Martínez, J.A. (2014). La influencia del primer cuarto en el resultado final en baloncesto / The influence of the first quarter on the final result in basketball. Revista Internacional de Medicina y Ciencias de la Actividad Física y el Deporte vol. 14 (56) pp. 755-769 Http://cdeporte.rediris.es/revista/revista56/artinfluencia513.htm

ORIGINAL

THE INFLUENCE OF THE FIRST QUARTER ON THE FINAL RESULT IN BASKETBALL

LA INFLUENCIA DEL PRIMER CUARTO EN EL RESULTADO FINAL EN BALONCESTO

Martínez, J.A.

Profesor Contratado Doctor. Departamento de Economía de la Empresa. Universidad Politécnica de Cartagena. España. http://joseantoniomartinez.weebly.com/, josean.martinez@upct.es

Código UNESCO: 5899 Educación Física y Deporte
Clasificación Consejo de Europa: 17 (Otras: Análisis cuantitativo del deporte)

Spanish-English translator: Martínez, J.A. josean.martinez@upct.es

RESUMEN

Esta investigación analiza la influencia que tiene el resultado en el primer cuarto de un partido de baloncesto sobre el marcador final, con el fin de detectar precozmente situaciones problemáticas para los equipos y dilucidar la probabilidad de victoria desde los primeros minutos de juego. Así se han implementado varios modelos estadísticos basados en la modelización del patrón temporal de resultados parciales: autorregresivos, trayectorias latentes, autorregresivos de trayectorias latentes y *path* logit. Los resultados indican que es preferible comenzar ganando, ya que la dependencia del estado temporal anterior del resultado es muy alta. Ir ganando en el primer cuarto es fundamental para obtener la victoria en el partido, si bien ello también depende, aunque en mucha menor medida, de si el equipo juega o no en casa y de la diferencia de potencial entre los contendientes.

PALABRAS CLAVES: baloncesto, detección precoz, predicción resultado, datos longitudinales

ABSTRACT

This research analyses the influence of the outcome in the first quarter of a basketball game on the final result. The aim is to early detect alarm states for teams and to compute the probability of win from data obtained from the first minutes of the game. To achieve this goal, several statistical models based on modeling change in partial results have been tested: autoregressive, latent growth, autoregressive latent growth and path logit. Results show it is desirable to start winning the game, because the dependence of each partial outcome from the early outcome is noticeable. Therefore, to win the first quarter is essential to win, although to a lesser extent this also depends on the home advantage and the difference of quality of the teams.

KEYWORDS: basketball, early detection, predicting results, longitudinal data

1. INTRODUCTION

In almost all scientific disciplines early detection is crucial, since it allows anticipating future events, issue diagnosis and implementing actions needed to improve the consequences of that projected state. In sports science talent identification probably is the area where more emphasis has been placed on early detection (e.g. Falk, Lidor, Lander & Lang, 2004), although there are also multiple investigations that attempt to predict the performance of teams and athletes, competitions and results through statistical models (e.g. Brown & Sokol, 2010; Caudill & Godwin, 2002).

Of particular interest are studies that aim to determine the influence of initial results on the final standings of teams, as the research performed by Lago & Casáis (2010). This research (in the field of professional football), shows that, for teams with lower budgets the performance achieved in the first matches of the competition has a strong impact on the final classification. Therefore, teams should try to start the season strong as possible; therefore the pre-season should be designed according to these objectives.

But what about predicting the outcome of a match based on performance in the first minutes of play? Lago & Casáis (2010) recommend "start winning" in the first games of the season, but could we establish similar conclusions for the final score of a match? These are the questions that this research attempts to answer. To achieve this aim 3103 basketball games from three seasons in the NBA were analyzed, using the results of the teams in the first quarter of each game as the predictor variable. Thus the aim of this study is to analyze the relationship between the performance of teams at the beginning of each match and the final result, in order to elucidate the probability of victory from the first minutes of playing.

