, as compared with the OLS estimate.

Figure 7: The relationship between relative aggregate MPG in 2008q1 and growth of relative aggregate MPG; OLS and bootstrapping quantile regression



Source: Own elaboration.

Notes: The shadow area indicates the 95% confidence intervals of quantile regression; the solid line presents coefficients of quantile regression; "---" stands for coefficient estimated by using OLS and "…" marks the 95% confidence interval of OLS. The number of bootstrapping replications is 500.

Turning to the Markov analysis, the municipalities, in each quarter, are grouped into ten groups and the group bounds are based on the initial level relative aggregate MPG, as shown in Table 10 below. Group 1 contains the municipalities whose relative aggregate MPG are smaller than (or equal to) 10% quantile of initial level relative aggregate MPG (0.8228), indicating the lowest vehicle energy efficiency group. Group 2 then consists of the municipalities that have a relative aggregate MPG between 10% and 20% quantile of the initial level, and so on and so forth. Group 10 the highest vehicle energy efficiency group. All the ten groups are then sorted from the lowest to the highest vehicle energy efficiency. Note that only in the initial period, the groups will be equal in terms of the number of municipalities.

Table 10: Group bounds based at municipality level relative aggregate vehicle energy efficiency in 2008q1, relative to the state level

Group 1: 0% - 10% quantile, (-∞, 0.8228]	Group 6: 50% - 60% quantile, (0.9795, 1.0011]
Group 2: 10% - 20% quantile, (0.8228, 0.8944]	Group 7: 60% - 70% quantile, (1.0011, 1.0363]
Group 3: 20% - 30% quantile, (0.8944, 0.9314]	Group 8: 70% - 80% quantile, (1.0363, 1.0625]
Group 4: 30% - 40% quantile, (0.9314, 0.9622]	Group 9: 80% - 90% quantile, (1.0625, 1.1057]
Group 5: 40% - 50% quantile, (0.9622, 0.9795]	Group 10: 90% - 100% quantile, [1.1057, +∞)

Source: Own elaboration.

Table 11 provides the estimated one-period transition matrix of relative aggregate MPG over the period 2008q1 – 2011q4. The row header indicates the group of relative aggregate MPG in time t and the column header is the group of relative aggregate MPG in time $t+1$. The last row shows the steady-state distribution of municipalities in the long run, that is, the Ergodic distribution. The last column shows the total number of municipalities in a certain group over the entire period. In the 10*10 one period transition matrix, the value in (i, j) indicates the probability that a municipality in Group i in time t ends in Group j in time $t+1$, $1 \leq i, j \leq 10$ and $i, j \in N^*$. For example, 0.7765 in (1, 1) means the estimated probability that a municipality in Group 1 will still stay in Group 1 in the next quarter is 0.7765, while 0.1647 in (1, 2) means the estimated probability that a municipality in Group 1 will move up to Group 2 in the next quarter is 0.1647.

In the 10*10 one period transition matrix the diagonal ranges from 0.5149 to 0.7972, which means more than half municipalities will stay in the same group in the next quarter. The municipalities in Group 5 are the most dynamic, showing the lowest persistence of staying in the same group. The group with the highest persistence is found to be Group 10 that is the most vehicle energy efficient. In one period, transition municipalities have a small probability to move up or move down to neighboring groups (ranging from 0.0706 to 0.2319), but have a very low probability to transit to the groups far away from their original groups. For example, municipalities put in the group with the lowest vehicle energy efficiency have a probability of 0.1647 to upgrade their vehicle energy efficiency to the second group, but only have a probability of 0.0235 to be successful in moving up to Group 3; less than 1% of them manage to have a better improvement.

