

HOW EFFICIENTLY DO U.S. CITIES MANAGE ROADWAY CONGESTION?

Anthony J. Glass*, Karligash Kenjegalieva[†] and Robin Sickles^{‡§}

Abstract

We estimate efficiency and TFP growth for two measures of congestion and two measures of the monetary value of congestion for the largest 88 contiguous cities in the U.S. over the period 1982 – 2007. Using Stochastic Frontier Analysis we find that the efficiency scores for congestion and the associated ranking of cities is sensitive to the measure of congestion. In contrast, the efficiency scores and rankings are robust for the two measures of the monetary value of congestion. Most importantly, for the most valid measure of congestion and both measures of the monetary value of congestion, we find that average TFP growth over the study period is characterized by an upward trend. This is an encouraging sign even though in all three cases growth is only zero or slightly less than zero at the end of the study period. We therefore conclude that policies which have been used towards the end of the study period such as providing incentives to carpool and encouraging employers to offer flexi-time and telecommuting arrangements appear to have been effective and should be implemented more widely.

Keywords: Congestion; Monetary Value of Congestion; Panel Data; Stochastic Frontier Analysis, Input Distance Function; TFP

JEL Classification: C23; R41

*Department of Economics, School of Business and Economics, Loughborough University, Leics, UK, LE11 3TU. Email: A.J.Glass@lboro.ac.uk.

[†]Department of Economics, School of Business and Economics, Loughborough University, Leics, UK, LE11 3TU. Email: K.A.Kenjegalieva@lboro.ac.uk.

[‡]Department of Economics, Rice University, Houston, Texas, USA. Email: rsickles@rice.edu.

[§]Corresponding author.

1 Introduction

Roadway congestion is an acute problem in urban areas in the U.S. In 2007 the negative impact of urban roadway congestion on the U.S. economy was valued at \$87.2 billion (Texas Transportation Institute, TTI, 2009).¹ Over the period 1997 – 2007 the value of urban roadway congestion in real terms increased by over 50%. To limit the impact of urban roadway congestion on U.S. GDP in the future, efficient management of the monetary value of congestion is key. Accordingly, in this paper we examine how efficiently two measures of congestion and two measures of the monetary value of congestion are managed in urban areas in the U.S. To the best of our knowledge this is the first efficiency analysis of roadway congestion.

Pollution and congestion are the most widely cited negative externalities from economic activity. There are a large number of non-parametric studies with multiple outputs which estimate environmental efficiency by incorporating at least one pollutant as an undesirable output (e.g. Färe *et al.*, 1989; 1996, Tyteca, 1997, Hernandez-Sancho *et al.*, 2000, Reinhard *et al.*, 2000 and Zaim and Taskin, 2000). We do not discuss the non-parametric literature because we undertake a stochastic frontier analysis (SFA). Two approaches have been used in the literature to estimate environmental efficiency using SFA. The first approach is simple and involves estimating input-oriented environmental efficiency by assuming that the relationship between pollution and output resembles the traditional input-output relationship (e.g. Cuesta *et al.*, 2009, Reinhard *et al.*, 1999; 2000 and Atkinson and Dorfman, 2005).² The second method involves using an inverse transformation of the bad output to obtain a good output (e.g. Fernández *et al.*, 2005 and Koop and Tole, 2008). Output-oriented environmental efficiency can then be obtained by manipulating the output-oriented efficiency associated with the transformed output.³

Using output-oriented SFA with emissions for a pollutant as an input the shadow price of emissions can be estimated. Any subsequent policy recommendations, however, could be erroneous because as Cuesta *et al.* (2009) note, the shadow price of sulphur dioxide emissions from the U.S. electricity generation industry is not a good estimate of the market price. This is because the theoretical assumptions on which duality theory is based do not reflect the market price of emissions. Ideally one would make policy recommendations based on a model which is estimated using data on the monetary value of emissions for DMUs i.e. obtain input or output-oriented efficiency associated with the monetary value of pollution. The availability of data on the monetary value of pollutants for DMUs, however, is poor. In contrast, rich data on the monetary value of congestion for U.S. cities is available from the TTI. Using this data for the 88 largest contiguous U.S. cities for 1982 – 2007, we estimate input distance functions and report efficiency scores for the monetary value of congestion. Putting the monetary value of congestion into context, Small and Verhoef (2007) estimate various incremental social values associated with an additional urban vehicle mile in the U.S. for an average commute (12.1 miles and 22.5 minutes) in a medium-sized

¹We use urban areas and cities interchangeably. Urban areas are determined by U.S. Census demographic criteria and typically tend to be associated with a distinct city.

²Strictly speaking Atkinson and Dorfman (2005) model pollution as a "technology shifter". It nevertheless affects the technology in much the same way as an input (Koop and Tole, 2008).

³See Fernández *et al.* (2005) for a detailed explanation on how to manipulate the output-oriented efficiency for the transformed output to obtain environmental efficiency.

automobile. They estimate the marginal social values of travel time, accidents and environmental externalities at 2005 prices to be \$0.388, \$0.178 and \$0.016, respectively. The large difference between the marginal social values of travel time and environmental externalities suggests that efficiency analysis of roadway congestion is a very promising area for research.

In short, we estimate four input distance functions. For each model there are two inputs and three outputs. Public transit passenger miles per head is an input in all four models. The other input in each model is an early measure of congestion in model 1, a more recent measure of congestion in model 2, a measure of the monetary value of congestion based on TTI (2009) in model 3 and a more valid measure of the monetary value of congestion in model 4. The three outputs in all four models are average daily vehicle miles per head per lane mile, a proxy for suburbanization and the ratio of average daily peak travellers to city population.⁴ The key empirical findings are as follows. At the sample mean the input and output elasticities for all four models have the expected signs and are significant thereby satisfying the monotonicity conditions. We also find that the efficiency scores and the associated ranking of cities varies according to which measure of congestion is used. The mean efficiency score is much higher when we use an early measure of congestion compared to when we use a measure of congestion which is calculated using a more recent methodology. In contrast, the efficiency scores and rankings are very robust for the two measures of the monetary value of congestion. At the sample mean we find that an increase in public transit passenger miles per head leads to a very small fall in the early measure of congestion and non-negligible falls in the recent measure of congestion and both measures of the monetary value of congestion. Having said this, all the public transit elasticities outside the sample mean for quintiles of the population size distribution decline over the study period.⁵ This suggests that increasing public transit ridership was a much more effective way of managing congestion and its associated monetary value in the early years of the study period.

For the more recent measure of congestion and both measures of the monetary value of congestion we find that average TFP growth is characterized by an upward trend. In all three cases, there is a marked decline in average TFP in the early years but by the end of the study period average TFP growth is slightly above or is approaching zero; which is a very encouraging sign for the future. The decomposition of TFP growth into technical change, efficiency change and scale change suggests that for both measures of congestion and both measures of the monetary value of congestion, average TFP growth primarily depends on the scale change. What then is the reason for the upward trend in average TFP growth for the most recent measure of congestion and both measures of the monetary value of congestion? It is most probably because as the scale effect gets progressively bigger over the study period, urban transportation policymakers will be more concerned about congestion and its associated monetary value and will adopt more innovative and productive policies: e.g. (i) ramp metering which involves using signals to restrict

⁴We justify the choice of inputs and outputs in the data section using, among other things, the results of Hausman-Wu tests of whether endogeneity has a significant effect on the consistency of the estimated parameters. We also provide in the data section details of the measure of congestion which the TTI initially used along with details of the measure of congestion which they now favor. Why we favor one measure of the monetary value of congestion over the other is explained in the data section.

⁵We thank an anonymous referee for recommending that we calculate public transit elasticities outside the sample mean for quintiles of the population size distribution.

the flow of traffic joining a freeway to prevent entering vehicles disrupting the mainlane flow of traffic; (ii) incentives to carpool; (iii) encouraging employers to offer flexi-time and telecommuting arrangements. Such innovations have been more prevalent towards the end of the study period when the scale effect is bigger, and on the basis of our TFP growth results these innovations have been effective and should therefore be implemented more widely in the short to medium term. It is questionable, however, whether implementing these relatively conservative innovations more widely will be sufficient to maintain the upward trend in average TFP growth in the medium to long term.

The remainder of this paper is organized as follows. In section 2 the methodology is set out. Specifically, we describe the approach which we use to model the technology and we explain how we use each of the fitted models to compute TFP growth. The data set is described in section 3, focusing on how the two measures of congestion and both measures of the monetary value of congestion are computed. In section 4, the estimation results and the efficiency scores from the SFA are presented and discussed. In section 5, for both measures of congestion and both measures of the monetary value of congestion, we present and analyze TFP growth and its constituent parts. We conclude in section 6 by using our results to provide some policy advice on managing congestion and its associated monetary value in the short to medium term.

2 Methodology

2.1 Modelling the Technology and Relative Efficiency

The stochastic frontier framework originates from Aigner *et al.* (1977) and Meeusen and van den Broeck (1977). The distinguishing feature of stochastic frontier analysis (SFA) is that it eliminates random shocks in the estimation of (in)efficiency. To estimate (in)efficiency SFA incorporates a one-sided error term in addition to the traditional symmetric random noise term. Using SFA we construct a best-practice frontier and evaluate the degree to which a city could decrease the level of a measure of congestion or the level of a measure of its associated monetary value, relative to other cities in the sample, holding the outputs constant. We thus employ an input-oriented method.

The input requirement set $I(y)$ represents the set of K inputs $x \in \mathbb{R}^+$ which can produce a set of R outputs $y \in \mathbb{R}^+$ i.e. $I(y) = \{x \in \mathbb{R}^+ : x \text{ can produce } y\}$. We represent the technology at time t by the input distance function, $D_I(y, x, t)$, à la McFadden (1978). Since the value of the input distance function equals one if a city is on the efficient frontier and exceeds one if a city is inefficient $D_I \geq 1$ and so

$$\ln D_I(y, x, t) - u = 0, \quad (1)$$

where $u \geq 0$.

