# EVALUATING ENERGY EFFICIENCY OF URBAN TRANSPORTATION SYSTEMS IN DEVELOPING CITIES USING A FOUR-WAVE PANEL DATA*

by Metin Senbil**, Junyi Zhang***, and Akimasa Fujiwara****

## 1. Introduction

Given the priority to economic transformation and development generally stacked in several ten-years, transportation systems in developing countries face paradoxical challenges created by motorization. In solid terms, increasing volume of vehicle fleet generally cannot be adequately met with the existing infrastructure; this causes long delays caused by the ensuing congestion. This might be surprising when one would hear that developing cities reportedly have fewer cars than developed cities (Gakenheimer, 1999). Moreover in developing cities, one can notice very diverse set of trip modes that find their way on the crowded streets, this can rarely be seen in the developed countries where everyday trips are almost conducted by one of several modes such as private car, bus or rail. Thus traffic in developing cities becomes highly mixed by different trip modes with variant speeds, lane usage, etc. When the city authority intervenes in the situation by, for example, building a new arterial road, the ensuing effects on the land use configuration are more striking, and can be seen after a shorter time lag than the developed countries where such kinds of investments stay within marginal changes on land use considering the developed transport infrastructure already put in place.

In addition to the above-mentioned differences, there are many other observable differences between developing and developed cities that seldom find their places in the statistical tables. In developing cities, one might encounter shortage of transport professionals or lack of trained personnel as well as shortage of relevant technical equipment for transport system management, poor driving habits and tendency not to obey rules by drivers, etc., existence of which might seriously decrease the efficiency of the transport system functioning. We might collect all of these inefficiencies under a latent variable in relation to capacity concept. Capacity to manage transportation system, which is related to transportation authority, capacity to obey transportation rules and respect others rights which are related to thousands if not millions of drivers on the streets.

Having put the very characteristics of the transportation system in developing cities, in this study our aim is to compare different cities- specially developing cities vs. developed cities, with respect to energy and fuel consumptions in their transportation sector. We hypothesize that after controlling certain aspects the differences between energy and fuel consumption between developing and developed cities are closely related to "unobserved" the inefficiencies that are mentioned above. To locate these kinds of differences, we use a panel data set of 46 cities around the world. The data set contains information on land use, transportation infrastructure, mobility, and energy consumption in these respective cities. The panel data set starts from 1960, and supplies decennial information until 1990. Significant problem of this panel data set is the existence of the missing values for different time points, which causes problems in the selection of the variables in the model analysis. This study consists of four sections including introduction. In the second section, we discuss the applicability of the economic concept of efficiency in transport energy consumption. We put forward hypotheses that lead us to propose frontier analysis and make use of its econometric analysis. With the hypotheses at hand, we supply details of the data set, along with a discussion on the missing data problem and the results of the econometric estimation in the third section. We supply our conclusions on efficiencies in the last section.

## 2. (In)Efficiency in energy consumption

In econometric applications of production economics literature, efficiency is calculated by utilizing some form of distance function between a production or a cost yield and the frontier beyond which all yields are impossible (see Aigner *et al.*, 1977; Kumbhakar and Lovell, 2003). Thus, it represents an absolute, though a latent value for a known process, i.e., production, factors of which are strictly known. In addition, it also known how much one can produce with a certain mix of amounts of factors. Thus, what distance function carries as information is the inefficiency, which can be attributable to the -very- specific conditions of the producer, which distinguishes the producer from other producers. When we turn our attention to urban transportation energy consumption, similar efficiency perspective used in the production economics literature might be

---

* Keywords: energy consumption, transportation system efficiency, stochastic frontier regression.

** Member of JSCE, Dr. Eng., Graduate School of International Development and Cooperation, Hiroshima University,
(1-5-1 Kagamiyama, Higashi-Hiroshima, 739-8529, Japan; Tel&Fax: 082-424-5997; E-mail: senbil@hiroshima-u.ac.jp)

*** Member of JSCE, Dr. Eng., Graduate School of International Development and Cooperation, Hiroshima University,
(1-5-1 Kagamiyama, Higashi-Hiroshima, 739-8529, Japan; Tel&Fax: 082-424-6919; E-mail: zjy@hiroshima-u.ac.jp)

