

Modelling Demand for Rail Transport with Dynamic Econometric Approaches

Na Chen*

Various dynamic models are described and applied in the rail transport literature because lags occur in the rail transport market. These dynamic models provide powerful tools for analyzing and forecasting rail transport demand, but they have some drawbacks that should be recognized. To clarify the problems arising from these drawbacks, this paper reviews and critically evaluates the various dynamic models. This paper has two novel aspects. The first novel aspect of the research is the division of the dynamic models into three categories. These three categories are models lagging the independent variables and the dependent variable, models lagging only the dependent variable and models lagging only the independent variables. The second novel aspect is the provision of fresh empirical evidence for estimating the elasticities by applying a static model based on annual data from the British Strategic Rail Authority.

Keywords: dynamic models, rail transport, lagged effects, transport demand.

Field of research: Economics

1. Introduction

A number of studies have been conducted during the past 20 years to estimate the impacts of economic factors on the demand for rail travel (e.g. Owen and Philips, 1987, NERA, 2003, and Wardman and Whelan, 2004) by the application of dynamic econometric models (e.g. a error correction model, and a partial adjustment model). However, the frequency of the data, lagged structures and the employment of various dynamic models can estimate different elasticities, the length of the long-run and even counterintuitive results. This increases the difficulty for researchers and decision makers to explain and forecast rail travel choices.

That dynamic models are important in modelling demand for rail transport is because they can portray the 'lag' in realizing the full effects of a change (e.g. fare and

*Na Chen, Marie Curie research fellow, Institute for Transport Studies, University of Leeds, Leeds, LS2 9JT. Email: tranc@leeds.ac.uk

generalized journey time) on demand for rail transport. Short-run elasticity, long-run elasticity, speed of adjustment and the length of the long-run can be estimated, or calculated by the application of some dynamic models. Oxera (2005) summarized five reasons for why the lag occurs in rail transport. Reasons include initial switching cost, lack of information, constraints on behaviour, initial teething problems, and changes in land use.

Although each dynamic econometric model is based on somewhat reasonable assumptions, problems and limitations of the models themselves may cause meaningless estimates (e.g. spurious regressionⁱ for the partial adjustment model may cause the estimate for speed of adjustment and therefore the length of the long-run to be totally meaningless). This may provide an explanation about several counterintuitive resultsⁱⁱ in demand for rail transport. However, to my knowledge, no research has addressed this issue so far in rail transport.

For the above reasons, two main objectives of the study are

- To review previous dynamic econometric models, including rationales, development history, shortcomings and limitations
- To provide fresh empirical evidence for estimating the elasticities by applying a static fixed effects model.

The structure of this paper is as follows. Section 2 critically evaluates previous studies. Section 3 presents data and research method. An innovative application of a static model based on annual data from the Strategic Rail Authority is provided in this section. The final section presents concluding remarks and recommendations for further analysis.

2. Literature review

2.1. The definition of dynamic models

Dynamic models are used to portray the dynamic characteristics of economic data. Changes or decision taken at time t , can affect the economy, choices, and behaviour at time $t+1, t+2, t+3$, etc. These effects are spread, over future time periods.

2.2. Why may lags in demand arise in practice?

The nature of lags in demand is that travelers cannot change their choices instantaneously. Therefore, changes in explanatory variables (for example fare and generalized journey time) at time $t-3$ may not only affect the demand at time $t-3$, but also affect the demand at time $t-2, t-1$, and t . Researchers (e.g. Goodwin, 1976 and Oxera, 2005,) have summarized many reasons for why lags in demand occur in the transport market.

Goodwin (1976) used the word “habit” to signify various sources of resistance to a change in the choice between transport modes. Besides the time travellers need to gather relevant information and notice the relative attractiveness of the modes, travellers are reluctant to “upset an ordered and understood routine”, which can contribute to peace of mind. He also argued that in the long run the habit will be lessened and the role of rational factors becomes dominant, and such effects are time-dependent. Although Goodwin did not discuss this time-dependence in detail, its importance was recognized.