The inverse of the input distance function D_I is a measure of Farrell input based efficiency. u corresponds to the inefficient slack in the use of inputs by a city relative to other cities in the sample i.e. it is the feasible contraction in inputs which will project an inefficient producer onto the

efficient frontier of the input requirement set. In econometric studies of inefficiency measurement u is treated as a random variable distributed across producers with a known asymmetric probability density function. The properties of the input distance function as stated by McFadden (1978) are as follows:

- (i) non-decreasing in inputs, x : $\partial \ln D_I / \partial \ln x_k \equiv ex_k \geq 0$ for $k = 1, \dots, K$, where ex_k is the k th input elasticity;
- (ii) homogeneity of degree one in x : $D_I(y, x/x_k, t) = D_I(y, x, t)/x_k$;
- (iii) concave in x ;
- (iv) non-increasing in outputs, y : $\partial \ln D_I / \partial \ln y_r \equiv ey_r \leq 0$ for $r = 1, \dots, R$, where ey_r is r th output elasticity;
- (v) the scale elasticity of the technology at time t is
$$E^t = - \left(\sum_{r=1}^{r=R} \partial \ln D_I / \partial \ln y_r \right)^{-1} \equiv - \left(\sum_{r=1}^{r=R} ey_r \right)^{-1}.$$

Applying the property in (ii) and normalizing the inputs yields a dependent variable in the regression analysis of $-\ln x_K$. Using the input distance definition this can be written as follows

$$-\ln x_K = \ln D_I(y, x/x_K, t) - u. \quad (2)$$

Three elements (1) $TL(y, x/x_K, t)_{it}$, (2) $\pi' z_{it}$ and (3) v_{it} are needed to make equation (2) operational for a panel data set ($i = 1, \dots, N$ and $t = 1, \dots, T$). The three elements yield

$$-\ln x_{Kit} \approx TL(y, x/x_K, t)_{it} + \pi' z_{it} + v_{it} - u_i, \quad (3)$$

where: $TL(y, x/x_K, t)_{it}$ represents the technology as the translog approximation to the log of the distance function containing the inputs normalized by the input on the left hand side; $\pi' z_{it}$ captures the inter-city heterogeneity that is separate from inefficiency where z_{it} denotes the exogenous characteristics; v_{it} denotes the conventional idiosyncratic error term incorporating sampling error, measurement error and specification error; u_i is the inefficiency component of the disturbance error.⁶

Using the notation $\tilde{x}_k \equiv x_k/x_K$, $ly' = (\ln y_1, \dots, \ln y_R)$ and $\tilde{l}\tilde{x}' = (\ln \tilde{x}_1, \dots, \ln \tilde{x}_{K-1})$, the translog input distance function $TL(y, \tilde{x}, t)_{it}$ is

$$TL(y, \tilde{x}, t) = \alpha_0 + \alpha' ly + \beta' \tilde{l}\tilde{x} + \frac{1}{2} ly' \mathbf{A} ly + \frac{1}{2} \tilde{l}\tilde{x}' \mathbf{B} \tilde{l}\tilde{x} + ly' \mathbf{\Gamma} \tilde{l}\tilde{x} + \delta_1 t + \frac{1}{2} \delta_2 t^2 + \mu' lyt + \eta' \tilde{l}\tilde{x}t, \quad (4)$$

where α' , β' , δ , μ' , η' , \mathbf{A} , \mathbf{B} and $\mathbf{\Gamma}$ are vectors and matrices of parameters to be estimated. The continuity property of the function requires symmetry restrictions on the elements of the matrices \mathbf{A} and \mathbf{B} : $\alpha_{rs} = \alpha_{sr}$ and $\beta_{jk} = \beta_{kj}$, respectively.

In short, the modelling exercise has two objectives. The first is to estimate four translog input distance functions, one for each measure of congestion and one for each measure of the monetary value of congestion. We then use the estimated functions to calculate relative efficiency.

⁶It is assumed that the efficiency scores for both measures of congestion and both measures of the monetary value of congestion are time invariant, hence why there is no subscript t in the final term of (3). This assumption is justified at the beginning of section 4.

The second objective is to calculate the change in TFP for each measure of congestion and each measure of the monetary value of congestion by calculating each of its three constituent parts and then summing.

2.2 Parametric Total Factor Productivity Growth

Productivity growth is the growth of output minus the growth in input. TFP is the rate of growth in a multiple output quantity index minus the rate of growth in a multiple input quantity index. Orea (2002) notes that a TFP index which is generalized from the case of one input and one output should satisfy four properties: (i) identity, (ii) monotonicity, (iii) separability and (iv) proportionality.

Identity requires that if inputs and outputs do not change the TFP index is unity. Monotonicity requires that the weighted output growth rates and input growth rates are chosen so that higher output and lower input unambiguously improve TFP. Separability, which is a property of the chosen technology set, permits the generalization to the multiple-output multiple-input case. Proportionality requires that the weights in the output and input growth indices sum to unity.

Coelli *et al.* (2003) demonstrate that a TFP index which satisfies the aforementioned properties can be constructed from the translog approximation to the input distance function. Since the negative log of the input distance is input based technical efficiency, TE_I , and by making use of the quadratic identity lemma (Caves *et al.*, 1982) the following expression for $\ln TFPC$ can be obtained

$$\begin{aligned} \ln TFPC = & [\ln TE_I(t+1) - \ln TE_I(t)] + \frac{1}{2} [(\partial \ln D_I(t+1)/\partial t) + (\partial \ln D_I(t)/\partial t)] \\ & + \left[\frac{1}{2} \sum_{r=1}^{r=R} ((ey_{rt+1} SF_{t+1}^I) + (ey_{rt} SF_t^I)) (\ln (y_{rt+1}/y_{rt})) \right], \end{aligned} \quad (5)$$

where:

$TFPC$ is TFP change;

ey_{rt} is the column vector of output elasticities in period t ;

SF_t^I is the input scale factor (see Saal *et al.*, 2007), $SF_t^I = \left(\left(\sum_{r=1}^{r=R} ey_{rt} + 1 \right) / \sum_{r=1}^{r=R} ey_{rt} \right) = 1 - E^t$.

The three terms in square brackets in (5) represent the familiar decomposition of $TFPC$ into efficiency change, EC , technical change, TC , and scale change, SC :

$$TFPC = EC + TC + SC. \quad (6)$$

Using the first order and second order elasticity and scale parameters from a fitted input distance function we calculate EC , TC and SC and then sum to obtain $TFPC$.

3 Data

The data for the inputs, outputs and z -variables, unless otherwise stated, are extracted from the data set which accompanies TTI (2009). The inputs, outputs and z -variables for the four models

together with the summary statistics for the raw data are presented in Table 1. It is evident from this table that the only difference between the four models is which measure of congestion or which measure of the monetary value of congestion is used as one of the inputs.

[Insert Table 1]

The roadway congestion index is the original measure of congestion that the TTI calculated and for this reason is the first measure of congestion which we use. For precise details on how the roadway congestion index is calculated see TTI (2009). Intuitively it is a measure of the density of traffic across an entire urban area which captures the intensity and duration of congestion. It is calculated using generally available data for an entire urban area on total lane miles and vehicle miles on freeways and major streets. To illustrate, if the roadway congestion index is equal to 1.0: (i) typical journey time in the morning and evening peaks would be 25% longer than in the off-peak; (ii) the longer commute in the peak periods would be because of slower moving traffic on the freeways than would be the case in the off-peak but travel on major streets in the peaks would be uncongested; (iii) there would be moderate congestion for 1.5 – 2.0 hours in each of the peak periods.

The roadway congestion index is a macroscopic measure of congestion as it does not operate at the level of individual sections of an area's network, some of which will be local bottlenecks. This is the principal reason for the development of other measures of congestion. Another measure of congestion that the TTI have since calculated, which is now much more widely used than the roadway congestion index is annual person hours of delay per peak traveller. Annual person delay for an urban area is basically calculated in two stages. Firstly, daily delay per vehicle is aggregated across individual sections of an area's network where delay is the difference between travel time in the morning and evening peaks (6 am–10 am and 3 pm–7 pm) and free-flow travel time, which is calculated assuming that free-flow speeds are 60 m.p.h. on a freeway and 35 m.p.h. on major streets. For precise details on how daily delay per vehicle for an entire urban area is calculated see TTI (2009). Secondly, daily delay per vehicle for an urban area is multiplied by 250 (average number of working days per year) and by 1.25 (the average number of vehicle occupants). Instead of using annual person delay per peak traveller as our second measure of congestion we use annual person delay per head because it is more representative of how well congestion is being managed in an urban area. This is because it incorporates the number of people who are choosing not to travel or who choose to travel in the off-peak instead, which could be a reflection of changes in travel behavior as a result of effective policies to manage congestion.

The first measure of the monetary value of congestion for an urban area which we use is the real monetary value of annual delay per head, where the monetary value of annual delay is taken from TTI (2009) and includes three elements: (i) the monetary value of wasted fuel by passenger vehicles as a result of travelling in congested conditions; (ii) the monetary value of person delay which is annual person delay multiplied by the value of person travel time; (iii) the monetary value of commercial delay which is annual delay per vehicle multiplied, firstly, by the percentage of vehicles in the peak periods which are trucks and, secondly, by the value of commercial vehicle time. The monetary value of annual delay from TTI (2009) is based on the

same value of person travel time and the same value of commercial vehicle time for each urban area, which is controversial. Along the lines of Winston and Langer (2006), we calculate another measure of the monetary value of congestion per head which is based on the convention that travel time is valued at 50% of the hourly wage (Small and Verhoef, 2007). To calculate our second measure we use: (i) annual person delay per head; (ii) average annual wage for the Metropolitan Statistical Area (MSA) in which the urban area is located, which we obtained from the Bureau of Economic Analysis; (iii) average annual hours worked per worker throughout the U.S. which was obtained from OECD StatExtracts.^{7,8}

In addition to a measure of congestion or a measure of the monetary value of congestion, there is an another input in each model (public transit passenger miles per head) and three outputs (average daily vehicle miles per head per lane mile, number of hours in the morning and evening peaks when there is congestion somewhere on the network and the ratio of average daily peak travellers to city population). The number of hours when there is congestion on any section of the city's network is a proxy for degree of suburbanization. This is because further suburbanization will increase the length of the commute and lead to congestion earlier in the morning further away from the central business district, thereby increasing the number of hours when there is congestion somewhere on the network. We tested for potential endogeneity bias from the right-hand-side input and outputs as follows. We estimated Hausman-Taylor models assuming each input or output is endogenous, with or without assuming that the squared terms and interaction terms pertaining to the relevant input or output are endogenous as well. Using the Hausman-Wu test for endogeneity the relevant estimates from each Hausman-Taylor model are compared against the within estimate of the model. More details on this approach are available in Adams *et al.* (1999). The tests fail to reject the null hypothesis of no endogeneity bias from y_1 , y_2 , y_3 and x_2 with the χ^2 equal to 0.46, 0.61, 3.45 and 0.92 for Models 1 – 4, respectively. More details on the tests are available from the corresponding author on request. We therefore conclude that our results are robust to endogeneity concerns.

Turning our attention to some of the z -variables. The two variables which we use to capture the climate in a city are the number of days per year when the temperature is greater than 90 deg. F and annual precipitation. The data for both variables was obtained from the National Atmospheric and Oceanic Administration.⁹ These variables were included because extreme weather conditions (i.e. more hot days and more rain) is detrimental to pavement condition, which may well lead to an increase in delays because there is likely to be more vehicle breakdowns and when the road surface is poor drivers will reduce their speed to reduce the risk of an accident. Also, more rain will increase delays because there will be an increase in the number of accidents and driving conditions will be more difficult so drivers will reduce their speed.

⁷The measures of the monetary value of congestion are at 1982 prices. The TTI values of person travel time and commercial vehicle time are not based on the prevailing wage rate but on the perceived valuation of delay. We therefore deflate the TTI measure of the monetary value of congestion using the CPI. Our second measure is based on the wage rate so we deflate it using the GDP deflator.

⁸Our second measure of the monetary value of congestion is therefore only based on the value of person travel time and overlooks the value of commercial vehicle time and the value of wasted fuel.

⁹Because of data availability issues the climatic data for 2007 is assumed to apply throughout the study period.

To capture the contribution to congestion from traffic travelling through a city en route we follow Winston and Langer (2006) and include a dummy variable which takes a value of 1 if the city lies on a major interstate (i.e. a highway ending in 0 or 5). We also include Census region dummies (West, Midwest, Northeast and South) and four population size dummies using the classification in TTI (2009): Small (population less than 500,000), Medium (population over 500,000 but less than 1 million), Large (population over 1 million but under 3 million) and Very Large (population in excess of 3 million). To take account of the effect of changes in speed limits two dummy variables are included. The first is 0 before 1987 and 1 thereafter (when states were permitted to raise the speed limit on rural interstate highways from 55 m.p.h. to 65 m.p.h.). The second takes a value of 0 before 1995 and 1 thereafter (when all federal speed limit controls were lifted, returning all authority on speed limit determination to the states).

