The Empirical Analysis of Multiproduct Pricing Using Principal Components: An Application to Major League Baseball[1]

Craig A. Depken, II
Department of Economics
University of Texas at Arlington
Arlington, TX 76013
Office: 817-272-3290
Fax: 817-272-3145
depken@uta.edu

Darren P. Grant
Department of Economics
University of Texas at Arlington
Arlington, TX 76013
Office: 817-272-3286
Fax: 817-272-3145
dgrant@uta.edu

[1] The authors acknowledge the helpful comments of Bill Crowder, Jahn Hakes, Courtney LaFountain, Steve Shmanske, Mike Ward, and seminar participants at the 2004 Western Economic Association Meetings.

The Empirical Analysis of Multiproduct Pricing Using Principal Components: An Application to Major League Baseball

Abstract

The theory of multiproduct pricing is well developed in stylized models, although a unified theory has yet to be developed. As such, empirical analyses are rarely guided by strong theoretical hypotheses and are therefore scarce. This paper analyzes the interactions of ticket, parking, and concession prices in Major League Baseball for the period 1991-2001 using a principal components methodology. The approach allows inferences to be formed about the factors underlying intertemporal price variation in the absence of information about costs and demand. The most important factor influencing prices in baseball is a general increase in the demand for baseball, but general demand shifts explain less than half of all price variation. The second most important factor is complementarity between required and voluntary purchases. The third most important factor is pricing interactions between frequently and infrequently purchased concessions that are consistent with theories of nonlinear multiproduct pricing. Secondary empirical analysis confirms these economic interpretations. The results show that the principal components methodology is an effective way to draw inferences about the economic forces underpinning pricing in a multiproduct context using data on prices alone.

JEL Classifications: D40, L11, L13

Keywords: multiproduct pricing, sports, structural model, complements, substitutes, principal components.

1. Introduction

Analyzing the pricing decisions of a multiproduct firm presents a daunting challenge for the researcher. To date the number of theoretical and empirical analyses seems small given the almost ubiquitous nature of the problem in practical business decisions. Rather than the canonical approach of independent demands, most multiproduct firms enjoy considerable complementarity and substitutability across the products they sell. Moreover, the ability to effectively price discriminate (or bundle) increases in a multiproduct environment. However, theoretical analyses of multiproduct pricing are either empirically intractable or rather stylized in order to keep the problem tractable. As a result, empirical analyses often suffer from a lack of strong theoretical guidance about the interaction of prices.

This fact alone complicates the empirical analysis of multiproduct pricing with a structural model, in which information about production costs and all own and cross-price elasticities for all goods is used to predict optimum pricing. Other practical issues also intercede. Data limitations often render a structural approach infeasible. And the results of a structural model, in which prices are a function of many costs and demand elasticities, might not be meaningful in that they provide “too much” information: one would lose the forest for the trees.

The questions we address in this paper are simple, as is the method we introduce for answering them. What factors, or economic forces, explain the variation in the prices changed by multiproduct monopolists? Are these forces consistent with those stressed in theories of multiproduct pricing? The method we present, which relies on principal components analysis, allows us to identify, or “visualize,” the forces underlying price variation. Then, taking advantage of the unique interpretation of each independent principal component, we can then heuristically ascribe a significant fraction of the variation in prices to a small number of underlying economic forces.

The use of principal components has three advantages. First, the technique has minimal data requirements, and therefore can be implemented in most cases. Second, the technique plausibly identifies the underlying independent sources of price variation in a meaningful and quantitative way that need not be specified *a priori*. Finally, further analysis of the principal components makes it possible to verify economic interpretations and to better understand the results.

After discussing theoretical models of multiproduct pricing, the pricing decisions of Major League Baseball (MLB) teams from 1991 through 2001 are analyzed using the methodology we describe. Most teams operate in a geographically isolated market, and thus are local monopolies, but all teams sell multiple products including tickets, parking, and concessions. Much of the overall variation in the prices of these products is explained by a small number of principal components, each of which can be interpreted in terms of fundamental economic theory. Thus we provide concrete, though circumstantial, evidence that the forces stressed in theories of multiproduct pricing are in fact the most important determinants of price variation in this market.

The most important influence on prices in professional baseball is a general demand effect. The second most important influence is a tradeoff between the prices of obligatory purchases—tickets and parking—and discretionary concessions. The third largest influence on prices is a tradeoff between non-food items (such as tickets and programs) and food concessions, perhaps more appropriately characterized as single-purchase and multi-purchase items. These latter two influences are consistent with theories of multiproduct pricing that stress demand interactions across products and nonlinear pricing in order to maximize the capture of consumer surplus. Therefore, it seems that prices at MLB games respond most strongly to overall demand effects, but that price adjustment consistent with theories of multiproduct pricing is also an important factor. In fact, less than one half of the overall variation in prices can be attributed to a general demand effect, which indicates that greater focus on the relationship between the various prices charged at professional sporting events is warranted.

The remainder of this paper is structured as follows. The next section discusses existing theoretical models of multiproduct pricing decisions, highlighting the difficulties involved in structuring econometric analyses around existing theory. Section 3 presents the existing literature on the pricing decisions of Major League Baseball teams and describes three fundamental economic influences on multiproduct pricing in this market. Section 4 describes how principal components analysis can be used in the context of multiproduct pricing choices. Section 5 describes the data and the empirical results. The final section provides concluding comments.

2. Multiproduct Pricing: Theory and Practice

The prices chosen by a multiproduct firm depend on the firm's costs, the own and cross-elasticities of demand for the firm's products, whether two-part tariffs, bundling, or price discrimination are feasible, and the type of competition present in the market. The complete theoretical solution for nonlinear monopoly prices was determined by Mirrlees (1976) and for competitive linear pricing by Bliss (1988). However, these solutions, which decompose prices into complex combinations of elasticities and costs, are, in the words of Sibley and Srinagesh (1997), "rather opaque as to intuitive content" and, in the words of Bliss, "difficult to apply empirically." Thus significant simplifications or approximations have been necessary to estimate structural pricing models. Reibstein and Gatignon (1984) estimate a structural model of grocery markets, but for only five products in a single product line, eggs, and without instrumenting price. Nonetheless, they find that "cross-elasticities have a substantial role in explaining the sales of various types of eggs." In studies of the automobile market, Bresnahan (1987) assumed that products were vertically differentiated, while Berry, Levinsohn, and Pakes (1995) utilize a Lancasterian demand framework and a "strong assumption on the orthogonality of observed and unobserved product characteristics." More recently, Guilietti and Waterson's (1997) study of grocery store pricing utilizes a linear expenditure demand system in which 31 products are aggregated into seven categories to permit estimation. This problem is understandable: precisely estimating the $31^2 = 961$ demand elasticities in a full structural model of pricing would require a very long time series of price, sales, and costs for each product.