This was a significant progress in understanding the lag effect in transport mode choice.

Oxera (2005) summarized five reasons for why the lag occurs in rail transport. Reasons include: initial switching cost, lack of information, constraints on behaviour, initial teething problems, and changes in land use.

2.3. The importance of modeling demand for rail transport with dynamic econometric approaches

That dynamic models are important in modelling demand for rail transport is because of the lag in realizing the full effects of a change (e.g. fare) for various reasons. The rationale of dynamic models (see section 2.1) justifies that its application in transport market should be appropriate. The application of dynamic models can help decision makers more precisely understand travellers’ preferences and behaviour because the assumption that travellers can not change their choices instantaneously is reasonable.

2.4. Dynamic models

Dynamic models can be divided into three categories: a. models lagging both independent variables and dependent variables (e.g. error correction model), b. models lagging only dependent variable (e.g. partial adjustment model), and c. models lagging only independent variable (e.g. Geometric distributed lag model, Arithmetic distributed lag model, and Almon distributed lag model).

2.4.1 Error correction model

Error correction model has a long tradition in time series econometrics dating back to Sargan(1964), although it has been more popularized after Engle and Granger(1987).

The purpose to develop an error correction model was that the co-integrating regression considers only the long-run property of the model, and does not deal with the short-run dynamics, and that the error correction model considers short-run dynamics and long-run equilibrium simultaneously.

The rationale of error correction model is that disequilibrium from one period is corrected in the next period because long-run equilibrium exists, as it is considered by co-integration relationship.

If both y_t and x_t are integrated on order 1, $I(1)$, but cointegrated, we define the error correction term by

$$u_t = y_t - \beta x_t \quad (1)$$

Where β is a co-integrating coefficient. u_t is the error from a regression of y_t on x_t .

An error correction model is specified as

$$\Delta y_t = \alpha_1 + \alpha_2 u_{t-1} + \alpha_3 \Delta x_t + \alpha_4 x_{t-1} + \alpha_5 y_{t-1} + v_t \quad (2)$$

$$\Delta y_t = y_t - y_{t-1} \quad (3)$$

$$\Delta x_t = x_t - x_{t-1} \quad (4)$$

The short-run adjustment behavior is partially but crucially captured by the error correction coefficient α_2 . That the error correction coefficient α_2 must be negative is because if y_{t-1} is above its equilibrium, then it will start falling in the next period and the equilibrium error will be corrected in the model. α_3 is the short-run effect and $-\frac{\alpha_4}{\alpha_5}$ is the

long-run response. This error correction model above explains how error correction model allows estimation of both short-run and long-run parameters simultaneously.

However, one of the limitations of error correction model is that if non-stationary time series data are not integrated of same orders, the employment of the error correction models is inappropriate.

2.4.2 Partial adjustment model

The partial adjustment model was first introduced by Nerlove. Gujarati (1995) illustrated partial adjustment model based on the flexible accelerator model of economic theory that assumes that there is an equilibrium, optimal, desired, or long-run amount of capital stock needed to produce a given output under the given state of technology, rate of interest, etc.

In the market for rail transport, we can assume that rail travelers have a 'desired' level of demand for rail journeys.

The desired level of demand for rail journeys, Y_t^* , which is not observable, is a linear function of fare X ,

$$Y_t^* = \beta_0 + \beta_1 X_t + u_t \quad (5)$$

The partial adjustment hypothesis that Nerlove postulates, is,

$$Y_t - Y_{t-1} = \delta(Y_t^* - Y_{t-1}) \quad (6)$$

$0 < \delta \leq 1$, is known as the coefficient of adjustment. The left side of the equation is the actual change and the right side of the equation is the desired change. After transformation, we get

$$\begin{aligned} Y_t &= \delta(\beta_0 + \beta_1 X_t + u_t) + (1 - \delta)Y_{t-1} \\ &= \delta\beta_0 + \delta\beta_1 X_t + (1 - \delta)Y_{t-1} + \delta u_t \end{aligned} \quad (7)$$

The speed of adjustment is calculated according to the formula

$$n = \frac{\text{Ln}(1-p)}{\text{Ln}\phi_1} \quad (8)$$

Where

n = number of periods taken for p % of demand to adjustment,

p = proportion of demand adjustment (eg.90%, 95%), and

ϕ_1 = absolute value of the coefficient associated with lagged first period volume. It is $(1 - \delta)$ in the partial adjustment model above.

