Lancaster University Management School
Working Paper
2006/004

Robust estimates of the impact of broadcasting on match attendance in football

Babatunde Buraimo, David Forrest and Rob Simmons

The Department of Economics
Lancaster University Management School
Lancaster LA1 4YX
UK

ROBUST ESTIMATES OF THE IMPACT OF BROADCASTING ON MATCH ATTENDANCE IN FOOTBALL

by

Babatunde Buraimo*

David Forrest **

and

Robert Simmons***

December, 2005

*Lancashire Business School, University of Central Lancashire, Preston, PR1 2HE, UK
**School of Accounting, Economics and Management Science, University of Salford, Salford, M5 4WT, UK
*** Department of Economics, The Management School, Lancaster University, Lancaster, LA1 4YX, UK.

Abstract

The paper employs data from 2,884 matches, of which 158 were televised, in the second tier of English football (currently known as The Football League Championship). It builds a model of the determinants of attendance that is designed to yield estimates of the proportionate changes in the size of crowds resulting from games being shown on either free-to-air or subscription based channels. The model has two innovatory features. First, it controls for the market size of home and away teams very precisely by including local population measures constructed from the application of GIS software and information on competition from other clubs. Second, it employs the Hausman-Taylor random effects estimator in order to take explicit account of the endogeneity of the television coverage variable and of other variables typically included in earlier studies based on ordinary least squares or fixed effects models of attendance. The Hausman-Taylor estimates of the impact of broadcasting are greater than those reported in such studies. In the case of free-to-air television, the negative impact is estimated as over 20 percent but for subscription television, which carried most of the transmissions, the negative effect was only of the order of 5 percent.

Keywords: football attendance; television; Hausman-Taylor Estimator

ROBUST ESTIMATES OF THE IMPACT OF BROADCASTING ON MATCH ATTENDANCE IN FOOTBALL

1. Introduction

From the inception of television, professional sports leagues in both Europe and America have been concerned that permitting matches to be broadcast live would detract from attendance at the stadium. As a result, in the case of English football, the governing body agreed to live telecasting of domestic league fixtures on a limited basis (initially for just ten matches per season) only in 1986 (Dobson and Goddard (2001) and Forrest, Simmons and Szymanski (2004) provide reviews of the relevant history of the often cool relationships between sports leagues and television).

It remains of interest to sports leagues and clubs whether and to what extent ticket revenue is threatened by licencing the coverage of a match. In fact, the issue is likely to grow in importance in English football if pressure from the European courts leads to the abandonment of collusive negotiation of rights at league level. In a market where competitive selling will erode the value of rights, individual clubs will require to know the minimum compensation for loss of gate revenue that they should build in to their demands. But the issue is of wider significance than narrow profit and loss accounting. If all clubs sell rights to all their home fixtures and if televising games systematically reduces the crowd at the stadium, then the appeal of the professional football match as a spectacle is likely to be reduced, with risk of long-run erosion of interest in the sport. From a societal perspective, football may become an example of an arena (film is another) where home screen based entertainment, consumed in isolation, is routinely substituted for an occasion involving social interaction and the mingling of groups, from disparate backgrounds, united in a common experience.

In fact, we know little about how likely this scenario is and how readily, if at all, fans are ready to substitute television for live viewing. In a comprehensive survey of influences on attendance demand in sport, Borland and Macdonald (2003) remark that "there is not strong evidence on how TV broadcasts affect attendance": the number of studies in the area is relatively small and they do not yield consistent findings. Borland and Macdonald note that there has been some tendency for European studies (on English and Spanish soccer and English rugby league) to find zero or small negative impacts on attendance while American work, for example on baseball and college football, has often found a positive relationship between crowd size and telecasting. It could of course be that the true effects are indeed heterogeneous over time and across sports, particularly since they may depend on the type of broadcasting platform involved. However, Borland and Macdonald suspect that the very mixed results can "as well be attributed to the difficulties in undertaking empirical analysis…One problem is potential joint endogeneity". In an attempt to improve on the reliability of previous results, it is this problem of endogeneity that we seek to address in our case study of English soccer.

