ACTA UNIVERSITATIS CAROLINAE,
KINANTHROPOLOGICA Vol. 46, 1 – 2010 Pag. 5–18

CHARLES UNIVERSITY IN PRAGUE,
FACULTY OF PHYSICAL EDUCATION AND SPORT,
DEPARTMENT OF KINANTHROPOLOGY, HUMANITIES,
AND MANAGEMENT[1],
DEPARTMENT OF FOREIGN LANGUAGES[2]

# ATTENDANCE AT BASKETBALL MATCHES: A MULTILEVEL ANALYSIS WITH LONGITUDINAL DATA[1]

ONDŘEJ PECHA[1], WILLIAM CROSSAN[2]

## SUMMARY

Multilevel analysis is used to evaluate the elements which compose fan attendance in Czech basketball. The data set analyzed is comprised of a ten year study of 18 teams which played in the highest Czech basketball league. This study differs from other demand studies which evaluate fan attendance in that a cultural secondary sport, basketball, is studied in a globally semi-periphery country, the Czech Republic. Previous studies have focused on primary sports in globally core countries. The study shows multilevel analysis to be a useful methodology for demand studies of fan attendance. It was shown that there is a slightly increasing linear tendency in attendance across time. The independent variables of final place and number of foreigners and are measured across time as the time-varying predictors of the dependent variable of fan attendance within teams. The independent variable of hall capacity was considered as the time-invariant predictor of growth rates and attendance initial statuses between teams. It was shown that final place is and number of foreigner players is not a good single predictor of attendance, respectively. Finally, the findings confirmed that the growth rate within teams is limited by the team's hall capacity.

**Key words:** fan attendance, multilevel analysis, longitudinal study, basketball

## INTRODUCTION

### Brief introduction of the significance of attendance

The battle for fans in the crowded sport marketplace is intense. Fans equal revenue. Fans come in the form of live spectators, media viewers, apparel purchasers and even video game players. The options available to the consumer for fandom, and to the brand for marketing, change and multiply constantly. In spite of this apparently unending change

[1] This paper was supported by the Czech Republic grant VZ MSM 0021620864.

live spectators remain at the core as a foundational element in building a sport brand (Davis, 2008). While live spectators equal less and less of a team's total revenue package, these spectators remain a necessary ingredient to all the other revenue sources.

Without live spectators sponsors will not advertise in the venue. Caslavova (2000) has outlined the path from live spectators to each line on the revenue side of a sport team's budget. Without spectators television media will not broadcast to the masses. Empty or half full venues do not broadcast well. This is true to such an extent that the NBA recently announced that in light of the current economic crisis they will offer free tickets to the unemployed in order to keep live fan attendance looking good for their media broadcasts around the world (Lombardo, 2008). Without live spectators teams cannot make money on parking, concessions, souvenir sales at the venue, filling luxury boxes, ect.

Demand research which attempts to quantify the elements which contribute to live spectator attendance has focused on the largest sport leagues in the world, or primary sports in core countries. Davis (2008) follows the most studied league in terms of demand studies with a VAR analysis of the USA's Major League baseball, a primary sport in a core country. Becker and Suls (1983) use a simple multiple regression model to measure the effect of four variables on attendance across 22 teams over a 10 year period; our data set most closely parallels theirs. The second most studied league in terms of demand studies is the USA's NBA; Morse et al. (2008) use a censored multiple linear regression model which is closest to the model we have chosen in this study. Much less research has been done on secondary sports in periphery and semi-periphery countries. Lawson et al. (2008) use a multivariate regression model to measure the effect of David Beckham on fan attendance in the USA's Major League of Soccer. It can be argued that soccer is a secondary sport in the USA. These sports have much smaller media following and make much less money on non-sport products such as parking, concessions, and souvenirs, as shown in a similar multiple regression study of the MLS by DeSchriver (2007). Thus in secondary sports, revenue is typically composed almost entirely of sponsorship revenue, a few television appearances a season, and live spectators.

There are many elements that contribute to the fan choice between one live event over another. The choices of entertainment in 2009 are virtually unlimited. A team's record, star performers, equality or inequality with opponents, marketing efforts, visibility in print, TV and internet media, event attractiveness at each game, venue attractiveness, other sport and non-sport event competition and tradition or history all contribute to a potential fan's choice to attend one event over another. Hansen and Gauthier (1989) have used factor analysis to identify 40 elements that contribute to fan attendance and divided them into the four categories of economic, demographic, attractiveness and residual preference factors. The sheer number of factors makes it a necessity to delineate between which factors are most significant and which make a smaller impact. The necessity for delineating primary and secondary sports, and core, periphery, and semi periphery in order to understand the interaction of these elements in the larger scope of globalization has been established in other studies (Crossan, 2007; Maguire & Pearton, 2000; Maguire & Poulton, 1999).