An alternative to full structural estimation is to use recent theoretical work on multiproduct pricing that has uncovered conditions under which optimal or near-optimal tariffs can be characterized simply. If the appropriate conditions apply, estimation could be structured around the postulated relationship, which could then be tested statistically. A number of theoretical studies exist that could be relied upon for this purpose. Armstrong (1999) finds that cost-based two part tariffs are nearly optimal when goods are independent in demand (neither substitutes nor complements) and the number of goods is large. Armstrong and Vickers (2001) and Rochet and Stole (2002) find that simple two-part tariffs pertain in duopoly when there is sufficient competition and certain other conditions apply.[2] Sibley and Srinagesh (1997) show that the "optimal nonlinear price schedule can be computed by finding optimal price schedules separately for each market" when preferences satisfy a strong condition called the "uniform ordering of demand curves."[3] Forbes (1988)

[2] For example, Armstrong and Vickers restrict the privately-known horizontal taste parameter to be distributed independently of the vertical taste parameter.

[3] This condition implies that there is an ordering of utility functions across consumers that is independent of prices. That is, for all goods, the quantity purchased by consumers of type 1 at any price exceeds that for consumers of type 2, which exceeds that for consumers of type 3, etc.

illustrates the relation between the monopoly linear prices of substitutes and complements for the two good case, showing that the prices of substitutes will tend to move in the same direction and prices of complements in opposite directions. Finally, Bliss (1988) finds that competitive retailing margins will be a constant percentage across all goods when consumers are "fixed budget shoppers." Unfortunately, these conditions are all fairly restrictive: pricing heuristics have not been formally developed for many realistic multiproduct pricing situations, such as those analyzed herein: choosing seven "quasi-linear" prices in a monopoly. Furthermore, Sibley and Srinagesh (1997), like Spence (1980) before them, show that optimality in multiproduct pricing can lead to a "variety of outcomes." Thus it would be rash to *impose* a simple pricing relationship in advance.

This state of affairs complicates the empirical analysis of multi-product pricing because it generally cannot be structured around simple relationships that can be imposed *a priori*. This motivates the third option: a heuristic approach. Even in the absence of strong theoretical guidance, the general forces that lead to relationships between the prices charged by a multiproduct monopolist can be identified and categorized. This exercise yields predictions about interrelationships between prices that do not depend on explicit knowledge of demand elasticities and costs, and which can be adapted to distinctive features of the market if necessary. Empirical implementation employs data reduction techniques that decompose the variation in prices into independent parts, or components, yet do not require a theoretical structure to be imposed *a priori*. If the economic forces under consideration are the primary sources of covariation in prices and are sufficiently independent in their operation, they are likely to be uncovered and quantified in this empirical analysis: the components yielded by the empirical analysis will be interpretable in terms of the economic forces identified in the theoretical discussion.

The well-developed principal component methodology described below is especially applicable to this problem. This methodology allows the prices of multiple products to be expressed in terms of a small number of components, each of which (in our case) admits a straightforward economic interpretation. Hypotheses concerning these components can be formed and tested given data about the market, and the fraction of the variance of any individual price attributable to any component can be estimated. Moreover, these insights can be related to theory *ex post* without depending on economic theory to provide the "correct" analytical structure *ex ante*. Altogether, this technique allows many insights about multiproduct pricing to be uncovered.

3. Pricing in Professional Baseball: An Industry Case Study of Multiproduct Pricing

A. Studies of Pricing in Major League Baseball

Major League Baseball presents a classic example of the multiproduct pricing problem. The game "package" purchased by most fans includes a combination of tickets, parking, food concessions, and other concessions such as programs and ball caps. These products are relatively homogenous across firms within the industry, facilitating empirical analysis, and most ball clubs are relatively isolated local monopolies. (The eight teams that play in the same metropolitan area, such as the New York Yankees and the New York Mets, have distinctly different fan bases (see Depken, 2000) and have significant monopoly power.) Furthermore, attendance serves as a clear measure of demand.

The *Team Marketing Report* uses ticket and concession prices to calculate a Fan Cost Index (FCI) that reflects the expenses incurred by a hypothetical family of four that attends a game, parks at the stadium, and consumes a typical mix of tickets and concessions. For 2004, the league average FCI was $155.52, of which ticket prices, at $78.98, accounted for barely more than half. Thus expenditures on concessions are quantitatively important.

Nonetheless, a comprehensive analysis of the relationship between the various prices in MLB (or any other sporting league) has yet to be undertaken. Previous studies of pricing have primarily focused on whether ticket prices are set consistent with profit maximization (for example Ferguson, et al., 1991). A primary point of contention, given the assumed near-zero marginal cost of seating additional fans, is whether prices are set at the revenue maximizing level, i.e., at unitary elasticity of demand (see Scully, 1989, pp. 111-113, and Zimbalist, 1992, p. 214).

A second focus is the relationship between ticket prices and attendance. Here, too, most attendance studies include some measure of ticket price but fail to include any other measures of the additional costs of attendance (for example Depken, 2001, Marburger, 1997, Scully, 1989 and Zimbalist, 1992).[4] Absent these measures, both sets of studies potentially suffer from omitted variable bias. The direction of the bias cannot be predicted in advance without a better

---

[4] Notable exceptions are Welki and Zlatoper (1994), who include the cost of parking in a study of NFL demand, Depken (2000) who includes average concession expenditures in a study of baseball demand, and Winfree et al. (2001), who proxy for the total cost of attending a game using a travel time measure. None of these measures, however, would be expected to completely reflect the full costs of attending a sporting event.

understanding of the relationship between ticket prices, parking prices, and concession prices. Thus an analysis of multiproduct pricing in MLB has the potential to significantly advance this strand of the sports economics literature in addition to its application to industrial organization.

B. Multiproduct Pricing in Major League Baseball

While the above-mentioned features of MLB make it suitable for analysis from an empirical perspective, other features of the market make it suitable for analysis from a theoretical perspective. In particular, economic theory has identified three general reasons that the prices of goods sold by a multiproduct monopolist would be related, and each of the three is well represented in MLB.

The first possible source of price correlations is a general change in cost or demand, stemming perhaps from a local increase in wage rates or from a surge in team popularity. This will exert price pressure in the same direction for all goods. Because of the relatively low marginal cost of tickets and many concessions and the high intertemporal variance in team popularity as their on-field success waxes and wanes, demand influences are likely to be fundamental: more popular teams will have greater demand and will be able to charge higher prices for tickets and for concessions. This would cause ticket and concession prices to be positively related.