The archetypical study in the recent literature (for example, Garcia and Rodriguez (2002), Forrest, Simmons and Szymanski (2004)) applies a fixed effects model to panel data describing attendances at each club's sequence of home games during one or more seasons. The dependent variable is crowd size at club *i*'s home game number *t.* Categorical variables representing each home club control for influences such as varying market size, historical tradition and ticket pricing policy. Additional controls include variables particular to each match such as the distance between the home and away stadia (to allow for the effect of travel cost on attendance by away fans) and indicators of team quality and form (such as the current league positions of the home and away teams). To such an all-purpose attendance model is added a categorical variable set equal to one where a game is televised live. The coefficient on this categorical variable provides the estimate of the impact from telecasting.

Such an estimate will be unreliable (biased) to the extent that the decision to televise is itself likely to be determined by the set of other variables included on the right hand side of the equation. Broadcasters will seek to maximise audience size and are likely therefore to select games with attractive characteristics, such as high team quality, that also drive stadium demand. Further, since a large crowd will itself be valued by the producers, because it adds to the spectacle, the probability that a game will be shown will be influenced by *any* variable legitimately included in the attendance equation. The endogeneity problems will also extend beyond the television variable. For example, the fixed effects will capture the influence of the size of population in a club's catchment area but this affects resources available to spend on talent and therefore variables such as current league position that are assumed exogenous in the model. We conclude that estimation from a fixed effects (or an ordinary least squares) model is an unsatisfactory basis for evaluating television impacts.

Another weakness of the standard fixed effects model, now almost always the technique of choice in match attendance studies, is that, if fixed effects are modelled to capture unobserved heterogeneity in club attributes, one cannot then separate out the impact of those time-invariant club characteristics, such as size of local population, that are in fact observed. Results are therefore less rich than they might be since the bulk of the variation in attendance across matches is then invariably simply attributed to which team happens to host the fixture. The underlying reasons for some clubs having higher intercept terms than others cannot be explored at all given the structure of the model.

We propose the application of the Hausman-Taylor random effects estimator, described in Section 2 below, both to allow the selection of matches for television to be modelled as endogenous and to permit isolation of the effects of time-invariant variables such as local population. Estimates of the impact of television on attendance will be more robust than those from either simple ordinary least squares or the fixed effects models now commonplace in the sports literature.

2. The Hausman-Taylor Estimator

Consider a general model in which the dependent variable $ln\ y_{it}$ is determined by:

$$ln\ y_{it} = \boldsymbol{Z_i}\alpha + \boldsymbol{X_{it}}\beta + \theta_i + \varepsilon_{it} \qquad (1)$$

where the subscript "i" denotes the cross-sectional unit ($i = 1,2,\ldots N$), the subscript "t" denotes the time period ($t = 1,2,\ldots T$), $\boldsymbol{Z_i}$ is a vector of fixed covariates, $\boldsymbol{X_{it}}$ is a vector of time-varying covariates, θ_i is a time-invariant fixed effect and ε_{it} is a well-behaved error term. If (1) is estimated as a conventional 'deviation from means' fixed-effects model the values of θ_i and $\boldsymbol{Z_i}$ are equal to their means and it is not possible to obtain an estimate of α. The same problem arises in a fixed effects model expressed in first differences. From:

$$(ln\ y_{it} - ln\ y_{it-1}) = (\boldsymbol{Z_i} - \boldsymbol{Z_i})\alpha + (\boldsymbol{X_{it}} - \boldsymbol{X_{it-1}})\beta + (\theta_i - \theta_i) + (\varepsilon_{it} - \varepsilon_{it-1}) \qquad (2)$$

it follows that

$$\Delta lny_i = (\Delta \boldsymbol{X_i})\beta + \Delta\varepsilon_i \qquad (3)$$

which again does not deliver an estimate of α. The problem is that the 'within' fixed effects estimator mean differences the data before generating a consistent estimate of β. The estimator removes θ and also $\boldsymbol{Z}$.

The Hausman-Taylor (1981) estimator proceeds by assuming that some of the covariates are correlated with the unobserved cross-section unit-level *random* effect and uses an instrumental variable method. The method is explained more fully in Baltagi (2005) and Wooldridge (2002). Here we offer just a brief summary. In the first stage the within fixed-effects estimator consistently estimates β and generates residuals ($ln\ y_{it}$ minus predicted values of $\boldsymbol{X}_{it}\beta$). These residuals are regressed on $\boldsymbol{Z}_i$ using a set of time-varying exogenous variables and time-invariant exogenous variables as instruments. This yields intermediate (consistent) estimates of α. Both overall and within residuals are obtained. Together, these residuals are used to estimate the components of variance of the dependent variable. The estimated variance components are used to undertake a General Least Squares transform on each of the variables in the second stage.

