# Public bus transport demand elasticities in India

---

Kaushik Deb, TERI University, New Delhi, India

Massimo Filippini, ETH, Zurich, Switzerland

## Address for Correspondence

TERI University, Plot No. 10 Institutional Area, Vasant Kunj, New Delhi - 110 070, India (kdeb@teriuniversity.ac.in)

Massimo Filippini is a Professor of Economics at the Department of Management, Technology, and Economics, ETH, Zurich, Switzerland

## Abstract

A number of static and dynamic specifications of a log linear demand function for public transport are estimated using aggregate panel data for 22 Indian states over the period 1990 to 2001. Demand has been defined as total passenger kilometers to capture actual market transactions, while the regressors include public transit fare, per capita income, service quality, and other demographic and social variables. In all cases, transit demand is significant and inelastic to the fare. Service quality is the most significant policy variable. Finally, social and demographic variables highlight the complex nature of public bus transit demand in India.

## Keywords

Demand Elasticities, Dynamic Panel Data, Bus Transport, India

## JEL Classification

R41

# 1.0 Introduction

Public bus transport is of great import in India. Not only is an efficient public bus system important for meeting the mobility needs in this rapidly growing economy, but a higher share of bus transport would also have a positive impact on pollution, both local and global, and energy demand. That apart, improved availability of public transport is critical for ensuring access to basic services such as education and health, and integrating rural communities into the economic mainstream. Hence, it is incumbent on governments in developing countries to institute appropriate policy initiatives to increase the share of public transport. Such interventions must be informed by research that identifies factors influencing the demand for public transport and quantifies the impact of environmental and policy variables. The most common method for characterizing the influence of such variables is by estimating the elasticity of demand with respect to each of these variables.

To this end, this paper estimates the elasticity of demand with respect to various policy and environmental variables that influence the demand for public bus transport in India. The role of both monetary and non–monetary variables in explored. All states with public bus transport in India are included in this study. The focus of the estimates is on price, income, and service quality while carrying out the estimations and presenting the elasticities. The approach here derives from neo–classical microeconomics, applied to data from a developing country to estimate the demand structure of the industry at the state level. The current research is possibly one of few studies that use panel data for a thorough analysis. In terms of methodology, this research compares the results from several econometric models, that have not all been applied in this context.

There are two major types of empirical transit demand studies, namely, those derived from the Random Utility Theory that analyze the choice of a transport mode (Winston (1983; Oum (1989)), and those derived from conventional analysis of consumer utility maximization that analyze continuous consumption patterns. In the former, the transport good is considered to be discrete, and demand is analyzed as a choice problem between competing modes such as bus and personal vehicles given a fixed level of aggregate travel (Oum et al. (1992)). In the latter, quantity changes in demand are analyzed using demand models that take quantity as a continuous variable (McCarthy (2001)). Demand analysis in the case of a continuous variable, in turn, follows one of two approaches. The first approach estimates a system of equations simultaneously for several commodities or commodity groups. The second focuses only on one commodity, or a commodity group, and hence essentially estimates the demand in a single market. In either case, with a complete systems approach that is theoretically more consistent, a more comprehensive dataset is required that includes demand for, or expenditures on, all commodity groups. In the absence of such an extensive dataset, equations are specified in a more ad hoc manner including cross–commodity influences from only close substitutes and complements (Thomas (1987)).

The empirical approach used for any estimation thus is dependent on the research objectives, and is constrained by the data available. This research uses an unbalanced aggregate panel dataset between 1990/91 and 2000/01 for 22 large states in India to assess the price and income effects on public bus transport demand. Here, direct price elasticities can be obtained after estimating an aggregate single equation demand model.

The paper is organized as follows. The next section discusses the relevant literature on number, timing, and spatial distribution of trips by mode in estimating travel demand, all of which are infinitely faceted and hence can result in a large variety of alternatives for each consumer, making travel demand modeling complex (Jovicic et al. (2003)). The specification used in this research is given in section 3.0. The estimation process and the data used are given in section 4.0. Section 5.0 presents the results of the analysis and discusses the implications therein, and finally, section 6.0 concludes.

# 2.0 Literature review

The literature has been reviewed in the context of the framework suggested by Berechman (1993) assessing the impact that different specifications and estimation approaches have on demand elasticities. In combination with the data available, the literature review allows identification of variables to be included and estimation of an appropriate specification to obtain the price elasticities of demand. The focus in this review is on aggregate demand estimations, ignoring the extensive literature estimating discrete modal choices. A summary of recent studies using either panel data or those estimating aggregate demand functions is presented in Table1. Following that section, issues in estimating travel demand using aggregate data, and how they have been addressed in the relevant literature, are presented and outlined.

Dargay et al. (1999) present a comprehensive review of the literature followed by demand estimations at the national, regional, and county levels in the United Kingdom, using annual time series data between 1970 and 1996. They use a dynamic specification of aggregate demand relating journeys per capita to real bus fares defined as revenue per journey, real per capita income, and service level defined as bus kilometers. By estimating a dynamic relationship, they distinguish between short run and long run elasticities, with the short run defined by the periodicity of the data, one year in this case. The estimated short run price elasticity for the entire country is –0.4 increasing to –0.9 in the long run. The regional price elasticities vary between –0.2 and –2.0 in the short run and –0.4 and –1.7 in the long run. Service quality elasticities at the national level are estimated to be 0.4 in the short run and 0.9 in the long run. The wealth effects are measured using the income elasticities and vary between –0.3 and –0.4 in the short run and –0.5 and –1.0 in the long run, making public transport an inferior good.

Romilly (2001) uses annual time series between 1953 and 1997 for the United Kingdom excluding London, to estimate a dynamic log linear demand function as a single equation Auto Regressive Distributed Lag model, after correcting for cointigrating relationships. Demand is defined as passenger journeys per person, with the regressors being bus fares and motoring costs, real personal disposable income, and service frequency proxied by vehicle kilometers per person. The fare elasticity is estimated to be –0.38 in the short run and –1.03 in the long run, the income elasticities are 0.23 in the short run and 0.61 in the long run, and finally, service elasticities are 0.11 in the short run and 0.30 in the long run.