However, demand interrelationships between tickets and concessions are also likely to be relevant. In particular, concessions enhance the game-viewing experience, and therefore tickets and concessions are likely to be complements. Forbes (1988) graphically illustrates how the prices of complements respond to changes in costs or demand in the two-good case. If the demand for each product increases or the production cost of each product increases, then both products' prices are likely to increase, as discussed above. An increase in the cost of or demand for just one of the products, however, will lead to a decrease in the price of the other. If we think of tickets and concessions as two composite, complementary goods, this economic force would lead to a negative correlation between the prices of these two composite goods.

Finally, product prices can be related because of nonlinear pricing—second degree price discrimination. Studies of nonlinear pricing emphasize bundling and menu pricing, in which the consumer can pay a fixed "entry fee" in order to purchase some range of quantities at a price below "list." For MLB, we do not observe product bundles (such as discounted "family fun packs") in the data. However, nonlinear pricing can generate price interactions across tickets and concessions because the entry fee can be factored into the ticket price and the prices of concessions correspondingly adjusted. In practice, most stadiums offer a range of seating options and ticket prices, and some teams may be more able than others to extract consumer surplus through ticket prices–perhaps because they use innovative marketing strategies or because the stadium permits a wider range of seating options. These teams might find it optimal to have higher average ticket prices and lower concession prices in consequence.

Once again, this leads to a negative relation between the price of tickets and the price of concessions, but with a twist: nonlinear pricing is operative only for multiple-purchase concessions such as food, not for single-purchase products such as programs. Indeed, prices for these latter products may increase if they are income elastic, as is likely the case. High-income fans will tend to purchase more food concessions and thus will gain more surplus if food prices are reduced. This surplus will not be fully extracted in ticket prices if teams need to satisfy the budgets of low-income consumers. Therefore, some of high-income consumers' surplus can be extracted by raising the prices of single-purchase concessions.[5] Non-linear pricing can lead to a negative relation between the prices of single-purchase items, including tickets and programs, and multiple-purchase food concessions.

In summary, pricing in MLB is governed by the general state of demand, demand interrelationships between goods, and second-degree price discrimination. Each yields different predictions for the relationships between prices. Interpretations of the principal components below will rely on these differences.

4. The Principal Components Analysis of Multiproduct Pricing

Traditionally in econometrics principal components are used to remove multicollinearity in a regression: the independent variables are transformed into a set of principal components, a subset of those (with low variance) are dropped, the regression is estimated, and estimates of the original regression coefficients are obtained utilizing the transformation matrix from which the principal components were originally obtained. *In this paper principal components are not used in this "traditional" way.*

[5] A formal model confirming this possibility is available from the authors.

The utility of the principal components methodology lies in its ability to reduce the dimensionality of data–to describe the movement of many variables in terms of a small number of independent underlying patterns that can often be interpreted in terms of simple heuristics. This can be useful in a variety of contexts, and principal components have been employed in many natural science and social science applications as a data reduction technique for "untangling complex patterns of association in multivariate data" (Green, 1978) independent of traditional regression analysis. Development and discussion of this technique can be found in Green (1978) and Johnson and Wichern (1982).

For example, Ahamad (1967) uses principal component analysis to investigate the relationship between eighteen crimes, from homicide to sodomy, in Scotland from 1950-1963 and "determine to what extent the variation in the number of crimes from year to year may be explained by a small number of unrelated factors." A single component associated with population change explained about 90% of the variance in these crimes. Nezlin and McWilliams (2003) use principal components to identify the effects of El Nino on sea surface temperature and sea surface height in the California Current in the Pacific Ocean during 1997 and 1998 and to determine whether these effects were due to "propagating coastal waves" or altered wind patterns, finding in favor of the former. Quant et al. (2003) employ principal components to discern the relationships between the concentrations of ten air pollutants and daily mortality rates in Holland. Five components explained almost all of the variation of ten major pollutants, and two of these, industrial pollution and "photochemical transformations," had strong associations with mortality.

The first – and, to our knowledge, only – use of principal components in this way in economics is Doll and Chin (1970). They had 19 years of annual market prices of shrimp at the retail, wholesale, and "ex-vessel" (the price received by the shrimper) level, and analyzed them using principal components. Not surprisingly, these prices moved very similarly across time; 96% of their joint intertemporal variation was explained by (essentially) the annual average of the three prices–the first component. Analysis of the second component, however, revealed "a lag between retail and ex-vessel prices does exist but occurs only when ex-vessel prices suffer a severe downward break." Our use of principal components mirrors Doll and Chin's, except that we apply it to firm prices, not market prices, and attempt to relate our findings not to competitive theory, but to the theory of multiproduct pricing.

Our empirical analysis of multiproduct pricing in Major League Baseball will generate three sets of results. First, we identify a small number of principal components that explain much, but not all, of the overall variation in the prices of tickets, parking, and concessions. These components admit straightforward interpretations in terms of the economic forces described in the previous section. We then present a decomposition of the variance in the price of each good into portions attributable to each principal component. Finally, we estimate regressions that relate the principal components to supply and demand factors. These help confirm the interpretations given to the components and offer additional insight into the forces driving price variation over our sample period.

The empirical analysis proceeds as follows. Consider a series of several prices, arranged in an [N x K] matrix P, where N is the number of observations and K is the number of variables included in the sample. Next, compute the eigenvalues of P^TP and place them in a diagonal matrix λ with associated orthonormal eigenvectors, E. These matrices satisfy the following equation:

$$E^TP^TPE = \lambda.$$

The eigenvalues and associated eigenvectors are conventionally put in descending order. The sum of the eigenvalues equals the sum of the variances of the prices in P.

Now, consider the transformed matrix Z = PE, with $Z^TZ=\lambda$. The i^{th} column of Z, Z_i, contains a series formed by multiplying the elements of P by the i^{th} eigenvector, leading to a transformed series that has variance λ_i, the i^{th} diagonal element of λ. This column Z_i is called the i^{th} principal component. Because the eigenvectors are orthogonal each principal component is uncorrelated with all others. The values of the i^{th} eigenvector, called factor loadings, can be examined to lend an interpretation to that principal component. If, for example, they are all similar in magnitude and equal in sign, the component might be interpreted as a simple period-by-period average of the prices in P, as in the first component of Doll and Chin (1970).

Any component can be treated as a variable worth explaining in its own right, if desired. In our application we do this by relating selected components to supply and demand factors and contextual information, contained in a matrix X. To do this for the component Z_i, conduct the following regression:

$$Z_i = \alpha_0 + \beta X + \varepsilon. \qquad \textbf{(1)}$$

Hypothesis tests on β can be formulated and tested to confirm the interpretation of the principal component.