In order to be implemented effectively, the Hausman-Taylor estimator requires several conditions. First, the unobserved cross-sectional level effect θ must indeed be random i.e. it has zero mean, finite variance and is independently and identically distributed over cross-section units. Second, we need to classify our explanatory variables into four types: time-varying and exogenous, time-varying and endogenous, time-invariant and exogenous, time-invariant and endogenous. Hence equation (1) should be re-written as

$$ln\ y_{it} = \boldsymbol{Z}_{1i}\alpha_1 + \boldsymbol{Z}_{2i}\alpha_2 + \boldsymbol{X}_{1it}\beta_1 + \boldsymbol{X}_{2it}\beta_2 + \theta_i + \varepsilon_{it} \tag{4}$$

where

$\boldsymbol{Z_{1i}}$ is a vector of exogenous, time-invariant variables that are not correlated with either θ_i or ε_{it};

$\boldsymbol{Z_{2i}}$ is a vector of endogenous, time-invariant variables that may be correlated with θ_i but are uncorrelated with ε_{it};

$\boldsymbol{X_{1it}}$ is a vector of exogenous, time-varying covariates that are not correlated with either θ_i or ε_{it};

$\boldsymbol{X_{2it}}$ is a vector of endogenous, time-varying covariates that may be correlated with θ_i but are uncorrelated with ε_{it};

Third, the order condition for identification requires that the number of variables in $\boldsymbol{X_{1it}}$ is at least as great as the number of variables in $\boldsymbol{Z_{2i}}$. Finally, there needs to be sufficiently strong correlation between instruments and $\boldsymbol{Z_{2i}}$.

Clearly, a major advantage of the Hausman-Taylor estimator is that it permits estimation of the impacts of time-invariant covariates in a panel data setting. Beyond this, the estimator economises on use of instruments. All instruments are derived from within the model. These are: $\boldsymbol{X_{1it}}$ and associated means, $\boldsymbol{Z_{1i}}$ and the deviations of $\boldsymbol{X_{2it}}$ from associated means. A search for external instruments, as would be required in fixed-effects models where covariates are potentially endogenous, is not required.

The Hausman-Taylor estimator has been applied in several settings. Among the questions addressed have been the impact of schooling on wages (Baltagi and Khanti-Akom, 1990), the impact of health on wages (Contoyannis and Rice, 2001) and the effects of distance on exports and foreign direct investment (Egger and Pfaffermayr, 2004). Here, we apply the Hausman-Taylor estimator to address the question of whether and by how much broadcasting of sports events reduces attendance at the stadium.

3. Context and data

The context for our analysis is English professional soccer where 92 clubs compete in four hierarchical divisions linked by a system of promotion and relegation. The top tier is known as The Premier League. The other divisions have been branded under various names over our study period and so we follow recent convention amongst sports analysts by referring to them as Tiers 2, 3 and 4. The current brand name of Tier 2 is the Football League Championship and it is this division that we choose for our case study. Tier 2, which comprised 24 teams, each playing 23 home games per season, is more amenable to analysis than Tier 1 because the proportion of sell-out games is so small (1.1% over our study period of seven seasons) that censoring of data raises no serious concerns. By contrast, capacity is filled regularly in the Premier League. While in principle, the tobit estimator is appropriate where some observations of the dependent variable are censored, the solution becomes untenable where certain clubs, as in the Premier League, sell all their seats every match. Further, the legitimacy of tobit estimation for examining attendance at other clubs is brought into question by the industry practice of restricting access to popular (sell out) games to those who have also purchased tickets for less attractive fixtures. Thus one does not observe ‘true’ demand even at games where the crowd is not capacity constrained. Tobit is incapable of estimating customer response to match characteristics if true demand is not observed at any game.

For matches in Tiers 3 and 4, sell-outs are virtually never observed. However, television exposure is very limited and there is an insufficient number of matches transmitted for any meaningful conclusions to be possible about broadcasting effects. We are left with Tier 2 as our preferred subject for analysis. Our approach should of course be applicable to other countries' football leagues and to competitions in sports such as baseball and American football that have been the focus of earlier published, but we believe flawed, studies.