4 Results from the Stochastic Frontier Analysis

4.1 Estimation Results

Urban transportation policy makers in the U.S. have opted to subtly manage roadway congestion by, for example, providing incentives to carpool by creating designated High Occupancy Vehicle (HOV) lanes which are free to use if there are a sufficient number of people in the vehicle. In March 2007 there were 345 HOV facilities in the U.S. (Booz Allen Hamilton and HNTB, 2008).¹⁰ Policymakers in the U.S. evidently have an aversion to the step change in the incentive mechanism which road pricing represents because only ten of the HOV facilities in March 2007 were also Single Occupancy Vehicle (SOV) buy-in lanes, otherwise known as High Occupancy Toll (HOT) lanes. Since over the study period there has not been a big change in the incentive mechanism for automobile users it is reasonable to assume that congestion efficiency and monetary value efficiency are both time-invariant. We therefore estimate the time-invariant efficiency model for panel data developed by Pitt and Lee (1981). Looking ahead, there would need to be a complete change of policy on HOT lanes to create a big enough change in the incentive mechanism to justify fitting an SFA model where efficiency is time-variant (e.g. Battese and Coelli, 1992, Cornwell *et al.*, 1990 and Greene, 2005).

The estimation results for the four distance functions are presented in Table 2. We first discuss the salient findings for the z -variables. In model 1, more of the coefficients on the z -variables are significant than in models 2 – 4. For example, the coefficient on population density (π_{10}) is only significant in model 1. Nevertheless, a small number of corresponding parameters have the same sign and are significant in all four models. For example, the second climate parameter (π_2) is negative and significant in all the models so we find robust evidence that more rainfall leads to an increase in congestion and its associated monetary value. As expected, *vis-à-vis* very large cities both measures of congestion and both measures of the monetary value of congestion are lower in medium-sized conurbations and, in particular, small cities. In large cities, the congestion index and the most valid measure of the monetary value of congestion (model 4) are significantly

¹⁰As of March 2007, there were 88 HOV facilities in California, followed by 83 in Minnesota, 41 in Washington State, 35 in Texas and 21 in Virginia.

lower than in very large cities. In contrast, for large and very large cities there is no significant difference between the most valid measure of congestion (model 2) or between the first measure of the monetary value of congestion.

Interestingly, π_{11} is negative and significant in models 2 – 4 which suggests there is a net increase in annual person delay per head and both measures of the monetary value of congestion from being part of the major interstate network. The implication is that the rise in annual person delay per head and both measures of the monetary value of congestion from being part of the major interstate network because more traffic passes through en route more than offsets the fall from improved highway links. The opposite is the case for model 1.

[Insert Table 2]

The input and output elasticities in all four models have the expected signs and are significant implying that the monotonicity conditions are satisfied at the sample mean. For quintiles of the population size distribution, the percentage of elasticities outside the sample mean which satisfy the monotonicity properties are reported in Appendix A.1. We can therefore conclude that, in general, the monotonicity conditions are satisfied in all four models for a large proportion of each quintile. This is not the case, however, for average daily vehicle miles per head per lane mile for the 5th quintile in models 2 – 4. This implies that when there is an increase in lane miles for the 5th quintile, the fall in annual person delay per head and both measures of the monetary value of congestion more than offsets the rise from the induced travel effect i.e. the increase in average daily vehicle miles per head because of the increase in the supply of road space (Duranton and Turner, 2011 and Cervero and Hansen, 2002). The opposite is the case for cities in the 1st–4th quintiles. This suggests that urban transportation policymakers for cities in the 1st–4th quintiles should not attempt to reduce annual person delay per head and both measures of the monetary value of congestion by increasing lane miles because it will have the opposite effect. And although models 2 – 4 suggest that increasing lane miles in cities in the 5th quintile would have the desired effect this does not take into account the cost of highway expansion. We revisit this issue in the concluding section. Furthermore, we note from the Hessian results for models 2 – 4 that the proportions of the sample for which the concavity condition is satisfied are quite large (76%, 76% and 69%, respectively). For model 1, the concavity condition is only satisfied for 33% of the sample but this is not a great concern because it is annual person delay per head which is the most valid measure of congestion. See Appendix A.2 for details of the test of the concavity condition.

Returns to scale at the sample mean is 1.06 for the index model and 0.39, 0.46 and 0.29 for models 2 – 4, respectively. This suggests that a reduction in the congestion index is characterized by mild increasing returns to scale, whereas a reduction in annual person delay per head and both measures of the monetary value of congestion are characterized by quite large diseconomies of scale. In Figure 1 we present the returns to scale elasticities outside the sample mean for quintiles of the population size distribution. For models 2–4 stable decreasing returns to scale are observed for cities in the 1st–3rd quintiles, whereas large fluctuations in returns to scale are observed for cities in the 4th and 5th quintiles.

[Insert Figure 1]

At the sample mean, vehicle miles per head per lane mile is not the source of the difference between returns to scale for model 1 and models 2 – 4 because the estimates of α_1 are similar in the four models. Rather the difference is because the proxy for suburbanization (α_2) and the ratio of peak travellers to city population (α_3) are bigger drivers of the most valid measure of congestion and both measures of the monetary value of congestion than is the case in model 1. Nevertheless, it is apparent from all four models that an increase in the proxy for suburbanization leads to a marked rise in both measures of congestion and both measures of the monetary value of congestion. This is presumably because an increase in suburbanization will lead to an increase in journey distance for the commute to work and so congestion is likely to set in further away from the central business district earlier in the morning. Similarly, in all four models the coefficient on the ratio of peak travellers to city population is large. This is almost certainly because the ratio is capturing the economic conditions in a city. This is borne out by the particularly large estimate of α_3 in model 4, where the monetary value of congestion is based on the average wage rate in the MSA.

It is evident from the estimation results that an increase in public transit passenger miles per head leads to a very small fall in the congestion index and non-negligible falls in annual person delay per head and both measures of the monetary value of congestion. From the elasticities in Figure 2 we can see that the impact of public transit ridership on both measures of congestion and both measures of the monetary value of congestion has declined over the study period for all quintiles of the population size distribution. This suggests that policymakers should now be placing less emphasis on increasing public transit ridership to manage congestion and its associated monetary value than in previous decades. Finally, we briefly discuss the implications of the estimates of δ_1 . The estimates of δ_1 in models 1 – 4 suggest that for a hypothetical average U.S. city with unchanged characteristics in the mid-year of the sample, the annual decrease in technical change ranges from 0.4% – 2.0%. One possible reason for the annual decline in technical change is the reluctance of policymakers to make widespread use of the available technology to internalize the congestion externality.

[Insert Figure 2]

4.2 Efficiency Results

An efficiency score of 100% for annual person delay per head, for example, would indicate that a city is doing the best that it can to tackle person delay per head, relative to the other cities in the sample. The efficiency scores by quintile of the population size distribution are presented in Appendix A.3 and the distributions of the efficiency scores are plotted in Figure 3.¹¹ The average efficiency score for the congestion index across the 88 cities is 91% and the average efficiency score for annual person delay per head and both measures of the monetary value of congestion

¹¹To plot the densities in Figures 3 and 5 we use the Gaussian density and obtain the bandwidth h using the Sheather and Jones (1991) solve-the-equation plug-in-approach. When estimating the kernel densities to avoid bias problems near the boundary, the reflection method, as described by Silverman (1986) and Scott (1992), is used.

is 72%. It is apparent that the distributions of efficiency scores from model 1 and models 2 – 4 differ greatly.¹² For models 2 – 4 there is moderate multi-modality in the density of the efficiency scores, whereas for model 1 the density of the efficiency scores is unimodal.

[Insert Figure 3]

A univariate distribution of efficiency scores à la Figure 3 does not provide any information about the relative performance of the cities. To shed some light on the relative performance of the cities we use the contour plot in Figure 4 of the efficiency scores from models 1 and 4.¹³ If, for example, in Figure 4 the cities were concentrated on the diagonal line we could conclude that, relative to the mean performance of the sample, cities' efficiency scores for the congestion index are no better or worse than cities' efficiency scores for the second measure of the monetary value of congestion. Cities are instead highly concentrated to the right and left of the diagonal so we can infer that some cities are relatively more efficient at tackling the congestion index, whereas some others are relatively more efficient at managing the second measure of the monetary value of congestion.

[Insert Figure 4]

Moving on to discuss the results for individual conurbations. In the following discussion, the income category (top third, middle third and bottom third), the categorization according to population size and Census region are in parentheses.¹⁴ The best three efficiency scores for the congestion index are Laredo TX –100% (bottom income-small-Southern city), Buffalo NY –99% (bottom income-large-Northeastern city) and Akron OH –99% (middle income-medium sized-Midwestern city). The three highest efficiency scores for annual person delay per head are Beaumont TX –98% (bottom income-small-Southern city), Lancaster-Palmdale CA –98% (top income-medium sized-Western city) and Oxnard-Ventura CA –97% (top income-medium sized-Western city). For model 3, the two most efficient cities are the same as for model 2 and the third best performer is Memphis TN-MS-AR –97% (middle income-large-Southern city). The best performers in model 4 are Salem OR –98% (bottom income-small-Western city), Buffalo NY –98% (bottom income-large-Northeastern city) and Sarasota-Bradenton FL –98% (top income-medium sized-Southern city).

The most inefficient cities in the congestion index model are Los Angeles-Long Beach-Santa Ana CA –75% (top income-very large-Western city), Cape Coral FL –73% (middle income-small-Southern city) and Knoxville TN –71% (bottom income-small-Southern city) whereas the most

¹²The correlation and Spearman rank correlation coefficients for the efficiency scores from models 1 and 2 are both around 0.46. For the efficiency scores from models 2 – 4, all the correlation and Spearman rank correlation coefficients are above 0.91. The Wilcoxon rank-sum (Mann-Whitney) test rejects the null of equality of the efficiency distributions for models 1 and 2 at the 1% level with a z -statistic of 7.68. This suggests that the two efficiency distributions are drawn from different populations. As was expected, the null for the Wilcoxon rank-sum test is not rejected for pairs of efficiency distributions from models 2 – 4.

¹³To construct Figures 4 and 6 we use bivariate Gaussian kernels and the bandwidths are calculated using the solve-the-equation plug-in approach for a bivariate Gaussian kernel à la Wand and Jones (1994). In Figure 4, the efficiency scores are normalized relative to the mean. As was expected, the contour plots of the efficiency scores from model 1 against the efficiency scores from model 2, 3 or 4 are indistinguishable so only the contour plot of the efficiency scores from model 1 against the scores from model 4 is presented.

¹⁴The basis for the income categorization is simply the ranking of cities according to mean real personal income per head over the study period.

inefficient cities in models 2 and 3 are Pittsburgh PA -40% in model 2 and 42% in model 3 (middle income-large-Northeastern city), Charleston-North Charleston SC -38% in model 2 and 37% in model 3 (bottom income-small-Southern city) and Spokane WA -36% in model 2 and 35% in model 3 (bottom income-small-Western city). The worst performers in model 4 are once again Pittsburgh PA -41% and Spokane WA -38% as well as San Jose CA -35% (top income-large-Western city). It is evident from these results that good and poor performance across the four models is not confined to cities which are located in a particular region or have particular income and population size characteristics.

