Oxera

Agenda

Advancing economics in business

Consumer responses: how quick are they?

Ignoring the potential delayed reactions of consumers when thinking about investment, pricing, or competition cases, or when evaluating changes in consumer behaviour, could lead to biased policy conclusions. This article highlights these limitations and presents ways in which they may be overcome. It also discusses implications for undertaking future modelling exercises

The use of econometrics is increasingly prevalent in competition analysis and policy evaluation, and is a crucial part of any assessment of the drivers of demand for goods and services. This article analyses some of the approaches to modelling consumer demand, and suggests possible limitations associated with current practice. Such limitations relate to the fact that the short-run response of consumers to a particular shock (eg, a price increase) may differ from their long-run response, and to understanding how long it takes to move from the short run to the long run. In an attempt to overcome these limitations, methods from recent Oxera work examining the demand for rail transport are discussed.

Why care about delayed responses?

The fact that consumers may take differing lengths of time to respond to changes in market conditions could have a significant impact on understanding demand behaviour. In competition cases, for example, the SSNIP test is used to define the relevant market, by asking whether a hypothetical monopolist would be able to increase prices profitably by 'a small but significant and non-transitory' amount. In answering this question, competition authorities are typically concerned with the effect that the price rise would have over a one- or two-year horizon, regardless of the market being considered. In some markets, this may easily capture the effect of the hypothetical price rise and would lead to correct conclusions regarding the appropriate market definition. However, in other markets there may be delayed responses that extend beyond the two-year horizon. In such cases, and where these longer-term effects are likely to have an impact on whether the hypothetical monopolist increases its prices, failure to take these effects into account could lead to an inaccurate delineation of the relevant market.

The importance of understanding delayed ('lagged') responses is not restricted to competition policy. For example, a full assessment of the effects of a policy change, such as the introduction of a new tax or an increase in existing taxes, should ideally take into account lagged effects as well. Indeed, it is far from obvious how long-run effects will compare with short-run effects. Conventionally, economic theory has argued that long-run elasticities will be greater than their short-run equivalents, as consumers will be more able to change their behaviour in the long run. However, in some markets, the opposite may be true, with short-run elasticities being greater than their long-run equivalents since they capture 'hoarding' (or stocking) behaviour by consumers.

Given the importance of appropriately capturing the delayed responses of market participants, this article suggests ways in which dynamic responses can be modelled, focusing on the determination of lagged effects, the estimation of short- and long-run responses, and the calculation of speeds of adjustment—ie, how long it takes to move from the short to the long run.

These factors are discussed in the context of recent work undertaken by Oxera for the Passenger Demand Forecasting Council, which looked at demand for passenger rail services in the UK and captured passengers' lagged responses.

Modelling passenger demand for rail services

A useful starting point is to consider why lagged effects should form part of the modelling process. Delays in response may arise as a result of the slow spread among potential users of information about changes in market conditions, such as an increase in the length of delays. In addition, passengers may be unable or unwilling to switch to alternative modes of transport in the short run.

To capture these effects in the rail context, it was first necessary to consider the factors that affected demand for rail services. It can be assumed that demand, as measured by recorded ticket sales data, is adequately explained by the following variables:

- real fares—ie, fares adjusted for inflation;
- generalised journey time (GJT)—a measure of journey time that takes into account both the perceived disadvantage of having to change services and the frequency of services;
- delay minutes per train-mile;
- gross domestic product (GDP).

The purpose of the study was to calculate the time taken for demand to reach its long-run equilibrium and to describe the profile of the change in demand, given a change in any of these explanatory variables. Modelling was based on both four-weekly and annual ticket sales data, with the four-weekly analysis covering the period between 1995/96 and 1999/2000, and the annual analysis covering 1992–2001.