Our data period extends over several more seasons than has been customary in this literature in order to capture an adequate number of televised games. It extends from season 1997/8 to season 2003/4. Over this period, the Football League entered into a number of contracts for its television rights and live coverage was variously relayed through three channels: the mainstream, terrestrial, free-to-air ITV; and two subscription channels, Sky Sports and the now defunct ITV Digital, accessible through cable and satellite. The variety of arrangements will permit separate estimation of the effect of telecasting according to whether the platform is free-to-air or pay but, since a large majority of screenings were on Sky, it is the effects of subscription television that will be estimated most precisely. Note that there were no examples of matches shown on pay-per-view television where viewing of each event is billed in addition to subscription charges.

Not all of the 3,864 matches played over the seven year period were included in the estimation (this number refers to 'regular season' fixtures; the small number of play-off games held at the end of each season to determine the final promoted club are not considered here). We deleted the opening round of matches from each season because two of our control variables required information on previous league form in the current season. There were also 21 cases of a club failing to declare its wage bill for a particular season. Since wage bills were used as one of our measures of the quality of teams on show in a game, we deleted all observations involving those clubs in those seasons. This left a final sample size of 2,884 matches.

Attendance across the games in the sample ranged from 3,436 to 44,135, with the mean 14,988 (standard deviation 7,237). Table 1 displays means (and standard deviations) by category of game: non-televised, televised on ITV, televised on ITV Digital and televised by Sky Sports.

The striking feature of the data in the table is that attendance was, on average, much higher (by over 17 percent) at televised than non-televised matches, regardless of broadcasting platform. In the early days of television, perhaps people may have attended an event just because they were drawn by the novelty of the cameras. But now it would be implausible to attribute higher crowds to television. We take it that broadcasters and attendees alike were attracted to games that had particularly strong characteristics and this is why large numbers in the stadium are observed when the cameras are present.

The point underlines the importance of multivariate analysis, with a full set of control variables to capture match characteristics and account taken of the relationship between those characteristics and the decision to broadcast or not. Without careful specification and appropriate technique, there is a danger that the greater attendance indicated for televised matches in the raw data will be reflected in coefficient estimates on television variables that are biased upwards (i.e. that underestimate any propensity for home viewing to be substituted for going to the stadium). The failure of previous studies to account for endogeneity of the television variable may have resulted in misleading findings. The positive effects of television coverage reported in American sport may in fact have been converted to zero or negative effects had endogeneity been taken into account. And findings of zero or negative impacts in Europe may conceivably mask a more substantial diminution in crowds when games are transmitted live on television.

4. Model

We have unbalanced panel data. The cross-sectional unit is a club playing home matches in a particular season (there are 147 such groups). The time unit is the match (observations per group varied, between 18 and 22, because some observations had been deleted due to missing information). The dependent variable is the natural logarithm of attendance.

Here we give details of the covariates included in the model. Table 2 presents a complete list (with summary statistics), grouped according to whether they are classified as exogenous time-invariant, endogenous time-invariant, exogenous time-varying or endogenous time-varying. Where allocation of a covariate to the appropriate vector might require justification, this is provided at the end of this section of the paper.

We hypothesise that the size of crowd at a given game will be influenced by: factors affecting the size of the market of the home club; by factors influencing the number of away supporters who will travel to the game; by scheduling issues; by the quality of the teams and players on show; and by television coverage of the match and of other football taking place at the same time.

Home club market size variables

A majority of attendees at a game will normally be local supporters and a considerable influence on the size of the crowd will therefore be the size of the market from which the home club draws its customers. Clubs in a large metropolitan area, so long as their advantage is not eroded by competition from other clubs nearby, would be expected to attract larger crowds for a typical match than those based in smaller centres. Some measure of local population should therefore be included in the model. Dobson and Goddard (1995), in a study of determinants of long-term levels of football club support, find positive effects from the population of the town (as recorded in the 1961 Census) in which the club is located; but this is a somewhat imprecise measure of market size to the extent that it is related to arbitrary administrative boundaries. Schmidt and Berri (2001) suggest that, in the context of the match level attendance literature, if a measure of market size is employed in preference to fixed effects, then “a common proxy for size of a team’s market is the size of its metropolitan statistical area (SMSA)” . They accordingly enter this in linear form in a baseball demand equation. But such use of the American SMSA population or its equivalent in other countries represents a misspecification. If one club is located in a city with twice the population of another, it cannot be considered as having double the market size. The bigger city will cover a wider area and the mean travel cost for residents to reach the stadium will be higher, implying that ticket demand will not be increased by as much as the population figures alone might suggest. SMSA population is therefore an inadequate proxy for market size.