Finally, a "reverse regression" can be used to explain the proportion of variance in the i^{th} price, P_i, attributable to any component. Because the components are statistically independent, the R^2 value in the following regression equals the fraction of variance in price *i* attributable to component *j*:

$$P_i = \alpha_1 + \gamma Z_j + v, \qquad \textbf{(2)}$$

where γ is a vector of coefficients.

The utility of the principal components technique relies on its ability to uncover relations between prices that have plausible interpretations in terms of economic theory and which can be confirmed with supplementary hypothesis tests that can be conducted given data about the market. Our discussion of MLB has identified three factors--overall demand, demand interrelationships across products, and second degree price discrimination--that influence pricing. For our principal components analysis of MLB to succeed, each must contribute independently to the variation in prices, and the amount of price variation explained by these factors must dominate the variation explained by other, idiosyncratic factors that are beyond the reach of theory. While we cannot be sure of this in advance, we argue that these assumptions are at least plausible, at least to the first order: there is no obvious, strong connection between the three economic factors of primary interest. Further confirmation of these assumptions, however, awaits empirical analysis. If the principal components are interpretable in terms of economic theory and satisfy supplementary hypothesis tests, we can have confidence in their validity despite the atheoretical nature of the methodology.

5. Pricing in Major League Baseball: Empirical Results

In our case study of multiproduct pricing we employ prices from Major League Baseball. The overall variation in the real prices of seven relatively homogenous products is decomposed into seven independent principal components. We heuristically investigate these principal components in the context of the prevailing theories of multiproduct pricing described above in Section 2. The three most important components correspond closely with existing theory and allow us to determine the magnitude that each influence contributes to the overall variation of each real price series. Subsequent regression analysis of the form of equation (1) attempts to "confirm" these interpretations. The analysis advances the understanding of pricing in Major League Baseball and demonstrates how the principal components methodology advances the understanding of multiproduct pricing.

A. Principal Component Analysis

To operationalize the principal components analysis, data describing the prices charged by all Major League Baseball teams, as reported by the *Team Marketing Report*, from 1991-2001 are employed. Over this period, baseball expanded by two teams in 1994 (Denver and Miami) and by two more teams in 1998 (Phoenix and Tampa Bay), and also dramatically realigned the divisions within the American and National Leagues. Over the same time period, each league expanded the post-season playoffs to include an additional wild card team in 1995 and inter-league play was introduced in 1997. Therefore, the price data describe relatively homogenous products across firms within an industry that has continued to evolve even while the individual firms have remained relatively isolated local monopolies.

The *Team Marketing Report* reports the average per-game season ticket price, the price of official stadium parking, and the prices of beer, soda, hotdogs, ballcaps and programs. Beer and soda prices are reported for different size drinks and are therefore normalized to 20 ounces. Finally, all prices are converted to 2000 dollars using the Consumer Price Index reported by the Bureau of Labor Statistics.

The prices for each good are reported for all teams at the beginning of the season and therefore promotional price changes are not included in the sample. While these changes might have a short-run impact on attendance, i.e., for a particular game, these data limitations are considered acceptable given the other desirable properties of the data. Primarily, the advantage of these data is that they describe all the teams in the (U.S.) professional baseball industry, the products the teams offer are fairly homogenous, and the teams are, for the most part, geographically separated, thereby limiting strategic pricing. These features make the data conducive to testing hypotheses about multiproduct pricing.

Table 1a presents the descriptive statistics for the prices used in the sample. The upper panel presents the prices used in the principal components analysis; the lower panel reports the variables (described in more detail below) used in the secondary regression analysis. The real price of tickets averages \$13.61 over the sample period, whereas the average real price of parking was \$7.18. Prices within the stadium averaged \$5.21 for a 20oz beer, \$2.57 for a 20oz soda, \$2.23 for a hotdog, \$3.57 for a program, and \$12.04 for a ball cap. Of these prices, the greatest variance was displayed in ticket prices, which is not surprising given the different local market and stadium characteristics across teams, and the smallest variance was in the price of hotdogs.

Table 1b reports the correlation matrix of the real ticket prices. As can be seen, the correlation between the prices of any two goods in the sample is generally positive, but never greater than 0.50. Prices are not so uncorrelated that the goods can be (unrealistically) viewed as having independent demands, nor are they so correlated that they can be viewed as a single "composite good." The positive correlations suggest that the dominant influence on price is fluctuations in the general demand for baseball, but the multiproduct pricing considerations discussed above, which tend to introduce negative correlations between prices, cannot be ruled out given the modest magnitude of these correlations.

Before calculating the principal components of the matrix of real prices, we normalize the data as much as possible. Principal component analysis can be sensitive to the scaling of the variables, especially in circumstances like ours, in which some prices are much more variable than others and there are significant scale differences across prices. As our main interest is to better understand the interplay of prices, principal components are calculated for the correlation matrix of the prices, not the covariance matrix, implicitly scaling each price by its standard deviation. This scaling is not uncommon in principal component analyses of this type.

Table 2a and Table 2b present the results of the principal component analysis. Table 2a reports the seven eigenvalues (elements of λ) that solve the characteristic root and the proportion and cumulative amount of overall variation associated with each eigenvalue. As can be seen, the first eigenvalue accounts for approximately 33% of the total variation in real prices, with the subsequent two eigenvalues accounting for approximately 16% each. The remaining four eigenvalues account for 35% of the overall variation in real prices. Thus the first three components account for nearly two-thirds of all price variation.

Although there are seven principal components, common practice is to focus attention on those that are most important. Alternative methods have been suggested to determine which factors should be of focus. One popular test is the so-called Scree Test, in which the eigenvalues of the extracted factors are plotted and the number of factors analyzed is determined by when the consecutive gain in explained variance approaches zero or the Scree plot flattens out. This analysis suggests the first four factors merit subsequent analysis. Another criterion is to focus on those eigenvalues greater than one. From Table 2a, the first three eigenvalues have a value greater than one. Finally, one can focus on those components that admit a natural (in our case, economic) interpretation. As discussed shortly, this is possible also for the first three components. The weight of the evidence supports a focus on the first three principal components.

Table 2b reports the estimated eigenvectors (rows of the matrix E), the elements of which represent weights that are placed on each price in calculating each principal component. For example, the first component would be calculated as 0.51xRPTIX + 0.33xRPARK + 0.45xRPBEER + 0.47xRPSODA + 0.43xRPHOTDOG + 0.08xRPPROGRAM + 0.08xRPCAP. Because any individual price series is a linear combination of the principal components (specifically, $P = ZE^T$), these factors represent independent contributions to the overall variation of ticket, parking, and concession prices over the period investigated. The first three eigenvectors can be interpreted in the context of existing economic theory of multiproduct pricing.