Given the elusive nature of gaining fans and the necessity of live fans as a cornerstone of the revenue picture for a sport brand, a quantitative model is needed to help the sport administrator know where to invest the limited resources he/she has in this secondary sport in order to survive.

## Brief introduction to multilevel analysis

Multilevel analysis (also known as hierarchical linear modeling) was originally developed as a tool for analyzing hierarchically structured data, where individuals are nested within organizations (Bryk & Raudenbush, 1992). There are both theoretical and statistical reasons for using multilevel models (Luke, 2004). Individuals share the same social context which influences them and, thus, the observations of individuals belonging to the same group tend to be more similar than the observations of individuals which come from different groups. Therefore, some dependencies can be found in the data (Snijders & Bosker, 1999). Hence, the commonly used single-level statistical tools like multiple regression or ANOVA cannot be used, because the assumption that observations were sampled independently from each other is violated in this two-level case.

Two-level analysis is the simplest case of multilevel or hierarchical models. Some typical examples are students nested within schools or children nested within families (Hox, 2002). Students or children are considered as the level-1 units and schools or families are called the level-2 units. A special kind of nesting is defined by longitudinal data as measurements (level-1 units) within subjects (level-2) units. The dependence of the different measures for a given subject is of primary interest in all longitudinal studies (Snijders & Bosker, 1999). Laird and Ware (1982) were among the first to apply the multilevel models to the analysis of change. To date, the analysis of longitudinal data using multilevel modeling is a common part of the majority of fundamental textbooks on multilevel analysis (e.g., Bryk & Raudenbush, 1992; Goldstein, 1995; Raudensbush & Bryk, 2002; Hox, 2002; Singer & Willett, 2003). In kinesiology and exercise sciences, Zhu (1997) conducted one of the first two-level analysis in physical fitness of students nested within schools. The first application of this methodology for longitudinal data was accomplished by Zhu and Erbaugh (1997) who analyzed the developments of swimming skills among pre-school children. In the Czech Republic, multilevel analysis was not applied to kinesiology until quite recently (Pecha, 2004, 2006).

The multilevel approach has several advantages compared to the traditional methods for analyzing change such as repeated measures ANOVA or MANOVA. From our point of view, the first advantage mentioned by Wood and Zhu (2006) seems to be the most essential. ANOVA requires a balanced design, where each subject (level-2 unit) is required to have the same number of time points or repeated measurements. However, our design is unbalanced because the basketball teams turned over in the first league which is typical in European leagues as compared to most of the US leagues where demand studies have previously been conducted. Only 7 of 18 teams which appeared in the first league remained there for the whole 10 year period. Fortunately, multilevel analysis can easily handle such a data set by fitting a regression curve of the same type for each team separately (Hox, 2002). This is possible for much lower and not necessarily the same number of repeated measures. Hence, all teams are taken into account and little information is lost due to dropouts.

## PROBLEM

The present study serves two purposes: the first is demonstrating the importance of attendance as a measurable dependent variable in demand studies in secondary sports for semi-periphery and periphery countries. The second purpose is to introduce multilevel regression analysis as an appropriate methodology for use when longitudinal data are unbalanced. Thus, this paper should assist investigators interested in longitudinal research and especially in attendance development and serve as a guideline for understanding when and how to apply the discussed methodology to their data.

## METHODS

### Research design

The current study is composed of 18 teams across a 10 year period which played in the Czech Extraleague of professional basketball. The rules of play in this league state that each year the last place of 12 teams will drop down to the 2nd league of Czech basketball

**Table 1.** Data collection design

| Team | Year of a season start | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 |
| 1 | x | x | x | x | x | x | x | x | x | x |
| 2 | x | | | x | | | | | | |
| 3 | x | x | x | x | x | x | x | | | |
| 4 | | | | | | | | | | x |
| 5 | x | x | x | x | x | x | x | x | x | x |
| 6 | | | | | | | | | x | x |
| 7 | x | x | x | x | x | x | x | x | x | x |
| 8 | | | | | | | | x | x | x |
| 9 | x | x | x | x | x | x | x | x | x | x |
| 10 | | | | | | | x | x | x | x |
| 11 | x | x | x | x | x | x | x | x | x | x |
| 12 | | | x | x | x | x | x | x | x | x |
| 13 | x | x | x | x | x | x | x | x | | |
| 14 | | x | x | | x | | | | | |
| 15 | x | x | x | x | x | x | x | x | x | x |
| 16 | x | x | x | x | x | x | x | x | x | x |
| 17 | x | x | x | x | x | x | x | x | x | |
| 18 | x | x | | | | x | | | | |

Note. x – a team played in the first league this season and the winner of that league will move up to the Czech 1st league. This explains our "messy design" in which only 7 of the 18 teams remained in the 1st league all 10 years. This rule was broken in the year 2005 when the last place team was allowed to stay due to the bankruptcy of the 8th place team. This is explained in Table 1. Each 9 month season spanned two calendar years and is denoted below by the beginning year of the season. The independent variables of final place, number of foreigners and hall capacity are measured across time against the dependent variable of fan attendance within and between teams. Final place was chosen because there is much variance within teams over the 10 year time period. Number of foreigners was chosen because they have increased from 11 to 56 out of the average 200 players in the top Czech league. Hall capacity was chosen because it remains relatively stable within teams throughout the 10 year period.