Our solution, new in the match attendance literature, is to exploit modern GIS software to measure population within certain distances of the stadium, with the distances defined sufficiently tightly that travel costs from each part of a zone within a club's catchment area will be of the same order of magnitude. In their study of the travel behaviour of Premier League fans, Forrest, Simmons and Feehan (2002) found that the bulk of attendees resided within 10 miles of the stadium. Accordingly we defined a club's catchment areas by a radial distance of ten miles from its stadium and divided this area into two zones, 0-5 and 5-10 miles from the ground, to ensure rough homogeneity of travel costs from each zone. We measured population in each zone at each club, employing 2001 Census microdata for 175,000 Output Areas, and manipulating them using stadia Ordnance Survey map references and the MapInfo software package.

In the event, home club population within 5-10 miles of the ground proved statistically insignificant (though positive in sign) in our initial estimation of the model and so the model whose results are reported here includes just one population variable for the home club, (the natural logarithm of) the population within 5 miles distance of the ground.

The impact of population density on crowd support will be mitigated to the extent that a club has to share its market with one or more rivals. Dobson and Goddard (1995) sought to evaluate the effect on long-run average attendance at each club by measuring the number of other Football League members within 30 miles of the ground; but this fails to reflect the spatial distribution of those other clubs relative to the subject club. We constructed, again using MapInfo software, an index, termed *market overlap*, to measure the degree of competition faced by each club in a more precise way and this also features as a variable in our model. *Market overlap* is the proportion of the catchment area population that also lies within the catchment area of another club. Where there is more than one neighbouring club, these intersections of population are aggregated and *market overlap* may then exceed one. Indeed it often does and the the highest value amongst the clubs here is 7.62 (for Fulham).

History, as well as geography, may influence the effective size of a club's market. In football, tradition is important and support may build up over time because interest is passed between generations. Older clubs may therefore have a larger following. We include as an additional covariate the duration in years of the club's membership of the Football League. This proved highly significant in Dobson and Goddard (1995).

Variables affecting away support

In contrast to American leagues, distances between clubs in European domestic competitions are small enough to ensure the presence of some, and often a significant number of, travelling supporters. It is standard therefore for European attendance models to include distance between clubs as a proxy for travel costs. It is customarily entered as a quadratic in the expectation that increasing distance will deter away fans' attendance but at a diminishing rate. We also include *distance* and *distance squared* in our specification. But we innovate by including in addition measures of the size of market from which away support will be drawn because it appears illogical to measure a deterrent factor without account also taken on the number of supporters on which the deterrent will work. Away market size is proxied by the same variables as in the case of home clubs: population within 5 miles of the stadium, market overlap and duration of League membership.

Scheduling

The season extends from mid-August to early May. We include a series of categorical variables to represent each individual month from October on (April and May are combined as only a small number of fixtures took place in May). Time of year could influence attendance because of weather conditions and competition from alternative sports and activities; but a particular factor identified in previous work is that interest peaks late in the season as many games become significant in the settlement of promotion, playoff and relegation issues. We also include categorical variables to allow for the effects of scheduling on bank holidays or on midweek evenings (where *midweek* refers to any day from Monday to Friday that is not a bank holiday). Most matches are set for weekends. However, the size of division and the reservation of some Saturdays for cup competition forces some rounds of fixtures to be allocated to midweek and other games are moved to midweek because of bad weather on the original date or because teams were still engaged in the knock-out Cup.

Quality variables

We expect more people to buy tickets for matches when the quality of the two teams is higher. Since the Bosman ruling effectively made players in Europe free agents, the labour market in European football has become competitive and wages for players should therefore reflect talent. Hence we include for both the home and away team the club wage bill for the season as a proxy for the quality of its playing squad. This follows the use of ‘budget’ as a variable in Garcia and Rodriguez (2002). However, we make an adjustment to account for player wage inflation over the long period described by our data. A club’s *relative wage* is its wage bill over a particular season divided by the mean wage bill for the division in that same season. By construction, our wage variable has a mean of one but its range was very wide, from 0.27 to 3.06, reflecting the difference in resources available between clubs aspiring to be in the Premier League and those struggling to avoid relegation to Tier 3.

