Econometric models for the forecast of passenger demand in Greece

Vassilios A. Profillidis*

Section of Transportation
Democritus Thrace University
671 00 Xanthi
Greece

George N. Botzoris
Transportation Engineer
Tantalou 30
546 29 Thessaloniki
Greece

Abstract
The structure of models of forecast of passenger demand is first discussed. Parameters affecting modal split and mobility are analyzed. After many trial and errors procedures, three econometric models have been developed for the forecast of passenger demand in Greece: one for total demand, one for rail demand and one for private car demand. The validity of each model is tested by means of statistical and diagnostic tests, which are: Estimation of Coefficient of Determination ($R^2$), Collinearity test of independent variables, Statistical test of the $F$-statistics, Statistical test of the standard error, Model function form test, First degree self-correlation test to residuals, Residuals correlation, Heteroscedasticity and normality test, Model stability test and Forecasting ability of the models through U-Theil Statistics. The validity of the models has been also corroborated through the study of elasticities of independent variables. Once checked the forecasting ability, the models can be used for the forecast of future demand and modal split for passenger demand in Greece.

Keywords: Demand, forecast, mobility, econometric, rail, private car.

*E-mail: info@profillidis.gr

Journal of Statistics & Management Systems
Vol. 9 (2006), No. 1, pp. 37–54
© Taru Publications
1. An overview of econometric models for the forecast of passenger demand

Forecast of passenger demand has been the subject of a number of analyses and papers. Jones et al. (1983) adapted a time-series model for traffic between 17 English cities and London, while Fowkes et al. (1985) adapted a similar model for traffic between English cities excluding London. The independent variables of the latter model were: traveled distance, employment ratio, revenue, fare and car ownership index. Wilson et al. (1990) described the adaptation of an interurban traffic model in Canada. In their analysis, they concluded that a trip utility relationship was dependent on travel time, trip distance, service frequency and household income. A multinomial logarithmic model was developed by Regianni and Stefani (1989) in which trips on the Bergamo-Milan axis by car, bus and rail were studied. A mode selection model for cars and rail was adapted on the Lyon-St. Etienne axis, (Plassard 1996). The variables that were used were income per capita, trip cost, travel times and service frequency. The French Railways quantified passenger mobility, adapting econometric models, with independent variables the trip cost with various modes, the household income and the fuel cost, (Transport 1992). The French Ministry of Transport quantified passenger demand by adapting an econometric model with principal independent variables the car ownership index and the average price of fuel, (ECMT 2001). A multinomial model was adapted for the forecast of private car and rail passenger demand for 22 city pairs in USA, (Stopher and Lee-Gosselin 1996). The independent variables were travel time, trip cost and service frequency. A recent application of a model was developed in Italy within the Expedite Research Program of the E.U., (Coppola and Carteni 2001). The model included all transport modes: car, bus, train and airplane.

2. Parameters affecting passenger demand and modal split

It is established that transport is closely related to the economic activity, (ECMT, 2001), (Profillidis, 2004). Both passenger and freight transport follow generally the rate of economic development. From a literature survey, we conclude that the principal parameters are: Gross Domestic Product (GDP), the car ownership index, as well as the car’s use cost, (Bhat and Pulugurta 1998), (Noland et al. 2003).
The market share of each transport mode in passenger sector has changed during the last decades, (Figure 1). However, crucial parameters affecting the market share of each transport mode vary with the mode:

- The market share of private car depends on car ownership index and cost of fuel.
- The market share of railways and buses depends on car ownership index, G.D.P., rail fares, fares of competitive modes, (e.g., bus, railways and airplane), and travel time.