Dargay et al. (2002) estimate a partial adjustment model relating per capita bus patronage to bus fares, income, and service level, using a panel dataset of 46 counties in England for the period 1987–1996. Two specifications are estimated, namely, log linear and semi log, with only the transit fare in levels in the latter. The models estimated include Fixed Effects, Random Effects, and Random Coefficients, where again only the coefficients on transit fare vary between counties. Interestingly, demographic variables are not found to be significant in the estimation. The results are similar to Dargay et al. (1999).

Bresson et al. (2003) estimate demand as a function of fares, service supply, and income using separate panels of 46 counties in England over 1987 and 1996, and 62 French urban areas over 1986 and 1995, with a partial adjustment specification. They estimate Fixed and Random Effect models and compare the results with a Random Coefficients approach, suggesting that the latter provide improved elasticity estimates. The English dataset is the same used by Dargay et al. (2002) while the French panel consists of 62 urban areas between 1987 and 1995. The fare elasticities for the two countries lie in the interval of –0.2 to –0.5 in the short run and –0.5 to –0.8 in the long run.

A small number of studies have used panel data that combine cross sectional and time series data. The current research is possibly one of the few studies estimating travel demand using aggregate data for developing countries. Unlike other studies that estimate demand functions for India using datasets comprising a limited number of firms or cities, this study uses panel data from almost all states in India and hence provides a comprehensive analysis of public transit demand in India. In terms of methodology, this research compares the results from several econometric models detailed in section 4, that have not all been applied in this context.

Most travel demand models use the number of trips or passengers as the dependent variable (Hanly et al. (1999; Romilly (2001; Dargay et al. (2002)). A trip, comprised of a combination of an origin with a destination, can be definitely defined as a commodity, and hence can be priced. Dargay et al. (1999) is the only study in the literature review undertaken that uses passenger kilometers as the measure of demand for their aggregate national analysis of travel demand. However, using only the number of trips or passengers as a measure of travel demand ignores an important characteristic of demand, the length of each trip. This is clearly an important parameter that also reflects the motivation for the supply and pricing of public transit services.

The objective of this research is to identify factors that influence public bus transit demand from the perspective of the bus transport industry. Hence, the definition of demand needs to reflect actual market transactions. Using passenger kilometers as an output measure allows transit demand to be related to a supply measure and can then be used to analyze public transit markets, as is the objective of this research. The two measures of transit demand, the number of passengers and passenger kilometers, are highly correlated in the dataset used in this analysis, with a correlation of over 90%. Hence, passenger kilometers are taken as the output measure. Moreover, a measure of the extent of operation, proxied by the population of the state that the firm is based in, could be suitably used as an indicator of market size.

In terms of independent variables, the studies listed in Table 1Table show that the empirical estimation of a demand function is determined by monetary and non–monetary variables. Monetary variables include the price of the product, prices of available alternatives, and wealth or income levels. Non–monetary variables include non–price product attributes such as quality and other characteristics, and consumer tastes. Consumer tastes are represented by non–income characteristics of households such as demographic or cultural attributes such as occupation, lifestyle, age, and gender (Wabe (1969; Kemp (1973)). Matas (2004) uses the level of suburbanization and employment levels to explain demand changes in Madrid during 1979–2001. The empirical estimation of the effect of these variables on transit demand is not always straightforward since many of them are highly correlated with income or other socioeconomic variables.

A public transit demand model should include some variables representing the quality of the service. Some studies use output measures such as vehicle kilometers as service quality measures (Goodwin et al. (1985; Fitzroy et al. (1993; Balcombe et al. (2004)). Such measures, however, result in an identification problem between the variable defining demand, and the variable defining service quality. In addition, service quality changes due to changes in capacity, such as larger buses resulting in more seat kilometers, would be ignored in such a measure (Balcombe et al. (2004)). Other aggregate service quality measures use the ratio of network length to area size or population as a proxy of access to transit services to avoid such identification issues (Romilly (2001; Dargay et al. (2002)). Bresson et al. (2003) estimate a log linear specification with income, price, and network density as variables for quality. FitzRoy et al. (1997) argue that journey time is an important quality parameter and use average frequency and route density as proxies.

There are various functional forms that have been used in the literature to estimate aggregate transit demand, namely, linear functions, semi–log or log linear, and generalized non–linear models (de Rus (1990); Appelbaum et al. (1991)). The most common functional form used is the log linear (Romilly (2001)). Only a handful studies have estimated a semi–log functional form where only transit price is included in levels and all other explanatory variables are in logs (Dargay et al. (2002; Bresson et al. (2003)). Statistically, a log linear specification significantly reduces the number of coefficients to be estimated. In terms of the estimates, the coefficients can be readily interpreted as elasticities. Finally, the log linear form also allows for non–linear interactions between demand and the various parameters, hence capturing more complex relationships than simple linear effects (Oum (1989; Clements et al. (1994)). Since the focus of this study is to estimate direct price elasticities for transit demand, a log linear specification is estimated.

## 3.0 Model specification

The model specification presented in this section is based on the review of the literature presented above and the issues discussed therein. Since the study assesses public bus transit price elasticities in the context of actual market transactions, passenger kilometers have been taken as the output measure ( $pkm$ ). Public bus transit fares ( $p$ ) and per capita income ( $w$ ) are the monetary variables. Service quality is characterized by the density of coverage ( $q$ ). The total population ( $pop$ ) of the state is included to isolate the effect of size of the market. The demographic and socioeconomic variables in the model are the proportion of population in the labour force ( $work$ ) and literacy rate ( $lit$ )[1].

[1] The proportion of population living in urban areas and the sex ratio were also included in early specifications on the model. However, these variables did not significantly improve the goodness of fit. In addition, in terms of the elasticities obtained for the key variables of interest, these were not found to have any significant influence.

Unfortunately, data on the prices of substitutes and complements is not available in this study. The only significant transport service here is personal vehicle usage. The impact of changes in personal vehicle usage can be approximated using another socioeconomic variable, per capita private vehicle ownership ( $s$ ).