**** Member of JSCE, Dr. Eng., Graduate School of International Development and Cooperation, Hiroshima University,
(1-5-1 Kagamiyama, Higashi-Hiroshima, 739-8529, Japan; Tel&Fax: 082-424-6921; E-mail: afujiw@hiroshima-u.ac.jp)

relevant. However, we know only some of the cause-effect relationships behind energy consumption of the urban transportation sector, we cannot say that we are aware of all. In other words, factors of urban transportation energy consumption are very diverse, and usually they cannot be reduced to relationships like economical production process, which is intrinsically an optimization problem for producers themselves. But nevertheless, given its latency, concept of frontier might be used in the context of urban consumption similar to the ways it is used in economic efficiency studies. In this respect, one might argue that given transportation system, the urban form and the technology used, the frontier of energy consumption is constrained from above by a latent energy frontier. This might be one way and the estimation of any regression can be done with the usual stochastic frontier analysis that locates both idiosyncrasy and inefficiency errors separately. On the other hand, another argument is equally acceptable which highlights the issue of the relativity between cities. In other words, with no clear information about the upper boundary for energy consumption, putting aside its latency- the only relevant information about the energy consumption might be the information collected about other cities. In this setting, one city becomes the most efficient- the frontier city, the other cities' efficiency can be computed with respect to the most efficient city. Estimation of energy efficiency can be achieved by a properly designed linear regression would estimate different parameter values via dummy variables in the case of cross-section data or different constant terms for different cities in the case of panel data- i.e., fixed effects panel data. Thus, there are two ways to handle the question of energy efficiency: i.) estimation of the efficiency by stochastic frontier regression ii.) estimation of the energy efficiency by fixed effects linear regression model. In the part of this section, we supply a short introduction to the stochastic frontier regression model with both random and fixed effects in the case of panel data. Note that the estimation routines will not be reproduced here (see Kumbhakar and Lovell, 2003).

The stochastic frontier model may be written

$$y_{it} = f(\mathrm{x}_{\mathrm{it}}, \mathrm{z}_{\mathrm{i}}) + \mathrm{v}_{\mathrm{it}} \pm \mathrm{u}_{\mathrm{it}} = \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\mu}'\mathbf{z}_{i} + v_{it} \pm u_{it}, \tag{1}$$

where the sign of the last term, $u_{it}$, depends on whether the frontier describes costs (positive) or production (negative). The functional form might be either linear or nonlinear, besides the error term in the model has two parts. The function $f(\cdot)$ denotes the theoretical production function. In our case, this function will be used for evaluating urban transportation sector energy consumption based on the urban form, transportation network and the travel characteristics. As we will deal with the energy consumption in cities around the world, the subscript $i$ stands for individual cities and $t$ stands for time of observation. When the number of observations for every city is equal to one, then the model given in Eq. 1 structurally reduces to a cross sectional model, when the number of observations is more than one than the model becomes either fixed-effects model or random-effects model with respect to the implicit assumptions made on error terms based on the discussions given above. The production function might consist of time variant, **x**, and time invariant variables, **z**. The city and time specific idiosyncratic and stochastic part of the frontier is $v_{it}$, which could be either positive or negative. The second component, $u_{it}$ represents technical or cost inefficiency, and must be positive. The stochastic frontier model as originally proposed by Aigner *et al.* (1977) adds the error terms- $v_{it}$, $u_{it}$ in the base case given in (1), and names the sum as the "composed error"- $\varepsilon$. This sum is composed of a symmetric, normally distributed variable (the idiosyncrasy)- $v_{it} \sim N[0, \sigma_v^2]$ and the absolute of a normally distributed variable (the inefficiency)- $u_{it} = |U_{it}|$ where $U_{it} \sim N[0, \sigma_u^2]$. The model is usually specified in (natural) logs, so the inefficiency term, $u_{it}$ can be interpreted as the percentage deviation of observed performance, $y_{it}$ from the city frontier performance, $y_{it}$*: $y_{it}^* = \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\mu}'\mathbf{z}_{i} + v_{it}$. To denote the full model (by subsuming time invariant characteristics-**z** into **x)**, we have the city performance $y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} \pm u_{it}$.