In generating the first component all seven prices enter in a similar qualitative fashion: the weights have the same sign. This suggests that the predominant source of price variation is an overall demand effect. Given the increased popularity of MLB, notwithstanding the momentary decrease in attendance after the 1994 players' strike, a general increase in demand seems a natural interpretation of this factor.

The second component is hard to interpret in full generality in terms of the theory, as the theory does not yield simple heuristics for a system of seven prices. However a natural interpretation is permitted if we temporarily group goods into "obligatory" purchases, i.e., tickets and parking, and "discretionary" purchases, i.e., concessions. The second component, then, is generated by differencing the prices of the obligatory goods from those of the discretionary goods. Forbes' (1988) analysis indicates that if these two "composite" goods are complements, which is likely, any increase in costs or demand for one good will raise its price and lower the price of the complement. Thus we can consider this second component as reflecting the multiproduct pricing interplay between complementary goods. The second principal component contributes only 16%

to the overall variation in prices and therefore can be considered somewhat minor compared to the primary influence of overall demand for baseball, but, still, is not trivial.

The third principal component suggests an inverse relationship between the prices of beer, soda, and hotdogs and the prices of caps, programs, tickets, and parking. The former goods can be classified as possible repeat purchases--any fan may purchase several beers in the course of a game. In contrast, the others are "fixed purchases"--a fan will likely purchase no more than one ticket or program. Studies of nonlinear multiproduct pricing indicate that, if possible, it can be profitable to offer an "entry fee," coupled with cost-plus pricing for goods that consumers may purchase in multiple units. This fee can be captured in the ticket price, but it may also be optimal to raise the price of single-purchase goods as well, as discussed above. For this reason one can expect to find an inverse relation between the prices of "fixed purchases" and those of "repeat purchases." This is our interpretation of the third component.

It is of interest to determine the contribution of these components to the overall variation in the price of each good. From these estimations, it is possible to decompose the overall variation in each real price into proportions attributed to the general demand effect, embodied in the first principal component, the interaction between prices, embodied in the second and third principal components, and the "idiosyncratic variation" associated with the remaining four components.

Table 3 reports the results of these decompositions. The first column indicates that the general demand effect, reflected in the first principal component, accounts for 61% of the overall variation in ticket prices, which is not an unexpected result. Looking at the remaining prices, the general demand effect accounts for approximately 50% of overall variation for soda, beer, and hotdogs, but considerably less for parking (25%), hats (10%) and programs (1%).

The second column of Table 3 reports the percentage of overall variation in each price that is attributed to the interaction between the prices charged by baseball teams *qua* multiproduct firms (that is, to the second and third components). In the case of tickets, this interaction accounts for 11% of real ticket price variation. Among the other prices, these interactions account for a considerable amount of the variation in the prices of parking (50%), hats (44%), and programs (76%). However, these interactions account for relatively modest amounts of overall variation in the cases of soda (6%), beer (10%), and hotdogs (19%). Therefore, multiproduct pricing considerations are an important determinant of the prices of some concessions, and a modest influence on prices of other concessions and of tickets.

The third column of Table 3 reports the percentage of overall variation in each price attributed to the remaining four principal components, which by construction are the values in column 1 and column 2 subtracted from unity. These values reflect the percentage of overall variation that is attributed to the remaining four principal components that do not have easily discerned economic content. Fortunately, the first three principal components consistently account for more than 50% of the total variation in each real price, while idiosyncratic factors usually account for only about one-third of the variation in the price of tickets or concessions.

The analysis thus far has focused on the overall variation of real prices charged by MLB teams, including both cross-team and cross-time variation. We believe this is appropriate; there is no reason to exclude either source of variation in advance. However, one may legitimately wonder if the dynamics of prices within teams across time exhibit similar covariance patterns. In order to examine this question, we replicated the principal components analysis on a set of prices that were purged of team and year effects. Specifically, each price was regressed on a full set of team and year fixed effects, and the residuals from these regressions were used in the principal components analysis in the place of the original price data.[6]

This analysis, available from the authors upon request, indicates the principal components derived from these scaled prices reveal the same qualitative relationships between the prices charged by MLB teams as shown in Tables 2a and 2b. In particular, the largest influence on prices remained a general demand effect, which explained about one-third of the within-team variation in prices over time. Two additional components, each having half the variance of the first, were interpretable as a price tradeoff between single and multiple purchase goods and between required and discretionary purchases, respectively. However, the results were not quite as “clean”: in the second component, one factor loading was zero, while in the third component, one factor loading was positive instead of the expected negative number. Still, these results show that our original conclusions are reasonably robust and suggest that the factors generating price variation across teams are similar to those generating price variation within teams over time.

[6] Incidentally, a structural dynamic analysis of multiproduct pricing would be several orders of magnitude more complex than a static or cross-section analysis. To proceed it would probably be necessary to set up a quasi-structural VAR, as in Blanchard (1989), though identification of the model would probably be very difficult.

In summary, most of the variation in the prices of multiple products set by MLB teams can be attributable to a small number of factors. These factors--demand interrelationships across goods and nonlinear pricing, in addition to general shifts in demand--are the same as those emphasized by economic theories of multiproduct pricing. It appears, therefore, that multiproduct pricing considerations contribute meaningfully to price variation in MLB.

However, while the analysis suggests relationships between the various prices charged by MLB teams not heretofore illuminated, the principal component analysis is, by necessity, heuristic in nature. In essence, the components are reduced form, which, while often conducive to material economic interpretation, are not independently confirmed without subsequent analysis. Therefore, the next step in the analysis is to relate the three independent components to a common set of city, team, and stadium specific characteristics that might provide confirmation of the posited economic interpretations.

B. Secondary Econometric Analysis: Are the Heuristics Confirmed?

Here, the three factors that have economic content are related to several variables commonly included in other studies of attendance models in professional baseball (and other sports). Specifically, season attendance, city per-capita income, city population, once-lagged team win percentage, the age of the team's stadium, whether the team's stadium is a dome (or retractable roof), and whether the stadium is single purpose are used as a common set of explanatory variables for the first three components reported in Table 2b. Attendance and team quality data were obtained from Major League Baseball, city income from the Bureau of Economic Analysis, city population from the Bureau of the Census, and stadium characteristics from Munsey and Suppes at www.ballparks.com. The descriptive statistics for these variables are reported in the bottom panel of Table 1 and are similar to those reported in other studies.