Table 11: One period transition matrix of municipality level relative aggregate vehicle energy efficiency; relative to the state level; 2008q1 – 2011q4

Group 1: 0% - 10% quantile, (-∞, 0.8228], the lowest energy efficiency group;

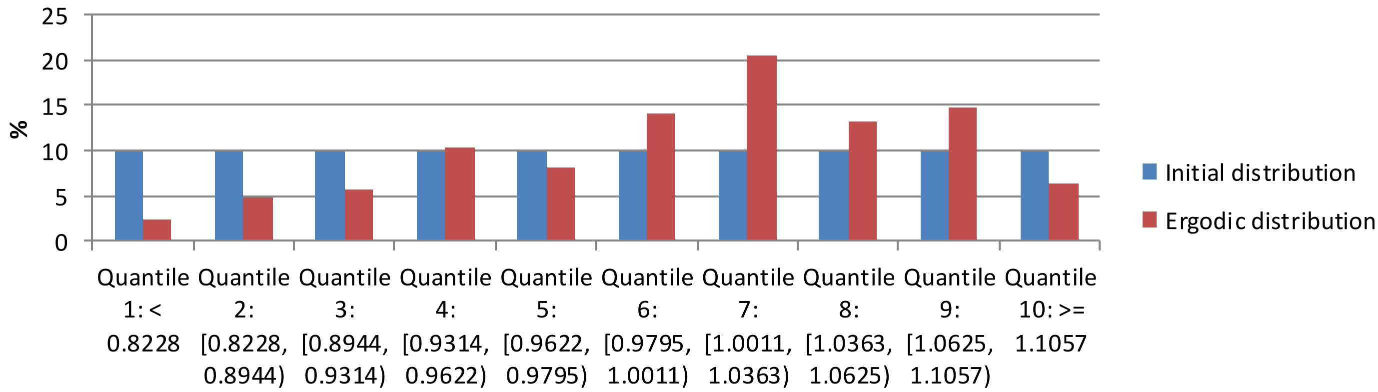
Group 10: 90% - 100% quantile, [1.1057, +∞), the highest energy efficiency group.

	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9	Group 10	Total number
Group 1	0.7765	0.1647	0.0235	0.0078	0.0078	0.0078	0.0039	0.0039	0.0039	0.0000	255
Group 2	0.0776	0.7153	0.1624	0.0353	0.0024	0.0024	0.0000	0.0024	0.0000	0.0024	425
Group 3	0.0137	0.1167	0.6339	0.2014	0.0206	0.0023	0.0092	0.0023	0.0000	0.0000	437
Group 4	0.0047	0.0187	0.0950	0.6573	0.1573	0.0436	0.0171	0.0031	0.0016	0.0016	642
Group 5	0.0000	0.0064	0.0170	0.1574	0.5149	0.2319	0.0638	0.0021	0.0043	0.0021	470
Group 6	0.0000	0.0014	0.0028	0.0408	0.1254	0.6437	0.1704	0.0099	0.0056	0.0000	710
Group 7	0.0011	0.0011	0.0034	0.0114	0.0182	0.1243	0.7184	0.1060	0.0125	0.0034	877
Group 8	0.0000	0.0018	0.0000	0.0054	0.0018	0.0054	0.1720	0.6900	0.1147	0.0090	558
Group 9	0.0000	0.0000	0.0017	0.0017	0.0017	0.0017	0.0202	0.1109	0.7916	0.0706	595
Group 10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0107	0.0036	0.1886	0.7972	281
Ergodic	0.0232	0.0475	0.0566	0.1025	0.0814	0.1410	0.2053	0.1311	0.1482	0.0630	

Source: Own elaboration.

Figure 8 provides a comparison between the initial distribution and the Ergodic distribution. According to the grouping criteria all the municipalities are put into 10 equal-sized groups, so that in the initial distribution each group contains 10% municipalities. In the long run steady state, however, the distribution changes significantly and provides evidence of improvement in vehicle energy efficiency. The shares of municipalities in the three lowest vehicle energy efficiency groups greatly reduce from 10% to around 5% or much lower than 5%. The largest reduction of the share of municipalities is found in Group 1, the lowest vehicle energy efficiency group: the share of municipalities reduces to 2.32%. On the other hand, most municipalities concentrate in high vehicle energy efficiency groups. In the long run steady state, about 20.53% of total municipalities are in Group 7 whose vehicle energy efficiency is slightly higher than the state average level.[17] About 14.1% municipalities are in Group 6 whose group bounds cover 1, the state average level. Convergence mainly takes place in these two groups (6 and 7), i.e., this is where most municipalities move. For the higher efficiency Groups 8 and 9, the shares of municipalities increase to 13.11% and 14.82%, respectively, while a small decline in the share of municipalities is observed in the highest vehicle energy efficiency group. Thus, long-run convergence does not seem to be complete. For high vehicle energy efficiency groups (Groups 7, 8, 9 and 10 that are absolutely more energy efficient than the state average level), the total share of municipalities is 54.76% in the long run steady state. As compared to the total share, 40%, in the initial distribution, it is evident that there is an improvement in the vehicle energy efficiency since more than half municipalities' vehicle energy efficiency is higher than the state average level.

