Available online at www.sciencedirect.com
ScienceDirect
Progress in Natural Science 19 (2009) 1643–1649
Progress in Natural Science
www.elsevier.com/locate/pnsc

Short communication

PLS path model building: A multivariate approach to land price studies – A case study in Beijing

Wenjie Wu [a,b], Wenzhong Zhang [a,*]

[a] *Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China*
[b] *Graduate University of Chinese Academy of Sciences, Beijing 100049, China*

Received 20 March 2009; received in revised form 20 April 2009; accepted 22 April 2009

Abstract

Previous studies have revealed that statistical methods can be used to analyze land-leasing parcel data. However, the conventional statistical methods used in land analysis have some limitations, especially in cases of limited observational data. In this paper, with the help of geographic information system (GIS) techniques, a partial least squares (PLS) path model is applied to study the relationship between residential land prices and various determinants through a case study of Beijing in China. From a preliminary analysis, four latent variables are selected: accessibility of the workplace center, livability, traffic, and environment facilities. The results show that the observation variables have a strong explanatory power for their corresponding latent variables, and the four latent variables have varying impacts on residential land prices. Of the latent variables, accessibility to the workplace center has the strongest impact on the residential land price.

Keywords: Land; PLS path model; GIS; Beijing

1. Introduction

The partial least squares (PLS) path model proposed by Wold et al. [1–3] can be used to study data with multiple regression problems where the number of observations is limited and correlations between the variables are high. In a PLS path model, the latent variable (LV) is an unobservable variable indirectly described by a block of observable variables X, referred to as measurement variables (MV) or indicators in this paper. A PLS path model is a novel multivariate data analysis model, which was developed in recent years and has been widely used in scientific fields. However, the model has not been widely applied to study the complex relationship between land price and its determinants. Detailed presentations on the PLS path model can be found in Wold [3] and Tenenhaus [4].

In relation to land price research, empirical statistical analysis has been proposed for describing the land price in quantitative terms and for testing the importance of its determinants [5–11]. Various statistical methods, such as multiple linear regression (MLR), canonical correlation analysis (CCA), and principal component analysis (PCA), have been adopted in recent studies [12–15]. However, problems associated with these methods include difficulties in dealing with the highly complex nature of the variable systems and the issue of limited data availability. Specifically, the CCA method aims to find the maximized correlation coefficient between the random variables, but it requires sufficient observation data and cannot calculate the coefficient when the number of observations is less than the number of variables. MLR is a frequently-used method in land price studies, but it is very susceptible to inter-correlations between variables. Additionally, MLR has a stringent requirement for observations; in order to ensure

* Corresponding author. Tel.: +86 10 64889097; fax: +86 10 64889302.
E-mail address: zhangwz@igsnrr.ac.cn (W. Zhang).

1002-0071/$ - see front matter © 2009 National Natural Science Foundation of China and Chinese Academy of Sciences. Published by Elsevier Limited and Science in China Press. All rights reserved.
doi:10.1016/j.pnsc.2009.04.016

statistical validity with a small number of observations, some explanatory variables have to be eliminated, and this could result in inaccurate results. By maximizing the covariance between the LV scores, PLS optimizes the usefulness of the analysis and provides an effective way to deal with the CCA and MLR limitations. PCA allows the transformation of a set of correlated explanatory variables into an equal number of uncorrelated variables, called principal components (PCs), which are linear combinations of the original correlated variables. The difference in the optimizing method employed in PLS and PCs is that the PCs formed are based on the X-variables only, whereas the PLS method forms the LV using the covariance between the X-variables and Y-variables, which can better explain variance in the Y-variables than PCR.

The objective of this paper is to summarize our initial attempts to apply a PLS path model to land price research in Beijing. The remainder of the paper is organized as follows: in the first part of this paper, we briefly review the PLS path model. In the second part, we apply the PLS path model to a study of the determinants of residential land prices in Beijing. The paper quantifies and accounts for static relationships between the residential land price and its determinants.

2. PLS path model

The PLS path model is described by two sub-models: (1) a measurement model relating the MV to its own LV and (2) a structural model relating some endogenous LVs to other LVs. Both models are described in this section.