The model specification used is the following:

$$pkm = f\left(p, w, q, s, pop, work, lit\right) \tag{1}$$

From the studies reviewed in Table 1, the functional forms most commonly used in the literature are log linear and semi–log. Since the log linear form is easily interpretable, and simple for computing elasticities, the log linear function has been estimated[2]. The demographic variables are already in percentages. These have not been converted into logs and are included as reported. In this case, the coefficients can be readily interpreted as elasticities. Thus, the static model is the following,

$$\begin{aligned}\mathrm{Ln}pkm = {} & \alpha_o + \alpha_p \mathrm{Ln}p + \alpha_w \mathrm{Ln}w + \alpha_q \mathrm{Ln}q + \alpha_s \mathrm{Lns} + \alpha_{pop} \mathrm{Ln}pop \\ & + \alpha_{work} work + \alpha_{lit} lit + \varepsilon_t\end{aligned} \tag{2}$$

The dynamic structure of demand has been captured using a partial adjustment model. This implies that given an optimum, but unobservable, level of transit demand, $pkm^*$ , demand only gradually converges towards the optimum level between any two time periods. Hence,

$$\mathrm{Ln}pkm_t - \mathrm{Ln}pkm_{t-1} = \delta(\mathrm{Ln}pkm^* - \mathrm{Ln}pkm_{t-1}) + \eta_t \tag{3}$$

where $1-\delta$ is the adjustment coefficient indicating the rate of adjustment of $pkm$ to $pkm^*$ and $\varepsilon_t$ is random disturbance (Kmenta (1978)). Substituting $pkm^*$ in the dynamic adjustment equation gives:

[2] The coefficient estimates obtained from using the log linear and the semi–log functional forms were compared and found to be similar.

$$\text{Ln}pkm_t = \alpha'_o + \alpha'_p \text{Ln}p + \alpha'_w \text{Ln}w + \alpha'_q \text{Ln}q + \alpha'_s \text{Ln}s + \alpha'_{pop} \text{Ln}pop$$
$$+ \alpha'_{work} work + \alpha'_{lit} lit + (1-\delta)\text{Ln}pkm_{t-1} + \varepsilon_t' \tag{4}$$

where $\alpha_i' = \alpha_i \delta$ and $\varepsilon_t' = \delta \varepsilon_t + \eta_t$. This dynamic specification is estimated.

This is possibly one of the few studies estimating public bus transit demand in developing countries. The specification being used also attempts to capture actual market transactions to relate these with firm behaviour using passenger kilometers as a measure of demand. In addition, using density of coverage provides a clear indicator of service quality in terms of access to the transit network, and hence avoids simultaneity with the measure of demand and output. Finally, the use of demographic and social characteristics is expected to reveal the import of such non–monetary variables in the context of a developing country.

## 4.0 Data and econometric approaches

An unbalanced panel of 22 states in India between 1990/91 and 2000/01 has been used in the analysis with 206 observations. The panel ranges from 21 states in 1993/94 to 16 in 1997/98. This data set is characterized by a relatively small number of cross-sectional units and a relatively long time period. Data on public bus transit demand has been taken from CIRT (Various years)[3]. Public bus transit fares have been estimated as the ratio between traffic revenue and total demand, with the information obtained from CIRT (Various years). Thus, non–traffic revenue such as advertising revenue or interest accrued, has been excluded from the definition of public transit fares. Unfortunately, user costs and external costs are not available for this study and hence only public bus transit fares are included. Hence, the price elasticities obtained are only for public bus transit fares and not generalized transportation costs for the public bus users as in Mohring (1970).

[3] For the six states with more than one operator, data has been summed across all the operators to obtain state level aggregates.

Density of coverage has been estimated as the ratio between vehicle kilometers reported in CIRT (Various years) and the area of each state. Demographic and social variables have been obtained from Census of India 2001 (2001). The per capita income series is based on total State Domestic Product reported in EPWRF (2003) and population totals from Census of India 2001 (2001). Private vehicles in the analysis have been defined as cars, two–wheelers, and jeeps with the data from MTS (Various Issues). This has been divided by the population of each state to obtain the per capita private vehicle ownership. The two monetary variables, namely public bus transit fares and per capita incomes, have both been deflated to 1989/91 prices using the Wholesale Price Index for All Commodities reported by the Government of India (2005) to carry out the estimations in terms of real values. Table 2 describes the dataset and the variables used in the analysis. Each observation of each variable, $x_{it}$ ,has also been decomposed into two separate series of between observations $\left(\bar{x}_i = \sum_t x_{it} / T\right)$ and within observations $\left(x_{it} - \bar{x}_i + \sum_i \bar{x}_i / I\right)$ to examine the cross section and time series behaviour in terms of the Between and Within standard deviations (STATA (2005)). The Between estimates reflect the cross section variation in the dataset, while temporal changes can be observed through the Within Variation. For most variables, the overall variation in the dataset comes from the Between Variation. For instance, the variation in per capita income, density of coverage, and private vehicle ownership is almost completely due to the Between Variation. In addition, there is a large variation in the dataset for most variables as can be observed from the minimum and maximum values. Hence, it is important to include a variable that reflects the size differences between states. This size effect is captured by using the total population of each state.

Early estimations using cross section datasets are usually considered to reflect long term relationships with the implicit assumption that all variables are at their long term equilibrium levels and consumers have adjusted to these values completely (Kmenta (1978)).

While cross sectional analysis can clearly identify the importance of inherent individual variation between different observations and hence isolate the impact of the variables under consideration from general heterogeneity, they are unable to identify the dynamics of adjustment. Estimations that use time series datasets, on the other hand, generally focus on transitions in variables over time. Hence, the values obtained from time series estimations are considered to reflect short run values, with variablesnot being at their long term equilibrium values. However, heterogeneity impacts often cannot be separated from other variables in time series datasets as can be done with cross section datasets (Hsiao (2003)).

With panel datasets, it is possible to distinguish between the short run and long run characteristics and address heterogeneity issues in parallel, hence combining the advantages of both cross section and time series analysis. This allows the quantification of effects that are not identified in time series and cross section analysis independently (Hsiao (2003)).

With regard to the choice of econometric technique, it should be noted that in the econometric literature we can find various types of models focusing on cross-sectional variation, i.e. heterogeneity across units. Moreover, we can distinguish between static and dynamic approaches.

