Production functions for sporting teams

Jeff Borland
Department of Economics
University of Melbourne

Introduction

Formal analysis and measurement of organizational production functions has been a significant field of research in recent years. Bartel et al. (2004, p.217) describe the new research methodology of ‘insider econometrics’ that seeks to test the effects of organization-specific determinants of firm productivity following:

“…two broad principles. First it uses field work to generate a detailed understanding of a specific production process, its technology, and the nature of work in a particular industry. This field work in turn provides valuable insights about how to model production in that industry and what data to collect to estimate those models. Second, detailed operating data from the industry are used to estimate econometric productivity models that permit convincing tests of hypotheses about the determinants of productivity.”

While this research approach may be relatively novel within the fields of industrial economics and labour economics, for sports economists it should sound very familiar. Estimation of production functions for sports teams – drawing on both a detailed understanding of the sporting competition, and team and player performance data – now extends over three decades.

In his seminal article on the labour market in professional sporting competitions Rottenberg (1956, p.255) recognised that:

“A baseball team, like any other firm, produces its product by combining factors of production. Consider the two teams engaged in a contest to be collapsed into a single firm, producing as output games, weighted by the revenue derived from admission fees. Let the players of one team be one factor and all others (management, transportation, ballparks, and the players of the other team), another. The quantity of the factor – players – is measured by making the appropriate adjustment for differential quality among players.”

This idea was first implemented empirically by Scully (1974) as part of his comparison between salaries and the marginal revenue product (MRP) of players in Major League Baseball (MLB). The production function estimated uses team season winning percentage in 1968 and 1969 as the ‘output’ variable:

\[
PCTWIN_{it} = \alpha + \beta TSA_{it} + \chi TSW_{it} + \delta NL_{it} + \phi CONT_{it} + \phi OUT_{it} + \varepsilon_{it}
\]
where TSA = team slugging average; TSW = team strikeout-to-walk ratio; NL is a dummy variable for National League; CONT is a dummy variable for pennant and divisional winners in the previous season; and OUT is a dummy variable for teams that at the end of the season were 20 or more games out of placing. Scully also estimates a team revenue equation that specifies revenue as a function of team winning percentage and market size. The two equations are then used to estimate the MRP of individual players, and those MRPs compared with player salaries in order to evaluate the extent of monopsonistic exploitation.

Scully’s paper set the tone for subsequent research on team production functions that has been almost exclusively empirical – estimating production functions for a range of sports, generally with the objective of testing economic theory or describing some aspect of the operation of sporting labour markets. By contrast, theoretical studies of sporting competitions have primarily adopted a simplistic approach assuming that team winning percentage depends on the ‘units of talent’ owned by a team relative to its competitors (see Fort and Quirk, 1995, and Szymanski, 2003 for recent overviews).

**Specification**

A general form of production function for a sporting team might be specified as:

$$ P_i = \gamma(Q_{1i}, \ldots, Q_{Ji}, M_{it}, X_{it}, T_{it}, \ldots, T_{Nit}) ; \quad i = 1, \ldots, I; \ t = 1, \ldots, T $$

Denoting $P$ as the performance or output measure, the model in equation (2) has performance of team $i$ at time $t$ dependent on the quality ($Q$) of the $J$ players in the team, the quality of the manager ($M$), other factors ($X$), and quality of the other $N$ teams in the competition.

To move from this very general specification, to a form of production function that can be empirically estimated, requires several decisions:

*What measure of output?*

Most commonly, studies that estimate team production functions have adopted a season as the time period over which to measure performance, and winning percentage has been used to represent output (see for example Espita-Escuer and Garcia-Cebrian, 2004, Table 1). However, in sports where draws are common using a win percentage measure may give a misleading impression of performance, so that ‘points’ won has been seen as more appropriate (for example, Schofield, 1988, and Dawson et al., 2000a). Some other studies have used alternative measures of team output such as attendance (Gustafson et al., 1999), revenue from TV broadcasts (Hausman and Leonard, 1997), and number of players drafted from a college team to the major league (Brown, 1994); or used a different time unit by analysing performance of a team in individual matches (Carmichael et al. 2000).

Most studies do not provide any explicit motivation for the choice of team output measure. In practice the choice seems to involve an implicit judgement about the objective of sporting teams, and the purpose of the study. Performance measures such as winning percentage or points won can be seen as either directly relevant to a team’s objective (win-maximizer hypothesis), or indirectly relevant as one factor influencing...
team revenue and profits (profit-maximizer hypothesis). But where a study has some explicit purpose, such as measuring determinants of TV revenue, then it may be most appropriate to use that variable directly as a measure of team output.