Figure 8: Comparison between initial distribution and Ergodic distribution



Source: Own elaboration.

4.4 Number of vehicles

The purpose of this section is to characterize how socio-economic patterns influence the distribution of vehicles across municipalities over the entire period. The results can provide useful information to deepen the knowledge regarding the role of socio-economic factors in shaping the distribution of vehicles. Unlike the cluster analysis employed in Section 4.2 that shows the cross-sectional distribution patterns by collapsing the dynamic process into several decomposition factors, the focus of the analysis in this section is on the dynamic process of the distribution of vehicles. Following the analysis in the previous section that studies the distribution of aggregate vehicle energy efficiency, this section studies the distribution of vehicles measured by the number of vehicles across municipalities. The movement of vehicles across municipalities and states (due to migration or trade) causes changes in the distribution of vehicles across municipalities, that is, upsizing (due to net inflow of vehicles) or downsizing (due to net outflow of vehicles) of the vehicle population at the municipalities level, which eventually shapes the distribution of aggregate (state level) vehicle energy efficiency, as shown in the decompsitions.

[17] The state level vehicle energy efficiency is 1.

Panel data over the period 2008 to 2011 are used. The dependent variable is the number of vehicles in the municipality in 2008q4, 2009q4, 2010q4 and 2011q4.[18] Three basic socio-economic variables are used as explanatory variables: municipality level income (measured by median household income and mean household income), population and unemployment rate over 2008 – 2011. The preferred empirical method employed in this section is the random-effects negative binomial regression, since the dependent variable is a count variable. Two commonly used econometric approaches allow researchers to deal with count dependent variables, that is, negative binomial regression and Poisson regression. The negative binomial regression with panel data is used because the mean and variance of the dependent variable are too different, which violates the basic assumption of Poisson regression that requires that the mean and variance of the dependent variable are equal.

Table 12 summarizes the regression results. In Model 1 the municipality level mean household income is used, while it is replaced with the municipality level median household income in Model 2 as a comparison specification. $\ln\_r$ and $\ln\_s$ present the estimation of r and s, respectively, which are parameters used in the regression: the assumption of the negative binomial regression is that the inverse of one plus the dispersion follows a $Beta(r, s)$ distribution.

Table 12: Results for random-effects negative binomial regression, using number of vehicles as dependent variable, 2008 – 2011

	Model 1	Model 2
Median household income (in logs)		0.4743***
		(-0.0354)
Mean household income (in logs)	0.5740***	
	(-0.0464)	
Population (in logs)	0.9874***	1.0049***
	(-0.0107)	(-0.0087)
Unemployment rate	-1.7077***	-1.4456***
	(-0.2449)	(-0.2483)
Constant	-10.9621***	-10.0453***
	(-0.4838)	(-0.393)
ln_r	2.8937***	3.2830***
_cons	(-0.1289)	(-0.1102)
ln_s	6.7784***	7.3204***
_cons	(-0.164)	(-0.1347)
Number of observations	1404	1404
Number of municipalities	351	351

Standard errors in parentheses; * p<.05, ** p<.01, *** p<.001

Source: Own elaboration.

[18] Because the vehicle census data are recorded quarterly, while the municipality level socio-economic data are recorded yearly.

Both models are significant in the Wald test at 1% significance level (all the p-values < 1%) and both panel estimators are significantly different from the pooled estimators in the likelihood-ratio test (p-values < 0.1%). Population size is the variable that picks up the effect of differences in the scale of municipalities. In both models the coefficients of income (mean household income and median household income) and population are positive while the coefficients of unemployment rate are negative. The results clearly support that the municipalities that are richer, more populated and have lower unemployment rates are more likely to have a larger vehicle population.