Although there are other studies that have attempted to explain the outcome of a game based on the result in the half time (e.g. Cooper, DeNeve & Mosteller, 1992) and found a positive association between performance at both times, this is probably the first study that addresses this issue in an earliest moment: in the first quarter, which may shed more light on the importance of the first few minutes of playing in the final outcome of a match. Thus, knowing the model that governs the pattern of partial results in a match, the factors associated with change in the outcome, and its influence on the probability of victory, teams can early detect problematic situations, and they may try to adjust its performance to such events.

2. METHOD

2.1. DATA AND VARIABLES

Data were acquired at www.nbastuffer.com, where information is available on the outcome each game in the NBA in each quarter. Data about the regular seasons of 2006/07, 2007/08 and 2008-09 were obtained, so a total of 3690 games made up the initial data base.

The variables were as follows: Firstly the score of each game was recorded in 4 different moments of time: first quarter (y1), second quarter (y2), third quarter (y3) and fourth quarter (y4), in order to take a picture of the various stages of the game result depending on the time of playing. The y4 variable refers to the final score when there is no overtime. These four variables are continuous, to which was added the registration of a binary variable (y5) that determines if the home team had won the game or not, i.e., it reflected if y4> 0 with "1" and "0" otherwise.

Secondly covariates that could affect the outcome of the match were identified. To do this, the studies Arkes & Martinez (2011) and Martinez (2012) were taken as reference. These authors modeled the outcome of basketball games using various covariates, being the most important the home/away advantage, the difference of quality between teams and the type of game (depending on the quality of the teams). Other variables such as rest days, or momentum also had an effect on the result, but the incidence was much lower than the three variables mentioned. Therefore, and in order to simplify the analysis, they were disregarded in this study.

As the registration of y1-y4 was always performed with reference to the home team, then the home/away advantage was implicitly considered. The difference of quality between the teams (x1) was calculated using the difference in the winning percentage at the end of the season, according to Martínez (2012). It is a variable bounded in the interval [0,1]. This author also describes the characteristics of the variable that refers to the type of game (x2): The absolute value of the difference between the winning percentage of teams transformed as a function of an exponential parameter. As there is interaction between the home/away advantage and the quality of teams, then teams with less potential are relatively stronger at home than those with the greatest potential. Therefore, the parameter value is the sum of the potential of both teams (from 0 to 2). Thus, for teams with similar quality, the value of x2 increases if the two teams have a high winning percentage compared to if they have a low winning percentage. For example, x2 = 0.63 when two teams with winning percentage 1 and 0.5, respectively face. While x2 = 0.31 if both teams have rates of 0.6 and 0.1 respectively. That is, for the same difference of quality, x2 corrects for the quality of both teams (something like a factor of "quality" of the game). It is a variable ranged in a [0,1] interval.

Having identified the variables of the analysis, we proceeded to perform a depuration of data. The recommendations of Wilcox (2010) were followed to cut 5% from both tails of the distribution of the ordered data; the outcome of the game was chosen as the reference variable. The goal was to eliminate those games whose difference in the score was extraordinarily high for the home team or the away team depending on the distribution of the data. This could mean the presence of outliers. Thus, the range of difference were limited to (-18, +24), leaving out of the analysis of games in which the visiting team won by more than 18 points and the t home team won by more than 24. A second criterion of exclusion was related with the overtime games. These games had to be eliminated from the analysis because they had different features breaking up the homogeneity of the context of analysis. So, they lasts longer (there were several games even with 3 overtimes), so that the effect of partial score on the end result loses the temporal homogeneity. A total of 221 matches were excluded because of the two filtering criteria. Therefore, the final database consisted of 3103 games.

2.2. STATISTICAL MODELS

To study to what extent the result of the first quarter may influence the final result of a game, different statistical approaches were performed.

Firstly, the temporal pattern of results y1-y4 was modeled. The aim was to identify the best model to represent the empirical data. For this purpose, three types of models were candidates for the analysis: the autoregressive (AR), latent trajectory models (LT) and autoregressive latent trajectory models (ALT).