It is evident from Figure 3 that the distribution of the efficiency scores for the most valid measures of congestion and its associated monetary value (models 2 and 4) are almost identical. Having said this, there are a 38 cities where a higher efficiency score is observed for model 2 *vis-à-vis* model 4. The most striking cases are Bridgeport-Stamford CT-NY -75% in model 2 and 50% in model 4, San Jose CA -51% in model 2 and 35% in model 4 and to a lesser extent Portland OR-WA -84% in model 2 and 70% in model 4, New Orleans LA -86% in model 2 and 76% in model 4 and Seattle WA -76% in model 2 and 66% in model 4. It is the monetary value of congestion which will potentially have implications for the growth of cities so for the above cities where the efficiency score from model 2 is greater than the score from model 4, we conclude that increasing how efficiently the most valid measure of the monetary value of congestion is managed so that it is similar to the efficiency score for annual person delay per head is an achievable target. How might these cities use the results reported here to achieve this target? They may analyze the traffic management projects in comparable cities (i.e. in terms of population size, income, etc.) with a higher efficiency score from model 4. It may also be worthwhile analyzing the approach to congestion management used in cities where the efficiency score for model 4 is substantially greater than the score for model 2 e.g. Riverside-San Bernardino CA, Phoenix AZ, Bakersfield CA, Brownsville TX, Laredo TX and Sarasota-Bradenton FL. Furthermore, very often case studies of urban transportation policies in other countries are used to inform domestic policy. The efficiency scores for, in particular, the most valid measure of the monetary value of congestion can be used by policymakers in other countries to identify which U.S. cities are good candidates for detailed case studies of traffic management policies.

5 TFP Growth Results

TFP growth rates for the 88 cities over the study period for models 1 – 4 are obtained from the input distance functions using (5).¹⁵ From the kernel densities of the TFP growth rates in Figure 5 we can see that TFP growth for the congestion index is clearly unimodal with a mode around zero. In contrast, the modes of the TFP growth rates from models 2 – 4 are positive and

¹⁵The correlation and Spearman rank correlation coefficients for the rates of TFP growth from models 1 and 2 are 0.01 and -0.04 , respectively. The correlation and Spearman rank correlation coefficients for the rates of TFP growth from model 1 and models 3 – 4 range from 0.00 – 0.01 and -0.05 – (-0.07) , respectively. The correlation and Spearman rank correlation coefficients for the TFP growth rates from models 2 – 4 are much higher and range from 0.40 – 0.47 and from 0.86 – 0.94, respectively.

The Wilcoxon rank-sum tests of the null of equality of TFP growth distributions from models 1 – 4 are rejected if the test involves the distribution from model 1. The other tests accept the null.

there is some evidence of multi-modality. To get an insight into the cities' relative productivity performance we draw on Figure 6 which is a contour plot of the rates of TFP growth for the most valid measures of congestion and its associated monetary value. Since cities are concentrated slightly to the right of the diagonal we can conclude that relative to the sample means, cities' TFP growth for annual person delay per head is slightly higher than TFP growth for the second measure of the monetary value of congestion.

[Insert Figures 5 and 6]

For models 1 – 4, average TFP growth and its three constituent parts (efficiency change, technical change and scale change) are plotted in Figure 7. Figure 8 graphs average TFP growth for quintiles of the population size distribution.¹⁶ It is evident from Figure 7 that the paths of average TFP growth for models 2 – 4 are similar, which is not surprising because annual person delay per head and both measures of the monetary value of congestion are based on the same delay time data. In all three cases average TFP growth is characterized by an upward trend. This is a very encouraging sign even though at the end of the study period there is either no change in TFP or it is falling slightly. The decomposition of average TFP growth for models 2 – 4 indicates that in all three cases the scale change is the principal driver of TFP growth. For the moment we observe this is the case and postpone any analysis of this finding until we make some policy recommendations in the final section of the paper.

[Insert Figures 7 and 8]

It is evident from the decomposition of average TFP growth for models 2 – 4 that, although the technical change component is negative throughout the study period, it is characterized by an upward trend. The upward trend is possibly because over the study period cities have increasingly made use of the technology which is available to help tackle congestion. There still remains plenty of opportunity to make further use of the available technology and the extent to which policymakers do so in the longer term is likely to have a big bearing on TFP growth for the most valid measures of congestion and its associated monetary value. Finally, we note that there are short cycles in average TFP growth for the congestion index (see the top panel of Figure 7). This is not surprising because the congestion index is calculated using data on urban vehicle miles which will depend on the business cycle. From the path of average TFP growth for the congestion index we can see that there is a regular cyclical pattern up until 1995. The cyclical pattern is much more irregular from thereon which is attributed to the repeal of the National Maximum Speed Limit in 1995, returning all authority on speed limits to the states. Short cycles in average TFP growth for some of the quintiles of the population size distribution are also observed for models 1 – 4 (see selected quintiles in all four panels of Figure 8). It is evident for some of the quintiles that some of the short cycles in average TFP growth are not synchronized because average TFP growth in Figure 7 is relatively smooth for models 2 – 4.

¹⁶Figures 7 and 8 were constructed having omitted the outliers, which were taken to be the top and bottom 2.5% of the standardized distributions.

6 Policy Recommendations

Although over the study period both measures of congestion and both measures of the monetary value of congestion have increased significantly, which is very concerning, our TFP growth results for the more recent measure of congestion (annual person delay per head) and both measures of the monetary value of congestion suggest that the outlook is quite positive. This is because over the study period for annual person delay per head and both measures of the monetary value of congestion, we find that average TFP growth is characterized by an upward trend. In the early years, average TFP declined sharply for annual person delay per head and both measures of the monetary value of congestion but in the latter years average TFP growth was zero or approaching zero in all three cases.

The input and output elasticities outside the sample mean for quintiles of the population size distribution can be used to shed some light on policies which can further improve TFP growth performance. From the models for annual person delay per head and both measures of the monetary value of congestion, the average daily vehicle miles per head per lane mile elasticities are negative for the vast majority of the cities in the 1st–4th quintiles but for nearly all the cities in the 5th quintile the elasticities are positive. This suggests that for cities in the 5th quintile, there is a net fall in annual person delay per head and both measures of the monetary value of congestion when lane miles are increased i.e. the fall in congestion and its associated monetary value from increasing road space more than offsets the rise from induced travel. The opposite is the case for urban areas in the 1st–4th quintiles. This suggests urban transportation policymakers for cities in the 1st–4th quintiles should not tackle annual person delay per head and both measures of the monetary value of congestion by increasing lane miles because it will not have the intended effect.

Although our results suggest that increasing lane miles would reduce annual person delay per head and both measures of the monetary value of congestion for cities in the 5th quintile, Winston and Langer (2006) find that state investment in highways is not a cost effective way of reducing road users' congestion costs. This suggests that other policy tools should be used in the largest cities. It is entirely consistent with average TFP growth for the most recent measure of congestion and both measures of the monetary value of congestion to conclude that recent innovative policies (e.g. ramp metering, incentives to carpool and encouraging employers to offer flexi-time and telecommuting arrangements) have been effective. At the very least, therefore, we recommend that these innovations are implemented more widely in the short to medium term to try and maintain the trend in average TFP growth for the most valid measures of congestion and its associated monetary value. The decomposition of TFP growth into technical change, efficiency change and scale change suggests that for both measures of congestion and both measures of the monetary value of congestion, the principal driver of average TFP growth is the scale change. How does the scale change explain the upward trend in average TFP growth for the more recent measure of congestion and both measures of the monetary value of congestion? One possibility is that as cities grow and the scale effect gets bigger, annual person delay per head and its associated monetary value will rise and will become a bigger concern for urban transportation policymakers. Policymakers will therefore be faced with using more innovative and productive policies. Having

said this, ramp metering, incentives to carpool and encouraging employers to offer flexi-time and telecommuting arrangements are all relatively conservative innovations and it is likely that more innovative policies such as widespread road pricing will be needed to maintain the trend in average TFP growth in the medium to long term. Because of the apparent success of relatively conservative innovations policymakers may be more willing to commit to more ambitious innovations. And it would not necessarily be very difficult to implement road pricing more widely and could simply involve introducing more HOT lanes by changing the status of a lot of the existing HOV lanes where buy-in is not permitted.

7 Acknowledgements

The authors would like to thank two anonymous referees and Tom Weyman-Jones for constructive comments on an earlier draft of this paper. We also acknowledge the comments from participants at the 2011 European Workshop on Efficiency and Productivity Analysis (EWEPA) and the 2011 Kuhmo Nectar Conference on Transportation Economics.

References

- ADAMS, R. M., A. N. BERGER AND R. C. SICKLES (1999): ‘Semiparametric approaches to stochastic panel frontiers with applications in the banking industry’. *Journal of Business and Economic Statistics*, Vol. 17, pp. 349-358.
- AIGNER, D., C. A. K. LOVELL AND P. SCHMIDT (1977): ‘Formulation and estimation of stochastic frontier production function models’. *Journal of Econometrics*, Vol. 6, pp. 21-37.
- ATKINSON, S. E. AND J. H. DORFMAN (2005): ‘Bayesian measurement of productivity and efficiency in the presence of undesirable outputs: Crediting electric utilities for reducing air pollution’. *Journal of Econometrics*, Vol. 126, pp. 445-468.
- BATTESE, G. E. AND T. J. COELLI. (1992): ‘Frontier production functions, technical efficiency and panel data: With application to paddy farmers in India’. *Journal of Productivity Analysis*, Vol. 3, pp. 153-169.
- BOOZE ALLEN HAMILTON AND HNTB (2008): *A compendium of existing HOV lane facilities in the United States*. Report prepared for the U.S. Federal Highway Administration.
- CAVES, D. W., L. R. CHRISTENSEN AND W. E. DIEWERT (1982): ‘The economic theory of index numbers and the measurement of input, output and productivity’. *Econometrica*, Vol. 50, pp. 1393-1414.
- CERVERO R. AND M. HANSEN (2002): ‘Induced travel demand and induced road investment: A simultaneous equation analysis’. *Journal of Transport Economics and Policy*, Vol. 36, pp. 469-490.
- COELLI, T., A. ESTACHE, S. PERELMAN AND L. TRUJILLO (2003): ‘A primer on efficiency measurement for utilities and transport regulators’. Washington, D.C: World Bank Institute.
- CORNWELL, C., P. SCHMIDT AND R. C. SICKLES (1990): ‘Production frontiers with cross-sectional and time-series variation in efficiency levels’. *Journal of Econometrics*, Vol. 46, pp. 185-200.
- CUESTA, R. A., C. A. K. LOVELL AND J. L. ZOFIO (2009): ‘Environmental efficiency measurement with translog distance functions: A parametric approach’. *Ecological Economics*, Vol. 68, pp. 2232-2242.
- DIEWERT, W. E. AND T. J. WALES (1987): ‘Flexible functional forms and global curvature conditions’. *Econometrica*, Vol. 55, pp. 43-68.
- DURANTON, G. AND M. A. TURNER (2011): ‘The fundamental law of road congestion: Evidence from US cities’. *American Economic Review*, Vol. 101, pp. 2616-2652.
- FARE, R., S. GROSSKOPF, C. A. K. LOVELL AND C. PASURKA (1989): ‘Multilateral productivity comparisons when some outputs are undesirable: A nonparametric approach’. *Review of Economics and Statistics*, Vol. 71, pp. 90-98.
- FARE, R., S. GROSSKOPF AND D. TYTECA (1996): ‘An activity analysis model of the environmental performance of firms- application to fossil-fuel-fired electric utilities’. *Ecological Economics*, Vol. 18, pp. 161-175.
- FERNÁNDEZ, C., G. KOOP AND M. F. J. STEEL (2005): ‘Alternative efficiency measures for multiple-output production’. *Journal of Econometrics*, Vol. 126, pp. 411-444.
- GREENE, W. (2005): ‘Reconsidering heterogeneity in panel data estimators of the stochastic frontier model’. *Journal of Econometrics*, Vol. 126, pp. 269-303.