The three most widely used static approaches are: the fixed-effects model (FE), the random effects model (RE) and the Kmenta approach.[4] Concerning the Kmenta approach, Beck et al. (1995) recognize that this method depends on knowing the true error process and in the absence of this knowledge, leads to a downward bias in the estimates of the standard errors and recommend using PCSE (Panel Corrected Standard Errors). PCSE uses Ordinary Least Squares parameter estimates but replaces Ordinary Least Squares standard errors with

[4] For a detailed presentation of the econometric methods that have been used to analyze panel data, see Greene (2003) and Baltagi (1995). The Kmenta approach is also technically known as the cross-sectionally heteroskedastic and timewise autoregressive model (Kmenta, 1986). This approach is attractive when N, the number of units, is lower than T, the number of periods, or when the within variation of many explanatory variables is very low.

panel corrected standard errors. In this paper we estimated the static version of the demand model using the the FE, RE and PCSE models.

In a dynamic specification of the model, there is autocorrelation between subsequent periods leading to persistence over time. Here the within estimator for Fixed Effects is biased and inconsistent, especially if the number of periods is not large (Nickell (1981; Kiviet (1995)). Similarly, the Random Effects estimator is also biased and not efficient (Sevestre et al. (1985; Baltagi (2002)). Often the individual effects are assumed to be Fixed and not Random to address the non–orthogonality issues (Bun et al. (2001)). Using Monte Carlo simulations, Doel van den et al. (1995) report that static panel models usually underestimate long run effects if the true specification is dynamic.

The commonly used technique to estimate panel data models with unobserved heterogeneity is to transform the model into first differences and then use sequential moment conditions to estimate parameters using Generalized Method of Moments. Arellano–Bond (Arellano et al. (1991)) present a Generalized Method of Moments estimator for panels with a dynamic specification that removes individual effects by carrying out estimation in differences. This is estimation with the instruments in levels while the regressors are in differences. While the lagged variable is still endogenous, deeper lags are assumed orthogonal to the error term and hence are used as instruments. The prerequisite for this model is that the number of periods should be larger than the number of regressors in the model, and the number of instruments should be less than the number of cross sectional units. Dargay et al. (2002) and Bresson et al. (2003) estimate the Arellano–Bond model using panel data from counties in England only in the case of the former, and counties both in England and France, in the latter, to distinguish between the short run and long run elasticities.

However, with highly persistent data, the first differenced Generalized Method of Moments estimators may suffer a small sample bias due to weak instruments. Here Arellano–

Bover (Arellano et al. (1995)) suggest an alternative transformation to the Arellano–Bond differencing of the dependent variable and the regressors. By carrying out estimations in first differences, the Arellano–Bond approach drops more observations in unbalanced panels. The Arellano–Bover approach uses differences from the mean of all future observations to reduce the loss of observations arising from unbalanced panels. Blundell–Bond (Blundell et al. (1998)), using the Arellano–Bover approach, present a Generalized Method of Moments estimator that uses differences of instruments to obtain orthogonality instead of differencing the regressors in the Arellano–Bond estimator. The principle used here is that even if the regressors used are endogenous to the model, as long as they are independent of the individual effects, the first differences of the regressors can be used as valid instruments and hence improve the efficiency of the estimates. Blundell–Bond use extra moment conditions that rely on stationarity of the initial observations. Abrate et al. (2007) is possibly the only application of the Blundell–Bond approach to estimating public transit demand yet.

The choice of the number of instruments here is an issue that needs to be addressed. On one hand, enough instruments are required so that the finite sample properties in such estimations are satisfactory. On the other hand, each additional instrument over and above the number of explanatory variables bias the estimates (Kennedy (2003)). Arellano et al. (1991) suggest the Sargan test which tests the joint hypothesis that the model is correctly specified and that the instruments used are valid. Hence, the Sargan test can be used to evaluate the performance of the Generalized Method of Moments based dynamic panel data models by assessing the use of instruments in obtaining consistent estimates. Bun et al. (2007) argue that with a highly persistent series, with a small sample of cross section and time series observations, the Blundell–Bond approach may lead to weak instruments. This is in part due to the high variance in the individual effects dues to variance in transitory shocks.

Kiviet (1995) proposes a Bias Corrected LSDV (Least Squares Dummy Variables) estimate, or a Fixed Effects estimate, by estimating the sample bias from an uncorrected LSDV estimate and using this to remove the inconsistency in the parameter estimates. This has been refined and simplified in Bun et al. (2003). The approximation depends on not just the conditioning variables but also on the unknown true parameter values. However, Monte Carlo experiments have shown that approximations arising from such a bias correction are 'very accurate for a wide range of parameterizations' (Bun et al. (2003)). Due to the small variance of the LSDV estimator, much smaller than the Generalized Method of Moments estimators, the parameter estimates are also very efficient. Again Abrate et al. (2007) is one application of this approach to public transit demand. In this paper we estimated the dynamic version of the demand model using the following models: Arellano-Bond, Blundell–Bond and Corrected-LSDV.

## 5.0 Analysis and results

The data has been analyzed and the estimations carried out in STATA Intercooled Version 10.0. Three models each for both the static and dynamic specifications have been estimated. Dynamic models allow a distinction between long run and short run effects. A comparison with the static models demonstrates the importance of persistence in demand, and the difference between the short run and long run equilibrium behaviour. In the static specification, the first type of models are the conventional static one way panel data models, namely, Fixed Effects and Random Effects. Both these models have been estimated with a first order Autoregressive specification of their error structure. The second type, the PCSE method as proposed by Beck et al. (1995), is an alternative to the conventional panel data models. The PSCE is appropriate for pooled datasets with low within variation as is the case with our dataset (Table 2), and in the presence of heteroscadesticity and autocorrelation. In terms of the dynamic specification, the two Generalized Method of Moments based models, Arellano–Bond and Blundell–Bond, have been estimated. The two models are distinguished by the way instruments are constructed for each system. The Corrected LSDV estimations provide an alternative estimate to the Generalized Method of Moments models for the dynamic specification.