**Analysis methods**

A two-level hierarchical linear model is used here, because the present data set is partitioned into two levels. They are defined as follows:

Level-1 (lower level) – time points (1 to 10 seasons)

Level-2 (higher level) – teams (18 basketball teams)

The dependent variable – *attendance* – is measured at the lower level. Two independent time-varying variables – *number of foreigners* and *final place* – are situated also at the lower level, because they are measured repeatedly in time at the same occasions as the *attendance*. Finally, the hall *capacity* is an independent time-invariant variable which is observed at the higher (team) level. Let us note that the time-invariant variable remained stable across all measurement occasions.

The analytic procedure of multilevel analysis with longitudinal data involves several steps. The *unconditional model* is the very first model which should always be fit (Singer & Willett, 2003). This model includes no independent variables at any level and can be written as:

$$(attendance)_{ij} = \pi_{0j} + e_{ij}, \qquad (1)$$
$$\pi_{0j} = \beta_{00} + r_{0j},$$

where $\pi_{0j}$ is the mean *attendance* of team $j$; $e_{ij}$ is a time-specific deviation of *attendance* at time $i$ within team $j$ from the team-specific mean $\pi_{0j}$; $\beta_{00}$ is the grand mean of *attendance* taken across all teams and all time points; and $r_{0j}$ is a deviation of the team-specific mean $\pi_{0j}$ from the grand mean $\beta_{00}$. The unconditional model allows us to evaluate the relative magnitude of the within-team and between-team variance components. Their proportion is usually expressed by the *intraclass correlation coefficient*, ICC, which describes the proportion of the total variance of *attendance* that lies between teams. This coefficient is expressed as:

$$\text{ICC} = \frac{Var(r_{0j})}{Var(r_{0j}) + Var(e_{ij})}. \qquad (2)$$

The second step is the *unconditional growth model.* The variable of *time* is introduced to equation (1) as the only independent variable in this model at the lower level. Thus, the model can now be written in the following form:

$$(attendance)_{ij} = \pi_{0j} + \pi_{1j}(time)_{ij} + e_{ij},$$
$$\pi_{0j} = \beta_{00} + r_{0j}, \qquad (3)$$
$$\pi_{1j} = \beta_{10} + r_{1j},$$

where $\pi_{0j}$ is the intercept of the regression line within team $j$; $\pi_{1j}$ is the slope of this line; and $e_{ij}$ is now a level-1 regression residual of team $j$ at time $i$. Intercepts and slopes are subsequently allowed to vary randomly across teams. $\beta_{00}$ is a grand mean intercept and $\beta_{10}$ is a grand mean slope of the overall regression line. The coefficients $r_{0j}$ and $r_{1j}$ are deviations of the team-specific means from the grand mean intercept and slope, respectively. Special attention was paid to the coding of the *time* variable. This variable ranged from 0 (season 1998/1999) to 9 (season 2007/2008). As a result, the intercept $r_{0j}$ is the estimation of *attendance* in the season 1998/1999 (also known as the *initial status*). Moreover, 1 year is the measurement unit of the independent variable *time*.

The third step consists of the introduction of the time-varying independent variable, $X$, into the level-1 sub-model. Thus, we obtain:

$$(attendance)_{ij} = \pi_{0j} + \pi_{1j}(time)_{ij} + \pi_{2j}X_{ij} + e_{ij},$$
$$\pi_{0j} = \beta_{00} + r_{0j}, \qquad (4)$$
$$\pi_{1j} = \beta_{10} + r_{1j},$$
$$\pi_{2j} = \beta_{20} + r_{2j}.$$

In our case, the X variable constitutes either the number of foreigner players (*foreigners*) or *final place* variables. They should be inserted into the equation (4) in the group-centered form (i.e., as deviations from the team-specific means) due to the meaningful interpretation of the intercepts and *time* slopes (Hox, 2002). If the time-varying variables are group-centered, the remaining regression coefficients remain unchanged. Obviously, the generalization of the model (4) to more than one time-varying variable is straightforward.