Table 4 reports the results of the auxiliary regression for each of the first three principal components from Table 2b. For each component, two models are presented. In the first (Model I), the explanatory variables include the population of the host city, the lagged per-capita income, the lagged winning percentage, the stadium age (and its square). These variables are included instead of attendance because they are highly correlated with attendance and avoid possible endogeneity problems (see Coates and Humphreys, 2003, and Depken, 2004). Model II replaces these variables with the season-total attendance.

The remaining explanatory variables, common to both Model I and Model II, are a dummy variable that takes a value of one if the team's stadium is less than six years old, a dummy variable for whether the team's stadium is a dome (or retractable roof), a dummy variable for whether the team's stadium is a single-purpose venue, and a general time trend. The intuition for including the stadium characteristics is to test whether new and unique stadiums allow team owners to price discriminate more than in older, generally more generic stadiums. When estimating Model I for each of the components, the sample size is restricted to only U.S. based teams because data limitations on income preclude the inclusion of the two Canadian teams. However, season attendance is available for every team and therefore for Model II the sample size is expanded to include the Canadian teams. Estimation uses the random effects estimator. This estimator is deemed appropriate vis-à-vis the fixed effects estimator or pooled OLS using Hausman specification tests.

Taking the three principal components in turn, the first component was associated with a general demand effect. The auxiliary regression, therefore, should support this intuition by indicating a positive and statistically significant relationship between the component and the explanatory variables that are correlated with attendance. As the general demand effects are expected to influence prices in the same direction as they influence attendance, those explanatory variables that are positively (negatively) related to attendance should also be positively (negatively) related to the first principal component. In both models, the results confirm this intuition. Population, lagged income, and lagged winning percentage positively influence demand and are all positively correlated with the first principal component. On the other hand, stadium age is generally found to reduce attendance (either because an older stadium is less attractive to fans or because an older stadium tends to have lower capacity) and is also found to contribute negatively to the first principal component. In Model II the demographic and team quality variables are replaced by season attendance. The positive and statistically significant parameter estimate on attendance confirms the findings in Model I.

Economic interpretations of the principal components can be further examined by exploring the effect of demand influences on the second and third components. Just as we expect the first component to be strongly related to demand, there is little reason to expect a strong demand influence on the second and third components.

The economic context of the second component is heuristically derived as a subtle inverse relationship between tickets and parking, considered required purchases to attend a baseball game, and the remaining concessions, considered discretionary purchases. There need not be a general demand (attendance) effect on this component--there is no reason to expect one theoretically. Or, if one exists, it need not be large. The same logic applies to the third component as well.

The third and fourth columns of Table 4 report the estimation results for Model I and Model II for the second principal component; the fifth and sixth columns do the same for the third component. There is indeed no strong relation between demand influences and the second component. In Model I, income, winning percentage, and population are all insignificant. Some desirable stadium characteristics (such as newness) are positively related to the second component, while others (such as a single purpose stadium) are negatively related. Model II puts this even more plainly: the coefficient on attendance is not statistically significant. Regarding the third component, the individual demand proxies in Model I are jointly insignificant, but attendance is significant. On balance, these results provide reasonable though not unanimous support for our interpretation of these components.

The economic context of the third principal component was a tradeoff between single and multiple-purchase goods, as a form of second degree price discrimination. In discussing this aspect of pricing we emphasized that this type of price discrimination might be more feasible for teams who market innovatively or whose stadiums better permit a range of seating options and ticket prices, to better extract surplus from consumers. One such variable exists in our data: a dummy for a new stadium (under six years old). Newer stadiums, such as Camden Yards (Baltimore), Safeco Field (Seattle), PETCO Field (San Diego), and Ameriquest Field in Arlington (Arlington, TX), provide a wider variety of sight lines as reflected in their greater number of different ticket prices and are therefore more conducive to this pricing strategy. The new stadium coefficient is indeed highly significant in Model I for component three, the only significant coefficient other than that on the time trend. The coefficient is almost as strong in Model II as well. If new stadiums are constructed so that team owners can more effectively extract consumer surplus in the ticket price, for example, by having a wider range of seating options and ticket prices, then it will be optimal to raise ticket prices and the prices of other infrequently purchased concessions and lower the price of repeat-purchase concessions. Thus, the economic interpretation of the third component is supported.

These auxiliary regressions provide a statistical test of the economic intuition applied to the first three principal components reported in Table 2b, the components of primary interest, explaining nearly two-thirds of the total variation in real ticket, parking, and concession prices in Major League Baseball over the sample period. The regressions offer statistical support for the economic context provided the first three principal components, and suggest that future research on the relationship between ticket and concession prices (and demand) might prove fruitful. More generally, the analysis of MLB prices provides a case study of how principal components, a little used methodology in economics, can provide valuable empirical insights into the pricing decisions of multiproduct monopolists.

The implications specific to MLB are therefore two-fold. First, from the second and third principal components, it appears that team owners do engage in price discrimination of some form by allowing a trade off in the prices of tickets and concessions, consistent with complementarities between the two, and allowing tradeoffs in the prices of single and multiple purchase products, reminiscent of a two-part tariff. Second, these relationships suggest that attendance models that fail to account for the price of concessions may suffer an upward omitted variables bias, perhaps to the extent that ticket price elasticities are biased towards unitary elasticity or inelasticity.

6. Conclusions

Principal components analysis can be a useful substitute for a more traditional, and more complicated, structural analysis of multiproduct pricing. To illustrate the approach, pricing in Major League Baseball is used as a case study. In the context of microeconomic theory, professional baseball teams are easily characterized as multiproduct firms. Baseball (and other sports) teams sell access to a “game” or “experience” but simultaneously sell various concessions that contribute to the overall enjoyment of the “experience.” Furthermore, tickets, unlike concessions, are required purchases. Therefore it is possible that professional baseball (and other sports) teams maximize profits by using sophisticated pricing schemes that take advantage of the complementarity amongst the different products sold by the team (firm) and incorporate second-degree price discrimination in food concessions, in which the “entry fee” is factored into the ticket price. While a structural model might be first-best to analyze these relationships, the data requirements to identify such a system of equations and use them to explain these price interactions can be insurmountable. The alternative employed herein, the principal components approach, decomposes the overall variation in the real prices charged by baseball teams into independent components conducive to interpretation consistent with economic theory.