The AR models exemplify the transition of a variable from one time point to another, where the state at a point in time depends only on the previous state plus a random error term. By contrast, the LT models can model the trend in each instance in time, through specifying an intercept parameter and a slope, that is, focusing on the trajectory change of the individual over time, whereas in the AR model the effects of a period of time on other are equal for all individuals. On AR models we speak about fixed effects, while in LT models we speak about random effects. The integration of both models are the ALT models that cover the disadvantages of previous models, but in return they have disadvantages of identifying conditions for estimation (Bollen & Curran, 2004), which makes, for example, to specify the intercept and slope parameters correlated with y1, that is, the first time measurement is taken out of the LT of the ALT model. These models allow the integration of covariates, although some authors as Kline (2011) recommend first choosing the model that fits the data, before adding covariates model. This procedure is criticized by Hayduk (1996) who defends the estimation of the models in only one step.

In any case, the strategy of analysis was the comparison of different models via chi-square, in order to determine which was the best model fitting the data, and thus to understand the pattern of change in the outcome of a match in function of time. A more extensive and technical explanation of these type of models can be found in Bollen & Curran (2004) and Morin, Maïano, Marsh, Janosz & Nagengast (2011).

Secondly, and once understood a model that fits the temporal evolution of the score, y1 was related to y5, i.e. the scoring in the first quarter with the outcome of the game, controlling by the covariates x1 and x2. Through the implementation of logistic regressions the marginal effect of y1 on y5 was studied, in order to get a clearer conclusion about the influence of the score at the first quarter of a game on the final result.

3. RESULTS

To test the different candidate models reflecting the pattern of change in the result the MPlus 4.21 software was used, which allows estimation of models with multiple dependent and independent variables, provides fit indices and it is flexible when used to establish causal constraints . It also supports the inclusion of latent variables.

Table 1 shows the results of the different estimated models, starting with the simplest AR and LT models, following by adding covariates to these models, and ending with the implementation of the ALT model and its corresponding extension with covariates. The syntax of the MPlus programming can be obtained from the author of this study upon request, and the graphical specification of relationships among the variables is showed in the Apendix 1.

Table 1. Results of the models estimated

Model	Chi^2 (df)	p-value*	R^2 y1	R^2 y2	R^2 y3	R^2 y4
AR	0.98 (3)	0.806	-	0.429	0.531	0.583
AR_Cov	2.72 (3)	0.437	0.053	0.448	0.561	0.614
Change in R^2			0.053	0.019	0.030	0.031
LT	627.8 (5)	<0.001				
LT_Cov	801.7 (10)	<0.001				
ALT**	1.45 (2)	0.482	-	0.467	0.564	0.609
ALT_Cov**	3.03 (5)	0.694	0.053	0.464	0.573	0.625
Change in R^2			0.053	-0.003	0.009	0.016

** Non-significant values favor the null hypothesis, i.e., the fit of the model. When the model does not fit, the estimated parameters should not be interpreted. This is the reason why R-square is not showed.*
*** Estimated models with the variance of the growing factor fixed to 0, in order to keep the covariance matrix positive definite.*

As it is shown in Table 1, both the AR model as its extension with covariates obtained an adequate fit to the empirical data (p = 0.806 and p = 0.437, respectively). However, the LT models did not fit, so that modeling individual trajectories of the games does not correctly reflect the temporal pattern of change in the result, which do make the AR models, so it follows that there is a high dependence on the previous state to determine the outcome of the game in the later stage.

However, the ALT models were also adjusted after the variance of the slope of the curve path was fixed to zero. This restriction is not too problematic, and has performed in other estimates of similar models, as performed by Morin et al. (2011). In this case, we used the fixation of the variance because of the problems of convergence of the estimate, which as indicated, often arise more profusely as models are growing in size. The effect of covariates on the endogenous model is uneven; the difference in quality between the teams (x1) always has a significant effect on the variables in all models, while the quality factor of the game (x2) has not. In fact, the models were re-estimated ignoring the variable x2, and the results were virtually identical with respect to the fit and the explanatory power.