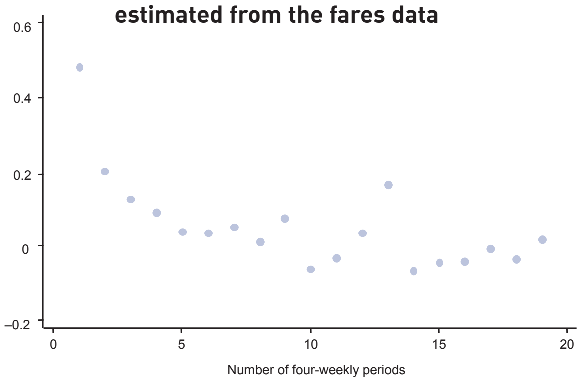
The determination of lagged effects

To model lagged effects, the first question that should be addressed is which lags—and of which variables—have the greatest ability to predict current ticket sales? The most suitable way to answer this question is through the development of a so-called partial autocorrelation function (PACF). This is an econometric tool that assesses the extent to which the value that a particular variable takes today, say at time t, is correlated with (broadly speaking, is linked with) the value that the same variable took at a previous point in time, say t – 2. The reason for it being termed a *partial* autocorrelation function is that this correlation is calculated holding the values of the variable fixed between time t and t – 2. For example, the value that a particular variable took two years ago could be expected to have an effect on the value the same variable took one year ago, which in turn would have an effect on that variable today. However, the PACF adjusts for all these 'linkage' effects and simply measures the correlation between the value that the variable took two periods ago and the value it takes today.

As the PACF removes the effects of intervening variables, it makes it easier to judge those lags that are worth including in the model. First, it can be assumed that the current values of all the explanatory variables mentioned above will help to explain ticket sales. Beyond this, if the previous values of a particular explanatory variable are not strongly related to the present value of that variable (ie, they have a low or zero partial autocorrelation), it is unlikely that these lags will have significant explanatory power in a model determining ticket sales. That said, highly positive or negative partial autocorrelation coefficients indicate that previous values of the explanatory variables are either directly or inversely correlated with the current value of the explanatory variable of interest, and therefore indicate lags that should be included in the model.

The PACF in Figure 1 shows the partial autocorrelation coefficients for one of the explanatory variables: real fares. The partial autocorrelation coefficients that do not follow the declining trend, that are noticeably different from zero, and that are therefore worth including in the initial model, are at periods one and two, and one year (13 periods ago). Further analysis of the fares data demonstrated that the two-year (26th-period) lag of fares was also significant. Similar results were found for GJT and delay minutes per train-mile, while analysis for GDP suggested that the three-year lag should also be assessed in the econometric modelling.

Figure 1 Partial autocorrelation coefficients estimated from the fares data



Source: Oxera calculations.

Estimating short- and long-run responses

After the lags to be considered have been identified from analysis of the PACFs, an error correction model, which distinguishes between short- and long-run responses, can be constructed. Error correction models seek to explain the movement in the dependent variable (ie, the volume of ticket sales) by how far that variable is from what is estimated to be its long-run 'equilibrium' value.

Initially, a general model should be estimated which considers a wide range of possible explanatory variables and their respective lags. From this model, the least statistically significant variables are removed sequentially until the remaining coefficients are statistically robust.[1] The error correction model, based on the four-weekly ticket sales data, was estimated according to Equation 1 (see below). Equation 1 notes that the change in ticket sales from the previous period can be explained by the change in fares, journey length and delays between today's value and the value that these variables took in the previous period, as well as the number of ticket sales, actual real fares, actual GJT, actual delay minutes and actual GDP in certain, previously identified, important periods.

$$\Delta v_t = \sum_i \beta_{vi} v_{t-i} + \beta_1 \Delta p_t + \beta_2 \Delta gjt_t + \beta_3 \Delta delay_t + \sum_i \beta_{pi} p_{t-i} + \sum_i \beta_{gjti} gjt_{t-i} + \sum_i \beta_{di} delay_{t-i} + \sum_j \beta_{gi} gdp_{t-j} + s_{2,\dots,13}$$

Equation 1

Δv_t	change in ticket sales from the previous period
$\sum_i v_{t-i}$	previous ticket sales from 1, 2, 13 and 26 periods ago
Δp_t	change in real fares from the previous period
Δgjt_t	change in total GJT from the previous period
$\Delta delay_t$	change in delay minutes per train-mile from the previous period
$\sum_i p_{t-i}$	previous real fares from 1, 2, 13 and 26 periods ago
$\sum_i gjt_{t-i}$	previous total GJT from 1, 2, 13 and 26 periods ago
$\sum_i delay_{t-i}$	previous delay minutes per train-mile from 1, 2, 13 and 26 periods ago
$\sum_i gdp_{t-j}$	previous GDP from 13, 26 and 39 periods ago
$s_{2,\dots,13}$	seasonal dummy (binary) variables

From the error correction model, short-run estimates of the elasticities can be directly interpreted as the coefficients associated with the change in the parameters from one period to the next. For example, the short-run elasticity of demand with respect to fares is simply the estimated coefficient on the change in fares between time t and time t – 1.