The formula above also applies to the error correction model.

Because Y_{t-1} is included as one of the explanatory variable for Y_t , estimates may indicate a significant relationship between Y_t and Y_{t-1} but in fact there is not. If Y_t is nonstationary, this regression may be a spurious regression.

Spurious regression is one of the most important limitations of this model. The estimates may indicate a significant relationship between Y_t and Y_{t-1} but in fact there is not.

2.4.3 Geometric distributed lag model, Arithmetic distributed lag model, and Almon distributed lag model and applied studies

2.4.3.1. Geometric distributed lag model

The idea of the model was first introduced by Koyck.

A distributed-lag model can be written as:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \dots + \beta_k X_{t-k} + u_t \quad (9)$$

Koyck assumes that all the coefficients ($\beta_0, \beta_1, \beta_2, \dots, \beta_k$) are all of the same sign and they decline geometrically

$$\beta_k = \beta_0 \lambda^k \quad k = 0, 1, \dots \quad (0 < \lambda < 1) \quad (10)$$

In which λ is the rate of decline, $1 - \lambda$ is the speed of adjustment.

Koyck transformation converts a distributed lag model to an autoregressive model.

$$Y_t = \alpha(1 - \lambda) + \beta_0 X_t + \lambda Y_{t-1} + v_t \quad (11)$$

$$v_t = u_t - \lambda u_{t-1} \quad (12)$$

The assumption of geometric lag model is quite plausible because as one goes back into the distant past, the effect of that lag on Y_t becomes progressively smaller. Tremendous simplification has accomplished and multicollinearity is resolved by replacing all the lags of X_t by a single variable Y_{t-1} . Same as the problem of partial adjustment model, because Y_{t-1} is included as one of the explanatory variable for Y_t , estimates may indicate a significant relationship between Y_t and Y_{t-1} but in fact there is not if it is a spurious regression.

2.4.3.2. Arithmetic Lag model

A distributed-lag model can be written as:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \dots + \beta_k X_{t-k} + u_t \quad (13)$$

The arithmetic lag model assumes that all the coefficients ($\beta_0, \beta_1, \beta_2, \dots, \beta_k$) are all of the same sign and they decline arithmetically (on a straight line).

$$\beta_i = (k + 1 - i)\lambda \quad (14)$$

The assumption of the arithmetic lag model is plausible because as one goes back into the distant past, the effect of that lag on Y_t can be expected to be smaller.

2.4.3.3. Almon distributed lag

The idea of polynomial distributed lag was first introduced by Almon(1965) to predict quarterly capital expenditures in manufacturing industries from present and past appropriations.

Data used by Almon were quarterly data on appropriations and expenditures from the survey conducted by the National Industrial Conference Board (NICB) among the thousands largest manufacturing companies.

It is reasonable to assume that the immediate impact might be less than the impact after several periods (e.g. months, quarters) and the effects will diminish for the remainder of the lag after. In polynomial distributed lag models, we restrict the lag weight to fall on polynomial. This is one to combat the ill-effects of collinearity by the use of restricted least squares.

2.5 Summary

This section critically evaluated previous work on the theories of dynamic models.

The applied studies by the employment of dynamic (or static) models summarised are briefly summarised in this section.