*What inputs?*

Deciding how to represent player performance is the common starting point for estimating a team production function. One approach is to include measures of player performance that are regarded as important explanatory variables for team performance. For example, in the Scully (1974) study measures relating to team batting and team pitching performance are included as inputs. The alternative approach is to include as inputs measures of player ability. For example, in their study of English soccer, Dawson et al., (2000b) include measures such as player career league experiences, player age, and goals scored by players in the previous season. Of course it is also possible to have a two-stage approach that treats player ability as an input to player performance, and player performance as the determinant of team performance (see Carmichael and Thomas, 1995, for application of this type of approach to rugby league in England). As well, other factors apart from ‘exogenous’ player ability may affect player performance – for example, Krautmann (1990) studies how the time to next contract negotiation may affect player performance in MLB.

The other main issue is the range of inputs to be included in the production function. Apart from player quality, the main input that many studies have sought to include has been managerial or coach quality. In some studies this has involved developing measures of managerial quality to incorporate together with player performance or quality measures into the team production function (see for example, Pfeffer and Davis-Blake, 1986 and Kahn, 1993); in other studies a fixed effect is used to represent a manager, with the effect on team performance being identified by managers changing between teams (for example, Borland and Lye, 1996, and Dawson et al., 2000a). Where both player and manager or coach quality measures are used as inputs, one difficulty that arises is disentangling between direct and indirect effects of the manager. For example, managers might affect team performance directly by the way that they are able to combine a set of players of given quality into a team performance such as through their role in determining team tactics in a match; but there may also be indirect effects of a manager on team performance such as through the quality of players at a team due to the manager’s role in recruiting, or on player performance through the role of a coach as motivator. A range of studies have found evidence that managerial quality and experience is positively related to team and player performance (for example, Kahn, 1993, and Singell, 1993); although managerial efficiency certainly does not have a perfect correlation with team performance (for example, Dawson et al., 2000a).

Generally studies have not included the quality of opposing teams as an input to team performance. This may be appropriate where the output measure is seasonal winning percentage and teams play each other the same number of times during a season. But where the output measure was for example at the match-level, or teams have different playing rosters through a season, then controlling for the quality of teams played would seem to be important.
What functional form?

Determining the appropriate functional form of a team production function requires a judgment to be made about the way that inputs combine to produce team output, and how increases in the ‘quantity’ of inputs will affect team output.

As Scully (1995, p. 64) notes “Players interact with one another in team sports. The degree of interaction among player skills determines the nature of the production function”. At one extreme it is possible to envisage a sporting competition where the output of individual players contributes additively to team output. Here there is no effect of interaction between players on team output; instead, output of individual players is perfectly substitutable. (This does not necessarily mean that individual players will be perfect substitutes. With a fixed number of players in a team there will still be a preference for higher ability players.) In sports such as baseball and cricket it has been argued that there is sufficient separability between the activities of hitting and pitching, and batting and bowling, for an appropriate representation of the team production function in those to be additive in the separate activities. At the other extreme would be a sporting competition where outputs of individual players are perfect complements in contributing to team output. That is, in a sporting competition where for example positional specialisation means that a team is ‘only as good as its weakest link’, a Leontief-type production function would be appropriate. Sports such as American and Australian football, that have a high degree of specialisation of tasks, and involve substantial interaction between players, would probably correspond most closely (although not exactly) to this case.

In general it will not be possible to increase the ‘quantity’ of inputs in a team production function by adding extra players. Instead the idea behind increasing inputs would be that the quality of inputs might become higher. (Although it may be possible to increase the pool of players that a team can draw on by increasing the roster size, or to increase the size of coaching staff.) How an increase in the quality of an individual input will affect team output depends on the degree of substitutability of inputs, but as well, it likely to depend on the nature of the competition and the output measure. For example, where output is measured using team winning percentage it might be reasonable to think that there would be decreasing returns to player quality since a bigger winning margin makes no difference to winning percentage; but where team output was measured by goal difference then it may be more sensible to represent constant returns to player quality.

The main approaches in empirical studies of team production functions have been to estimate fairly simple models that specify either a linear or log-linear relation between output and inputs. Linear models (for example, Scully, 1974) implicitly assume additive separability of inputs. Log-linear models (for example, Gustafson et al., 1999) assume that individual inputs interact multiplicatively to determine team output. A more ‘structural’ approach is taken in the study by Atkinson et al. (1988) of American football (National Football League) where specific interaction terms between inputs are chosen on the basis of expected interdependencies between positions. And in a study of team production functions for cricket in Australia and New Zealand (Bairam et al., 1999) a more general functional form – the CES production function – is estimated. This approach allows the degree of
substitutability between batting and bowling performances in determining team performance to be assessed.