The last step is a two-level model with independent variables at both levels. In principle, a time-invariant variable, $Z$, is joined to the equation (4) at level-2. For example, this model can be expressed as follows:

$$(attendance)_{ij} = \pi_{0j} + \pi_{1j}(time)_{ij} + \pi_{2j}X_{ij} + e_{ij},$$
$$\pi_{0j} = \beta_{00} + \beta_{01}Z_j + r_{0j}, \qquad (5)$$
$$\pi_{1j} = \beta_{10} + \beta_{11}Z_j + r_{1j},$$
$$\pi_{2j} = \beta_{20} + r_{2j}.$$

In the present study, $Z_j$ is the hall capacity of team $j$. It is assumed in model (5) that *capacity* is a predictor of the intercept $\pi_{0j}$ and time slope $\pi_{1j}$ only. It is not neccesary here to explain the variability of the *foreigners* or *final_ place* slope $\pi_{2j}$. Nevertheless, if an investigator wishes to estimate this relationship as well, the time-invariant variable $Z_j$ can also be added to the last row of equation (5).

All analysis has been conducted in HLM 6 (Raudenbush et al., 2004), which is a specialized package for hierarchical linear modeling or multilevel analysis.

## RESULTS

The estimated parameters of the unconditional model are presented in Table 2. The grand mean of attendance is 649. Let us remember that this mean is calculated across all time points and all teams.

**Table 2.** Estimation of the fixed and random effects of the unconditional model

| Random effect | Variance | df | $\chi^2$ | P-value | Variance decomposition (*Percentage by level*) |
|---|---|---|---|---|---|
| $e$ | 56350 | – | – | – | 58 (level-1) |
| $r_0$ | 40995 | 17 | 87 | 0.00 | 42 (level-2) |
| **Fixed effect** | **Coefficient** | **S.E.** | **T-ratio** | **P-value** | |
| $\beta_{00}$ | 649 | 53 | 12 | 0.00 | |

*Note.* df = degrees of freedom, S.E. = standard error, $\chi^2$ = chi-square

Our main interest is focused on the variances of $e$ and $r_0$. By substituting these findings in equation (2), the ICC can be computed as:

$$\text{ICC} = \frac{Var(r_{0j})}{Var(r_{0j}) + Var(e_{ij})} = \frac{40995}{40995 + 56350} = 0.42.$$

This means that about 42% of the attendance variability is situated between teams. This value is quite high and gives us support for accomplishing consequential steps, because there is a substantial amount of attendance variability at both levels.

The results of the unconditional growth model are presented in Table 3. The fixed effects are the regression coefficient of the mean growth line averaged over all teams. The most interesting finding is that the variances of both intercepts and slopes are significantly different from zero. This means that all teams have different team-specific regression lines in terms of intercepts (i.e., different initial status) and different slopes (i.e., lines are not parallel). This fact is graphically demonstrated in Figure 1. The regression growth lines of the first ten teams are depicted here. As we can see, there are many intersections and, thus, the data cannot be easily described by one mean growth line averaged over all teams. This finding is confirmed by the estimated slope of this mean regression line $\beta_{10}$, which is not significantly different from zero ($p$ = 0.11). In other words, *attendance* increased on average by 22 persons per year, but there is simultaneously a large standard error (13) in this estimation.

**Table 3.** Estimation of the fixed and random effects of the unconditional growth model

| Random effect | Variance | Df | $\chi^2$ | P-value |
|---|---|---|---|---|
| level-1, $e$ | 37831 | – | – | – |
| level-2, $r_0$ | 50521 | 16 | 53 | 0.00 |
| level-2, $r_1$ | 1954 | 16 | 47 | 0.00 |
| **Fixed effect** | **Coefficient** | **S.E.** | **T-ratio** | **P-value** |
| intercept, $\beta_{00}$ | 544 | 69 | 7.9 | 0.00 |
| *time* slope, $\beta_{10}$ | 22 | 13 | 1.7 | 0.11 |

*Note.* df = degrees of freedom, S.E. = standard error, $\chi^2$ = chi-square

Another important question is how well the growth of attendance can be described by a linear function of time at the lower level. The degree of how appropriate the linear function is to fit the time points within teams is expressed by the percentage of explained variance at level-1. This index is similar to the coefficient of determination ($R^2$) in ordinary single-level regression and is expressed as:

$$\text{\% of variance explained } \frac{56350 \text{ (see Table 2)} - 37831 \text{ (see Table 3)}}{56350 \text{ (see Table 2)}} = 33\%.$$

This result means that about 33% of the variability within teams (across all time points) is explained by linear growth curves (lines). One way of increasing this percentage of variance explained would be to fit a different curve type (polynomial, logarithmic, etc.) to the data. However, there is currently no known theoretical support for fitting such functions. Moreover, these functions are less parsimonious (have more parameters) than a line with only two parameters.