### *5.1 Comparing the models*

As previously mentioned, three static models and three dynamic models have been estimated. The static and dynamic specifications cannot be directly compared in terms of statistical performance except in terms of general goodness of fit and significance of key variables. Overall, only general remarks comparing the models are possible. The empirical results are presented in Table 3.

The Fixed and Random Effects models can be directly compared. The Hausman test comparing the coefficients on the regressors in the Fixed Effects and Random Effects rejects the null hypothesis that the Random Effects Coefficients are consistent ( $\chi^2_{(7)} = 152.27$ ). However, as pointed out by Cameron et al. (2005), the low Within Variation for several of the regressors could result in imprecise coefficients in the Fixed Effects model since it relies on Within Variation to carry out the estimation. Moreover, Random Effects estimates can be applied outside the sample for predictions, which is not appropriate for estimates obtained from the Fixed Effects models (Cameron et al. (2005)). This is important in the context of the current research since the objective is to identify general policy directions.

The results reported for the PCSE are similar to those of the Random Effects model as discussed earlier. The hypothesis of independently and identically distributed errors, homoskedasticity, cannot be rejected ( $\chi^2 = 0.8769$ for the Breusch–Pagan test). Hence, following Baltagi (1986), the Random Effects model provides more efficient estimates.

Within the dynamic models, the null hypothesis in the Sargan test that the over–identifying restrictions are valid is not rejected in the Arellano–Bond model (Table 3). The model cannot reject the null hypothesis of no first order autocorrelation. In addition, the model also rejects the null hypothesis of second order autocorrelation. Hence, the estimates in the Arellano–Bond model are consistent.

Estimates were also obtained using the Blundell–Bond model for the dynamic specification. However, the Sargan test for over–identifying restrictions is not satisfied even if only the last lag of only one variable is used as an instrument. This problem probably arises from the small dataset that is available (Bun et al. (2007)).

The Corrected LSDV has been estimated with coefficients from the Arellano–Bond estimation as the starting values since these were the only consistent and statistically significant dynamic estimates available. The estimates are not very sensitive to the initial values assumed. Initial values from the Blundell–Bond estimates result in coefficient values comparable to the Arellano–Bond initial values. The bootstrapped errors have been estimated based on 300 replications. In this case, the estimates are robust to the number of replications. Since this model cannot be directly compared with any of the other estimations, the results are reported only for interest.

In comparing the static and the dynamic specifications, the parameter of interest is the coefficient on the persistence variable, $1-\delta$ , since this denotes the importance of the dynamic component in the model. Observing the estimated value in Table 3, the coefficient of adjustment is significant in the Blundell–Bond and Corrected LSDV models, though it is not significant in the Arellano–Bond model. Hence, the benefits from using a dynamic specification are not evident.

### 5.2 Regression results

The regression results from all the models are presented in Table 3. Transit price has the correct sign and is significant in all the models. The confidence interval is smaller in the dynamic models indicating a change in transit price is mostly reflected in travel demand immediately and only to a much smaller degree over time through lagged values of transit demand.

Income is negative but not significant in any of the models. As reported in some of the literature, the negative sign indicates that income is an inferior good. Even with the distinction between the direct income effect on demand and the indirect effect through higher vehicle ownership, a negative income effect is obtained. However, since the coefficient is not significant in any of the models, a negative income effect is not definite.

Related to wealth, private vehicle ownership is negatively correlated with demand. The coefficient is significant in the models where the individual effects are random but is insignificant in the Fixed Effects and the Corrected LSDV models. This is probably due to the low within variation observed for this variable (Table 2).

Service quality has the highest elasticity values. Clearly, this is the most significant policy variable and has the largest impact on travel demand as expected from the literature (Cervero (1990)). Following Lago et al. (1981), this likely reflects the low coverage of public transit services in India.

As expected, population has a positive and significant impact on demand in all the static models and the Blundell–Bond model. Surprisingly, there is a negative correlation between population and passenger transit demand in the Arellano–Bond and the Corrected LSDV models. This could probably be due to the coefficient of adjustment in the models already capturing some population increase effects.

Literacy rate is negatively correlated with demand. The negative correlation with literacy rate indicates the low social acceptance of public transit. The impact of a large working population is positive and significant. Thus, with a larger proportion of population in the workforce, travel demand is higher and resulting in a larger demand for public transit. In general, the significance of social variables such as the proportion of working population and literacy rates indicated the importance of non–monetary factors in determining travel demand.

### *5.3 Price and Income Elasticities*

Given the model specification as log linear in transit price, income, and service quality, the coefficients on these variables can be interpreted as elasticities. However, arising from the log linear specification, elasticity values do not vary with the level of demand. The long run elasticities have been approximated around their mean values using the Delta method (Oehlert (1992)) to obtain significance levels as well. Since the Blundell–Bond does not satisfy the Sargan test, elasticities are not estimated for this specification. In addition, since the dynamic component in the Arellano–Bond model is not significant, elasticity estimates are not presented for this model as well. The estimated price and income elasticities are reported in Table 4.

The reported price elasticity is significant in all models and less than unity. The estimates lie between –0.354 and –0.523 in equilibrium or the long run. In all cases, transit demand is inelastic to fare changes. Also, as predicted by Doel van den et al. (1995), the static panel models report lower price elasticity values than the long run estimates using dynamic models, though the difference is not large. The price elasticity values are very much in consonance with the literature reported in section 2.0. The lower long run values compared to the literature could be perhaps explained by the fact that, most demand elasticity estimates in the literature have been obtained using datasets from developed countries, while this study is based in India. The low elasticity values, therefore, may represent the state of economic development in India vis–à–vis estimates in other studies. The inelastic demand may also arise from the fact that only public transit fares are included in this analysis since estimates for user costs and external costs are not available for this study. As a result, these estimates do not reflect the elasticity of demand with respect to the generalized transportation costs for the public bus users.

The literature reports negative income elasticities and characterizes public transit as an inferior good. Even though the estimates presented about report a negative income elasticity, since the coefficients are not significant, public transit cannot be characterized as an inferior good in India. These results are similar to Maunder (1984) where again income effects are insignificant above a minimum threshold of income. Dargay et al. (1999) report that the negative income elasticity during the period of analysis in their study of the United Kingdom between 1970 and 1998 coincided with a rapid increase in personal vehicle ownership. This may be the case in this study as well, given the rapid increase in personal vehicle population in India during the period under consideration and the significant negative coefficient obtained for personal vehicle ownership in most models.