- HERNANDEZ-SANCHO, F., A. PICAZO-TADEO AND E. REIG-MARTINEZ (2000): 'Efficiency and environmental regulation- an application to Spanish wooden goods and furnishings industry'. *Environmental and Resource Economics*, Vol. 15, pp. 365-378.
- KOOP, G. AND L. TOLE (2008): 'What is the environmental performance of firms overseas?: An empirical investigation of the global gold mining industry'. *Journal of Productivity Analysis*, Vol. 30, pp. 129-143.
- McFADDEN, D. (1978): 'Cost, revenue and profit functions'. In M. Fuss and D. McFadden (eds.) *Production Economics: A Dual Approach to Theory and Applications*. Amsterdam: North-Holland.
- MEEUSEN, W. AND J. VAN DEN BROECK (1977): 'Efficiency estimation from Cobb-Douglas production functions with composed error'. *International Economic Review*, Vol.18, pp. 435-444.
- OREA, L. (2002): 'Parametric decomposition of a generalized Malmquist productivity index'. *Journal of Productivity Analysis*, Vol. 18, pp. 5-22.
- PITT, M. M. AND L-F. LEE (1981): 'Measurement and sources of technical inefficiency in the Indonesian weaving industry'. *Journal of Development Economics*, Vol. 9, pp. 43-64.
- REINHARD, S., C. A. K. LOVELL AND G. J. THIJSEN: (1999): 'Econometric estimation of technical and environmental efficiency: An application to Dutch dairy farms'. *American Journal of Agricultural Economics*, Vol. 81, pp. 44-60.
- REINHARD, S., C. A. K. LOVELL AND G. J. THIJSEN: (2000): 'Environmental efficiency with multiple environmentally detrimental variables; estimated with SFA and DEA'. *European Journal of Operational Research*, Vol. 121, pp. 287-303.
- SAAL, D., D. PARKER AND T. WEYMAN-JONES (2007): 'Determining the contribution of technical, efficiency and scale change to productivity growth in the privatized English and Welsh water and sewerage industry: 1985-2000'. *Journal of Productivity Analysis*, Vol. 28, pp. 127-139.
- SCOTT, D. W. (1992): *Multivariate Density Estimation: Theory, Practice, and Visualization*. New York: Wiley.
- SHEATHER, S. J. AND M. C. JONES (1991): 'A reliable data-based bandwidth selection method for kernel density estimation'. *Journal of the Royal Statistical Society: Series B*, Vol. 53, pp. 683-690.
- SILVERMAN, B. W. (1986): *Density Estimation for Statistics and Data Analysis*. Chapman and Hall: London.
- SMALL, K. A. AND E. T. VERHOEF (2007): *The Economics of Urban Transportation*. Abingdon, Oxfordshire: Routledge.
- TEXAS TRANSPORTATION INSTITUTE (2009): *2009 Annual Urban Mobility Report*. Texas Transportation Institute: College Station, Texas.
- TYTECA, D. (1997): 'Linear programming models for the measurement of environmental performance of firms- concepts and empirical results'. *Journal of Productivity Analysis*, Vol. 8, pp. 183-197.
- WAND, M. P. AND M. C. JONES (1994): 'Multivariate plug-in bandwidth selection'. *Computational Statistics*, Vol. 9, pp. 97-116.
- WINSTON, C. AND A. LANGER (2006): 'The effect of government highway spending on road users' congestion costs'. *Journal of Urban Economics*, Vol. 60, pp. 463-483.
- ZAIM, O. AND F. TASKIN (2000): 'A Kuznets curve in environmental efficiency: An application on OECD countries'. *Environmental and Resource Economics*, Vol. 17, pp. 21-36.

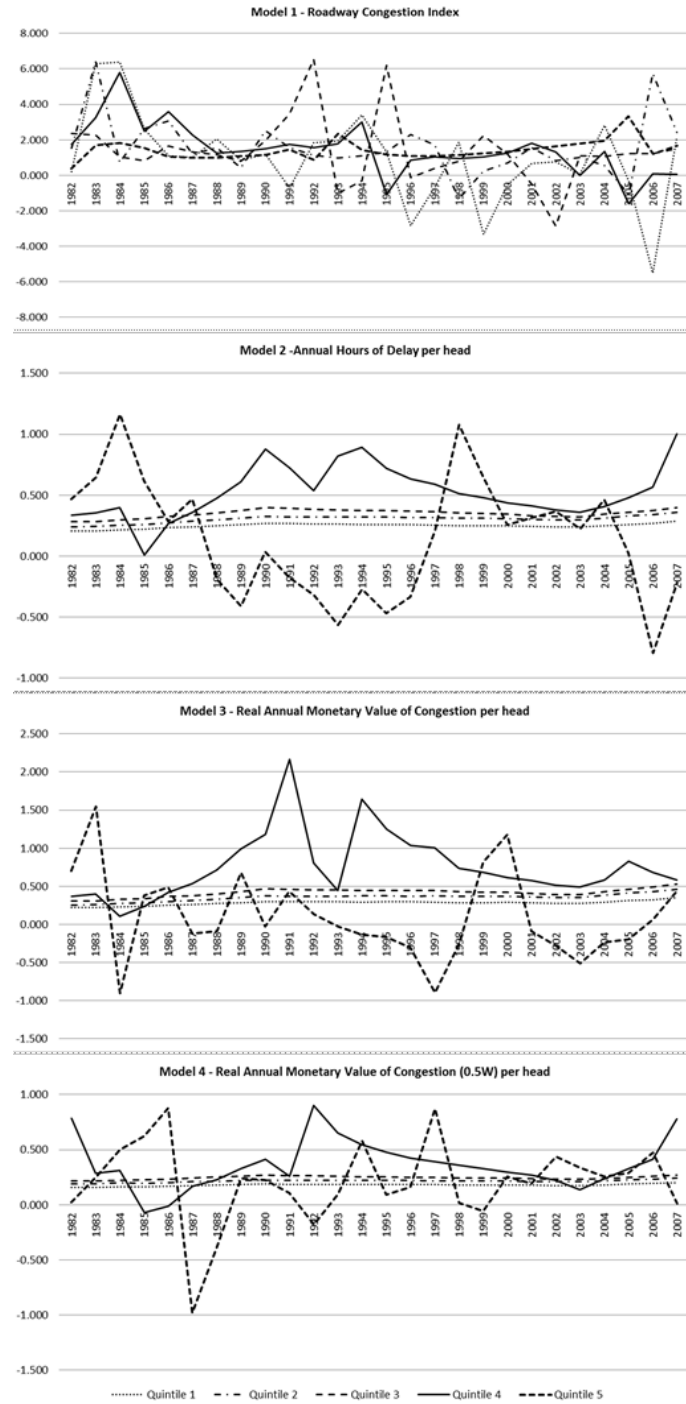


Figure 1: Returns to scale elasticities by quintile: 1982-2007

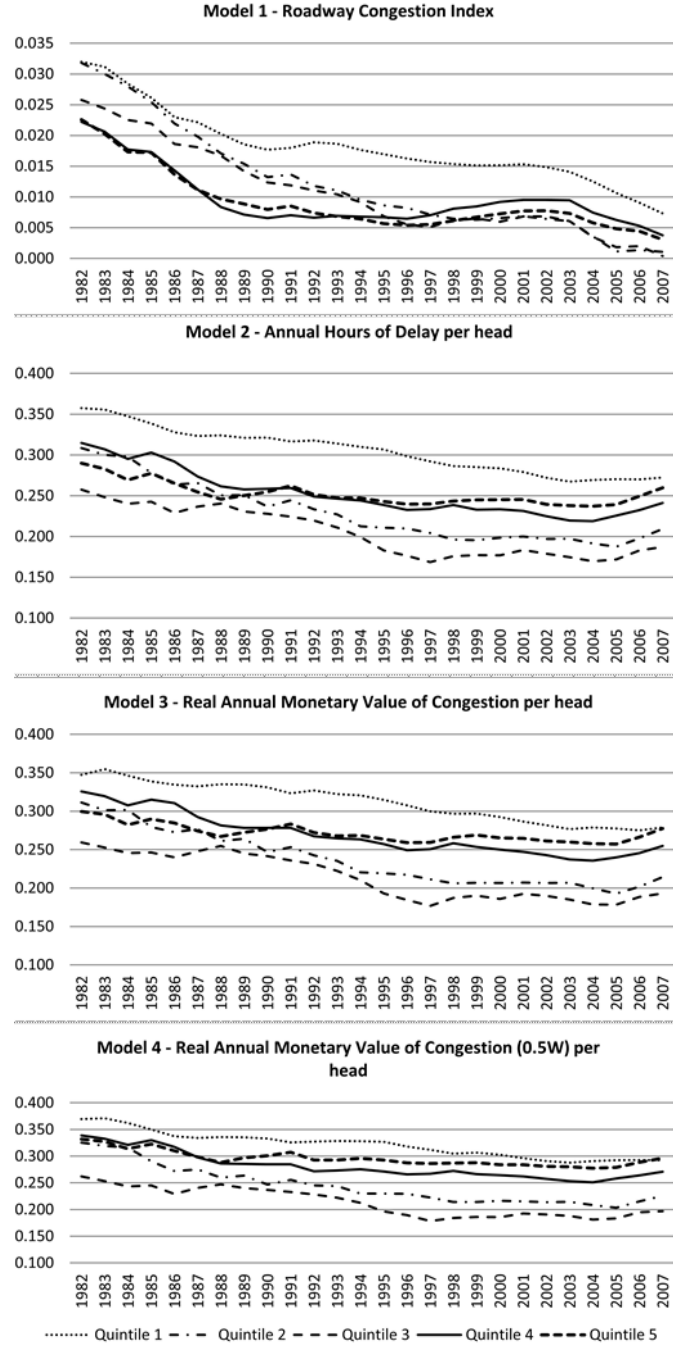


Figure 2: Public transit passenger miles per head elasticities by quintile: 1982-2007