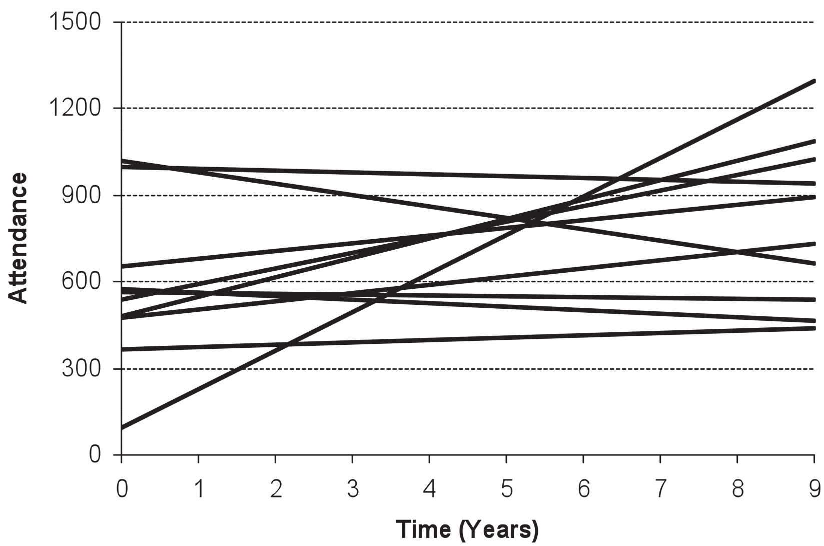


**Figure 1.** Regression lines within the first ten teams

A better approach is to introduce a time-varying independent variable at level-1 and examine the effect of this variable on the proportion of explained variance at the lower level. The *foreigners* variable was the first variable introduced to the model (4). The estimated parameters of this model are presented in Table 4. The results indicate that the *foreigners* variable has negligible positive effect on attendance (parameter $\beta_{10}$) with a large standard error of estimation ($p$ = 0.67). Furthermore, the *time* slope $\beta_{20}$ becomes non-significant ($p$ = 0.13) when the *foreigners* variable is included.

**Table 4.** Estimation of fixed and random effects of the model with *time* and *foreigners* as independent variables at level-1

| **Random effect** | **Variance** | **Df** | $\chi^2$ | **P-value** |
|---|---|---|---|---|
| level-1, $e$ | 34450 | – | – | – |
| level-2, $r_0$ | 53734 | 14 | 38 | 0.00 |
| level-2, $r_1$ | 1783 | 14 | 38 | 0.00 |
| level-2, $r_2$ | 1532 | 14 | 21 | 0.10 |
| **Fixed effect** | **Coefficient** | **S.E.** | **T-ratio** | **P-value** |
| intercept, $\beta_{00}$ | 549 | 72 | 10.63 | 0.00 |
| *time* slope, $\beta_{10}$ | 21 | 13 | 1.6 | 0.13 |
| *foreigners*, $\beta_{20}$ | 8 | 18 | 0.4 | 0.67 |

*Note.* df = degrees of freedom, S.E. = standard error, $\chi^2$ = chi-square

Both variables explain together only a slightly higher percentage of the attendance variability at level-1. This index is calculated analogically to the previous one:

$$\text{\% of variance explained} = \frac{56350 \text{ (see Table 2)} - 34450 \text{ (see Table 4)}}{56350 \text{ (see Table 2)}} = 39\% .$$

Hence, we may conclude that the *number of foreigners* is not an acceptable predictor of attendance, because it improves the percentage of the variance explained only by about 6%.

The next logical step is to incorporate the *final place* variable instead of *foreigners* into the model (4). The results of this model are given in Table 5. We have obtained consistent estimations of all regression coefficients. It was found that only the variance of the *final place* residuals is not significantly different from zero. This means that there is a negative relationship between *final place* and attendance at the lower level and this rule remains stable across all teams. The interpretation of this is that an improvement of about 1 place at the end of a season will gain 43 more spectators (on average) to a team.

Subsequently, the percentage of variance explained at level-1 was calculated based on the results presented in Table 5.

$$\text{\% of variance explained} = \frac{56350 \text{ (see Table 2)} - 28865 \text{ (see Table 5)}}{56350 \text{ (see Table 2)}} = 49\%$$

Both variables (*time* and *final place*) explain now about 49% of the total variance. All these findings suggest that *final place* is an important predictor of attendance and should remain in the final two-level model.

**Table 5.** Estimation of fixed and random effects of the model with *time* and *final place* as independent variables at level-1

| Random effect | Variance | Df | $\chi^2$ | P-value |
|---|---|---|---|---|
| level-1, $e$ | 28865 | – | – | – |
| level-2, $r_0$ | 40985 | 14 | 41 | 0.00 |
| level-2, $r_1$ | 1223 | 14 | 34 | 0.00 |
| level-2, $r_2$ | 501 | 14 | 14 | >0.50 |
| **Fixed effect** | **Coefficient** | **S.E.** | **T-ratio** | **P-value** |
| intercept, $\beta_{00}$ | 528 | 62 | 8.5 | 0.00 |
| *time* slope, $\beta_{10}$ | 26 | 11 | 2.4 | 0.03 |
| *final place*, $\beta_{20}$ | –43 | 10 | –4.3 | 0.00 |

*Note.* df = degrees of freedom, S.E. = standard error, $\chi^2$ = chi-square

The results of the final two-level model with *time* and *final place* as the time-varying variables at level-1 and *capacity* as the time-invariant variable at level-2 are presented in Table 6. Estimated coefficients are the parameters of model (5). With respect to the previous result, variance of the *final place* regression coefficient was not explained by capacity, because there was a little variability in this parameter.