The analysis of inefficiency in this modeling framework consists of two stages. At the first stage, we obtain parameter estimates, $\boldsymbol{\beta}$. This estimation step also produces estimates of the parameters of the distributions of the error terms in the model, $\sigma_u$ and $\sigma_v$. With the parameter estimates in hand, it is possible to estimate the composed deviation: $\varepsilon_{it} = v_{it} \pm u_{it} = y_{it} - \boldsymbol{\beta}'\mathbf{x}_{it}$. By "plugging in" the observed data for a given city in year $t$ and the estimated parameters. But, the objective is usually estimation of $u_{it}$, not $\varepsilon_{it}$, which contains the city specific heterogeneity. Jondrow *et al.* (1982) have devised a method of disentangling these effects. Their estimator of $u_{it}$ is

$$\mathrm{E}[u_{it} \mid \varepsilon_{it}] = \frac{\sigma\lambda}{1+\lambda^2}\left[\frac{\phi(a_{it})}{1-\Phi(a_{it})} - a_{it}\right] \tag{2}$$

where

$\sigma = [\sigma_v^2 + \sigma_u^2]^{1/2}$, $\lambda = \sigma_u / \sigma_v$, $a_{it} = \pm\varepsilon_{it}\lambda/\sigma$,

$\phi(a_{it})$ = the standard normal density evaluated at $a_{it}$,

$\Phi(a_{it})$ = the standard normal CDF evaluated at $a_{it}$.

For the random-effects, assumption is made on the invariability of the efficiency error term and estimation is straightforward with the method devised by derivation of the log-likelihood function is first proposed by Pitt and Lee (1981). For the fixed-effects, the estimation can be done equally by fixed-effects panel estimation of multiple linear regression and the estimated constant values are used to derive efficiency values (Kumbhakar and Lovell, 2003).

## 3. Data and estimation results

We have used the data compiled by Kenworthy and Laube (1999) to disseminate information about the automobile dependence in the 46-cities around the world. The dataset contains both developed and developing cities, and it is compiled four-time decennial panels starting from 1960. The model parameter values are estimated using the pooled database, comparisons are presented for developing countries. Cities in the database consist mostly of developed cities along with a few of developing cities of Asia-Pacific region. Developed cities in the dataset are from North America- seven from Canada, thirteen from USA, Europe- twelve altogether, Australia- six, and Asia-Pacific- three. Developing cities are only from Asia-pacific, they are six in altogether- Bangkok, Kuala Lumpur, Seoul, Manila, Jakarta and Surabaya. The dataset contains many missing values (Table 1). Especially, for the developing cities, almost all of the time points before 1990, i.e., 1960, 1970, and 1980, are missing for energy and fuel consumption. Thus data imputation for these variables is technically impossible with the available data set. For this reason, we have reduced the time points for developing cities to only one, i.e., 1990, while keeping all of the time points for developed cities.

Table 1: Missing data analysis

| # of cases= 166 | | N | Mean | Std. Deviation | Missing | |
|---|---|---|---|---|---|---|
| | | | | | Count | Percent |
| Developed city | Private transport fuel consumption in private transportation (joules) | 120 | 38.79 | 1.16 | 40 | 25.00 |
| | Puclic transport energy consumption (joules) | 120 | 35.09 | 1.29 | 40 | 25.00 |
| | # of jobs in CBD | 147 | 11.85 | 1.08 | 13 | 8.13 |
| | Population in CBD | 158 | 9.84 | 1.49 | 2 | 1.25 |
| | # of jobs in the inner area | 141 | 12.96 | 1.08 | 19 | 11.88 |
| | Inner area population | 157 | 13.20 | 1.19 | 3 | 1.88 |
| | Lenght of road network (km) | 144 | 8.81 | 1.31 | 16 | 10.00 |
| | Motor vehicles on register | 160 | 13.40 | 1.17 | 0 | .00 |
| | Public transport vehicle kilometers | 131 | 23.21 | 1.29 | 29 | 18.13 |
| | Private transport vehicle kilometers | 149 | 18.22 | 1.38 | 11 | 6.88 |
| | Modal share of public transport | 130 | 2.87 | .89 | 30 | 18.75 |
| | Modal share of private transport | 131 | 4.08 | .62 | 29 | 18.13 |
| | Private transport average road network speed | 92 | 3.60 | .51 | 68 | 42.50 |
| | CBD parking spaces | 134 | 10.38 | .78 | 26 | 16.25 |
| Developing city | Private transport fuel consumption in private transportation (joules) | 8 | 38.61 | .88 | 16 | 66.67 |
| | Puclic transport energy consumption (joules) | 18 | 35.87 | 1.79 | 6 | 25.00 |
| | # of jobs in CBD | 11 | 12.82 | .81 | 13 | 54.17 |
| | Population in CBD | 12 | 13.25 | 1.16 | 12 | 50.00 |
| | # of jobs in the inner area | 9 | 13.79 | .71 | 15 | 62.50 |
| | Inner area population | 13 | 14.51 | .78 | 11 | 45.83 |
| | Lenght of road network (km) | 12 | 8.13 | .91 | 12 | 50.00 |
| | Motor vehicles on register | 17 | 12.98 | 1.05 | 7 | 29.17 |
| | Public transport vehicle kilometers | 13 | 22.83 | 1.09 | 11 | 45.83 |
| | Private transport vehicle kilometers | 9 | 19.78 | 1.49 | 15 | 62.50 |
| | Modal share of public transport | 12 | 3.65 | .43 | 12 | 50.00 |
| | Modal share of private transport | 12 | 3.59 | .66 | 12 | 50.00 |
| | Private transport average road network speed | 10 | 3.22 | .26 | 14 | 58.33 |
| | CBD parking spaces | 6 | 10.43 | .74 | 18 | 75.00 |