Therefore, globally, which is derived from the estimation of the six models is that the pattern of change is highly dependent on the previous state and the only influential covariate is the difference of quality between the teams, although its explanatory power is small. Thus, as shown in the change in R^2, the addition of covariates does not significantly increase the explained variance, once controlled by the result in the previous quarter. Moreover, the greatest variation is obtained in the first quarter, according to the AR_cov model, and also the ALT_cov. The difference between the two models is that the latter also models the individual differences of the games in the intercept factor, i.e., in the initial state of the first time period (when the slope is 0). As the variance of the slope were also fixed to zero, then the intra-game variability was homogenous in terms of the slope of the straight path, that is, the only differences are from where the line (intercept) begins.

Once understood the model that governs the empirical data, where it has been shown that the difference in quality between teams importantly influences the outcome of the first quarter, we have proceeded to implement a model that explicitly indicates the effect of the first quarter results on the probability of victory. To do this, we have implemented a logistic regression model path (see Muthén & Muthén, 2008), which includes the potential difference between the teams (x1) as an exogenous variable, whose estimated coefficients are shown in Table 2, and the specification diagram in Appendix 2.

Table 2. Results form the path model of logistic regression. Coefficient estimation.

	x1	y1	*Intercept*	Cases correctly classified
y1	7.706*		1.153*	
y5	4.197*	0.081*	0.496*	71.64%**

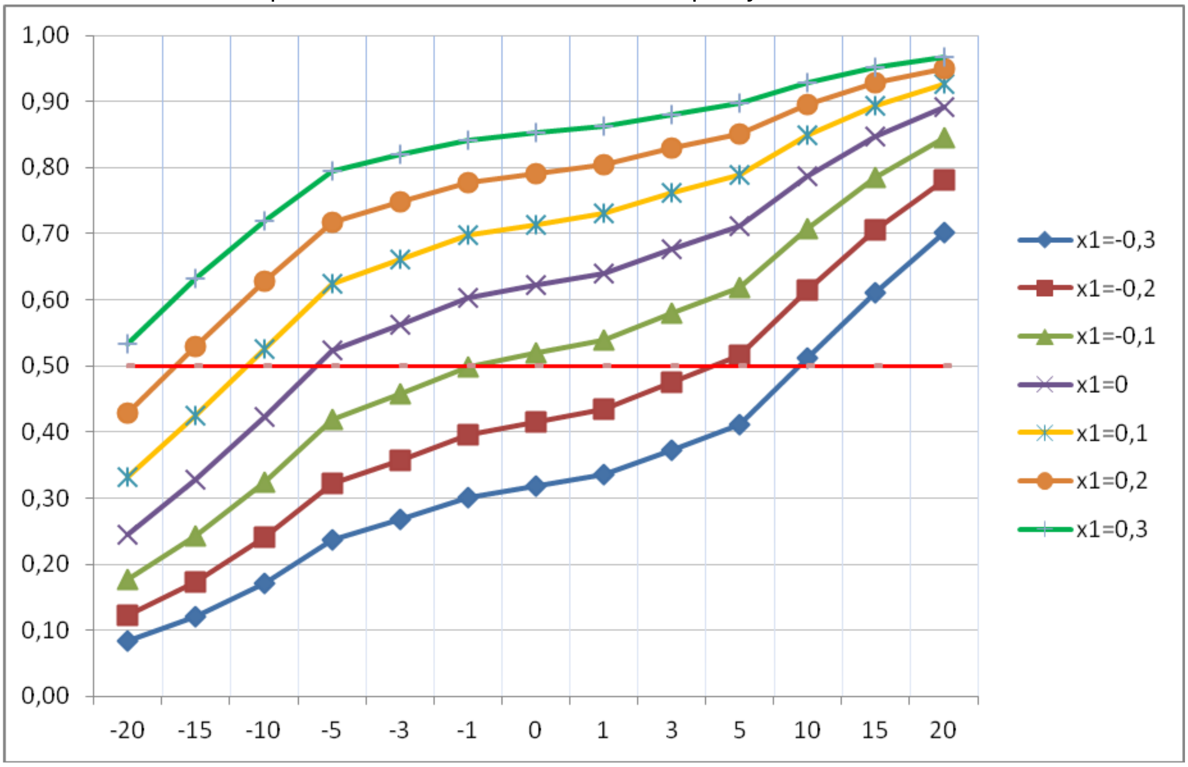
*p<0,05

** Improvement of a 5% of classification of corrected cases with respect to a model not including the covariate x1, and of a 10% with respect to a baseline model without independent variables. .