Types of studies

Empirical analyses of team production functions in sporting competitions have predominantly studied Major League Baseball in the United States, and soccer in Europe. In part this may be due to (especially for baseball) the perception that the appropriate model of team output is relatively simple and hence easily amenable to econometric analysis; and as well the availability of detailed data on individual player and team performance. However, these sports also account for the bulk of empirical analysis on other topics such as demand for attendance at sporting competitions (see Borland and Macdonald, 2003) so that it seems an explanation must also be the general interest in and importance of baseball and soccer. Other sports where team production functions have been studied are cricket, American professional and college football, rugby league in England, and basketball in the United States.

There appear to have been a variety of objectives in studies that have estimated team production functions. The common link between most studies however is the use of sporting competitions as a laboratory for testing economic theories (Kahn, 2000). The original study that estimated team production functions in MLB by Scully (1974), and a range of subsequent research, have sought to use sports labour markets to test the extent of monopsonistic exploitation in a type of market where there is generally a single employer (or set of employers) in the highest quality competition, and significant restrictions on mobility of labour. While Scully’s study did find evidence of monopsonistic exploitation, later studies that have examined time periods with different wage bargaining institutions in MLB, or used alternative methods for imputing the ‘value’ of a player to a team, have not found evidence of exploitation (for example, Marburger and Reynolds, 1994). More recently, several studies have used estimates of sporting team production functions to assess managerial efficiency (for example, Dawson et al., 2000a). And team production functions have also been used as the basis for tests of whether ‘CEO’ succession adversely affects organizational performance (Pfeffer and Davis-Blake, 1986), whether turnover of coaches is consistent with labour market theories of matching (Borland and Lye, 1996), and incentive effects on player performance (Krautmann, 1990).

Empirical methodology

The approach for estimating team production functions most often applied has been to use the OLS method together with either of the following models:

\[(3a) \quad WPCT_{it} = \alpha + \beta X_{it} + \varepsilon_{it}; \quad i = 1,\ldots,I; \quad t = 1,\ldots,T\]

\[(3b) \quad \ln(WPCT_{it}) = \alpha + \beta \ln(X_{it}) + \varepsilon_{it}; \quad i = 1,\ldots,I; \quad t = 1,\ldots,T\]

where \( WPCT_{it} \) and \( X_{it} \) are respectively winning percentage (or alternative output measure) and a vector of input variables for team \( i \) in season \( t \). This model can be supplemented with fixed effects for team or season, although especially in early
studies, this was rarely done. The OLS approach provides estimates of the average effect of the set of input variables on the team performance measure.

The main alternative estimation approach, applied in particular in studies that have sought to estimate managerial inefficiency, is stochastic frontier analysis (see Dawson et al., 2000a, for a detailed overview). This approach estimates:

\[
WPC_{it} = \gamma + \phi X_{it} + (\delta_{it} - \lambda_{it}); \quad i = 1, \ldots, I; \ t = 1, \ldots, T
\]

where \( \delta_{it} \) and \( \lambda_{it} \) are respectively a two-sided error term representing the effect of ‘noise’ on team output and a one-sided error term representing inefficiency effects. Stochastic frontier models can be estimated either using a maximum likelihood method on cross-section data with an assumption on the distribution of the inefficiency error term, or using panel data that allows the inefficiency effect to be identified from repeated observations on the same manager.

The major econometric issues to be concerned about in interpreting findings from studies of team production functions would seem to be omitted variable bias, mis-specification of the production function, and selection effects. Knowing all relevant inputs to team output, and finding appropriate measures for those inputs, is likely to be a difficult exercise in most sporting competitions. However, not doing this may cause biased estimates. As a simple example, where managerial quality and player quality are likely to be correlated, then measuring one of those variables with error will cause upward-biased estimates of the effect of the other variable. As soon as some degree of inter-dependence between players in a team exists in how their individual performance will affect team output, a significant degree of complexity is introduced to modelling the team production function. But again, not doing this may introduce bias into estimates of how those individual inputs affect team output. Thus far the literature on team production functions appears to assume that all teams in a competition will have the same production function. But it does not seem unreasonable to think that different teams in the same competition could choose different ways to combine inputs, and that they would do this in a way that would cause selection effects to arise – such as according to their comparative advantage in following particular playing strategies.

References