We have applied a data imputation in three phases. Firstly, we have used annual urbanization and GDP rates between 1960 and 1990 to interpolate the missing population and employment[1] data with the assumption that total population and employment in cities with missing data has changed at the same rate with the country average. Afterwards, we have grouped Australian and North American cities, European and Asian-developed, Asian-developing countries under different categories. Lastly, further imputation has been carried out by using series means of different categories separately for each time point (Table 2).

Table 2: Descriptive statistics of the imputed data

| | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| Private transport fuel consumption in private transportation (joules) | 166 | 36.05 | 41.37 | 38.73 | 1.01 |
| Public transport energy consumption (joules) | 126 | 31.44 | 38.01 | 35.16 | 1.34 |
| # of jobs in CBD | 166 | 9.29 | 14.67 | 11.90 | 1.06 |
| Population in CBD | 166 | 6.69 | 14.29 | 9.96 | 1.59 |
| # of jobs in the inner area | 166 | 10.17 | 15.72 | 12.99 | 1.03 |
| Inner area population | 166 | 10.55 | 15.99 | 13.25 | 1.20 |
| Lenght of road network (km) | 166 | 4.57 | 11.72 | 8.79 | 1.24 |
| Motor vehicles on register | 166 | 9.80 | 16.40 | 13.42 | 1.16 |
| Public transport vehicle kilometers | 166 | 20.69 | 30.68 | 23.16 | 1.16 |
| Private transport vehicle kilometers | 166 | 14.52 | 21.77 | 18.30 | 1.39 |
| Modal share of public transport | 166 | .47 | 4.30 | 2.93 | .83 |
| Modal share of private transport | 166 | 1.19 | 6.78 | 4.03 | .61 |
| Private transport average road network speed | 166 | .92 | 4.18 | 3.61 | .40 |
| CBD parking spaces | 166 | 7.90 | 12.14 | 10.36 | .71 |

With respect to the hypotheses given in the previous section, we have estimated two different regressions. Linear regression with fixed effects locates the relative efficiency in consumption; stochastic frontier regression locates a latent frontier from

[1] Relevant data have been compiled and organized using Global Environmental Outlook portal maintained by the United Nations. http://geodata.grid.unep.ch. Accessed on 2004.05.06.

which each city deviates by efficiency term. Estimations of the regression equations leave some of the independent variables insignificant. However, city specific fixed effects as well as parameters of the efficiency terms turn out to be significant (Table 3).