Note: the probability value of the Chi-square of the Pearson and Hosmer-Lemeshow tests were not significant (0.76 and 0.28, respectively), which supports the fit of the model.

As expected, after the estimation of the models shown in Table 1, although now the response variable is not the outcome of the game quantified in point differential, but categorized in victory or defeat, the coefficients are significant. Of interest is, therefore, to know the marginal effects of the result in the first quarter and the difference of quality of teams on the probability of victory. Figure 1 shows different probability curves based on different values for the score difference (between -20 and 20) and the teams quality difference (between -0.3 and 0.3). The red line marks the threshold probability at which the case is classified as a victory.

Figure 1. Probability curves of winning as a function of the difference in the score in the first quarter and as the difference in the quality of teams.



Nota: Curves of probability for differents values of the difference of quality between teams (x1), in function of the score difference in the first quarter (X axis). The home team is considered as the reference team, so that negative values in the X axis means that the home team was losing the first quarter. All these are probability curves for the home team; to obtain the probability curves for the away team 1- probability of home team wins has to be computed.

Given that all the data are taken as a reference considering the home team, the interpretation of Figure 1 is clear. For example, for games where the quality of the teams is similar (x1 = 0), when the score reaches tied to the first quarter, the home team has over 60% chance of winning. And that probability is still above 50% but is losing up to 5 points. For cases in which the difference of quality between the home team and the away team is more evident (x = 0.2), losing up to 15 points in the first quarter is not synonymous of higher probability of losing against win. In the opposite case (x = -0.2), when the visiting team is more powerful than the local, that threshold is 5 points. That influence is most evident to the extent that the difference of quality between teams grows.

But most interesting is even if the analysis of changes in the probability of victory, through the marginal effects of the logit model. Thus, when the home team is more powerful than the visitor (x1> 0), the slope of the curve is much steeper in the negative stretch (especially up to -5 and -3 points) than in the positive. However, when the visitor team is the more powerful a priori (x1 <0), the shape of the curves is slightly different, and symmetry exists between the ends thereof.

4. DISCUSSION AND CONCLUSSIONS

This research has tried to clarify the influence of the result in the first quarter of a basketball game on the final score, in order to early detect problematic situations for teams. Through the implementation of various statistical models based on the modeling of the temporal pattern of partial results, we can discuss the most important contributions of this study as follows:

First, the data are best explained by autoregressive models, where the state at a point in time depends on the previous state. The influence of covariates to explain the variation in these temporary states is only significant in the case of the quality difference between the teams. This variable adds explanation to the dependent variables but its effect is about 4 times lesser than the effect of the partial result in the previous temporary state. However, the potential difference between teams has its most significant effect on the scoreboard in the first quarter.

Second, the change yielded by the evolution of the game can be considered unchanged between these games, i.e. homogeneous, being the only source of heterogeneity the value of the point spread. Thus, differences between the games are best explained considering that evolution is the same for all games, with the only difference being the point where the difference, i.e., the result in the first quarter begins

Third, the home/away advantage plays a key role in determining the final result. Teams playing at home can afford to go losing in the first quarter because the probability of winning at the end of the game is higher than the probability of losing. Thus, to the extent that the quality difference between teams grows, the home team has more margin to recover the difference in the score. However, when the visiting team is significantly stronger than the home team, the home team must try to get ahead in the first quarter.

Fourth, the difference in the shapes of the probability curves indicates that the effort to cut distance on the scoreboard in the first quarter for the home team has a greater influence on the change in the probability of victory that the effort to expand the difference on the scoreboard when it is positive. However, when the visitor team is the more powerful team a priori (x1 <0), the shape of the curves is different, and symmetry exists between the ends thereof, thus indicating that the attempt to expand the difference in the score and to shorten it, brings similar increases in the probability of victory.