4.5 Determinants of Exit from the state

Finally, the analysis turns to the factors that determine the exit of vehicles from the state population. The results in Sections 4.2 and 4.4 illustrated some municipality-level distributional patterns correlated to socio-economic factors, but we do not know yet how these socio-economic factors are associated with vehicle level exit or entry. The decision of entry or exit, i.e., purchase of sale of a vehicle is the ultimate micro foundation of the aggregate level distributional patterns. The analysis will here be focused on the exit vs. stay decision, at the state level, and does not take into account the decision of moving to other municipalities within the same state. The results are of interest because the existing vehicle population is ageing over time, which leads to a decreasing aggregate vehicle energy efficiency, and the second hand market for vehicles can prolong the presence of these ageing vehicles in the state population. The crucial point is that the consequence of vehicle exiting is an update of population level vehicle energy efficiency.

The initially existing vehicles are defined as the vehicles that were in Massachusetts (that is, in the dataset) in 2008q4.[19] The exit event is defined on whether the vehicle is in Massachusetts or not in 2010q4. If the initially existing vehicle is still in the state in the end quarter, it is classified as a stayer, otherwise it exits. Three municipality characteristics for the year 2008 are used: household income (median household income and mean household income), population and unemployment rate. In addition, other control variables include some vehicle characteristics such as MPG2008, original retail price, age, vehicle type dummies, fuel dummies and a dummy for whether the vehicle is a commercial vehicle or a passenger vehicle.

The econometric approach used is Probit regression. A value of 1 for the dependent variable means that the vehicle is a stayer. The main results are shown in Table 13. In Models 1 and 3 the mean household income is used, while the income variable is replaced by the median household income in Models 2 and 4. Models 1 and 2 only use the observations that contain precise municipality information. Models 3 and 4 include additional observations that have a 0 in the municipality ID in the end quarter.[20] So there are more observations used in Models 3 and 4. All the four models are significant in the Wald test at 1%

[19] According to Figure 1, there is a sharp increase in the total number of vehicles in Massachusetts between 2008q1 and 2008q4 probably due to data collection. Many vehicles that already existed might enter the dataset for the first time in a later quarter. From 2008q4 onwards the total number of vehicles is less fluctuant. Additionally, only yearly data for socio-economic variables are available. 2008q4 is a proper choice of initial quarter.

[20] In the dataset if the municipality ID is 0, it means the vehicle is in the state but the location of the vehicle cannot be coded using a proper municipality ID.

significance level (P-values < 1%) and have a high rate of correct classification (84.9% and 85.45%). There are two crucial findings.

All municipality level socio-economic variables are very significant and all the four models provide consistent results. The coefficient of household income (including median household income and mean household income) is positive, which means the vehicles are more likely to stay are initially located in rich municipalities. The possible explanation is that rich municipalities buy more high-quality vehicles that last longer. This is a similar effect to the significant and positive coefficient of the vehicle's original retail price. The coefficient of unemployment rate is positive as well, which means the vehicles initially located in the municipalities with higher unemployment rates have a higher possibility of staying in the state. This is reasonable because in the municipalities with high unemployment rates, the prospect of consumer's well-being is lower, which would restrict the trade of vehicles and eventually the update of vehicle population. The coefficient of population is negative, indicating that the vehicles located in less populated municipalities tend to stay in the state, possibly as an indication of differences between urban and rural areas.

The coefficient of MPG2008 is also positive, while the coefficient of vehicle age is significant and negative. This means the vehicles that have higher initial energy efficiency and more recent model design are more likely to stay in the state. In addition, SUVs, trucks, vans and wagons tend to stay in the state as compared to other cars. Commercial vehicles are more likely to stay, compared to non-commercial passenger vehicles. The vehicles consuming flexfuel and gasoline tend to be more durable, compared to the ones using diesel.

Table 13: Probit regression results for initially existing vehicle's staying or exiting decision

Dependent variable: 1 means staying in the state; 0 indicates exiting to other states.