Table 3: Estimation results

| | Private transport fuel consumption | | | | Public transport energy consumption | | | |
|---|---|---|---|---|---|---|---|---|
| | FLR | | SFR | | FLR | | SFR | |
| | coefficient | t-stat | coefficient | t-stat | coefficient | t-stat | coefficient | t-stat |
| Constant | | | 28.02 | 24.13 | | | 18.68 | 16.56 |
| Population in CBD | 0.07 | 0.51 | -0.04 | -0.59 | 0.20 | 1.07 | 0.03 | 0.49 |
| Inner area population | 0.32 | 0.83 | 0.34 | 3.24 | -1.51 | -3.03 | -0.04 | -0.24 |
| # of jobs in CBD | 0.33 | 2.48 | 0.12 | 0.97 | -0.05 | -0.33 | 0.00 | 0.00 |
| # of jobs in the inner area | -0.15 | -0.66 | -0.23 | -1.49 | 0.59 | 1.46 | -0.06 | -0.21 |
| Lenght of road network (km) | 0.22 | 2.19 | 0.26 | 3.99 | -0.27 | -1.69 | 0.12 | 2.09 |
| Motor vehicles on register | 0.14 | 1.33 | 0.17 | 2.01 | | | | |
| CBD parking spaces | 0.11 | 1.01 | 0.10 | 1.12 | | | | |
| Private transport vehicle kilometers | 0.07 | 1.23 | 0.12 | 5.00 | | | | |
| Public transport vehicle kilometers | | | | | 1.02 | 6.10 | 0.96 | 8.11 |
| Modal share of private transport | -0.11 | -0.98 | 0.04 | 0.35 | 0.09 | 0.48 | -0.09 | -0.51 |
| Modal share of public transport | | | | | -0.14 | -1.27 | -0.20 | -0.80 |
| Public transport vehicle speeds | | | | | | | | |
| $\lambda$ | | | 1.11 | 2.66 | | | 0.88 | 1.90 |
| $\sigma_u$ | | | 0.50 | 3.29 | | | 0.40 | 1.98 |
| df | | | 1 | | | | 1 | |
| Restricted log-likelihood | | | -131.15 | | | | -93.56 | |
| Log-likelihood | | | -123.90 | | | | -93.26 | |
| $R^2$-Adjusted $R^2$ | 0.87-0.81 | | | | 0.94-0.90 | | | |

FLR: Fixed effects linear regression; SFR: Stochastic frontier regression

## 4. Conclusions

With the estimated efficiency term, u and the fixed terms we compare the cities with respect to private transport fuel consumption and public transport energy consumption. The results suggest that among developing cities the most efficient one in public transportation is Surabaya, the second to Surabaya changes when we switch regressions: the stochastic frontier regression yields Manila and Kuala Lumpur as the second and third efficient cities respectively while fixed effects regression yields Kuala Lumpur and Manila as the second and third best respectively. Following the first three efficient developing cities in public transport, Seoul, Bangkok and Jakarta are founded to be forth, fifth and sixth efficient developing city in public transport. Efficient developed cities in public transport are found to be Portland, Canberra, Montreal and Tokyo. The worst cases in the developed cities are Houston, Boston, Munich, and Winnipeg. For the private transport, the results of the regression models are contradictory. The stochastic frontier regression yields European and Australian cities, i.e., Perth, Copenhagen, Adelaide, Brisbane, London, as most efficient cities while fixed effects linear regression supports predominantly North American cities, i.e., Denver, San Diego, Toronto and Paris. Among developing cities, Surabaya again is found to be the most efficient city. Manila and Jakarta follow Surabaya consistently in both regression analyses. Seoul surpasses Kuala Lumpur in the fixed effects model while the reverse is observed in the stochastic frontier regression. Bangkok consistently is found to be the least efficient city in terms of private transportation fuel transportation in both of the regression analysis.

Immediate results of this study suggest that after controlling for the common factors of transportation system energy consumption, frontier approach might be useful as a capacity approach to the urban transportation system. In this study, we associate the capacity to the energy-efficient management of the public transport and the capacity of energy-efficient driving habits.

## References


1) Aigner, D., Lovell, K., and Schmidt, P.: Formulation and estimation of stochastic frontier function models, *Journal of Econometrics*, Vol.6, pp.21-37, 1977.
2) Gakenheimer, R.: Urban mobility in the third world, *Transportation Research Part A*, Vol.33, pp.671-689, 1999.
3) Jondrow, J., Materov, I., Lovell, K., and Schmidt, P.: On the estimation of technical inefficiency in the stochastic frontier production function model, *Journal of Econometrics*, Vol.19, pp.233-238, 1982.
4) Kenworthy, J. R., and Laube, F. B.: *An International Sourcebook of Automobile Dependence in Cities 1960-1990*, University Press of Colorado, Niwot, Colorado, 1999.
5) Kumbhakar, S. C., and Lovell, C. A. K.: *Stochastic Frontier Analysis*, Cambridge University Press, N.Y., 2003.
6) Pitt, M. and Lee, L.: The measurement and sources of technical inefficiency in Indonesian weaving industry, *Journal of Development Economics*, Vol.9, pp.43-64, 1981.