.

Therefore, as a final conclusion, we can say that it is better to start winning, because the temporal dependence of the result of a basketball game is very high. Go winning in the first quarter is crucial for the victory in the game, although it also depends, albeit to a much lesser extent, if the team is playing at home or away and the potential difference between the contenders. So, for visiting teams, an advantage of 10 or 15 points does not "guarantees" the victory at the end of the fourth quarter if the home team is stronger than them, while if the visitor is the strongest team, this visitor team should to continue working hardly, because short advantages favoring home teams may be unrecoverable.

The implications for the game of basketball at the strategic or tactical level are clear. Thus, teams must consider before every game the potential difference between the contenders. Although this study employs the value of the winning percentage at the end of the season, and that's a fact that obviously is not known before each game, teams can use the current winning percentage as a proxy to the real potential difference. To do this, they should wait until the season has progressed a bit, so that data is more reliable (Martínez, 2011). Once considered the potential between the teams, it seems advisable that teams go "strong" from the beginning, starting with their best players, i.e., they have to go out with the intention of opening a scoring gap as soon as possible because this difference is already very difficult to retrieve, especially for visitors even if the away team it is stronger than the home team.

Therefore, if the home team is stronger than the visitor, should act differently if, for example, in the middle of the first quarter has a differential of -8 or +8. For the first case, teams should focus to cut the difference at the end of the first quarter to increase it in the second case, since the change in the probability of victory by reducing the differential in, say 4 points, is relatively greater than increasing 4 points more the advantage. However, when the visiting team is stronger, both strategies yield a similar effect on the change in probability of victory. Thus, the coaches should have discretion to manage the rotation of players in the first quarter, making decisions regarding whether or not to keep their key players in the bench.

One limitation of this study is the possible presence of measurement error in the variable reflecting the potential difference between the teams. As indicated, before each game a number that is an approximation to the true value is obtained. Since it is well known that the measurement error attenuates the correlation between the variables, it would be expected that the true effect of the potential difference on the results were a little more important, but certainly not change significantly the results shown.

Finally, future research could delve into this subject considering the distribution of minutes between the players in the roster in every quarter. This would allow controlling some coaches' tactical decisions made regarding the distribution of playing time. These data are not available in any database, so it should be registered through the play by play data from sites like www.82games.com. If some researcher manage to make that hard work for every game, then we could obtain more specific conclusions about the variables that affect the change of the probability of victory depending on the partial score of the match.

REFERENCES

Arkes, J. & Martínez, J. A. (2011). Finally, evidence for a momentum effect in the NBA. *Journal of Quantitative Analysis in Sports*, 7 (3), Article 13.

Bollen, K. A. & Curran, P.J. (2004). Autoregressive Latent Trajectory (ALT) models: A synthesis of two traditions. *Sociological Methods and Research*, 32, 336-383.

Brown, M. & Sokol, J. (2010). An improved LRMC method for NCAA basketball prediction. *Journal of Quantitative Analysis in Sports*, 6 (3), Article 4.

Caudill, S. B. & Godwin, N. H. (2002). Heterogeneous skewness in binary choice models: predicting outcomes in the men's NCAA basketball tournament. *Journal of Applied Statistics,* 29 (7), 991-1001.

Cooper, H., DeNeve, K. M. & Mosteller, F. (1992). Predicting professions sports game outcomes from intermediate game scores. *Chance* 5, 18–22.

Falk, B., Lidor, R., Lander, Y., & Lang, B. (2004), Talent identification and early development of elite water-polo players: a 2-year follow-up study. *Journal of Sports Sciences*, 22, 347-355.