![Market share in passenger transport of various transport modes in European Union countries during last years, (European Commission, 2004)](image)

**Figure 1**

**3. The proposed econometric models**

**3.1 Determination of independent variables – method of adjustment**

Once considering the implementation of an econometric model, the first step would be to determine the dependent variable, which will be in our analysis the number of passenger-kilometers. Explanatory (independent) variables quantifying the phenomenon under study, which is modal split and evolution of traffic of each transport mode in Greece, (Figures 2, 3) and its relationship with dependent variable, have been surveyed and are described analytically for each case below.
In order to adjust the appropriate demand analysis and forecasting models, the General to Specific Approach method proposed by Hendry was selected. According to this method, the initial model includes all explaining (independent) variables. The addition or rejection of explaining variables to or from the model is based on the statistical check of the $F$ and the $t$ criterion, (Maddala, 1992), (Ortuzar and Willumsen 1994).
All variables are expressed in a logarithmic form, a fact that facilitates an immediate determination of elasticities constants. Variables expressed in monetary units have been deflated according to the annual consumer price index. All variables are incorporated into the model as indexes that have the value 100 for the median year of the analysis period.

After selection of the appropriate explanatory variables in each case, we first calculated the coefficients for each econometric model through regression analysis and then we conducted the appropriate statistical tests.

3.2 Econometric model for total passenger demand

The analysis period spans over the years 1980-2000. The proposed model $M_1$ for the analysis and forecasting of total passenger demand for Greece is (with $R^2 = 0.95$):

\[
\ln D_{tot} = 1,667 \cdot \ln GDP - 0,887 \cdot \ln C_{fuel} + 0,991 \\
\text{t-student} (4,88) (-15,65) (0,60)
\]

where:

- $D_{tot}$: total passenger demand/population,
- GDP: Gross Domestic Product (in constant prices of the year 2000).
  This was preferred over Gross Available Income, due to a better t-student and lower standard error,
- $C_{fuel}$: Cost of fuel (in constant prices of the year 2000).

Figure 4 illustrates the results of the econometric model compared to the real number of passengers. A satisfactory model adjustment to real data can be observed.
3.3 **Econometric model for passenger demand with private cars**

The analysis period spans over the years 1980 ÷ 2000. The proposed model $M_2$ for the forecast of private car passenger demand for Greece is (with $R^2 = 0.99$):

$$\ln D_{\text{car}} = 0.691 \cdot \ln I_{\text{co}} - 0.066 \cdot \ln C_{\text{fuel}} + 1.723$$  \hspace{1cm} (2)

\[(t\text{-student}) \quad (34.61) \quad (-2.40) \quad (8.03)\]

where:

$D_{\text{car}}$ : Private car passenger demand/population,

$I_{\text{co}}$  : Private car ownership index.

Figure 5 illustrates the results of the econometric model compared to the real number of passenger demand with private cars.

3.4 **Econometric model for rail passenger demand**

The analysis period spans over the years 1960-2000. The proposed model $M_3$ for the forecast of rail passenger demand is (with $R^2 = 0.89$):

$$\ln D_{\text{rail}} = -0.192 \cdot \ln C_r - 0.078 \cdot \ln I_{\text{co}} + 0.111 \cdot \ln C_{b,r}$$

\[(t\text{-student}) \quad (-2.01) \quad (-1.56) \quad (1.56)\]

$$+ 0.109 \cdot \ln GDP + 0.776 \cdot \ln D_{\text{rail}(-1)} + 1.273$$  \hspace{1cm} (3)

\[(1.09) \quad (7.54) \quad (2.11)\]
where:
\[ D_{\text{rail}} \]: rail passenger demand/population,
\[ C_r \]: rail use cost per passenger-kilometer,
\[ C_{b,r} \]: competition variable, expressed as the cost (for the passenger) to use the bus instead of the railway, per passenger-kilometer,
\[ D_{\text{rail}}(-1) \]: a time lag dependent variable, which is common in cases of aggregate models of transport demand analysis, (Maddala, 1992), and represents habitual inertia and constrains on supply (service frequency, rail capacity, services in stations and on trains, etc).

The results of the econometric model compared to the real number of rail passengers are given in Figure 6. We can observe the satisfactory model adjustment to real data.