As regards dynamic models, dynamic models include three categories: a. dynamic models lagging independent variables and dependent variables (e.g. error correction model); b. dynamic models lagging only dependent variables (e.g. partial adjustment model) and c. dynamic models lagging only independent variables (e.g. adaptive expectations model, Koyck model and polynomial distributed lag model). The origin, rationale, development history up to date and critical appraisal of each model have been reviewed. For dynamic models lagging only independent variables, because of the multicollinearity problem, which reduces statistical precision and also confounding various effects of the estimation, one solution to this is to impose some kind of lag structure which reduces the number of variables to be specified. For the partial adjustment model that lagging only the dependent variable, spurious regression and the meaningless result are the main problems. Moreover, compared with the partial adjustment model, the error correction model has a theoretical superiority because it

can include various lag structures. The lag structure can be identified by the partial autocorrelation function, as it has been used in Oxera(2005).

The applied studies by the employment of dynamic (or static) models summarised several studies from Owen and Phillips (1987) to Oxera(2005). As regards speed of adjustment, while the majority of studies, reported that over 90% of the demand adjustment occurred within a year after fares change, Phillips (1987) found evidence of a longer lag structure, suggesting that it would take 16-22 four-weekly period for 95% of demand to adjust, given a shock to a right-hand-side variable. An array of elasticities has been obtained by the employment of various models based on different variables and time periods. Estimated results indicate that long-run elasticities estimated are approximately double the size of the short-run elasticities. Short-term demand elasticities with respect to fare have ranged from -0.16 (NERA, 2003) to a high of -0.69 (Owen and Phillips, 1987). Long-term fare elasticities have ranged from NERA's estimated value of -0.70 for London long-distance flows to a high of -1.14 for non-London short-distance flows. Of particular significance is Oxera (2005), which carried out a literature review of applied studies by the employment of dynamic models and a meta-analysis of results from this review.

3. Data and research method

3.1. Data

We used the second class season tickets to London (annual data from 1995 to 2002) as a trial. Data from SRAⁱⁱⁱ include the number of second class season tickets from 46 origin stations to London, central London employment, fare and GVA^{iv}. We also construct an intercept dummy for Hatfield tragic accident on Oct. 17th, 2000 because a broken track derailed a high-speed train.

The 46 origin stations include: Banbury, Bath Spa, Birmingham International, Birmingham New Street, Birmingham Stns, Bournemouth, Bristol Parkway, Bristol Temple Meads, Bury St Edmunds, Cardiff Central, Chichester, Chippenham, Coventry, Derby, Ely, Folkestone Central, Folkestone Stns, Folkestone West, Gloucester, Grantham, Huntingdon, Ipswich, Kings Lynn, Leicester, Lincoln Central, Margate, Market Harborough, Moreton in Marsh, Newark North Gate, Newbury, Newport(South Wales), Northampton, Norwich, Nottingham, Peterborough, Portsmouth Harbour, Portsmouth Stns, Rugby, Salisbury, Southampton Central, Stafford, Swindon, Warwick, Wellingborough, Weston-Super-Mare, and Worcester Shrub Hill.

3.2. Research Method

3.2.1 A fixed effects model

Because 8-year annual data are not adequate to employ dynamic models for each flow, we apply the fixed effects approach to pool time-series and cross-sectional data to solve this problem. The main hypothesis is each coefficient of the model (15) is statistically

significant. The assumption is that the coefficients of fare, Gross value added, and central London employment are the same for all origin stations, and all behavioural differences between individual origin stations and over time are captured by the intercept.

The demand equation is:

$$J_{it} = \alpha + \beta_1 F_{it} + \beta_2 E_t + \beta_3 G_{it} + \beta_4 H_{it} + \beta_{d1} D_1 + \beta_{d2} D_2 + \dots + \beta_{d45} D_{45} + e_{it} \quad (15)$$

Where

J_{it} = number of second class season tickets between an origin station to London (46 stations in total),

F_{it} = average revenue per journey,

E_t = central London employment,

H_{it} = Hatfield dummy,

G_{it} = regional gross value added per head (2002 price),

D_1, D_2, \dots, D_{45} = origin stations dummy variables, and

e_{it} = random error term.

That we choose the explanatory variables above is because central London employment is an index of economic activity in London, GVA is an index of Britain as a whole, and fare is the most direct factor that commuters consider when they make travel mode decisions.