Hayduk, L. A. (1996). *LISREL Issues, Debates and Strategies*. Baltimore, MD: Johns Hopkins University Press.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford Press

Lago, C. & Casáis, L. (2010). La influencia de los resultados iniciales en la clasificación final de los equipos de fútbol de alto nivel. *Motricidad. European Journal of Human Movement*, 5(14), 107-122

Martínez, J. A. (2011). El uso del porcentaje de victorias en modelos predictivos en la NBA. *Revista Internacional de Derecho y Gestión del Deporte*, 13.

Martínez, J. A. (2012). Entrenador nuevo, ¿victoria segura? Evidencia en baloncesto. *Revista Internacional de Medicina y Ciencias de la Actividad Física y el Deporte*, 12 (48), 663-679.

Morin, A. J. S., Maïano, C., Marsh, H. W., Janosz, M. & Nagengast, B. (2011). The longitudinal interplay of adolescents' self-esteem and body image: A conditional autoregressive latent trajectory analysis. *Multivariate Behavioral Research*. 46, 157-201.

Muthén, L., & Muthén, B.O. (2008). *Mplus User's Guide (5.1)*. Los Angeles, CA: Muthén & Muthén.

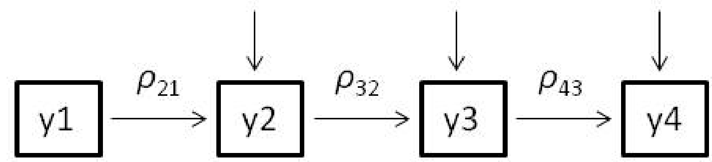
Wilcox, R. R. (2010). *Fundamentals of modern statistical methods*. Second Edition. New York: Springer.

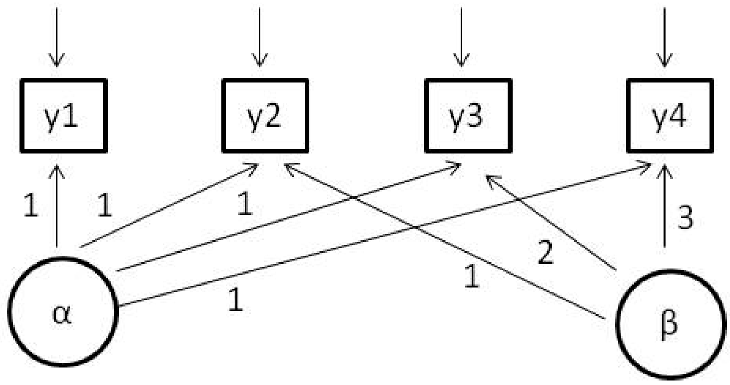
Number of total cites: 14

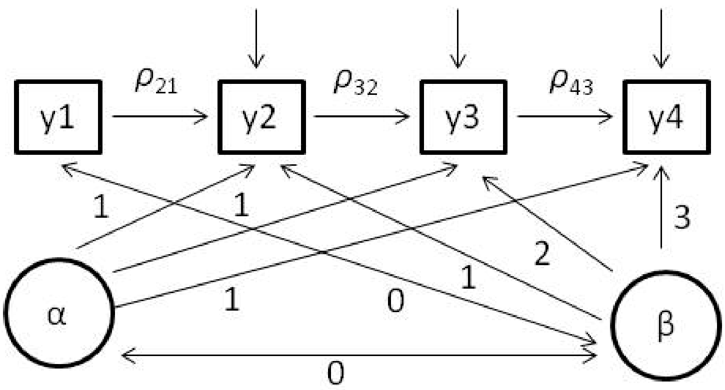
Number of citas of RIMCAFD: 1

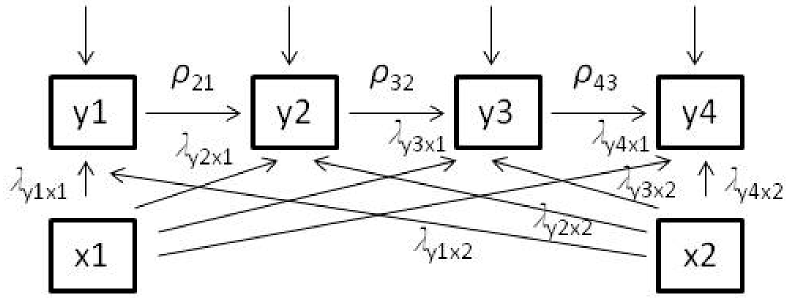
Referencias totales / Total references: 14 (100%)
Referencias propias de la revista / Journal's own references: 1 (7,14%)