Of course, players in a squad may work together more or less successfully than the market value of its players’ services might suggest. We therefore include, as additional covariates, actual measures of current season team productivity in the form of the points per game that had been won by the home and by the away team in the current season prior to the match taking place (three League points are awarded for a win and one for a draw). The cardinal measure is preferred to the ordinal measure of League position adopted by some authors.

For given team and player quality, certain matches will attract more public interest than usual. *Derby* is a categorical variable included to identify matches between local or regional rivals. Such games are often played with particular passion and the results may have an importance to supporters independent of their implications for positions in the League.

Television variables

Categorical variables are used to identify matches transmitted live on television. Each channel is treated separately because of the varying audience reach of the terrestial ITV and the satellite and cable outlets, ITV Digital and Sky Sports.

There is also a risk to attendance if television is showing a game from a higher level of football at the same time as a Tier 2 match is taking place. During the study period, Premier League matches that were televised were rescheduled to times not used by the rest of the domestic programme. But European matches in the Champions League were transmitted live on many midweek evenings and these will have been competitive with live attendance at matches included in our sample. We include a covariate which takes a value of one if a match was held on the same evening as terrestial television was relaying a European game featuring an English club. Sky Sports also showed (different) European fixtures but we dropped from our model a variable representing subscription television coverage because it failed to be statistically significant.

The effect of broadcasting some matches on attendance at others is hitherto unexplored in the sports literature (though Paton and Cooke (2005) found suggestive evidence that county cricket attendances were lower when the fixtures clashed with England playing in a test match; all England games were televised live). Whether sports fans actually substitute televised coverage of a, perhaps higher status, match for attendance at a local event has importance beyond soccer. For example, in the United States, NFL games screened nationally are blacked out within a 75 miles radius of any other game taking place at the same time while the NFL is actually prohibited by the

Sports Broadcasting Act from allowing any television coverage at all on Saturdays in case attendance at lower levels of the sport (college) should suffer. No evidence has ever been presented to justify these restrictions (Voluntary Trade Reports (2005)).

Exogenous and endogenous variables

Finally in this section, we explain the allocation of individual covariates to the four vectors defined in the Hausman-Taylor model (equation (4) above).

The fundamentals of geography and history, represented by the population and duration of membership variables, are exogenous. But *market overlap* is treated as endogenous because extra clubs may be spawned where population densities are high. Likewise, team quality and performance variables are endogenous because the resources available to build a squad of players will depend in part on market size. Scheduling variables are treated as exogenous as is coincident television coverage of a European game; but the decision to screen Football League matches themselves is endogenous as it will be influenced by other match characterisics included as covariates of the model.

The cross-sectional unit is a particular club playing home games in a particular season. All home club variables except points per-game are time-invariant. Away club variables are time-varying because each observation in a given group refers to a different visiting team.

5. Results

Table 3 displays results from our estimation. All control variables attract signs and significance consistent with prior expectations. For example, attendances build steadily over the season from December on. Using the formula for marginal effect of a change in categorical variable X from zero to one, $e^{\beta X} - 1$, where β is the estimated coefficient, we find that a bank holiday is associated with a 10.4% boost in attendance relative to a normal weekend. In contrast, a 7.0% contraction may be expected from scheduling midweek. 'Derby' games attract substantial extra interest with a 13.9% rise in the size of crowd in addition to effects from the distance variable taking a low value in such cases. Potential supporters respond readily where home or visiting teams can draw on expensive squads or where teams have performed well through the season.

An advantage from employing the Hausman-Taylor Estimator is that it permits evaluation of the contribution of home club market size to attendance. Coefficient estimates on all three indicators of market size are strongly statistically significant and of a magnitude consistent with the centrality accorded the issue by theoreticians who have analysed the way sports leagues are likely to work. With variables initially set equal to their means, a one standard deviation increase in the size of local population is predicted to increase attendance by 3,842 and a one standard deviation increase in our index of market overlap is expected to diminish attendance by 3,625. These are substantial impacts relative to a mean attendance in the sample of just below 15,000.