$(\alpha + \beta_{di})$ is origin stations intercepts. We take natural logarithm for number of journeys, central London employment, fare and GVA. Therefore, the estimates of these coefficients can be explained as elasticities. The table below summarizes estimates.

Table 1

Variables	One-year elasticities	Standard error	t -Statistic	Sig.
Fare	-0.767	0.189	-4.067	0.000
Central London Employment	2.267	0.797	2.843	0.005
GVA	0.899	0.432	2.083	0.038
Hatfield Dummy	0.112	0.062	1.825	0.069

$$\text{Adjusted } R^2 = 0.96$$

The signs of estimates of fare, central London employment, and GVA are consistent with expectation. Their t statistic of fare and central London employment show that they are significantly different from 0 (at 5% level). The t - statistic of GVA indicates that it is significantly different from 0 (at 10% level). However, the Hatfield dummy is not statistically significant at 10% level. The estimated fare elasticity indicates that if the price increases 1%, the demand for rail travel will decrease 0.767%. Compared with values of fare elasticities in previous studies, the value of the elasticity is also consistent with the British rail travel market as a whole. The estimates of elasticities of central London employment and GVA indicate that the demand for rail transport will increase

2.267% or 0.899% respectively, if central London employment (or GVA) increases 1%. Among the three explanatory variables, central London employment is the most important factor that affects the transport demand. The estimate of the Hatfield dummy is 0.112, indicating that people were more likely to choose rail transport as their travel mode after the tragedy. This is not consistent with the previous expectation.

However, these estimates may be unreliable because central London employment and GVA are highly correlated. The multicollinearity problem may give rise to meaningless results.

As regards limitations and future research directions, besides the multicollinearity, there are also two other problems in the model. First, this model cannot yield long-run elasticity, speed of adjustment, and the length of long-run. Second, there is a danger of getting spurious regression results by using non-stationary data. Future research is expected to solve all these problems, starting with removing outliers, testing non-stationarity, including more explanatory variables.

4. Conclusions and recommendations

Dynamic models have been widely used to model transport demand because of the lag effect in the market. A serious shortcoming of previous studies in railway market is that they did not address the issue how commuting market is different to business travel market and leisure travel market. The dynamics of these markets would be expected to be different. The further research is proposed to address this issue, conducting a comparative empirical study of different models in different markets. It is expected that the dynamics of these travels are different.

The research has applied a static fixed-effects model for estimation. Problems of preliminary analysis include multicollinearity, fixed effects assumption, and a danger of getting spurious regression results by using non-stationary data. Further research is expected to solve all these problems, starting with removing outliers, testing non-stationarity, including more explanatory variables, collecting four-weekly data and finally developing a new model.

Acknowledgements

I would like to thank Professor Mark Wardman, Dr Richard Batley, and Dr. Susan Grant-Muller, for valuable advice and feedback. This research is funded by a PhD fellowship award from the Marie Curie Association. Responsibility for the analysis and conclusions is solely my own.

References

Almon, S. 1965: 'The Distributed Lag between Capital Appropriations and Expenditures', *Econometrica*, 33, pp178-196.

Banerjee, A., J. Dolado, J. W. Galbraith, and D.F. Hendry 1993: Co-integration, Error Correction, and the Econometric Analysis of Non-Stationary Data, Oxford University Press, Oxford.

Box, G.E.P., and G.M. Jenkins 1976: Time Series Analysis: Forecasting and Control, 2nd Ed.. San Francisco: Holden Day.

Engle, R.F. and C.W.J. Granger 1987: 'Co-integration and Error Correction: Representation, Estimation, and Testing.' *Econometrica*, 55, pp251-276.

Fisher, I. 1925: 'Our Unstable Dollar and the So-Called Business Cycle', *Journal of the American Statistical Association*, 20, pp179-202.

Fisher, I. 1937: 'Note on a Short-Cut Method for Calculating Distributed Lags', *Bulletin of the International Statistical Institute*, 29, pp323-327.