**Table 6.** Estimation of fixed and random effects of the two-level model with *time* and *final place* as independent variables at level-1, and *capacity* as an independent variable at level-2

| Random effect | Variance | Df | $\chi^2$ | P-value |
|---|---|---|---|---|
| level-1, $e$ | 29264 | – | – | – |
| level-2, $r_0$ | 39634 | 13 | 31 | 0.00 |
| level-2, $r_1$ | 651 | 13 | 21 | 0.08 |
| level-2, $r_2$ | 290 | 14 | 14 | >0.50 |
| **Fixed effect** | **Coefficient** | **S.E.** | **T-ratio** | **P-value** |
| *intercept*, $\beta_{00}$ | 519.174 | 60.370 | 8.6 | 0.00 |
| *capacity*, $\beta_{01}$ | –0.045 | 0.026 | –1.7 | 0.10 |
| *time*, $\beta_{10}$ | 25.903 | 9.251 | 2.8 | 0.01 |
| *time* × *capacity*, $\beta_{11}$ | 0.009 | 0.004 | 2.2 | 0.04 |
| *final place*, $\beta_{20}$ | –40.481 | 9.183 | –4.4 | 0.00 |

*Note.* df = degrees of freedom, S.E. = standard error, $\chi^2$ = chi-square

It was shown that *capacity* is not a good predictor of *intercept*, because the regression coefficient $\beta_{01}$ is not significantly different from zero ($p$ = 0.10). Furthermore, *capacity* explains only about 3% of *intercept* variability at level-2. This finding is based on the formula given below, which is an analogy of the variance explained earlier.

$$\%\text{ of variance explained (of } intercept) = \frac{40985\text{ (see Table 5)} - 39634\text{ (see Table 6)}}{40985\text{ (see Table 5)}} = 3\%$$

On the other hand, the situation in the case of the *time* slope is quite different. Capacity is a good predictor of the time slope variability at level-2. A similar formula for the percentage of time slope variance explained is given by:

$$\text{\% of variance explained (of } time \text{ slope)} = \frac{1223 \text{ (see Table 5)} - 651 \text{ (see Table 6)}}{1223 \text{ (see Table 5)}} = 47\% .$$

The regression coefficient $\beta_{01}$ of *capacity* × *time* slope (0.009) can be interpreted as: if the capacity of a hall is about 1000 seats higher, then it increases the growth of attendance by about 9 people per year. The estimations in Tables 2–6 are slightly different, because the sample size of teams decreases when the number of model parameter increases.

By substituting the estimates from Table 6 to model (5), the final two-level model can be written in the following form:

$$\begin{aligned}(attendance)_{ij} &= 519 - 0.045(capacity)_j + 26(time)_{ij} - 40.48(final\_place)_{ij} + \\ &+ 0.009\left[(time)_{ij} \times (capacity)_j\right] + e_{ij} + r_{0j} + r_{1j}(time)ij + r_{2j}(final\_place)_{ij}.\end{aligned} \quad (6)$$

## DISCUSSION

This study with its focus on the secondary sport of basketball in the semi-periphery country of the Czech Republic provides an initial quantification of some of the most significant predictors that comprise live fan attendance in such an environment. This research is based on a 10 year study of Czech 1$^{st}$ league basketball. Over 10 years Czech 1$^{st}$ league basketball has undergone a number of significant changes. Average home game attendance has increased 29% from 557 to 721 fans per game across 12 teams for a 44 game season. The rules governing how many foreigners a team may use have increased from 2 per team to the current 8, with 4 rules changes in between. The two teams which lead the league in winning, budget, TV coverage and number of foreigners did not even play in the 1$^{st}$ league 10 years ago and are both in relatively smaller markets (14,000 and 46,000 inhabitants). The top Czech basketball league has gone from not being on TV at all 10 years ago to being on TV 53 times in the last regular season. And yet the battle to draw fans has not gotten any easier. The number of people registered with the Czech basketball federation has actually decreased by 5% (from 41,198 to 39,075) over this 10 year period as new sports such as floorball and baseball have gained a following.

A main purpose of the present study was to demonstrate the benefit of using multilevel modeling to quantifiably evaluate the components of live fan attendance. The results suggest that there is a substantial amount of the intra- and inter- attendance variability observed at both levels. Thereupon, the value of ICC (0.42) is quite high. High ICCs are typical for longitudinal studies where time is nested within teams or individuals (Singer & Willett, 2003), whereas the ICCs tend to be much lower when individuals are nested within organizations (Zhu, 1997; Luke, 2004; Pecha, 2004).