![Figure 6](image)

**Results of the econometric model for rail passenger demand for Greece and comparison with real data (Model \( M_3 \))**

4. **Statistical tests of the validity of the proposed models**

4.1 *Necessary statistical and diagnostic tests*

The model’s adjustment to actual data is satisfactory with a coefficient of determination \( (R^2) \) equal to 95% for the model \( M_1 \), 99% for the model \( M_2 \) and 89% for the model \( M_3 \). The model’s validity is tested by means of statistical and diagnostic tests, which are:

- Collinearity test of independent variables
- Statistical test of the $F$-statistics
- Statistical test of the standard error
- Model function form test
- First degree self-correlation test to residuals through Durbin’s-$h$ statistics
- Residuals correlation test
- Residuals normality
- Residuals heteroscedasticity test
- Model stability test
- Model stability test through Chow’s statistical test (predictive Failure Test)

4.2 Collinearity test of independent variables

A prerequisite for the use of the independent variables in a model is the absence of high correlation between the independent variables. Tables 1, 2 and 3 present the correlation matrix of the variables of the Models $M_1$, $M_2$ and $M_3$. We can note the low correlation between the independent variables and the high correlation of independent variables with the dependent variable for each model.

### Table 1
Correlation matrix of variables of the Model $M_1$

<table>
<thead>
<tr>
<th>Variables</th>
<th>$D_{tot}$</th>
<th>GDP</th>
<th>$C_{fuel}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_{tot}$</td>
<td>1,00</td>
<td>0,68</td>
<td>−0,93</td>
</tr>
<tr>
<td>GDP</td>
<td>0,68</td>
<td>1,00</td>
<td>−0,32</td>
</tr>
<tr>
<td>$C_{fuel}$</td>
<td>−0,93</td>
<td>−0,32</td>
<td>1,00</td>
</tr>
</tbody>
</table>

### Table 2
Correlation matrix of variables of the Model $M_2$

<table>
<thead>
<tr>
<th>Variables</th>
<th>$D_{car}$</th>
<th>$I_{co}$</th>
<th>$C_{fuel}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_{car}$</td>
<td>1,00</td>
<td>0,99</td>
<td>−0,95</td>
</tr>
<tr>
<td>$I_{co}$</td>
<td>0,99</td>
<td>1,00</td>
<td>−0,52</td>
</tr>
<tr>
<td>$C_{fuel}$</td>
<td>−0,95</td>
<td>−0,52</td>
<td>1,00</td>
</tr>
</tbody>
</table>
### Table 3
Correlation matrix of variables of the Model $M_3$

<table>
<thead>
<tr>
<th>Variables</th>
<th>$D_{rail}$</th>
<th>$C_r$</th>
<th>$I_{co}$</th>
<th>$C_{b,r}$</th>
<th>GDP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_{rail}$</td>
<td>1.00</td>
<td>0.79</td>
<td>-0.74</td>
<td>0.84</td>
<td>-0.72</td>
</tr>
<tr>
<td>$C_r$</td>
<td>0.79</td>
<td>1.00</td>
<td>-0.48</td>
<td>0.71</td>
<td>-0.49</td>
</tr>
<tr>
<td>$I_{co}$</td>
<td>-0.74</td>
<td>-0.48</td>
<td>1.00</td>
<td>-0.44</td>
<td>0.80</td>
</tr>
<tr>
<td>$C_{b,r}$</td>
<td>0.84</td>
<td>0.71</td>
<td>-0.44</td>
<td>1.00</td>
<td>-0.62</td>
</tr>
<tr>
<td>GDP</td>
<td>-0.72</td>
<td>-0.49</td>
<td>0.80</td>
<td>-0.62</td>
<td>1.00</td>
</tr>
</tbody>
</table>

#### 4.3 Statistical test of the $F$-statistics

For the model $M_1$, $F$-statistics (2,18) has been found to have the value 176.93, (Table 4), which is far greater than the critical value ($3.55$ for $n = 18$). In the case of the model $M_2$, $F$-statistics (2,18) has been found to have the value 5.399,30 which is far greater than the critical value ($3.55$ for $n = 18$) and for the model $M_3$, $F$-statistics (6,33) has been found to have the value 43.61 which is also greater than the critical value ($2.42$ for $n = 30$ and $2.34$ for $n = 40$). In this way the null hypothesis (for $\alpha = 0.05$) that all the explaining variables do not contribute to the explanation of the dependent variable of the three models is rejected.