Obtaining the long-run elasticities is more complicated. The statistically significant coefficients associated with the past values of ticket sales and those from the variable of interest, such as fares or journey time, need to be considered. Equation 2 shows the formula used to calculate the long-run fares elasticity (p_{LR}), with variables defined below. It shows that the long-run elasticity can be calculated by adding together all of the coefficients on the lagged values of fares that are statistically significant and then dividing this by the (absolute value) sum of the statistically significant coefficients associated with lags of ticket sales.[2]

$$p_{LR} = \sum_i \beta_{pi} / (-\sum_i \beta_{vi})$$

Equation 2

p_{LR}	long-run elasticity of demand with respect to fares
$\sum_i \beta_{pi}$	sum of the statistically significant coefficients associated with lags of fares from Equation 1
$-\sum_i \beta_{vi}$	absolute value of the sum of the statistically significant coefficients associated with lags of ticket sales from Equation 1

Calculating speeds of adjustment

After estimating short- and long-run elasticities, the time it takes to move between the short and long run, and the corresponding changes in the profile of demand over time, can also be inferred. The formula most commonly used to calculate the time taken for a specified percentage of demand to reach its long-run equilibrium level is shown in Equation 3. This formula relates the assumed proportion of demand adjustment to the coefficient associated with ticket sales one period ago.

$$T = \ln(1 - p)/\ln(1 - \varphi_1)$$

Equation 3

T	number of periods taken for x% of demand to adjust
ln	natural logarithm
p	proportion of demand adjustment (x%)
φ_1	absolute value of the coefficient associated with the first period lag of ticket sales

For example, if the coefficient on ticket sales from one period ago is –0.9, the time taken for 99% of demand to reach its long-run level, given a change in an explanatory variable, can be calculated according to Equation 4.

$$T = \ln(1 - 0.99)/\ln(1 - 0.9) = 2$$

Equation 4

This implies that, after two periods, 99% of demand would have adjusted to reach its equilibrium level, given a change in fares.

However, Equations 3 and 4 are only valid if one lag of the dependent variable is included in the model. As has been seen above, there are good reasons to believe that real-world behaviour is more accurately captured by a model that uses a more sophisticated lag structure. To obtain accurate results from such a model, an alternative calculation needs to be adopted, based on more advanced mathematical techniques.

Results from the rail industry

Applying these techniques to the UK passenger rail industry, it is interesting to note that Oxera's findings have diverged from other studies in two important ways:

- elasticities are larger than previously estimated;
- greater time is needed in order to reach the long run.

In terms of the first of these effects, the fare elasticities estimated by Oxera are provided in Table 1 below. They show that, for all of the different journey types examined, the long-run elasticities are in absolute terms greater than 1, with the aggregate long-run elasticity being –1.84. This compares with estimates in the *Passenger Demand Forecasting Handbook*,[3] which suggest that long-run fares elasticities are less than 1. Similarly, the Transport Research Laboratory estimated a long-run rail

Table 1 Fare elasticities estimated using the four-weekly data

	Non-London long distance	Non-London short distance	London to the rest of the country	London and the south-east	To and from airports	Aggregate
Long-run elasticity	–1.77	–1.99	–2.03	–1.48	–1.74	–1.84
One-year elasticity	–1.23	–1.56	–0.78	–1.27	–1.46	–1.26
One-period elasticity	–0.81	–1.54	–0.69	–1.47	–0.76	–0.80
90% speed of adjustment expressed in years (y) and periods (p)	1y 4p	2y 2p	4y 8p	1y	1y 1p	1y 6p

Source: Oxera calculations.