Another important issue was related to the shape of the growth curve linking time with attendance. Our study was restricted to the simplest model, where attendance is considered as a linear function of time. It was found, that lines within teams explain about 33% of the attendance variability. Obviously, more complex types of growth curves could be used to describe the development of attendance over time. For example, polynomial, exponential, logarithmic, or piecewise curves could be fitted to the data to improve the proportion of the explained variance at the lower level. However, such modifications are purely data driven and currently have unclear theoretical interpretation. This is very similar to Davis (2008) study of Major League Baseball.

Subsequently, the percentage of the variance explained at level-1 was increased by taking into account *final place* and *foreigners* as the time-varying predictors. It was found that *foreigners* is not an appropriate indicator of attendance. On the other hand, *final place* is an important predictor of attendance and together with time explained 49% of the total attendance variance observed at level-1. As we noted above, this percentage can be increased either by using a more complex growth curve or by analyzing some other time-varying variables which remain unmeasured or even unknown. For instance, team annual budget is one of these variables. Unfortunately, this variable is generally not accessible to the public in secondary sports. Although a direct effect of *foreigners* on attendance was not found, this effect can be moderated by the *final place* variable, because the correlation between these two variables is about –0.8. The theory about moderators and mediators is well-established in social sciences (Baron & Kenny, 1986) and the effect of such variables can be modeled, for example, within the path-analysis framework. Nevertheless, testing this type of relationship is beyond the scope of the present study.

Finally, the effect of the time-invariant variable, *capacity*, was investigated. It was found that *capacity* is not a good predictor for estimating variability in initial status (intercept). On the other hand, *capacity* is a substantial predictor of the time slope. The positive relationship between capacity and time slope indicates that the attendance within teams with higher hall capacity tends to grow faster than teams with low hall capacity. Obviously, this confirms the well-known fact from developmental research with individuals, where the growth of an individual is limited by a certain value which is represented by maximal personal possibilities (e.g., theoretical personal best in long jump). In our case, this upper limit is explicitly given by the hall capacity. However, the capacity was not full in any team within any year, as contrasted with Morse et al. (2008) who dealt with many full capacity halls in the NBA (a primary sport in a core country). Furthermore, the teams with lower initial status had higher dispositions for faster growth which was reflected by the negative correlation between the initial status and time slope (about –0.67). The variances of intercept and time slope can be further explained by some unmeasured time-invariant variables or by averaged time-varying variables. The negative *final place* slope remains stable across all teams, which means that the rule of better final place, higher attendance holds for all teams similarly.

## CONCLUSION

The present paper provided some new findings which allow us to formulate the following concluding remarks:

- This study shows the usefulness of multilevel analysis in demand studies of fan attendance. The multilevel approach is able to handle the type of incomplete data set that such a lengthy study generates by fitting a growth curve of the same type to each team.

- In studying a culturally secondary sport in a globally semi-periphery country this study differs from the majority of the previous demand studies on fan attendance which focused on primary sports (NBA and MLB) in core countries (the USA). The need for understanding the significance of the elements which compose fan attendance grows as we move away from primary sports because a larger portion of the revenue is generated from fan attendance.
- There is a need to further identify and define the significance of the elements which compose fan attendance. We have only focused on 5 of the elements identified by Hansen and Gauthier (1989), but the use of multilevel analysis appears to be valuable methodology for identifying the contribution of many additional elements.
- There is also a further need to conduct such demand studies across different cultural and sport settings to further differentiate between primary and secondary sports and core, semi-periphery and periphery countries.
- Finally, the effect of the number of foreign players on attendance moderated by the team's final place at the end of season should be further investigated. However, this is surely a challenge for some future follow-up research.

## REFERENCES


BARON, R., & KENNY, D. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182.

BECKER, M. A., & SULS, J. (1983). Take me out to the ballgame: The effects of objective, social, and temporal performance information on attendance at Major League baseball games. *Journal of Sport Psychology*, *5*(3), 302–313.

BRYK, A. S., & RAUDENBUSH, S. W. (1992). *Hierarchical linear models for social and behavioural research: Applications and data analysis methods*. Newbury Park, CA: Sage Publications.

CASLAVOVA, E. (2000). *Management sportu*. East West Publishing Company: East Publishing.

CROSSAN, W. (2007). *A preliminary categorization of sporting immigrants* (pp. 93–98). Prague: Charles University Press.

DAVIS, M. C. (2008). The interaction between baseball attendance and winning percentage: A VAR analysis. *International Journal of Sport Finance*, *3*(1), 58–73.

DeSCHRIVER, T. D. (2007). Much Adieu about Freddy: Freddy Adu and attendance in Major League soccer. *Journal of Sport Management*, *21*(3), 438–451.