The absence of away club market size in previous match level attendance studies is confirmed by the results to constitute an important omission. All three indicators of market size are again significant. The implication is that visiting fans contribute to crowd size and it is insufficient to recognise this merely by the inclusion of a proxy for travel costs.

But our focus of interest is the effect of televising games. All four coefficient estimates on the television variables are strongly significant. The point estimate indicates a negative impact of 23.6% in the case of matches shown on free-to-air television, which exceeds the size of any effect reported in previous studies; but the confidence interval is wide because only a small number of games were transmitted on this platform. Of more interest therefore is the impact from subscription television coverage. We find here much less ambiguous evidence, than in earlier models, of a negative impact from broadcasting. However, while the effect is well determined, the extent of cannibalisation of the live by the televised product is limited in magnitude. Sky Sports was the dominant broadcaster of Football League games over our sample period and is currently the sole provider of live telecasts. In the case of Sky Sports, the point estimate implies an impact of -4.8%. Of course, the negative figure contrasts sharply with the fact that matches shown on Sky had a mean crowd some 17% greater than non-televised games and this illustrates the extent to which Sky selected for screening fixtures which would have attracted large live audiences anyway.

Some supporters therefore appear to substitute home for stadium consumption of a match when the choice is available. There is also evidence in our results that some potential customers switch to home viewing of a higher status match when it is televised in competition with a Tier 2 fixture at the stadium. The estimate of the negative impact of a European game (with English involvement) being shown on terrestial television is 5.2%. In experimentation, no significant influence was felt from European games relayed on subscription television. This is further evidence that, at current levels of penetration by paid-for sports channels, mainstream television coverage is potentially much more damaging to football attendance. We note also that the willingness to stay at home to watch high-level football in preference to going out to watch a more routine game might be be more pronounced amongst lower league tiers than in Tier 2 whose attendances we analysed here.

6. Summary

We have illustrated an approach that permits evaluation of time-invariant but observable club characteristics when analysing pooled cross-sectional time series data on attendances in a sports league. From it, we were able to present estimates of the substantial importance of local population density and competition from other clubs. For these indicators of market size, we innovated by using GIS technology to derive more precise measures than those previously attempted in the sports literature.

The estimator employed also permits more robust estimation of the effects of television coverage than was possible in earlier studies because it allows television coverage itself to be modelled as endogenous. We found strong evidence that broadcast of games on paid-for television channels diminishes attendance at the games shown but only to a limited extent. There was evidence, albeit weaker, of more substantial inroads into the crowd at the stadium if the television medium was free-to-air. We also identified the potential of screenings of games from higher levels of competition in the same sport to detract from attendance.

References

Baltagi, B.H. (2005) *Econometric Analysis of Panel Data*, 3rd ed., Chichester: John Wiley.

Baltagi, B.H. and Khanti-Akom, S. (1990) On efficient estimation with panel data: An empirical comparison of instrumental variables estimators, *Journal of Applied Econometrics*, **5**, 401-406.

Borland, J. and Macdonald, R. (2003) Demand for sport, *Oxford Review of Economic Policy*, **19**, 478-502.

Contoyannis, P. and Rice, N. (2001) The impact of health on wages: Evidence from the British Household Panel Survey, *Empirical Economics*, **26**, 599-622.

Dobson, S.M. and Goddard, J.A. (1995), The demand for professional league football in England and Wales, 1925-92, *The Statistician*, **44**, 259-277.

Dobson, S.M and Goddard, J.A. (2001) *The economics of football*, Cambridge: Cambridge University Press.

Egger, P. and Pfaffermayr, M. (2004) Distance, trade and FDI: A Hausman-Taylor SUR approach, *Journal of Applied Econometrics*, **19**, 227-246.

Forrest, D., Simmons, R. and Feehan, P. (2002), A spatial cross-sectional analysis of the elasticity of demand for soccer, *Scottish Journal of Political Economy*, **49**, 336-355.

Forrest, D., Simmons, R. and Szymanski, S. (2004) Broadcasting, attendance and the inefficiency of cartels, *Review of Industrial Organization*, **24**, 243-265.

Garcia, J. and Rodriguez, P. (2002) The determinants of football match attendance revisited: empirical evidence from the Spanish Football League, *Journal of Sports Economics*, **3**, 18-38.

Hausman, J.A. and Taylor, W.E. (1981) Panel data and unobservable individual effects, *Econometrica*, **49**, 1377-1398.