fares elasticity of –0.65.[4] The fact that the demand for rail travel may, in the longer run, be much more sensitive to price than was previously thought is clearly a significant finding in the rail industry, not least in light of the decision by the UK Strategic Rail Authority to relax the fares cap from RPI – 1% to RPI + 1% in order to generate more passenger revenue.[5]

However, although the work suggests that absolute fares elasticities may be greater than previously thought, it also suggests that the time taken for consumers to respond to changes in fares, or indeed other variables, is likely to be significantly longer than previously envisaged. Reflecting the fact that previous studies did not use similar lag structures in their modelling approach, and indeed had often only looked at one lag, the industry consensus has been that the full effect of any fares change (or change to other explanatory variables) was worked through within one year.[6] By contrast, Oxera's work suggests that the speed of adjustment is often slower than this, and for some journey types, such as London to the rest of the country, can be slower still.

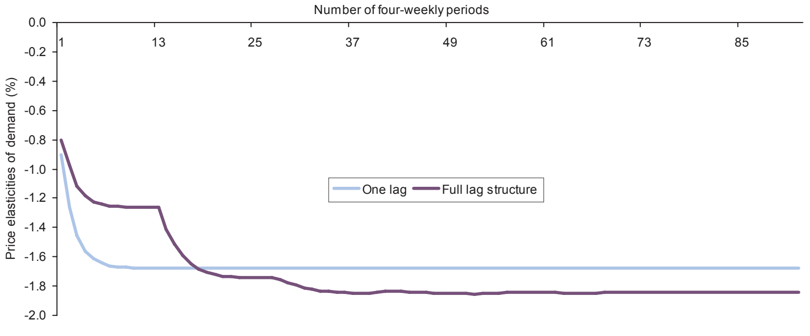
The modelling technique also allowed an estimate to be made of how the elasticities change over time towards the long-run 'equilibrium'. This is illustrated in Figure 2, which shows that the move towards the long-run elasticity was estimated to take place predominately in two large steps in years 1 and 2 after the shock, with only smaller adjustments after this time period had elapsed. The graph also illustrates the effect that artificially suppressing the number of lags in the model will have on the speed of adjustment estimates—the 'one-lag' line illustrates the (much quicker) speed of adjustment that would have been estimated by the same data, had a model with only one lag been used.

Conclusions

The analysis of the drivers of demand for rail transport suggests that it is important to take into account all lagged responses, particularly when seeking to model the behaviour of market participants. The time taken to reach the long run was found to be substantially greater than previous industry estimates, and the elasticities were also found to be larger than other results. This illustrates that artificially suppressing the lag structure could lead to a shorter speed of adjustment, and estimates of the coefficients may differ significantly from the 'true' estimates.

This has ramifications for many types of analysis, including policy appraisal, market definition in competition cases, and any demand modelling exercises, and could lead to differing policy conclusions—from revisions to the definition of 'non-transitory' when applying the SSNIP test, to changes in the length of rail franchises in order to increase the controllability of revenues for successful bidders.

Figure 2 Speeds of adjustment in response to changes in fare



Source: Oxera calculations.

[1] No inference can be drawn from estimates that are not statistically robust, as they may be significantly different from their 'true' values.
[2] In the long run, it is assumed that equilibrium is reached and there will be no change in real fares. Therefore, the coefficient associated with the change in real fares is not included in the calculation of the long-run fare elasticity.
[3] Passenger Demand Forecasting Council (2002), *Passenger Demand Forecasting Handbook*, version 4.
[4] Balcombe, R. (ed) (2004), 'The Demand for Public Transport: a Practical Guide', Transport Research Laboratory Report TRL593.
[5] Strategic Rail Authority (2003), 'Fares Review: Conclusions', June.
[6] Passenger Demand Forecasting Council (2002), op. cit.

If you have any questions regarding the issues raised in this article, please contact the editor, Derek Holt: tel +44 (0) 1865 253 000 or email d_holt@oxera.co.uk

Other articles in the May issue of *Agenda* include:

- **future secure? why asset managers locate in London**
- **reform of Article 82: where the link between dominance and effects breaks down**
- **water mergers: what are the prospects?**

For details of how to subscribe to *Agenda*, please email agenda@oxera.co.uk, or visit our website

www.oxera.com