### Table 4
Statistical test of the $F$-statistics through $F$ distribution

<table>
<thead>
<tr>
<th>Model</th>
<th>$M_1$</th>
<th>$M_2$</th>
<th>$M_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freedom degree: • numerator</td>
<td>2</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>• denominator</td>
<td>18</td>
<td>18</td>
<td>33</td>
</tr>
<tr>
<td>$F$-statistics calculated from models</td>
<td>176.9</td>
<td>5399.3</td>
<td>43.6</td>
</tr>
<tr>
<td>Critical values of $F$-statistics</td>
<td>3.55</td>
<td>3.55</td>
<td>~2.40</td>
</tr>
</tbody>
</table>

#### 4.4 Statistical test of the standard error

Another positive point regarding the model’s adjustment is its low Standard Error (Regression Standard Error for the Model $M_1$ is equal to 0.062, for the Model $M_2$ is equal to 0.011 and for the Model $M_3$ is equal to 0.062), in relation to the average of the dependent variable (4.572 for the Model $M_1$, 4.573 for the Model $M_2$ and 4.773 for the Model $M_3$).
4.5 Model function form test

The diagnostic test, which examines the model’s function form, does not reject the null hypothesis that residuals are in accordance with the normal distribution. The $X^2$ distribution is equal to 0.083 for the Model $M_1$, 0.690 for the Model $M_2$ and 0.360 for the Model $M_3$, whereas the critical value for one degree of freedom and a significance level of $\alpha = 0.05$ is 3.841, (Table 5).

<table>
<thead>
<tr>
<th>Model → Test ↓</th>
<th>Degree of freedom</th>
<th>$X^2$ Statistics</th>
<th>Critical values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M_1$</td>
<td>$M_2$</td>
<td>$M_3$</td>
</tr>
<tr>
<td>A: Serial correlation</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>B: Functional form</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>C: Normality</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>D: Heteroscedasticity</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

A: Lagrange multiplier test of residual serial correlation  
B: Ramsey’s-RESET test using the square of the fitted values  
C: Based on a test of skewness and kurtosis of residuals  
D: Based on the regression of squared residuals on squared fitted values

4.6 First degree self-correlation test to residuals

The absence of the three model residuals first degree self-correlation is also satisfactory. For the Model $M_1$, Durbin-Watson statistics was equal to 1.80, (which is very close to the optimum value of ±2.00) and for the Model $M_2$ was equal to 1.74. For the Model $M_3$ and due to the fact that the dependent variable $D_{rail}$ has been included as an explaining variable with a time lag ($D_{rail(-1)}$), Durbin Watson statistics is not employed for the self-correlation test (Guthberston et al. 1992). In such cases, the Durbin’s-$h$ statistics can be used, which has the value of −0.58, a result within the space $+1, 96 \div -1, 96$.

4.7 Residuals correlation test

The lack of residual correlation for the three Models is confirmed by the appropriate test, throughout the Lagrange multiplier test of residuals serial correlation, presented in Table 5. In the case of Model $M_3$, which
has a time lag dependent variable, the order of self-correlation check was extended beyond one (order of self correlation = 2), whereas the results do not reject the null hypothesis in all cases.

4.8 Residuals normality

The diagnostic test that examines the pattern of distribution and its parameters (residual normality) does not reject the null hypothesis that residuals are in accordance to the normal distribution. The $X^2$ distribution is equal to 0.820 for the Model $M_1$, 1.878 for the Model $M_2$ and 1.262 for the Model $M_3$, whereas the critical value for two degrees of freedom and a significance level of $\alpha = 0.05$ is 5.991, (Table 5).