Paton, D. and Cooke, A. (2005) Attendance at County Cricket: An economic analysis, *Journal of Sports Economics*, **6**, 24-45.

Schmidt, M.B. and Berri, D.J. (2001), Competitive balance and attendance: the case of Major League Baseball, *Journal of Sports Economics*, **2**, 145-167.

Voluntary Trade Reports (2005) *Fumbling the antitrust football: the NFL blackouts of 2004-5*, Report No. 6.

Wooldridge, J.M. (2002) *Econometric analysis of cross-section and panel data*, Cambridge MA: MIT Press.

Table 1. Mean attendances at televised and non-televised matches

	number of games	mean attendance	standard deviation
Non-televised	2,626	14,746	7,168
Televised on ITV	4	17,448	2,884
Televised on ITV Digital	17	19,433	6,817
Televised on Sky Sports	237	17,304	7,585

Table 2. Variables employed in estimation.

	sample mean	standard deviation
Dependent Variable		
Match attendance ❶	14,987.75	7,237.40
Exogenous time-invariant		
Duration in years of home club's League membership	95.33	18.39
Population within 5 miles of home club's stadium❶	442,044	342,016
Endogenous time-invariant		
Market overlap for home club	2.02	2.00
Home club's relative wage	0.999	0.568
Exogenous time-varying		
Derby match❷	0.012	0.110
Distance in miles between the home grounds of the two clubs❸	127.42	70.10
Midweek match (not on television)❷	0.260	0.439
Bank Holiday fixture❷	0.067	0.250
October❷	0.122	0.327
November❷	0.115	0.319
December❷	0.119	0.324
January❷	0.075	0.263
February❷	0.099	0.299
March❷	0.128	0.334
April/ May❷	0.154	0.361
Terrestrial t.v. coverage of European match with English club❷	0.054	0.226
Population within 5 miles of away club's stadium❶	444,144	344,112
Duration in years of away club's League membership	95.42	18.37
Endogenous time-varying		
Market overlap for away club	2.03	2.01
Points per game in season to date (home team)	1.38	0.479
Points per game in season to date (away team)	1.40	0.482
Match shown on ITV❷	0.001	0.037
Match shown on ITV Digital❷	0.006	0.077
Match shown on Sky Sports❷	0.082	0.275

❶ Variable expressed as a natural logarithm in estimation
❷ Categorical variable
❸ Variable also entered in squared form in estimation

Sources: fixture and attendance information collected or derived from the *Rothmans* and *Sky Sports Football Year Books*. Points per game calculated from League tables. Distances obtained from the RAC. Club wage data from editions of the *Deloitte and Touche* (formerly *Deloitte*) *Annual Review of Football Finance*. Population and overlap measures derived from the 2001 Census (see text). Television coverage from various issues of *TV Sports Markets*.

Table 3. Results from Hausman-Taylor Estimation

Dependent variable: natural log of attendance	coefficient	|z|
Exogenous time-invariant		
Duration in years of home club's League membership	0.004	2.22
Natural log of population within 5 miles of home club's stadium	0.401	2.14
Endogenous time-invariant		
Market overlap for home club	-0.159	2.13
Home club's relative wage	0.759	4.14
Exogenous time-varying		
Derby match	0.130	5.10
Distance in miles between the home grounds of the two clubs	-0.002	12.64
Distance squared	0.000005	9.39
Midweek match (not on television)	-0.068	9.06
Bank Holiday fixture	0.099	7.96
October	0.016	1.61
November	0.008	0.79
December	0.032	3.21
January	0.033	2.90
February	0.045	4.29
March	0.057	5.93
April/ May	0.108	11.84
Terrestrial t.v. coverage of European match with English club	-0.051	3.87
Natural log of population within 5 miles of away club's stadium	0.040	5.22
Duration in years of away club's League membership	0.0008	5.56
Endogenous time-varying		
Market overlap for away club	-0.016	6.36
Points per game in season to date (home team)	0.040	4.57
Points per game in season to date (away team)	0.035	6.02
Match shown on ITV	-0.212	2.93
Match shown on ITV Digital	-0.083	2.32
Match shown on Sky Sports	-0.047	4.58
constant	2.96	1.30
number of observations	2,884	
Wald chi-squared (27)	1338.65	