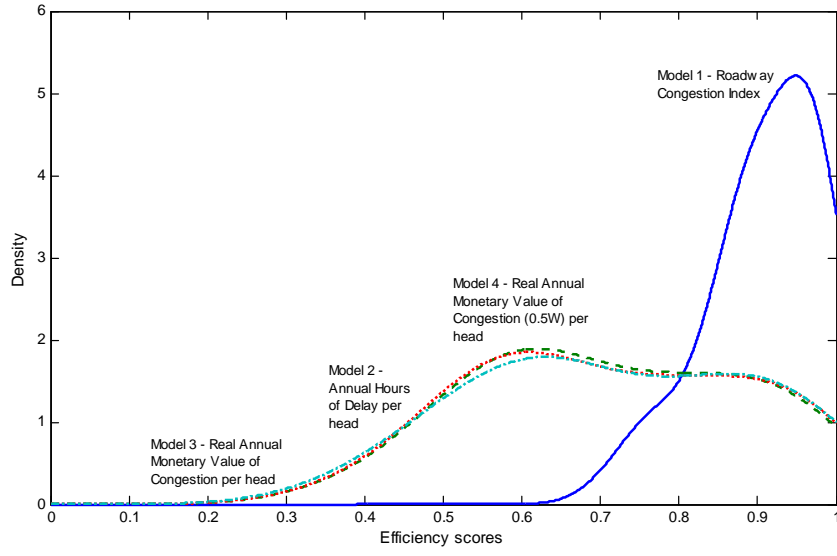


Figure 3: Kernel densities of the efficiency scores from the four estimated models

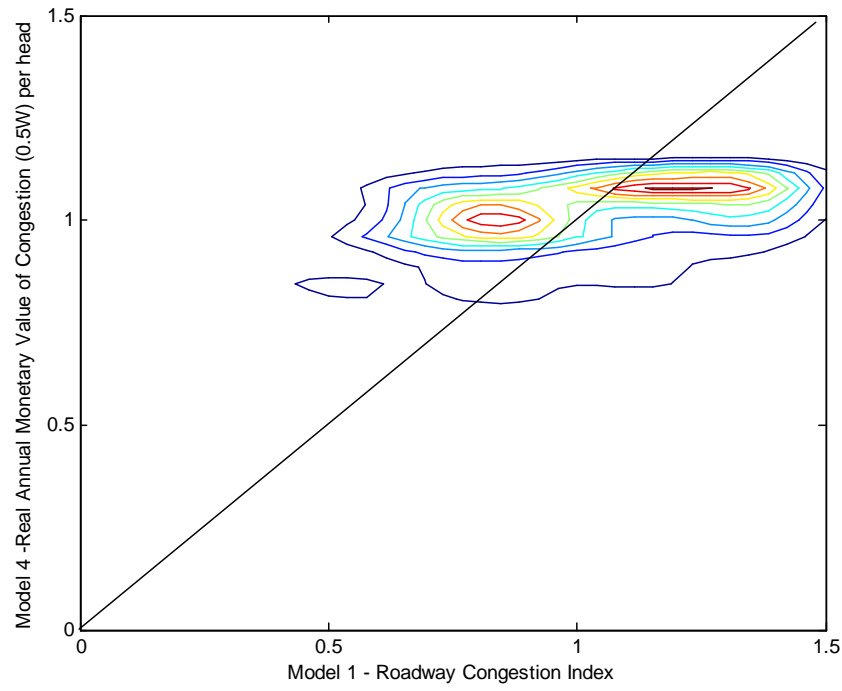


Figure 4: Contour plot of the efficiency scores for the congestion index and the real monetary value of congestion (0.5W) per head

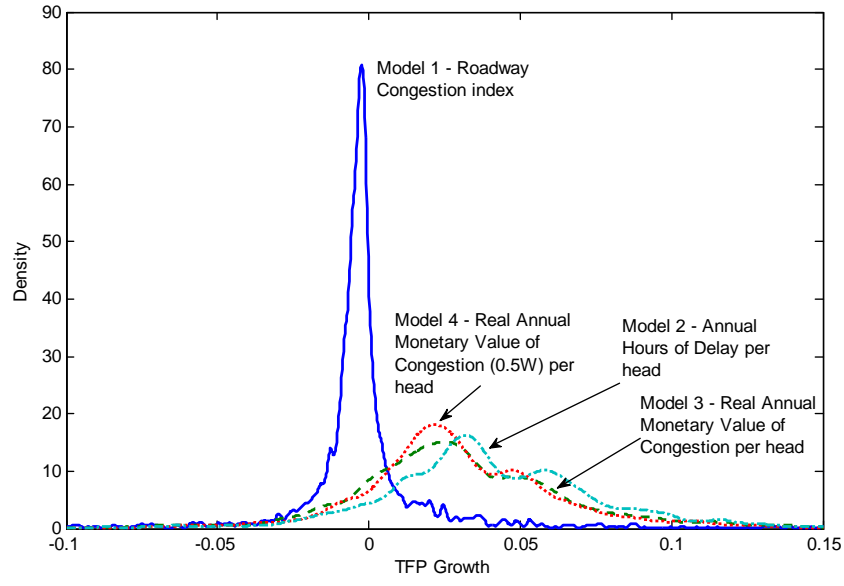


Figure 5: Kernel densities of TFP growth from the four estimated models

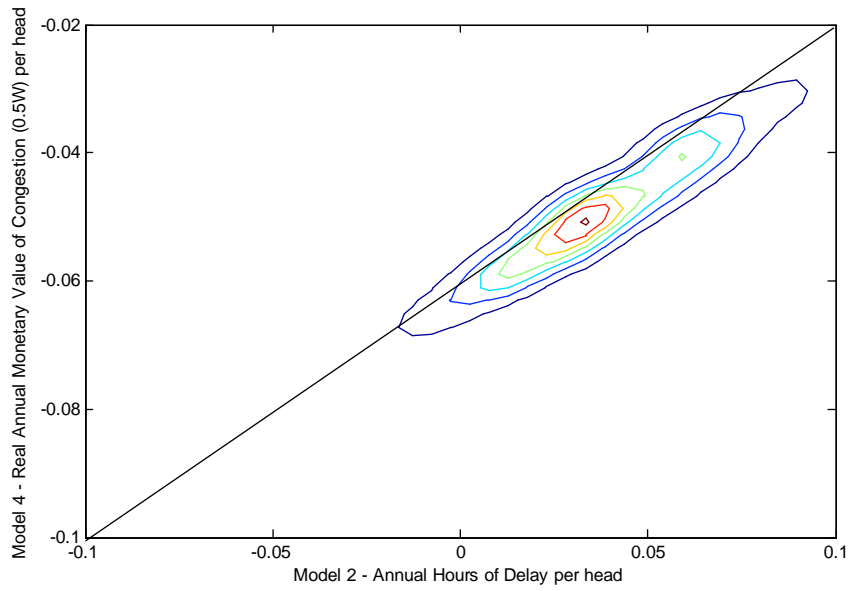


Figure 6: Contour plot of TFP growth for annual hours of delay per head and the real monetary value of congestion (0.5W) per head

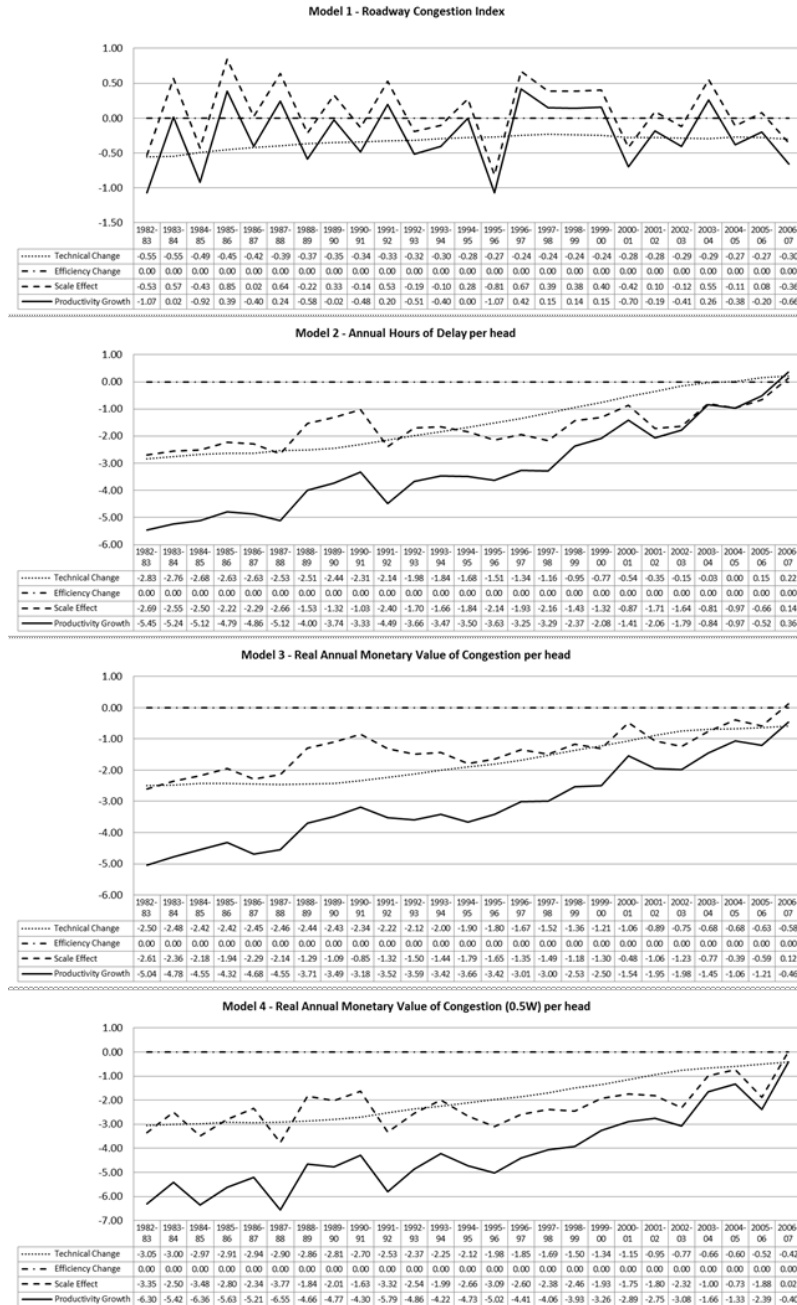


Figure 7: Average TFP growth: 1983-2007 .