Service quality remains the most significant policy variable for influencing transit demand. Again, this is as expected since the constraining factor for most infrastructure services in India, including public bus transit, is availability (Lago et al. (1981)). Fouracre et al. (1987) also report in their limited analysis of three Indian cities that a higher level of service results in a higher demand for public transit. They also report this to be a more significant policy variable for influencing demand. As a result, transit demand can be increased by making more services available.

# 6.0 Conclusions

This paper estimates transit price, income, and service quality elasticities for a direct aggregate demand function for public transport in India using an unbalanced panel between

1990/91 and 2000/01 for 22 states to assess the price, income, and service quality effects on bus transport demand using both static and dynamic specifications of a log linear model. Demand has been defined as total passenger kilometers to capture actual market transactions, while the regressors include public transit fare, per capita income, service quality, and other demographic and social variables. The measure of service quality used in the study is density of coverage, hence ensuring independence from demand and output measures.

The estimated price elasticity is significant in all models. In all cases, transit demand is inelastic to the fare level and comparable to those reported in the literature. The long run estimates, however, are lower compared to other studies. This is ascribed to the state of economic development in India since most studies emanate from developed countries. In particular, the low elasticity values indicate that public transit remains a necessity in India. In addition, all models report negative but insignificant income elasticity. This can be attributed to the transition in the Indian transport industry, with a rapid increase in the number of personal vehicles, which masks some of the direct wealth effects. Service quality is the most significant policy variable for influencing transit demand given the low availability of transit services. Finally, social and demographic variables highlight the complex nature of public bus transit demand in India.

There are two key policy implications that arise from the above analysis. First, the role of pricing is limited with public bus transit demand being price inelastic. Second, factors such as demographic changes and social variables have a larger influence on demand. Access to a public bus transport network has a much larger impact on aggregate demand and hence is possibly a more effective policy variable. If the policy objective is to raise public transit ridership to meet environmental or energy goals, service quality is clearly a much more important policy tool compared to transit prices. However, it is noted that service quality also depends on revenues to finance quality improvements, which in turn would lead to higher costs and hence fares (Cervero (1990)).

# 7.0 References


Abrate, G., M. Piacenza, et al. (2007). The Impact of Integrated Tariff Systems on Public Transport Demand: Evidence from Italy. XIX Riunione scientifica SIEP, Pavia.

Appelbaum, E. and J. Berechman (1991). "Demand conditions, regulation, and the measurement of productivity." Journal of Econometrics **47**(2–3): 379–400.

Arellano, M. and S. Bond (1991). "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." The Review of Economic Studies **58**(2): 277-297.

Arellano, M. and O. Bover (1995). "Another look at the instrumental variable estimation of error-components models." Journal of Econometrics **68**(1): 29-51.

Balcombe, R., R. Mackett, et al. (2004). The demand for public transport: a practical guide. Transportation Research Laboratory Report. London, Transportation Research Laboratory**:** 246.

Baltagi, B. H. (1986). "Pooling under Misspecification: Some Monte Carlo Evidence on the Kmenta and the Error Components Techniques." Econometric Theory **2**(3): 429-440.

Baltagi, B. H. (2002). Econometric Analysis of Panel Data. New York, John Wiley.

Beck, N. and J. N. Katz (1995). "What to do (and not to do) with Time-Series Cross-Section Data." The American Political Science Review **89**(3): 634-647.

Berechman, J. (1993). Public Transit Economics and Deregulation Policy. Amsterdam, North–Holland.

Blundell, R. and S. Bond (1998). "Initial conditions and moment restrictions in dynamic panel data models." Journal of Econometrics **87**(1): 115-143.

Bresson, G., J. Dargay, et al. (2003). "The main determinants of the demand for public transport: a comparative analysis of England and France using shrinkage estimators." Transportation Research Part A: Policy and Practice **37**(7): 605–627.

Bun, M. J. G. and J. F. Kiviet (2001). The Accuracy of Inference in Small Samples of Dynamic Panel Data Models. Tinbergen Institute Discussion Papers. Amsterdam, Tinbergen Institute.

Bun, M. J. G. and J. F. Kiviet (2003). "On the diminishing returns of higher-order terms in asymptotic expansions of bias." Economics Letters **79**(2): 145-152.

Bun, M. J. G. and F. Windmeijer (2007). The Weak Instrument Problem of the System GMM Estimator in Dynamic Panel Data Models. London, The Institute for Fiscal Studies, Department Of Economics, University College London**:** 32.

Cameron, A. C. and P. K. Trivedi (2005). Microeconometrics: Methods and Applications. New York, Cambridge University Press.

Census of India 2001 (2001). Census of India 2001, Office of the Registrar General India, Government of India.

Cervero, R. (1990). "Transit pricing research: A review and synthesis." Transportation **17**(2): 117-139.

CIRT (Various years). State Transport Undertakings: Profile and Performance. Pune, Central Institute of Road Transport.

Clements, K. W. and S. Selvanathan (1994). "Understanding consumption patterns." Empirical Economics **19**(1): 69-110.

Dargay, J. and M. Hanly (1999). Bus Fare Elasticities. Report to the UK Department of the Environment, Transport and the Regions. London, ESRC Transport Studies Unit, University College London**:** 132.

Dargay, J. M. and M. Hanly (2002). "The Demand for Local Bus Services in England." Journal of Transport Economics and Policy **36**(1): 73-91.

de Rus, G. (1990). "Public transport demand in Spain." Journal of Transport Economics and Policy **24**(2): 189–201.

Doel van den, I. T. and J. F. Kiviet (1995). "Neglected Dynamics in Panel Data Models: Consequences and Detection in Finite Samples." Statistica Neerlandica **49**: 343-361.

EPWRF (2003). Domestic Product of States of India: 1960-61 to 2000-01. Mumbai, Economic and Political Weekly Research Foundation.

Fitzroy, F. R. and I. Smith (1993). The Demand for Public Transport: Some Estimates from Zurich. Fife, Centre for Research into Industry, Enterprise, Finance and the Firm.