4.9 Residuals heteroscedasticity test

The residuals heteroscedasticity test showed that residuals have a standard fluctuation (homoscedastic error fluctuation) as the $X^2$ distribution’s value is 1.366 for the Model $M_1$, 2.271 for the Model $M_2$ and 0.301 for the Model $M_3$, whereas the critical value for one degree of freedom and a significance level of $\alpha = 0.05$ is 3.841, (Table 5).

4.10 Model stability test

In what regards the model’s stability and its robustness, the time period of the sample used for the model’s adjustment is checked regarding its importance on the model’s form, through recursive regressions. In order to check the degree to which the model has been appropriately specialized, the Cumulative Sum test (CUSUM), (Figures 7, 8 and 9) and the Cumulative Sum Squared test (CUSUMQ), (Figures 10, 11 and 12) were employed. Figures 7 to 12 present the two tests with the critical levels for a 5% significance level. If any of the two tests exceeds the critical values for a 5% significance level, then the null hypothesis that the model has been appropriately specialized is rejected, (Maddala, 1992).

4.11 Model stability test through Chow’s statistical test (Predictive Failure Test)

In the case of the model $M_3$, the presence of the dependent variable through a time lag, causes doubts about limits’ reliability, because a number of cases have been reported in which the ability of the two tests to detect the parameters’ stability was considered unsatisfactory (Guthberston et al., 1992). Therefore, Chow’s second test must also be employed, whereas the model’s robustness is tested through the Predictive
Cumulative Sum test of recursive residuals (with critical limits) for a 5% significance level of the econometric model $M_1$

Cumulative Sum test of recursive residuals (with critical limits) for a 5% significance level of the econometric model $M_2$

Cumulative Sum test of recursive residuals (with critical limits) for a 5% significance level of the econometric model $M_3$
Figure 10
Cumulative Sum Squared test of recursive residuals (with critical limits) for a 5% significance level of the econometric model $M_1$

Figure 11
Cumulative Sum Squared test of recursive residuals (with critical limits) for a 5% significance level of the econometric model $M_2$

Figure 12
Cumulative Sum Squared test of recursive residuals (with critical limits) for a 5% significance level of the econometric model $M_3$
Failure Test. The test was used for the period 1961 ÷ 1998 for which the $F(2, 32)$ statistics value was equal to 1,483. This value is less than the corresponding critical value (3,23 for 40 periods that is years from 1960 to 2000), which means that the null hypothesis for the model’s stability cannot be rejected.

5. Elasticities of the models

5.1 Total passenger demand elasticities

The validity of the model $M_1$ is corroborated, besides the diagnostic tests, with the signs of explaining variables and their elasticities, (Table 6):

- For the variable GDP (Gross Domestic Product of Greece) a positive sign was correctly calculated, since an income increase is expected to have a positive influence on mobility.
- For the variable $C_{\text{fuel}}$ (cost of fuel) a negative sign was correctly calculated, since an increase in the cost of fuel is expected to have a negative influence on mobility.

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>Elasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gross Domestic Product per capita (GDP)</td>
<td>1,667</td>
</tr>
<tr>
<td>Cost of fuel ($C_{\text{fuel}}$)</td>
<td>$-0.887$</td>
</tr>
</tbody>
</table>

5.2 Private car demand elasticities

The elasticities of the econometric model $M_2$ concerning the private car demand are given in Table 7. We can remark that:

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>Elasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car ownership index ($I_{\text{co}}$)</td>
<td>0.691</td>
</tr>
<tr>
<td>Cost of fuel ($C_{\text{fuel}}$)</td>
<td>$-0.066$</td>
</tr>
</tbody>
</table>

- For the variable $I_{\text{co}}$ (car ownership index) a positive sign was correctly calculated since an increase in the car ownership index has a positive influence on the use of car.
For the variable $C_{\text{fuel}}$ (cost of fuel) a negative sign was correctly calculated, since an increase in the cost of use of private car is expected to have a negative effect on its use.