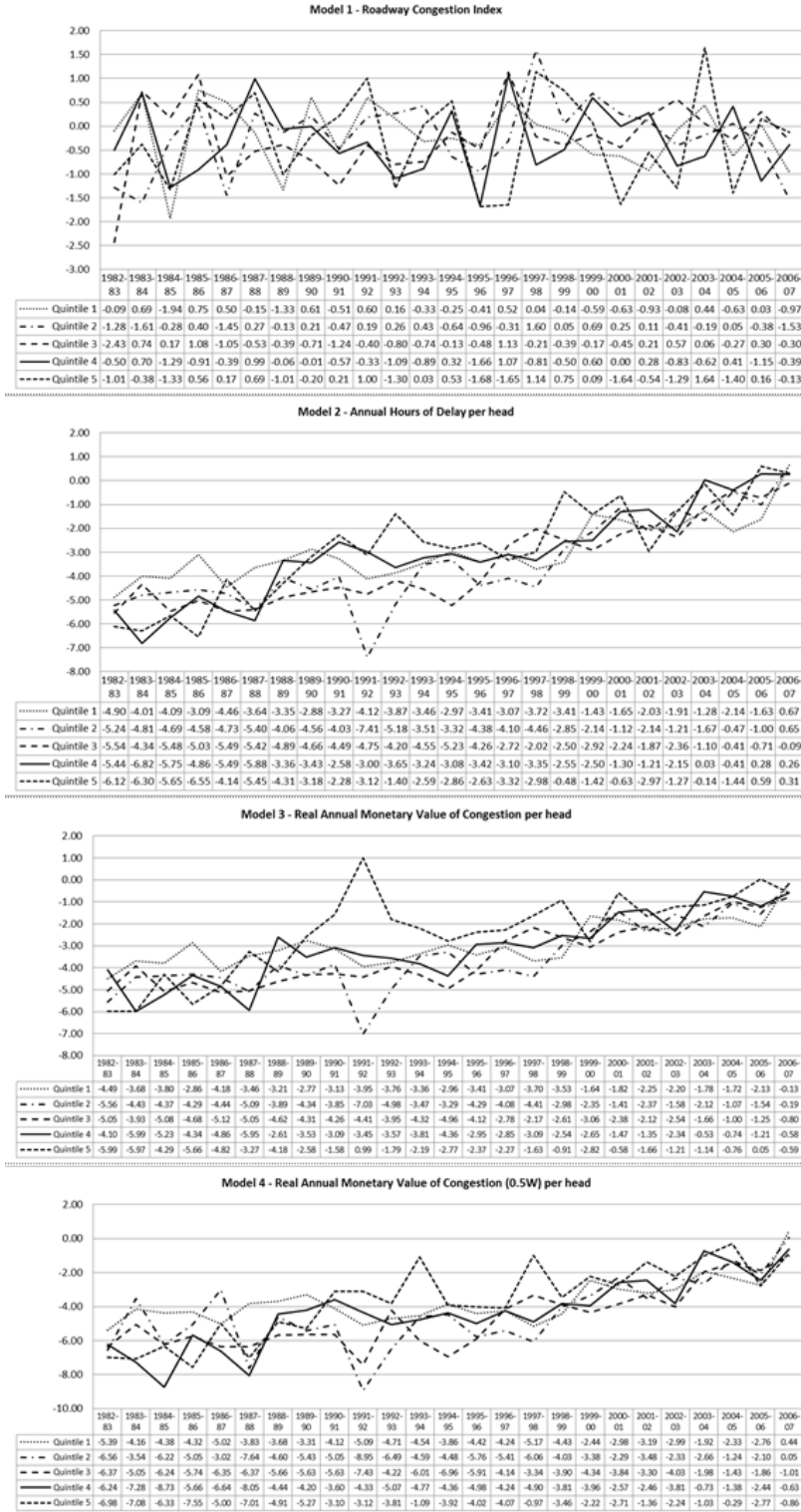


Figure 8: Average TFP growth by quintile: 1983-2007 .

Table 1: Summary statistics

	Variable	Mean	St.Dev.	Min	Max
<i>Variables minimized i.e. inputs</i>					
Roadway congestion index (Model 1)	x_1	0.91	0.21	0.41	1.58
Annual person hours of delay per head (Model 2)	x_1	11.02	7.63	0.64	39.13
Real annual value of congestion per head (Model 3)	x_1	102.24	71.69	6.19	384.48
Real annual value of congestion (0.5W) per head (Model 4)	x_1	64.56	56.01	2.13	373.17
Annual passenger miles on public transit per head	x_2	116.87	139.93	1.97	1175.1
<i>Variables held constant i.e. outputs</i>					
Average daily vehicle miles per head per lane mile	y_1	0.01	.008	.0003	.052
Number of hours per day when there is congestion somewhere on the city's network	y_2	5.02	1.74	2.00	7.99
Ratio of average daily peak travellers to city population	y_3	1.62	0.08	1.43	1.86
<i>Weather characteristics</i>					
Number of days per year when the temperature is greater than 90 deg. F	z_1	57.92	45.64	0.00	189.99
Annual precipitation	z_2	33.20	16.05	1.88	69.16
<i>Geographical location</i>					
Northeast (#)	z_3	0.16	0.37	0	1
Midwest (#)	z_4	0.19	0.40	0	1
South (#)	z_5	0.41	0.49	0	1
West (#)	z_6	0.28	0.45	0	1
<i>Population size and density characteristics</i>					
Small (#)	z_7	0.17	0.38	0	1
Medium (#)	z_8	0.34	0.47	0	1
Large (#)	z_9	0.33	0.47	0	1
Very large (#)		0.16	0.37	0	1
Population Density (persons per square mile)	z_{10}	2203.7	853.1	988.9	5664.0
<i>Roadway links and speed limits</i>					
City located on a major interstate highway (#)	z_{11}	0.77	0.42	0	1
States are given authority over all speed limits (#)	z_{12}	0.47	0.50	0	1
Rural interstate speed limit law change (#)	z_{13}	0.79	0.39	0	1

Notes: # denotes a dummy variable. The sum of the means for the regional dummies exceeds 100%. This is because a small number of cities lie in two regions such as Philadelphia (Northeast and South) and Cincinnati (Midwest and South).

Table 2: Estimated input distance function parameters

Variable	Parameter	Model 1— Congestion Index (x_1)		Model 2— Annual Person Hours of Delay per head (x_1)		Model 3— Real Annual Congestion Value per head (x_1)		Model 4— Real Annual Congestion Value (0.5W) per head (x_1)	
		Coef.	Std. Error	Coef.	Std. Error	Coef.	Std. Error	Coef.	Std. Error
$\ln y_1$	α_1	-0.108	0.009***	-0.131	0.030***	-0.107	0.030***	-0.110	0.030***
$\ln y_2$	α_2	-0.376	0.007***	-0.862	0.029***	-0.868	0.030***	-0.859	0.030***
$\ln y_3$	α_3	-0.462	0.133***	-1.579	0.461***	-1.223	0.464**	-2.517	0.475***
$\ln x_2$	β_1	0.012	0.003***	0.249	0.010***	0.261	0.010***	0.271	0.011***
$(\ln y_1)^2$	$A_{1,1}$	-0.021	0.004***	-0.100	0.014***	-0.110	0.014***	-0.109	0.014***
$(\ln y_2)^2$	$A_{2,2}$	0.000	0.013	0.215	0.062***	0.214	0.062***	0.306	0.066***
$(\ln y_3)^2$	$A_{3,3}$	7.662	1.857***	-25.753	6.991***	-25.809	7.017***	-25.436	7.301***
$\ln y_1 \ln y_2$	$A_{1,2}$	0.015	0.009	0.203	0.034***	0.203	0.034***	0.243	0.035***
$\ln y_1 \ln y_3$	$A_{1,3}$	-0.118	0.138	-1.412	0.477**	-1.330	0.481**	-1.987	0.479***
$\ln y_2 \ln y_3$	$A_{2,3}$	-1.791	0.253***	-0.633	1.040	-0.751	1.045	0.029	1.106
$(\ln x_2)^2$	B_1	0.009	0.001***	0.067	0.005***	0.070	0.005***	0.080	0.005***
$\ln y_1 \ln x_2$	$\Gamma_{1,2}$	0.008	0.003*	0.042	0.011***	0.037	0.012**	0.031	0.012*
$\ln y_2 \ln x_2$	$\Gamma_{2,2}$	-0.016	0.007*	0.041	0.032	0.046	0.032	0.116	0.033***
$\ln y_3 \ln x_2$	$\Gamma_{3,2}$	0.433	0.128***	-0.353	0.421	-0.460	0.427	0.165	0.456
t	δ_1	-0.004	0.001***	-0.015	0.002***	-0.017	0.002***	-0.020	0.003***
t^2	δ_2	-0.00001	0.000	-0.000	0.0002	-0.000	0.0002	-0.000	0.0002
$\ln y_1 t$	μ_1	0.001	0.001	0.004	0.003	0.004	0.003	0.007	0.003*
$\ln y_2 t$	μ_2	0.011	0.002***	-0.000	0.006	-0.000	0.006	-0.009	0.006
$\ln y_3 t$	μ_3	-0.030	0.021	0.303	0.075***	0.316	0.075***	0.316	0.077***
$\ln x_2 t$	η_1	-0.003	0.001***	0.005	0.002*	0.006	0.002*	0.005	0.002
z_1	π_1	0.004	0.005	0.012	0.009	0.015	0.009	0.020	0.010
z_2	π_2	-0.040	0.014**	-0.211	0.033***	-0.215	0.033***	-0.119	0.033***
z_3	π_3	0.015	0.024	0.068	0.096	0.044	0.090	0.002	0.098
z_4	π_4	-0.062	0.016***	0.018	0.074	0.006	0.071	-0.036	0.071
z_5	π_5	-0.066	0.021**	-0.148	0.077	-0.161	0.074*	-0.147	0.078
z_6	π_6	-0.089	0.033**	-0.280	0.102**	-0.292	0.100**	-0.201	0.103
z_7	π_7	0.503	0.034***	1.180	0.113***	1.116	0.105***	1.230	0.108***
z_8	π_8	0.244	0.023***	0.545	0.115***	0.480	0.101***	0.665	0.094***
z_9	π_9	0.131	0.022***	0.125	0.090	0.080	0.082	0.203	0.082*
z_{10}	π_{10}	-0.086	0.008***	-0.021	0.032	-0.016	0.031	0.018	0.031
z_{11}	π_{11}	0.037	0.017*	-0.123	0.052*	-0.131	0.047**	-0.090	0.044*
z_{12}	π_{12}	0.002	0.004	0.015	0.012	0.016	0.013	0.008	0.013
z_{13}	π_{13}	-0.010	0.005*	0.003	0.017	0.027	0.018	0.008	0.018
Constant		-0.063	0.027*	0.146	0.130	0.211	0.117	0.050	0.114
Log likelihood function		4253.9		1243.2		1209.0		1153.6	

Note: *, **, *** denote statistical significance at the 5%, 1% and 0.1% levels, respectively.

A Appendices

A.1 Monotonicity properties outside the sample mean

	Monotonicity of x_2 (property (i))	Monotonicity of y_1 (property (iv))	Monotonicity of y_2 (property (iv))	Monotonicity of y_3 (property (iv))
Model 1 - Congestion Index (x_1)				
Quintile 1	83.3	100.0	100.0	63.9
Quintile 2	77.6	100.0	100.0	69.2
Quintile 3	77.4	100.0	100.0	82.7
Quintile 4	75.3	100.0	100.0	75.8
Quintile 5	73.9	87.6	100.0	82.7
Model 2 - Annual Person Hours of Delay per head (x_1)				
Quintile 1	99.1	100.0	100.0	100.0
Quintile 2	100.0	100.0	100.0	100.0
Quintile 3	98.9	100.0	100.0	100.0
Quintile 4	100.0	66.3	100.0	84.2
Quintile 5	99.4	2.6	100.0	51.3
Model 3 - Real Annual Monetary Value of Congestion per head (x_1)				
Quintile 1	98.9	100.0	100.0	100.0
Quintile 2	100.0	100.0	100.0	100.0
Quintile 3	98.9	100.0	100.0	99.8
Quintile 4	100.0	48.9	100.0	75.6
Quintile 5	99.4	0.0	100.0	47.4
Model 4 - Real Annual Monetary Value of Congestion (0.5W) per head (x_1)				
Quintile 1	98.9	100.0	100.0	100.0
Quintile 2	100.0	100.0	100.0	100.0
Quintile 3	98.9	100.0	100.0	100.0
Quintile 4	100.0	48.9	100.0	88.9
Quintile 5	99.4	0.9	100.0	51.3

Notes: The numbers are percentages of elasticities which satisfy the monotonicity properties outside the sample mean. Quintile 1 refers to the smallest cities with average population size over the study period between 0 and 20 percentiles and Quintile 5 refers to the largest cities between 80 and 100 percentiles.