FitzRoy, F. R. and I. Smith (1997). Passenger Rail Transport: Demand and Quality in a European Panel. Fife, Centre for Research into Industry, Enterprise, Finance and the Firm.

Fouracre, P. R. and D. A. C. Maunder (1987). Travel demand characteristics in three medium-sized Indian cities. Transport and Road Research Laboratory Research Report 121. Crowthorne, Transport and Road Research Laboratory**:** 22.

Goodwin, P. B. and H. C. W. L. Williams (1985). "Public transport demand models and elasticity measures: An overview of recent British experience." Transportation Research Part B: Methodological **19**(3): 253-259.

Government of India (2005). Economic Survey 2004/2005. New Delhi, Ministry of Finance, Government of India.

Hanly, M. and J. Dargay. (1999). "Bus Fare Elasticities: A Literature Review. Report to the UK Department of the Environment, Transport and the Regions." Retrieved 17 July 2006, from www.cts.ucl.ac.uk/tsu/papers/BusElasticitiesLiteratureReview.pdf.

Hsiao, C. (2003). Analysis of Panel Data. Cambridge, Cambridge University Press.

Jovicic, G. and C. O. Hansen (2003). "A passenger travel demand model for Copenhagen." Transportation Research Part A: Policy and Practice **37**(4): 333–349.

Kemp, M. A. (1973). "Some evidence of transit demand elasticities." Transportation **2**(1): 25-52.

Kennedy, P. (2003). A Guide to Econometrics. Cambridge, MIT Press.

Kiviet, J. F. (1995). "On bias, inconsistency, and efficiency of various estimators in dynamic panel data models." Journal of Econometrics **68**(1): 53-78.

Kmenta, J. (1978). Some Problems of Inference from Economic Survey Data. Survey Sampling and Measurement. N. K. Namboodiri. New York, Academic Press**:** Chapter 8: 107-120.

Lago, A. and P. Mayworm (1981). "Transit service elasticities." Journal of Transport Economics and Policy **15**(2): 99-119.

Matas, A. (2004). "Demand and Revenue Implications of an Integrated Public Transport Policy: The Case of Madrid." Transport Reviews **24**(2): 195-217.

Maunder, D. A. C. (1984). Trip rates and travel patterns in Delhi, India. Transport and Road Research Laboratory Research Report 1. Crowthorne, Transport and Road Research Laboratory**:** 24.

McCarthy, P. S. (2001). Transportation Economics - Theory and Practice: A Case Study Approach. Malden, Blackwell.

Mohring, H. (1970). "The Peak Load Problem with Increasing Returns and Pricing Constraints." The American Economic Review **60**(4): 693–705.

MTS. (Various Issues). "Motor Transport Statistics." Retrieved 24 August 2005, 2005, from http://morth.nic.in/mts.htm.

Nickell, S. (1981). "Biases in Dynamic Models with Fixed Effects." Econometrica **49**(6): 1417-1426.

Oehlert, G. W. (1992). "A Note on the Delta Method." The American Statistician **46**(1): 27-29.

Oum, T. H. (1989). "Alternative Demand Models and their Elasticity Estimates." Journal of Transport Economics and Policy **23**(2): 163-187.

Oum, T. H., W. G. Waters, Jr. , et al. (1992). "Concepts of Price Elasticities of Transport Demand and Recent Empirical Estimates." Journal of Transport Economics and Policy **26**(2): 139–154.

Romilly, P. (2001). "Subsidy and Local Bus Service Deregulation in Britain: A Re-evaluation." Journal of Transport Economics and Policy **35**(2): 161-193.

Sevestre, P. and A. Trognon (1985). "A note on autoregressive error components models." Journal of Econometrics **28**(2): 231-245.

STATA (2005). Longitudinal/Panel Data Reference Manual. College Station, Texas, Stata Press.

Thomas, R. L. (1987). Applied Demand Analysis. London and New York, Longman.

Wabe, J. S. (1969). "Labour Force Participation Rates in the London Metropolitan Region." Journal of the Royal Statistical Society. Series A (General) **132**(2): 245-264.

Winston, C. (1983). "The demand for freight transportation: models and applications." Transportation Research Part A: General **17**(6): 419-427.

## Table 1. Recent studies estimating demand functions

| Paper | Variables | Functional Form & Estimation Method | Data | Price elasticity | | Income elasticity | | Service elasticity | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Short run | Long run | Short run | Long run | Short run | Long run |
| Dargay et al. (1999)5 | Two models: Bus passenger kilometers per capita and bus trips per capita bus fares, disposable income, car ownership and motoring costs used only in the structural models. | Two types of models: Error Correction Models and Structural Models. | Time series of annual observations between 1970 and 1996 for the United Kingdom. | From –0.33 to –0.40 for trips. –0.18 to –0.19 for passenger kilometers | From –0.62 to –0.95 for trips. –0.43 to –0.92 for passenger kilometers | From 0.18 to 0.41 for trips. 0.05 to 0.16 for passenger kilometers | From –0.45 to –0.80 for trips. –0.15 to –0.63 for passenger kilometers | | |
| Romilly (2001) | Bus journeys per capita. Personal disposable income, index of bus fares, index of motoring cost, service frequency measured by vehicle kilometers per person. | Log linear model, estimated as a single equation Auto Regressive Distributed Lag model after corrections for cointigrating relationships. | Time series of annual observations between 1953 and 1997 for United Kingdom excluding London. | –0.38 | –1.03 | 0.23 | 0.61 | 0.11 | 0.30 |
| Paper | Variables | Functional Form & Estimation Method | Data | Price elasticity | | Income elasticity | | Service elasticity | |
| Dargay et al. (2002) | Bus journeys per capita fare, service level, per capital disposable income, pensioners in population, motoring costs. | Partial adjustment Fixed Effects, Random Effects, and Random Coefficient models. Two specifications: log linear, semi log where only fares are in levels and not logs. | Panel data between 1987 and 1996 for 46 counties in United Kingdom. | From –0.33 to –0.44 | From –0.68 to –0.75 | From –0.39 to –0.60 | From –0.81 to –1.02 | 0.42 to 0.49 | 0.79 to 1.03 |
| Bresson et al. (2003) 6 | Journeys per capita. Mean fare defined as | Two specifications: Semi log with only | Panel data of 46 county | –0.53 for England | –0.73 for England | –0.48 for England | –0.66 for England | –0.71 for England. | –0.97 for England. |