	Model 1	Model 2	Model 3	Model 4
Median household income (in logs)		0.0885***		0.0768***
		(-0.0051)		(-0.005)
Mean household income (in logs)	0.0721***		0.0674***	
	(-0.0049)		(-0.0048)	
Unemployment rate	0.3307***	0.4666***	0.3013***	0.3627***
	(-0.0767)	(-0.0761)	(-0.0754)	(-0.0748)
Population (in logs)	-0.0277***	-0.0232***	-0.0301***	-0.0266***
	(-0.0009)	(-0.001)	(-0.0009)	(-0.001)
MPG2008	0.0111***	0.0112***	0.0102***	0.0104***
	(-0.0004)	(-0.0004)	(-0.0004)	(-0.0004)
Original retail price (in logs)	0.0900***	0.0908***	0.0953***	0.0965***
	(-0.0044)	(-0.0044)	(-0.0043)	(-0.0043)
Age	-0.0413***	-0.0412***	-0.0372***	-0.0371***
	(-0.0003)	(-0.0003)	(-0.0003)	(-0.0003)
Type of vehicles dummies (reference type: car)				
SUV	0.0577***	0.0581***	0.0495***	0.0499***
	(-0.003)	(-0.003)	(-0.003)	(-0.003)
Truck	0.1521***	0.1524***	0.1380***	0.1382***
	(-0.0042)	(-0.0042)	(-0.0041)	(-0.0041)
Van	0.0666***	0.0666***	0.0616***	0.0616***
	(-0.004)	(-0.004)	(-0.004)	(-0.004)
Wagon	0.3570*	0.3577*	0.3410*	0.3414*
	(-0.169)	(-0.1691)	(-0.1683)	(-0.1684)
Type of fuels dummies (reference type: diesel)				
Flexfuel	0.0942***	0.0943***	0.0796***	0.0798***
	(-0.0148)	(-0.0148)	(-0.0146)	(-0.0146)
Gasoline	0.0698***	0.0701***	0.0622***	0.0627***
	(-0.0135)	(-0.0135)	(-0.0133)	(-0.0133)
Other	-0.0812	-0.0805	-0.1005*	-0.0998*
	(-0.0429)	(-0.0429)	(-0.0425)	(-0.0425)
Commercial vehicles dummy	-0.7168***	-0.7159***	-0.7251***	-0.7243***
(1: commercial vehicle; 0: passenger vehicle)	(-0.0032)	(-0.0032)	(-0.0032)	(-0.0032)
Constant	-0.4011***	-0.6360***	-0.3475***	-0.4936***
	(-0.0788)	(-0.0824)	(-0.0773)	(-0.0808)
Number of observations	2676807	2676807	2776368	2776368
Correctly classified	84.90%	84.90%	85.45%	85.45%

Standard errors in parentheses; * p<.05, ** p<.01, *** p<.001

Source: Own elaboration.

5. Conclusions

This paper has manifested that longitudinal micro data can provide very useful information for obtaining a more accurate aggregate indicator of population-level vehicle energy efficiency, and for answering questions that are related to heterogeneity in the vehicle population. As one of the first few studies of energy efficiency built on micro level evidence, this paper benefits from the newly released big data from Massachusetts Vehicle Census that allows researchers to track the energy efficiency and location of almost every vehicle in Massachusetts over time. This paper is not only a recent application of extending the well-established approaches in productivity dynamics and growth theory to the field of energy efficiency dynamics, but also an example of providing comprehensive information that supports relevant policymaking in a data-intensive context.