A.2 Test of the concavity condition

Microeconomic theory assumes that the form of an input distance function, D_I , satisfies a particular curvature property i.e. D_I is linearly homogeneous and concave in inputs x . The symmetry property and homogeneity of degree one in x of D_I implies that only $K(K-1)/2$ elements of the Hessian matrix, $H(x)$, where $H(x) \equiv \frac{\partial^2 D_I}{\partial x \partial x'}$, are linearly independent. Applying the arguments of Diewert and Wales (1987), the Hessian of the input distance function with respect to x can be calculated as follows:

$$H(x) = \mathbf{B} - \widehat{ex} + exex',$$

where: \widehat{ex} is a diagonal matrix with estimated input elasticities, ex_k for $k = 1, \dots, K-1$, on the leading diagonal and zeros elsewhere; ex is a vector of estimated input elasticities, ex_k ; \mathbf{B} is the matrix of second order coefficients on the input terms in the translog function.

Concavity of the input distance function in x requires that the Hessian matrix is negative semidefinite. Whether the Hessian matrix is negative semidefinite can be verified from the sign pattern of the principal minors of the Hessian. The necessary and sufficient conditions for D_I to be concave are as follows: all the odd-numbered principal minors of the Hessian must be non-positive and all the even-numbered principal minors must be non-negative. At the sample mean with mean corrected data, the Hessian is given by

$$H(\bar{x}) = \mathbf{B} - \widehat{\beta} + \beta\beta',$$

where $\widehat{\beta}$ is a diagonal matrix with estimated input elasticities, β_k for $k = 1, \dots, K-1$, on the leading diagonal and zeros elsewhere, and β is a vector of estimated input elasticities, β_k . The Stata code for the concavity test is available from the corresponding author on request.

A.3 Efficiency scores

City	Model 1— Congestion Index		Model 2— Annual Person Hours of Delay per head		Model 3— Real Annual Value of Congestion per head		Model 4— Real Annual Value of Congestion (0.5W) per head	
	Eff. Score	Rank	Eff. Score	Rank	Eff. Score	Rank	Eff. Score	Rank
1st quintile of cities by population size								
Bakersfield CA	0.993	6	0.734	42	0.703	44	0.836	28
Beaumont TX	0.864	69	0.977	1	0.974	2	0.939	14
Boulder CO	0.976	18	0.914	16	0.948	13	0.857	25
Brownsville TX	0.967	24	0.791	33	0.751	40	0.958	10
Cape Coral FL	0.734	87	0.602	63	0.608	61	0.674	47
Colorado Springs CO	0.993	4	0.772	36	0.790	35	0.768	38
Columbia SC	0.772	82	0.587	67	0.557	70	0.642	57
Corpus Christi TX	0.834	76	0.506	80	0.506	81	0.543	73
Eugene OR	0.844	75	0.636	56	0.632	54	0.632	60
Indio-Cathedral City-Palm Springs CA	0.928	41	0.558	71	0.549	72	0.647	55
Knoxville TN	0.706	88	0.461	85	0.457	84	0.505	77
Lancaster-Palmdale CA	0.922	44	0.975	2	0.978	1	0.969	5
Laredo TX	0.996	1	0.589	66	0.567	66	0.696	44
Little Rock AR	0.802	79	0.676	47	0.675	46	0.701	41
Pensacola FL-AL	0.768	83	0.785	35	0.812	29	0.824	31
Poughkeepsie-Newburgh NY	0.972	19	0.935	14	0.947	14	0.938	16
Salem OR	0.906	51	0.964	6	0.967	4	0.975	1
Spokane WA	0.755	84	0.359	88	0.351	88	0.376	87
<i>Average</i>	<i>0.874</i>		<i>0.712</i>		<i>0.710</i>		<i>0.749</i>	
2nd quintile of cities by population size								
Akron OH	0.993	3	0.957	9	0.952	12	0.970	4
Albany-Schenectady NY	0.928	42	0.565	70	0.559	69	0.545	72
Albuquerque NM	0.899	56	0.575	69	0.566	67	0.599	64
Allentown-Bethlehem PA-NJ	0.923	43	0.647	51	0.660	52	0.663	49
Charleston-North Charleston SC	0.750	85	0.376	87	0.372	87	0.424	85
El Paso TX-NM	0.971	20	0.809	31	0.800	32	0.859	24
Fresno CA	0.963	27	0.506	79	0.479	82	0.580	67
Grand Rapids MI	0.978	17	0.520	76	0.521	77	0.542	74
New Haven CT	0.876	62	0.737	41	0.761	38	0.686	46
Omaha NE-IA	0.953	32	0.630	57	0.664	51	0.637	58
Oxnard-Ventura CA	0.912	47	0.971	3	0.962	8	0.958	9
Raleigh-Durham NC	0.989	9	0.935	13	0.906	20	0.939	15
Sarasota-Bradenton FL	0.857	74	0.889	20	0.943	15	0.972	3
Springfield MA-CT	0.868	67	0.746	40	0.795	34	0.755	40
Toledo OH-MI	0.981	14	0.964	8	0.963	7	0.954	12
Tucson AZ	0.865	68	0.475	83	0.444	85	0.489	80
Wichita KS	0.883	60	0.641	55	0.673	48	0.663	50
<i>Average</i>	<i>0.917</i>		<i>0.703</i>		<i>0.707</i>		<i>0.720</i>	
3rd quintile of cities by population size								
Austin TX	0.983	13	0.929	15	0.911	18	0.852	26
Birmingham AL	0.884	59	0.613	59	0.607	62	0.605	61
Bridgeport-Stamford CT-NY	0.812	78	0.746	39	0.749	41	0.497	79
Charlotte NC-SC	0.979	15	0.909	17	0.876	22	0.862	22
Columbus OH	0.986	11	0.826	28	0.812	28	0.834	29
Dayton OH	0.904	55	0.606	61	0.617	57	0.579	68
Hartford CT	0.859	72	0.507	78	0.524	76	0.453	84
Indianapolis IN	0.944	36	0.642	54	0.601	63	0.650	53
Jacksonville FL	0.958	31	0.819	30	0.831	26	0.804	34
Las Vegas NV	0.909	49	0.556	72	0.551	71	0.602	63
Louisville KY-IN	0.937	38	0.530	75	0.529	74	0.532	75

City	Model 1— Congestion Index		Model 2— Annual Person Hours of Delay per head		Model 3— Real Annual Value of Congestion per head		Model 4— Real Annual Value of Congestion (0.5W) per head	
	Eff. Score	Rank	Eff. Score	Rank	Eff. Score	Rank	Eff. Score	Rank
3rd quintile of cities by population size (cont.)								
Memphis TN-MS-AR	0.971	21	0.968	4	0.969	3	0.963	7
Nashville-Davidson TN	0.875	63	0.499	81	0.510	80	0.488	81
Oklahoma City OK	0.907	50	0.895	19	0.928	16	0.865	21
Richmond VA	0.911	48	0.604	62	0.625	55	0.548	70
Rochester NY	0.874	64	0.663	49	0.682	45	0.633	59
Salt Lake City UT	0.858	73	0.494	82	0.511	79	0.465	82
Tulsa OK	0.965	25	0.646	52	0.674	47	0.603	62
<i>Average</i>	<i>0.918</i>		<i>0.692</i>		<i>0.695</i>		<i>0.658</i>	
4th quintile of cities by population size								
Buffalo NY	0.994	2	0.965	5	0.964	5	0.973	2
Cincinnati OH-KY-IN	0.989	10	0.951	12	0.952	11	0.952	13
Cleveland OH	0.933	39	0.954	10	0.959	10	0.895	19
Denver-Aurora CO	0.886	58	0.613	60	0.610	59	0.547	71
Kansas City MO-KS	0.968	23	0.851	23	0.841	24	0.860	23
Milwaukee WI	0.968	22	0.667	48	0.668	49	0.700	43
New Orleans LA	0.992	7	0.855	21	0.820	27	0.764	39
Orlando FL	0.964	26	0.474	84	0.475	83	0.500	78
Pittsburgh PA	0.871	66	0.401	86	0.419	86	0.413	86
Portland OR-WA	0.905	53	0.837	26	0.834	25	0.700	42
Providence RI-MA	0.914	45	0.844	24	0.892	21	0.923	17
Riverside-San Bernardino CA	0.781	81	0.822	29	0.784	36	0.959	8
Sacramento CA	0.876	61	0.703	44	0.717	43	0.643	56
San Antonio TX	0.959	30	0.789	34	0.799	33	0.788	36
San Jose CA	0.906	52	0.508	77	0.516	78	0.351	88
Tampa-St. Petersburg FL	0.819	77	0.852	22	0.909	19	0.923	18
Virginia Beach VA	0.950	33	0.752	38	0.760	39	0.830	30
<i>Average</i>	<i>0.922</i>		<i>0.755</i>		<i>0.760</i>		<i>0.748</i>	
5th quintile of cities by population size								
Atlanta GA	0.863	70	0.652	50	0.623	56	0.653	52
Baltimore MD	0.905	54	0.590	65	0.565	68	0.530	76
Boston MA-NH-RI	0.895	57	0.617	58	0.586	65	0.586	65
Chicago IL-IN	0.942	37	0.827	27	0.802	31	0.841	27
Dallas-Fort Worth-Arlington TX	0.990	8	0.906	18	0.915	17	0.891	20
Detroit MI	0.948	35	0.678	46	0.667	50	0.690	45
Houston TX	0.978	16	0.840	25	0.846	23	0.779	37
Los Angeles-Long Beach-Santa Ana CA	0.748	86	0.583	68	0.612	58	0.657	51
Miami FL	0.960	28	0.954	11	0.963	6	0.966	6
Minneapolis-St. Paul MN	0.863	71	0.601	64	0.608	60	0.584	66
New York-Newark NY-NJ-CT	0.950	34	0.964	7	0.961	9	0.956	11
Philadelphia PA-NJ-DE-MD	0.984	12	0.799	32	0.808	30	0.810	33
Phoenix AZ	0.871	65	0.688	45	0.599	64	0.816	32
San Diego CA	0.800	80	0.544	73	0.543	73	0.575	69
San Francisco-Oakland CA	0.959	29	0.706	43	0.769	37	0.791	35
Seattle WA	0.993	5	0.761	37	0.719	42	0.664	48
St. Louis MO-IL	0.913	46	0.643	53	0.637	53	0.647	54
Washington DC-VA-MD	0.932	40	0.541	74	0.528	75	0.458	83
<i>Average</i>	<i>0.916</i>		<i>0.716</i>		<i>0.708</i>		<i>0.716</i>	
Sample Average								
<i>Sample Average</i>	<i>0.909</i>		<i>0.715</i>		<i>0.716</i>		<i>0.718</i>	
Sample Standard Deviation								
<i>Sample Standard Deviation</i>	<i>0.071</i>		<i>0.168</i>		<i>0.172</i>		<i>0.174</i>	