In the received theoretical literature on multiproduct pricing, one can identify three predominant influences on prices: overall demand or cost changes common to all goods, changes in the demand/cost of one good or a subset of goods, and the ability for firms to extract surplus through two-part tariffs, bundling, and other forms of price discrimination. In our analysis each influence is revealed to be an important determinant of price variation. The most important component, accounting for one-third of overall price variation, is a general demand effect which contributes to a general increase in all prices for teams that are successful on the field. The second most important component is a tradeoff between the real prices of tickets and parking and the real prices of concessions within the stadium. This relationship is consistent with the pricing of complements by multiproduct firms. The third most important component seems to be a negative relationship between multiple-purchase goods (beer, soda, and hotdogs) and single-purchase goods (hats, programs, tickets, and parking), consistent with theories of nonlinear multiproduct pricing. These interpretations are verified in a supplementary analysis that relates these components to various supply and demand characteristics of professional baseball markets. Thus we confirm that the economic forces stressed in the theoretical literature on multiproduct pricing are the most important influences on pricing in MLB.

From the point of view of the sports economics literature, the results suggest a) pricing decisions by team owners seem more sophisticated than previously modeled, and b) significant relationships between ticket, parking, and concession prices imply omitted variables bias in traditional attendance/demand models that do not include the additional prices as explanatory variables. Given the results presented herein, it is anticipated that the omitted variables bias may push estimated price elasticities of demand closer to zero, which may mistakenly indicate that teams price in the unitary or inelastic portion of their demand.

References

Ahamad, B. (1967) "An Analysis of Crimes by the Method of Principal Components," *Journal of Applied Statistics*, **16**(1), pp. 17-35.

Armstrong, Mark (1999) "Price Discrimination by a Many-Product Firm," *Review of Economics and Statistics,* **66**(1), pp. 151-168.

Armstrong, Mark, and John Vickers (2001) "Competitive Price Discrimination," *RAND Journal of Economics*, **32**(4), pp. 579-605.

Berry, Steven, James Levinsohn, and Ariel Pakes (1995) "Automobile Prices in Market Equilibrium," *Econometrica*, **63**(4), pp. 841-890.

Blanchard, Olivier Jean (1989) "A Traditional Interpretation of Macroeconomic Fluctuations," *American Economic Review* **79**(5), pp. 1146-1164.

Bliss, Christopher (1988) "A Theory of Retail Pricing," *Journal of Industrial Economics,* **36**(4), pp. 375-391.

Bresnahan, T. (1987) "Competition and Collusion in the American Automobile Oligopoly: The 1955 Price War," *Journal of Industrial Economics*, 35, pp. 457-482.

Coates, Dennis and Brad Humphreys (2003) "Novelty Effects of New Facilities on Attendance at Professional Sporting Events," *mimeo*, UMBC Economics Department Working Paper 03-101.

Depken, Craig A. II (2000) "Fan Loyalty and Stadium Funding in Professional Baseball," *Journal of Sports Economics,* **1**(2), pp. 124-138

Depken, Craig A. II (2001) "Fan Loyalty in Professional Sports: An Extension to the National Football League," *Journal of Sports Economics,* **2**(3), pp. 275-284

Depken, Craig A. II (2004) "The Novelty Effect of a New Stadium and Concession Prices" *mimeo,* Department of Economics, University of Texas at Arlington, working paper #04-003.

Doll, John P. and Sean B. Chin (1970) "A Use for Principal Components in Price Analysis," *American Journal of Agricultural Economics*, **52**(4), pp. 591-593.

Forbes, Kevin (1988) "Pricing of Related Products by a Multiproduct Monopolist," *Review of Industrial Organization,* **3**(3), pp. 55-73.

Ferguson, D.G., Kenneth G. Stewart, J.C.H. Jones, and Andre Le Dressay (1991) "The Pricing of Sports Events: Do Teams Maximize Profit?" *Journal of Industrial Economics*, **XXXIX**(3), pp. 297-310.

Giulietti, Monica, and Michael Waterson (1997) "Multiproduct Firms' Pricing Behavior in the Italian Grocery Trade," *Review of Industrial Organization*, **12**(5-6), pp. 817-832.

Green, Paul (1978) *Analyzing Multivariate Data*, Dryden Press: Hinsdale, Ill.

Johnson, Richard, and Dean Wichern (1982) *Applied Multivariate Statistical Analysis*, Prentice Hall: Englewood Cliffs, N.J.

Mirrlees, James (1976) "Optimal Tax Theory: A Synthesis," *Journal of Public Economics*, **6**(4), pp. 327-358.

Nezlin, N.P., and J.C. McWilliams (2003) "Satellite Data Empirical Orthogonal Functions Statistics, and the 1997-1998 El Nino off California," *Remote Sensing of Environment* 84, pp. 234-254.

Quant, C.M., Fischer, P.H., Buringh, E., Ameling, C.B., Houthuijs, D.J.M., and Cassee, F.R.(2003) "Application of Principal Component Analysis to Time Series of Daily Air Polution and Mortality," RIVM report 650010035/2003, available at <*www.rivm.nl/bibliotheek/rapporten/650010035.pdf*>, last accessed June 2004.

Reibstein, David, and Hubert Gatignon (1984) "Optimal Product Line Pricing: The Influence of Elasticities and Cross-Elasticities," *Journal of Marketing Research*, **21**(3), pp. 259-267.

Rochet, Jean-Charles, and Lars Stole (2002) "Nonlinear Pricing with Random Participation," *Review of Economic Studies* **69**(1), pp. 277-311.

Scully, Gerald W. (1989) *The Business of Major League Baseball*, The University of Chicago Press: Chicago.

Sibley, David and Padmanabhan Srinagesh (1997) "Multiproduct Nonlinear Pricing with Multiple Taste Characteristics," *The RAND Journal of Economics*, **28**(4), pp. 684-707.

Spence, A. Michael. "Multi-Product Quantity Dependent Prices and Probitability Constraints," *Review of Economic Studies* **47**, pp. 821-841.

Welki, Andrew M. and Thomas J. Zlatoper (1994) "US Professional Football: The Demand for Game-Day Attendance in 1991," *Managerial and Decision Sciences*, **15**(5), pp. 489-495.

Winfree, Jason A., Jill J. McCluskey, Ron C. Mittelhammer, and Rodney Fort (2001) "Location and Attendance in Major League Baseball," working paper, Department of Economics, Washington State University.

Zimbalist, Andrew (1992) *Baseball and Billions*, Basic Books: New York.