GOLDSTEIN, H. (1995). *Multilevel statistical models*. London: Edward Arnold.

HANSEN, H., & GAUTHIER, R. (1989). Factors affecting attendance at professional sport events. *Journal of Sport Management*, *3*(1), 15–32.

HOX, J. J. (2002). *Multilevel analysis. Techniques and applications*. Mahwah (NJ): Lawrence Erlbaum Associates.

LAIRD, N. M., & WARE, J. H. (1982). Random effects models for longitudinal data. *Biometrics, 38*, 963–974.

LAWSON, R. A., SHEEHAN, K., & STEPHENSON, E. F. (2008). Vend it like Beckham: David Beckham's effect on MLS ticket sales. *International Journal of Sport Finance*, *3*(4), 189–195.

LOMBARDO, J. (2008). Attendance down, TV up for NBA. *Sports Business Journal*.

LUKE, D. A. (2004). *Multilevel modeling*. Sage University Papers Series on Quantitative Applications in the Social Sciences. Thousand Oaks, CA: Sage.

MAGUIRE, J., & PEARTON, R. (2000). The impact of elite labour migration on the identification, selection and development of European soccer players. *Journal of Sports Sciences, 18*(9), 759–769.

MAGUIRE, J., & POULTON, E. (1999). European identity politics in EURO 96: invented traditions and national habitus codes. *International Review for the Sociology of Sport, 34*(1), 17–29.

MORSE, A. L., SHAPIRO, S. L., McEVOY, C. D., & RASCHER, D. A. (2008). The effects of roster turnover on demand in the National Basketball Association. *International Journal of Sport Finance, 3*(1), 8–18.

PECHA, O. (2004). Víceúrovňová analýza hierarchicky strukturovaných dat v kinantropologickém výzkumu. [Multilevel analysis of hierarchically structured data in kinanthropological research]. *Česká kinantropologie, 8*, 65–80.

PECHA, O. (2006). Multilevel extension of the factor analysis: Important aspects for physical educational research. *Acta Universitatis Carolinae Kinanthropologica, 42*(1), 125–139.

RAUDENBUSH, S. W., & BRYK, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd Edition). Thousand Oaks, CA: Sage.

RAUDENBUSH, S. W., BRYK, A. S., CHEONG, Y. F., CONGON, R., & duTOIT, M. (2004). *HLM 6: Hierarchical linear and nonlinear modeling.* Lincolnwood, IL: Scientific Software International, Inc.

SINGER, J. D., & WILLETT, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.

SNIJDERS, T. A. B., & BOSKER, R. J. (1999). *Multilevel analysis.* London: Sage.

WOOD, T. M., & ZHU, W. (Eds.). (2006). *Measurement theory and practice in kinesiology*. Champaign, IL: Human Kinetics.

ZHU, W. (1997). A multilevel analysis of school factors associated with health-related fitness. *Research Quarterly for Exercise and Sport, 68,* 125–135.

ZHU, W., & ERBAUGH, S. (1997). Assesing change in swimming skills using the hierarchical linear model. *Measurement in Physical Education and Exercise Science, 1*, 179–201.

## NÁVŠTĚVNOST BASKETBALOVÝCH UTKÁNÍ: VÍCEÚROVŇOVÁ ANALÝZA S LONGITUDINÁLNÍMI DATY

ONDŘEJ PECHA, WILLIAM CROSSAN

SOUHRN

V této studii byla provedena víceúrovňová analýza návštěvnosti basketbalových utkání extraligy v České republice. Do analýzy bylo zahrnuto celkem 18 týmů, které alespoň jeden rok během desetiletého intervalu hrály v této nejvyšší soutěži. Jedním z cílů článku je představit víceúrovňovou analýzu jako vhodnou metodologii pro výzkumy dlouhodobé návštěvnosti v případech, kde design studie je nevyvážený. Výsledky naznačují, že během sledovaného období návštěvnost mírně rostla. Dále bylo sledováno, jaký vliv na růst návštěvnosti má celkové umístění týmu a počet zahraničních hráčů, kteří působí v týmu danou sezonu. Bylo zjištěno, že umístění týmu výrazně ovlivňuje návštěvnost, zatímco vliv počtu zahraničních hráčů na návštěvnost je zanedbatelný. Kromě těchto dvou časově proměnlivých nezávisle proměnných se vzala do úvahy i role kapacity haly při zkoumání vlivu na návštěvnost. Tato časově invariantní nezávisle proměnná plnila úlohu přirozené horní hranice návštěvnosti a limitovala míru růstu návštěvnosti uvnitř jednotlivých týmů.

**Klíčová slova:** divácká návštěvnost, víceúrovňová analýza, longitudinální studie, basketbal

Mgr. Ondřej Pecha, Ph.D.
onpecha@seznam.cz