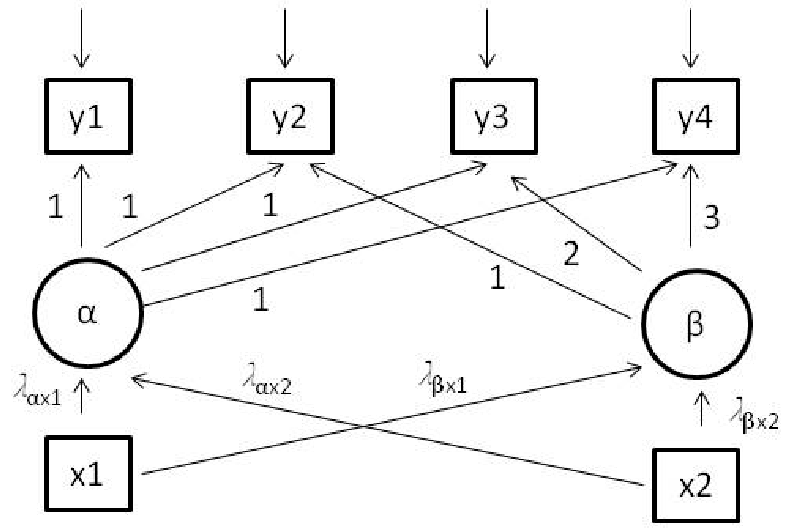
Apendix 1: Six models tested in the Table 1 (AR; LT; ALT; AR_cov; LT_cov; ALT_cov). Based on the notation of Bollen and Curran (2004).

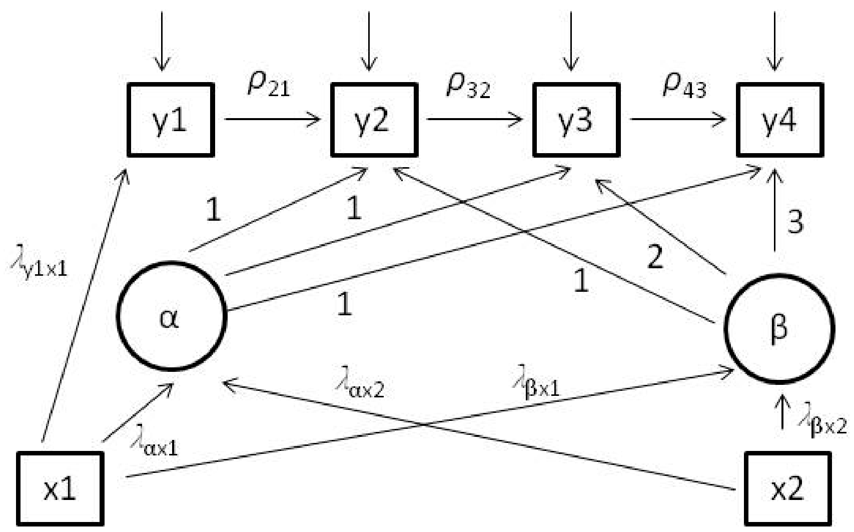







y1
y2
y3
y4
x1
x2
ρ21
ρ32
ρ43
λy1x1
λy2x1
λy3x1
λy4x1
λy1x2
λy2x2
λy3x2
λy4x2


y1
y2
y3
y4
α
β
x1
x2
1
1
1
1
1
2
3
λαx1
λαx2
λβx1
λβx2


y1
y2
y3
y4
α
β
x1
x2
ρ21
ρ32
ρ43
1
1
1
1
2
3
λy1x1
λαx1
λαx2
λβx1
λβx2

Apendix 2: Path logit model. Based on the notation of Muthén and Muthén (2008) and Kline (2011).