For the issues related to energy efficiency and sustainable development, this paper provides some evidence that the dynamics of aggregate vehicle energy efficiency in Massachusetts are mainly driven by the reallocation of high energy efficiency vehicles. Specifically, the main findings can be summarized as follows. Using the aggregate vehicle energy efficiency based on vehicle level micro data, this paper (1) decomposes the state level aggregate vehicle energy efficiency over the entire period into an unweighted average and a covariance term between vehicle level energy efficiency and share of miles per day; and (2) decomposes the growth of aggregate vehicle energy efficiency between 2008q4 and 2010q4 into a between effect, a within effect and a net entry effect aggregating two entry effects and two exit effects (at both state level and municipality level). The covariance term in the decomposition of time series is found to be positive in each quarter in the dataset, indicating that higher (lower) usage is found for higher (lower) energy efficiency vehicles. In other words, drivers choose energy-0efficient vehicles when they have to drive a lot. The growth decomposition at both state level and municipality level provides a consistent finding that the most important factor positively contributing to the growth of aggregate vehicle energy efficiency is the net entry effect, the reallocation due to the entering of high energy efficiency vehicles and the exiting of low energy efficiency vehicles.

The decomposition of changes (growth rates) of vehicle efficiency also suggests that the reallocation effect among continuing vehicles is important. These results show a positive between effect at the state level and in most municipalities, which implies more (less) miles per day are allocated in high (low) energy efficiency vehicles. The energy efficiency growth gain due to changes in consumer behaviors, however, is much less than the energy efficiency growth loss due to vehicle's ageing measured by the within effect. For instance, at the state level the between effect can only compensate slightly more than half of energy efficiency loss caused by the ageing of vehicles. The reallocation due to vehicle's entering and exiting then fills the energy efficiency gap and leads to growth.

Also, this paper provides an empirical framework to analyze the update process of the vehicle (sub)population(s). This is an important determinant of overall energy efficiency: as a non-production unit, a vehicle's energy efficiency is decreasing as time progresses. The old vehicles can be replaced by newer vehicles that have higher energy efficiency. Without product substitution, the update process can occur through migration of vehicles, causing either upsizing or downsizing of the vehicle population. Looking into all the four components of net entry effect, the interesting finding is that besides the two entry effects bringing energy efficiency gain, the exit effect can also be positive if inefficient vehicles exit the state (going out of service or moving to other states). However, an energy efficiency drain can result if high energy efficiency vehicles move.

In general, the aggregate vehicle energy efficiency at state level shows an upward sloping trend through 2008q4 and 2010q4, but this is only a small effect. A statistical investigation of the growth of vehicle energy efficiency and its decomposition at the sub-state level shows an imbalanced development across municipalities. Rich municipalities located in the eastern side of the state tend to have the smallest reallocation effect related to vehicle's entering and exiting. However, various econometric approaches provide robust evidence that the gap in aggregate vehicle energy efficiency across municipalities tends to narrow over time and converge to the state average level.

Finally, further investigation into the role of socio-economic factors in shaping the distribution of vehicles and the exit event (a vehicle leaving the state) provides several interesting findings. We find that vehicles are more likely to be distributed into the municipalities that are richer, more populated or have lower unemployment rate. An interesting finding is that the update process of vehicle population through product substitution or migration will take more time if the municipalities are richer, less populated or have higher unemployment rate, since the initially existing vehicles in these types of municipalities are more likely to stay in the state.

With regard to policy implications, this paper can help policymakers as well as entrepreneurs to obtain a better understanding of the life cycle of vehicles: as durable consumer products, existing vehicles decline in energy efficiency and eventually exit. Meanwhile new vehicles enter the market and replace the old ones. The timing and different impacts of this cycle will provide policy opportunities that can improve aggregate vehicle energy efficiency. The policymakers should be alert to the fact that benefits due to changes in consumer behaviors of vehicle usage cannot keep pace with the big energy efficiency loss caused by the ageing of vehicles. Another fact that all the policymakers should be aware of is that consumes are actually less rational in terms of fuel economy than economic assumptions suggest and they are actually not very sensitive to changes in fuel expenditure of the vehicle (Turrentine & Kurani, 2007). The most important contribution to the growth of aggregate vehicle energy efficiency comes from the reallocation due the entering and exiting of vehicle. This may imply a public finance policy scenario (including policy instruments like tax and subsidy) that can stimulate the update of vehicle population and match specific municipality socio-economic characteristics. Additionally, in the US market benefits of technological change and innovation have been often used to satisfy some specific consumer preferences of the vehicle, such as more power and larger size, but not necessary higher energy efficiency (Lutsey & Sperling, 2005). A policy that promotes innovation towards vehicle fuel economy would be needed as well.