[5] Only national level results reported.

| Paper | Variables | Functional Form & Estimation Method | Data | Price elasticity | | Income elasticity | | Service elasticity | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Short run | Long run | Short run | Long run | Short run | Long run |
| | revenue per trip, service measured by vehicle kilometers per capita, disposable income per capita | fares being in levels, and log linear. Estimated as Arellano and Bond fixed coefficients and random coefficient models | annual observations in United Kingdom and 62 urban areas in France during 1987 and 1996. | and –0.40 for France | and –0.70 for France | and –0.01 for France | and –0.02 for France | –0.19 for France | –0.33 for France |

[6] Only fixed coefficients' results reported.

## Table 2. Descriptive statistics of the variables included in the analysis

| Variable | | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| Passenger kilometers ($10^5$ km) | Overall | 230 194.300 | 324 248.600 | 112.570 | 2 236 124.000 |
| | Between | | 292 371.900 | | |
| | Within | | 127 504.700 | | |
| Public transit fare (Rupees[#] per passenger kilometer) | Overall | 0.089 | 0.046 | 0.031 | 0.384 |
| | Between | | 0.038 | | |
| | Within | | 0.033 | | |
| Per capita income (Rupees per person) | Overall | 6 073.593 | 3 430.798 | 164.383 | 19 191.890 |
| | Between | | 3 398.596 | | |
| | Within | | 1 470.881 | | |
| Density of coverage ($10^5$ vehicle km per $km^2$) | Overall | 0.257 | 0.778 | 0.0001 | 4.089 |
| | Between | | 0.868 | | |
| | Within | | 0.123 | | |
| Per capita private vehicle ownership (Vehicles per person) | Overall | 0.047 | 0.079 | 0.005 | 0.493 |
| | Between | | 0.089 | | |
| | Within | | 0.017 | | |
| Population (number) | Overall | 43 700 000.000 | 38 300 000.000 | 719 601.000 | 166 000 000.000 |
| | Between | | 38 300 000.000 | | |
| | Within | | 3 029 920.000 | | |
| Population the labour force (%) | Overall | 38.87% | 0.048 | 30.87% | 49.24% |
| | Between | | 0.047 | | |
| | Within | | 0.009 | | |
| Literacy rate (%) | Overall | 53.55% | 0.109 | 30.57% | 80.04% |
| | Between | | 0.106 | | |
| | Within | | 0.035 | | |

[#]1 Swiss Franc equaled approximately 36 Indian Rupees in February 2008.

## Table 3. Regression results

| | Fixed Effects | | Random Effects | | PCSE | | Arellano–Bond | | Blundell–Bond | | Corrected–LSDV | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Coefficient | Standard Error | Coefficient | Standard Error | Coefficient | Standard Error | Coefficient | Standard Error | Coefficient | Standard Error | Coefficient | Standard Error |
| $\alpha'_p$ | –0.460*** | 0.041 | –0.354*** | 0.050 | –0.359*** | 0.076 | –0.420*** | 0.035 | –0.262*** | 0.051 | –0.374*** | 0.043 |
| $\alpha'_w$ | –0.020 | 0.034 | –0.065 | 0.038 | –0.061 | 0.040 | –0.016 | 0.026 | –0.058 | 0.038 | –0.027 | 0.025 |
| $\alpha'_q$ | 0.834*** | 0.031 | 0.818*** | 0.027 | 0.754*** | 0.029 | 0.691*** | 0.048 | 0.160*** | 0.043 | 0.676*** | 0.052 |
| $\alpha'_s$ | –0.028 | 0.048 | –0.106*** | 0.052 | –0.212*** | 0.064 | –0.003*** | 0.052 | –0.118* | 0.047 | 0.037 | 0.054 |
| $\alpha'_{pop}$ | 0.662*** | 0.038 | 0.938*** | 0.043 | 1.026*** | 0.032 | –1.662*** | 0.291 | 0.099* | 0.049 | –0.500* | 0.248 |
| $\alpha'_{work}$ | 6.770*** | 0.618 | 6.798*** | 0.481 | 11.797*** | 0.564 | 3.711*** | 0.788 | 2.674** | 0.370 | 2.450* | 1.085 |
| $\alpha'_{lit}$ | –4.089*** | 0.138 | –3.665*** | 0.887 | –1.099*** | 0.862 | –4.140*** | 0.035 | –0.787* | 0.051 | –1.974*** | 0.506 |
| $(1-\delta)$ | | | | | | | 0.119 | 0.070 | 0.886 | 0.049 | 0.294*** | 0.056 |
| $\alpha'_o$ | 1.273*** | 0.041 | –3.350*** | 0.050 | –8.829*** | 0.076 | 0.042*** | 0.012 | –1.060 | 0.579 | | |
| F statistic | 366.57*** | | | | | | | | | | | |
| $R^2$ | 0.8939 | | | | 0.9874 | | | | | | | |
| Wald $\chi^2$ | | | 1635.03*** | | 3784.45*** | | 662.94*** | | 69,650.56*** | | | |
| Sargan $\chi^2$ | | | | | | | 47.40 | | 77.01*** | | | |
| AR (1) | | | | | | | –1.72 | | –5.73*** | | | |
| AR (2) | | | | | | | 0.16 | | 0.88 | | | |

*Variables significant at 95% confidence level, **Variables significant at 99% confidence level, ***Variables significant at 99.9% confidence level

## Table 4. Price and Income Elasticity estimates

| | Fixed Effects | Random Effects | PCSE | Corrected LSDV | |
|---|---|---|---|---|---|
| | | | | Short run | Long run |
| Price | –0.460*** | –0.354*** | –0.359*** | –0.374*** | –0.523*** |
| Income | –0.020 | –0.065 | 0.061 | –0.027 | –0.038 |
| Service quality | 0.834*** | 0.818*** | 0.754*** | 0.676*** | 0.957*** |

*** Significant at 99.9% confidence level