Table 1a: Descriptive Statistics of the Data

Variable	Description	Mean	Std. Dev.	Min	Max
RPTIX	Average per-game season ticket price	13.611	3.850	8.68	33.89
RPARK	Price of parking	7.185	2.639	2.91	17.87
RPBEER	Price of 20oz Beer	5.217	0.843	3.29	10.53
RPSODA	Price of 20oz Soda	2.572	0.487	1.48	4.11
RPDOG	Price of hotdog	2.230	0.516	0.79	3.87
RPPROGRAM	Price of program	3.578	1.059	0.69	7.74
RPHAT	Price of ball cap	12.054	2.194	4.84	20.00
ATTEND	Total season home attendance (100Ks)	22.168	7.227	9.05	44.83
INCOME	MSA Per-capita income (1000s)	30.594	3.618	21.56	47.18
POP	MSA population (millions)	6.276	5.493	1.60	21.31
LAGWIN	Previous season's win percentage	0.500	0.067	0.327	0.704
STAGE	Age of team's stadium	30.112	24.378	0.00	89.00
NEWSTAD	Team's stadium is less than six years old	0.154	0.361	0.00	1.00
DOME	Team's stadium is a dome or retractable roof	0.140	0.347	0.00	1.00
SPURP	Team's stadium is single purpose	0.615	0.487	0.00	1.00
TIME	Time trend (1=1990)	6.168	3.163	1.00	11.00

Price data (reported in upper panel) describe all Major League Baseball teams from 1991 through 2001 and were obtained from various issues of *Team Marketing Report*. All prices and income converted to 2000 dollars using the Consumer Price Index from the Bureau of Labor Statistics. Attendance and team win percentage obtained from Major League Baseball. Population and income obtained from Census Bureau. Stadium characteristics obtained from Munsey and Suppes at *www.ballparks.com.* The price data comprise a sample of 312 observations used in principal component analysis. Stadium, income, and population data are for 286 observations for U.S. baseball teams (two Canadian teams not included).

Table 1b: Correlation Matrix of Real Ticket, Parking and Concession Prices

	RPTIX	RPARK	RPBEER	RPSODA	RPDOG	RPPROGRAM	RPHAT
RPTIX	1.000						
RPARK	0.482	1.000					
RPBEER	0.348	0.225	1.000				
RPSODA	0.419	0.154	0.392	1.000			
RPDOG	0.365	0.024	0.365	0.409	1.000		
RPPROGRAM	0.104	0.083	-0.074	0.013	0.112	1.000	
RPHAT	-0.006	-0.032	0.078	0.061	0.122	0.109	1.000

Price data describe all Major League Baseball teams from 1991 through 2001, were obtained from various issues of *Team Marketing Report*, and were converted to 2000 dollars using the Consumer Price Index from the Bureau of Labor Statistics.

Table 2a: Principal Components

Component	Eigenvalue	Difference	Proportion of Variation Explained	Cumulative Variation Explained
One	2.323	1.164	0.332	0.332
Two	1.159	0.061	0.165	0.497
Three	1.098	0.233	0.157	0.654
Four	0.865	0.262	0.123	0.778
Five	0.602	0.044	0.086	0.864
Six	0.558	0.167	0.079	0.944
Seven	0.391	.	0.055	1.000

Table 2b: Eigenvectors

Variable	Eigenvector One	Eigenvector Two	Eigenvector Three	Eigenvector Four	Eigenvector Five	Eigenvector Six	Eigenvector Seven
RPTIX	0.51372	-0.25871	0.16756	0.00031	-0.26746	-0.27126	-0.70425
RPARK	0.32892	-0.55324	0.36601	0.38088	0.02159	-0.06954	0.54900
RPBEER	0.44997	0.03893	-0.30778	0.17808	0.76435	0.25256	-0.14679
RPSODA	0.47306	0.12265	-0.21149	-0.16757	-0.50148	0.63091	0.19706
RPDOG	0.43282	0.38158	-0.12489	-0.35187	0.03473	-0.62274	0.37235
RPPROGRAM	0.08441	0.23666	0.79547	-0.40646	0.25823	0.26617	-0.03687
RPHAT	0.08477	0.63939	0.22391	0.71142	-0.15619	-0.0332	-0.04735

Table 3: Real Price Variation Explained by Principal Components

Variable	General Demand Effect (Component 1)	Multiproduct Pricing Effects (Component 2 and 3)	Idiosyncratic or Unexplained Effects (Components 4–7)
RPTIX	0.61	0.11	0.28
RPARK	0.25	0.50	0.25
RPBEER	0.47	0.10	0.43
RPSODA	0.52	0.06	0.42
RPDOG	0.43	0.19	0.38
RPPROGRAM	0.01	0.76	0.23
RPHAT	0.10	0.44	0.46

Values reflect proportion of variation attributed to each effect(s). Values in column one reflect R^2 obtained from regressing each real price on the first component, and so on. Values in column two reflect the additional variation in real price explained by the second and third components. The values in the third column reflect the remaining variance of each real price, by construction totally attributed to components four through seven. Rows sum to one.

Table 4: Auxiliary Regressions
(Principal Components as Dependent Variables)

	Component One		Component Two		Component Three	
Variable	**Model I**	**Model II**	**Model I**	**Model II**	**Model I**	**Model II**
POP	0.066* (0.02)	---	0.034 (0.02)	---	-0.033 (0.03)	---
LAGINCOME	0.101* (0.03)	---	-0.039 (0.02)	---	0.025 (0.03)	---
LAGWIN	3.100* (0.82)	---	0.576 (0.69)	---	1.025 (0.79)	---
STAGE	-0.038* (0.02)	---	0.048* (0.01)	---	0.004 (0.02)	---
STAGESQ	0.0005* (0.0002)	---	-0.0005* (0.0002)	---	-0.000 (0.00)	---
ATTEND	---	0.064* (0.01)	---	-0.002 (0.01)	---	0.024* (0.01)
NEWSTAD	-0.081 (0.25)	0.264 (0.21)	0.394** (0.21)	0.102 (0.17)	0.472** (0.25)	0.457* (0.18)
DOME	0.272 (0.30)	-0.135 (0.38)	0.123 (0.27)	-0.674** (0.35)	-0.377 (0.38)	0.081 (0.31)
SPURP	0.419* (0.25)	0.673* (0.26)	-0.361** (0.22)	-0.637* (0.27)	0.080 (0.29)	-0.132 (0.22)
TIME	0.171* (0.02)	0.183* (0.01)	0.059* (0.02)	0.016 (0.01)	0.069* (0.02)	0.063* (0.02)
R^2	0.623	0.367	0.237	0.037	0.137	0.122
Wald (X^2_9)	378.73*	39.65*	41.95*	280.21*	54.42*	46.73*
Observations	286	312	286	312	286	312

A random effects estimator was applied after Hausman specification tests. The first three principal components from Table 2b are the dependent variables. For example, component one is calculated as 0.51xRPTIX + 0.32xRPARK + 0.45xRPBEER + 0.47xRPSODA + 0.43xRPDOG + 0.08xRPROGRAM + 0.08xRPHAT. Model I is based on 286 observations of U.S. baseball teams, Model II is based on 312 observations, including two Canadian teams. * indicates significance at the 5% level, ** at the 10% level.