References

Aw, B. Y., Chen, X. M., & Roberts, M. J. (2001). Firm-level evidence on productivity differentials and turnover in Taiwanese manufacturing. *Journal of Development Economics, 66*(1), 51-86. doi: 10.1016/s0304-3878(01)00155-9

Baily, M. N., Bartelsman, E. J., & Haltiwanger, J. (1996). Downsizing and productivity growth: Myth or reality? *Small Business Economics, 8*(4), 259-278. doi: 10.1007/bf00393276

Baily, M. N., Bartelsman, E. J., & Haltiwanger, J. (2001). Labor productivity: Structural change and cyclical dynamics. *Review of Economics and Statistics, 83*(3), 420-433. doi: 10.1162/00346530152480072

Baily, M. N., Hulten, C., & Campbell, D. (1992). PRODUCTIVITY DYNAMICS IN MANUFACTURING PLANTS. *Brookings Papers on Economic Activity*, 187-267.

Barro, R. J. (1991). Economic Growth in a Cross Section of Countries. *The Quarterly Journal of Economics, 106*(2), 407-443.

Barro, R. J., & Salaimartin, X. (1991). CONVERGENCE ACROSS STATES AND REGIONS. *Brookings Papers on Economic Activity*(1), 107-182.

Barro, R. J., & Salaimartin, X. (1992). CONVERGENCE. *Journal of Political Economy, 100*(2), 223-251. doi: 10.1086/261816

Bartelsman, E., Dobbelaere, S., & Peters, B. (2014). Allocation of human capital and innovation at the frontier: firm-level evidence on Germany and the Netherlands. *Industrial and Corporate Change*. doi: 10.1093/icc/dtu038

Bartelsman, E. J., & Dhrymes, P. J. (1998). Productivity dynamics: US manufacturing plants, 1972-1986. *Journal of Productivity Analysis, 9*(1), 5-34. doi: 10.1023/a:1018383629341

Bartelsman, E. J., & Doms, M. (2000). Understanding productivity: Lessons from longitudinal microdata. *Journal of Economic Literature, 38*(3), 569-594. doi: 10.1257/jel.38.3.569

Bartelsman, E. J., & Wolf, Z. (2014). FORECASTING AGGREGATE PRODUCTIVITY USING INFORMATION FROM FIRM-LEVEL DATA. *Review of Economics and Statistics, 96*(4), 745-755. doi: 10.1162/REST_a_00395

Basu, S., & Fernald, J. G. (1997). Returns to Scale in U.S. Production: Estimates and Implications. *Journal of Political Economy, 105*(2), 249-283. doi: 10.1086/262073

Berry, S., Levinsohn, J., & Pakes, A. (2004). Differentiated products demand systems from a combination of micro and macro data: The new car market. *Journal of Political Economy, 112*(1), 68-105. doi: 10.1086/379939

Brownstone, D., & Train, K. (1999). Forecasting new product penetration with flexible substitution patterns. *Journal of Econometrics, 89*(1-2), 109-129.

Buchinsky, M. (1994). Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression. *Econometrica, 62*(2), 405-458. doi: 10.2307/2951618

Christensen, L. R., Cummings, D., & Jorgenson, D. W. (1981). Relative productivity levels, 1947–1973: An international comparison. *European Economic Review, 16*(1), 61-94. doi: http://dx.doi.org/10.1016/0014-2921(81)90049-0

Daly, H., & Ó Gallachóir, B. P. (2011). Modelling private car energy demand using a technological car stock model. *Transportation Research Part D: Transport and Environment, 16*(2), 93-101. doi: http://dx.doi.org/10.1016/j.trd.2010.08.009

European Commission. (2010). *Europe 2020: A European strategy for smart, sustainable and inclusive growth*. Brussels.