The importance of the constant $c$ and its high value suggest that use of car is done regardless of other factors affecting private car demand.

5.3 Rail passenger demand elasticities

Elasticities of the econometric model $M_3$ for rail passenger demand are given in Table 8. We can remark that:

- For the variable $C_r$ (rail use cost per passenger-kilometer) a negative sign was correctly calculated, since an increase in a mode’s use cost has a negative effect on its transport demand. However, the appearance of the variable $C_r$ three times as an independent variable $(C_r, C_b, r, D_r)$ suggests a short-run price elasticity of $-0.303$ ($= 0.192 - 0.111$) and a long-run elasticity of $-1.353$ ($-0.303 / (1 - 0.776)$).

- For the variable $I_{co}$ (car ownership index) a negative sign was correctly calculated, since an increase in the car ownership index has an adverse effect on mass transport demand.

- For the variable $C_{b,r}$ which expresses competition, a positive sign was correctly calculated because an increase on the ratio (bus use cost)/(rail use cost) has a positive impact on rail demand.

- For the variable GDP per capita, a positive sign was correctly calculated, since an income increase is expected to have a positive effect on rail demand. The low value of the coefficient of the variable $(0.111)$ suggests that rail transport is considered as a usual and normal good (as opposed to luxury goods) and therefore its demand is not strongly affected by economic conditions.

- The variable which we should probably emphasize is $D_r(-1)$. Besides its high importance in explaining demand ($t$-student $= +7.54$), this variable has also the highest coefficient of all explaining variables $(0.776)$. If services of Greek Railways (transport quality, employee behavior, cleanliness, reliability) could be quantified, it could be said than a 1% increase in them would lead to 0.776% demand increase.
Table 8
Elasticities of the independent variables of the model $M_3$

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>Elasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rail use cost per passenger-kilometer ($C_r$)</td>
<td>$-0.192$</td>
</tr>
<tr>
<td>Car ownership index ($I_{co}$)</td>
<td>$-0.078$</td>
</tr>
<tr>
<td>Competition variable, expressed as the cost to use bus instead of railway per passenger-kilometer ($C_{b,r}$)</td>
<td>$0.111$</td>
</tr>
<tr>
<td>Gross Domestic Product per capita (GDP)</td>
<td>$0.109$</td>
</tr>
<tr>
<td>Time lag dependent variable $D_{rail}(-1)$</td>
<td>$0.776$</td>
</tr>
</tbody>
</table>

6. Forecasting ability of the proposed models

The forecasting ability of the proposed models is tested with the U-Theil Statistics method. This method allows the examination of residuals and appraisal of the forecasting ability by employing appropriate statistical tests. When U-Theil Statistics is calculated equal to zero for a model, then the model’s forecasting ability is perfect, whereas when U-Theil Statistics is calculated equal to one, the model lacks any forecasting ability (Jarret, 1987).

Through calculation of the U-Theil Statistics for the proposed econometric models, it is derived that the U-Theil Statistics has the value $0.125$ for the model $M_1$, the value $0.025$ for the Model $M_2$ and the value $0.244$ for the model $M_3$; these values are very close to the ideal value, which, as explained earlier, is zero.

7. Forecast of future modal split for Greece

Once the forecasting ability is checked, the models can be used for the forecast of future demand and modal split for passenger transport for Greece, (Figures 13, 14).

8. Concluding remarks

It is essential in transport planning to establish a causal relationship between the demand of each transport mode and the parameters affecting demand. Such a causal analysis has been presented in this paper. Three econometric models have been proposed for passenger demand in Greece,
relating total demand, rail demand and private car demand to factors affecting each mode. The necessary appropriate tests assure the validity of the proposed models, which can then be used for the forecast of future demand.

References


Received May, 2005