Goodwin, P.B. 1976: 'Habit and Hysteresis in Mode Choice', *Urban Studies*, 14, pp95-98.

Goodwin, P., J. Dargay and M. Hanly 2004: 'Elasticities of Road Traffic and Fuel Consumption with Respect to Price and Income: A Review', *Transport Reviews*, 24(3), pp 275-292.

Gujarati, D.N. 1995: *Basic Econometrics*, McGraw-Hill, New York.

Hendry, D.F. 1995: *Dynamic Econometrics*, Oxford University Press, Oxford.

Hill, R. C., W.E. Griffiths and G.G. Judge 2001: *Undergraduate Econometrics*, 2nd Ed. Wiley & Sons, New York.

Koyck, L. M. 1954: *Distributed Lags and Investment Analysis*, North Holland, Amsterdam.

Little, R. and D. Rubin 1987: *Statistical Analysis with Missing Data*, Wiley, New York.

Maddala, G.S. 1992: *Introduction to Econometrics*, 2nd Ed.. Macmillan Publishing Company, New York.

NERA 1999: *Analysis of Passenger Rail Demand. A Report for OPRAF*, National Economic Research Associates, London.

NERA 2003: *Research on Long Term Fare Elasticities, A Final Report for the Strategic Rail Authority*, National Economic Research Associates, London.

Nerlove, M. 1956: 'Estimates of Supply of Selected Agricultural Commodities', *Journal of Farm Economics*, 38, pp496-509.

Nerlove, M. 1958: *Distributed Lags and Demand Analysis for Agricultural and Other Commodities*. Agricultural Handbook No. 141, Government Printing Office, Washington, D.C.

Owen, A.D. and G.D.A. Philips 1987: 'The Characteristics of Railway Passenger Demand: An Econometric Investigation', *Journal of Transport Economics and Policy*, 21(3), pp231-253.

Oxera 2004: *Literature Review*. Report prepared for Passenger Demand Forecasting Council, Oxera Consulting Ltd, Oxford.

Oxera 2005: *How do Rail Passengers Respond to Change?* Report Prepared for Passenger Demand Forecasting Council, Oxera Consulting Ltd, Oxford.

Pokorny, M. 1987: *An Introduction to Econometrics*, Basil Blackwell, New York.

Philips, G.D.A. 1987: *An Econometric Investigation into the Characteristics of Railway Passenger Demand: Further Analysis*, Working paper, University of Manchester.

Philips, G.D.A. and R.A. Woodfin 1988: *An Econometric Investigation into the Characteristics of Railway Passenger Demand: The Forecasting Performance of the Model*, Working paper, University of Manchester.

Preston, J. 1998: *Public Transport Elasticities: Time for a Re-think?* Transport Studies Unit, University of Oxford

Sargan, J.D. 1964: 'Wages and Prices in the UK: A Study in Econometric Methodology, Econometric Analysis for National Economic Planning, pp.25-59.

Steer Davies Gleave 2003: *The Effect of Road Congestion on Rail Demand*. Final Report Prepared for Association of Train Operating Companies, Steer Davies Gleave, London.

Voith, R. 1997: 'Fares, Service Levels and Demographics: What determines commuter Rail Ridership in the Long Run?', *Journal of Urban Economics*, 41, pp176-197.

Wardman, M. and J. Shires 2003: *Review of Fares Elasticities in Great Britain*, Paper ITS: 573

Wardman, M. and G. Whelan 2004: *Generalised Journey Time Reformulation: Final Report*, Working paper, Institute for Transport Studies, University of Leeds.

ⁱ Spurious regression refers to when nonstationary time series are used in a regression model the results may indicate a significant relationship when there is none.

ⁱⁱ For example, in Oxera (2005), when the ratio of long-to short-run elasticities is calculated, both an error correction model and a partial adjustment model are applied. When elasticities are obtained from partial adjustment model, the ratio is 32% lower.

ⁱⁱⁱ SRA refers to Strategic Rail Authority

^{iv} GVA refers to Gross Value Added