Foster, L., Haltiwanger, J., & Krizan, C. J. (1998). Aggregate Productivity Growth: Lessons from Microeconomic Evidence. *National Bureau of Economic Research Working Paper Series, No. 6803*. doi: 10.3386/w6803

Foster, L., Haltiwanger, J., & Krizan, C. J. (2006). Market selection, reallocation, and restructuring in the US retail trade sector in the 1990s. *Review of Economics and Statistics, 88*(4), 748-758. doi: 10.1162/rest.88.4.748

Griliches, Z., & Regev, H. (1995). FIRM PRODUCTIVITY IN ISRAELI INDUSTRY - 1979-1988. *Journal of Econometrics, 65*(1), 175-203. doi: 10.1016/0304-4076(94)01601-u

Hsieh, C. T., & Klenow, P. J. (2009). MISALLOCATION AND MANUFACTURING TFP IN CHINA AND INDIA. *Quarterly Journal of Economics, 124*(4), 1403-1448.

Koenker, R., & Bassett, G. (1978). Regression Quantiles. *Econometrica, 46*(1), 33-50. doi: 10.2307/1913643

Koenker, R., & Bassett, G. (1982). Robust Tests for Heteroscedasticity Based on Regression Quantiles. *Econometrica, 50*(1), 43-61. doi: 10.2307/1912528

Koenker, R., & Hallock, K. F. (2001). Quantile Regression. *The Journal of Economic Perspectives, 15*(4), 143-156.

Lee, Y., & Mukoyama, T. (2015). Entry and exit of manufacturing plants over the business cycle. *European Economic Review, 77*, 20-27. doi: 10.1016/j.euroecorev.2015.03.011

Li, Q., & Racine, J. S. (2007). *Nonparametric econometrics: theory and practice*: Princeton University Press.

Lutsey, N., & Sperling, D. (2005). Energy efficiency, fuel economy, and policy implications *Energy and Environmental Concerns 2005* (pp. 8-17). Washington: Transportation Research Board Natl Research Council.

Maasoumi, E., Racine, J., & Stengos, T. (2007). Growth and convergence: A profile of distribution dynamics and mobility. *Journal of Econometrics, 136*(2), 483-508. doi: 10.1016/j.jeconom.2005.11.012

Olley, G. S., & Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica, 64*(6), 1263-1297. doi: 10.2307/2171831

Patterson, M. G. (1996). What is energy efficiency? Concepts, indicators and methodological issues. *Energy Policy, 24*(5), 377-390. doi: 10.1016/0301-4215(96)00017-1

Quah, D. (1993). EMPIRICAL CROSS-SECTION DYNAMICS IN ECONOMIC-GROWTH. *European Economic Review, 37*(2-3), 426-434. doi: 10.1016/0014-2921(93)90031-5

Quah, D. (1996). Regional convergence clusters across Europe. *European Economic Review, 40*(3-5), 951-958. doi: 10.1016/0014-2921(95)00105-0

Quah, D. (1997). Empirics for Growth and Distribution: Stratification, Polarization, and Convergence Clubs. *Journal of Economic Growth, 2*(1), 27-59. doi: 10.1023/A:1009781613339

SalaiMartin, X. X. (1996). The classical approach to convergence analysis. *Economic Journal, 106*(437), 1019-1036. doi: 10.2307/2235375

Salgado-Ugarte, I. H., Shimizu, M., & Taniuchi, T. (1994). Exploring the shape of univariate data using kernel density estimators. *Stata Technical Bulletin, 3*(16).

Salgado-Ugarte, I. H., Shimizu, M., & Taniuchi, T. (1996a). ASH, WARPing, and kernel density estimation for univariate data. *Stata Technical Bulletin, 5*(26).

Salgado-Ugarte, I. H., Shimizu, M., & Taniuchi, T. (1996b). Practical rules for bandwidth selection in univariate density estimation. *Stata Technical Bulletin, 5*(27).

Turrentine, T. S., & Kurani, K. S. (2007). Car buyers and fuel economy? *Energy Policy, 35*(2), 1213-1223. doi: 10.1016/j.enpol.2006.03.005

UNITED NATIONS
INDUSTRIAL DEVELOPMENT ORGANIZATION

Vienna International Centre · P.O. Box 300 9 · 1400 Vienna · Austria
Tel.: (+43-1) 26026-0 · E-mail: info@unido.org
www.unido.org