2.1. The structural model

Eq. (1) is called the structural model, and it expresses the hypothesized relationships among the LVs.

$$\eta = B\eta + \Gamma\xi + \zeta \tag{1}$$

An LV, which never appears as a dependent variable, is also called an exogenous variable (LEXV). Otherwise, it is referred to as an endogenous variable (LENV). The $m \times 1$ vector η contains the latent endogenous constructs and the $n \times 1$ vector ξ consists of the latent exogenous constructs. The coefficient matrix $\boldsymbol{B}$ shows the effect of the endogenous constructs on each other, and the coefficient matrix Γ signifies the effects of the exogenous constructs on the endogenous constructs, which is called the path coefficient. The vector of disturbances ζ represents errors in the equation.

2.2. The measurement model

Eqs. (2) and (3) are factor-analytic measurement models tying the constructs to observable indicators (MVs).

$$y = \Lambda_y\eta + \varepsilon \tag{2}$$

$$x = \Lambda_x\xi + \delta \tag{3}$$

The $p \times 1$ vector $\boldsymbol{y}$ contains measurements of the endogenous constructs, and the $q \times 1$ vector $\boldsymbol{x}$ consists of the exogenous indicators' measurements. The coefficient matrices Λ_y and Λ_x show how y relates to η and x relates to ξ, respectively (η and ξ are standardized here). ε and δ represent errors in the variables (or measurement error). Generally, the measurement model comprises a simple structure such that each observed variable relates to a single latent variable.

3. Empirical research

3.1. Study area and data source

Beijing, China's capital city, serves as a key area to study the internal determinants of urban land prices in a Chinese metropolis undergoing market reforms and globalization. The city covers a land area of 16,808 km^2 and is divided into 18 districts (Fig. 1). The central city comprises four districts, namely Dongcheng, Xicheng, Chongwen, and Xuanwu. The inner suburb also comprises four districts, Chaoyang, Fengtai, Shijingshan, and Haidian. The outer suburb contains the remaining eight districts (Mentougou, Fangshan, Tongzhou, Shunyi, Changping, Daxing, Pinggu, Huairou) and two counties (Miyun and Yanqing).

However, the transfer of residential land-leasing parcels remains concentrated on the central and inner suburb urban areas (A and B regions in Fig. 1), which account for only 8.3% of the city's land area but 62.0% of the total population [16]. Therefore, we have selected the A and B regions for this study.

In our land parcel data set, 1045 open-auction residential land parcels were collected from the Beijing Land Resources Bureau's statistical yearbook for the period

Fig. 1. Sketch map of the study area in Beijing.

2004–2008. Fig. 2 shows the residential land parcels' spatial locations plotted by ArcGIS 9.2.

In addition, data for public service facilities such as government organizations (GOV), schools (SCH), hospitals (HOS), shopping centers (SHC), public transportation hubs (PTH), metro stations (MES), public squares (PUS), and parks (PAR) were collected from the statistical bureaus in each district and offered by the internal data sets. In addition, data for large- and medium-sized enterprises (ENT) were collected from the Beijing economic census data of 2004. Fig. 3 shows the spatial locations in Beijing for all of this data, plotted by ArcGIS 9.2.

3.2. Variables

Table 1 provides a basic description of the model's variables, showing the LEXV (workplace center, livability, traffic, environment) and LENV (residential land price) from the PLS path model, and illustrates the potential determinants (MVs) used in the analysis. Factors that potentially influenced the price of residential land were selected from previous literature and the specific situation in Beijing [11].

In this study, computation of the distance from each point in an MV dataset to the nearest point in the land parcel datasets, together with the calculation process for the workplace centers, was conducted using ArcGIS9.2 software and its "Proximity" and "Density" procedures, respectively.

3.3. Frame of analysis

Fig. 4 shows the relationship between two sub-models, including a measurement model and a structural model. The measurement model indicates the relationship between the MVs (Table 1) and the four LEXVs. The structural model reflects the relationship between the four LEXVs and the LENV (Table 1).

The main purpose of the PLS path model is to evaluate both the measurement and structural models. In particular, we focus on assessing the path coefficient (Γ) value in the structural model, which reflects the impact of the LEXVs on the LENV.

The PLS procedure from SmartPLS 2.0 software [17] is used in this study.

3.4. Results

3.4.1. Analysis of the measurement model

The parameter, standardized factor loading reflecting the explanatory power of MVs to their corresponding LVs, is calculated (Fig. 4) based on the MVs datasets.

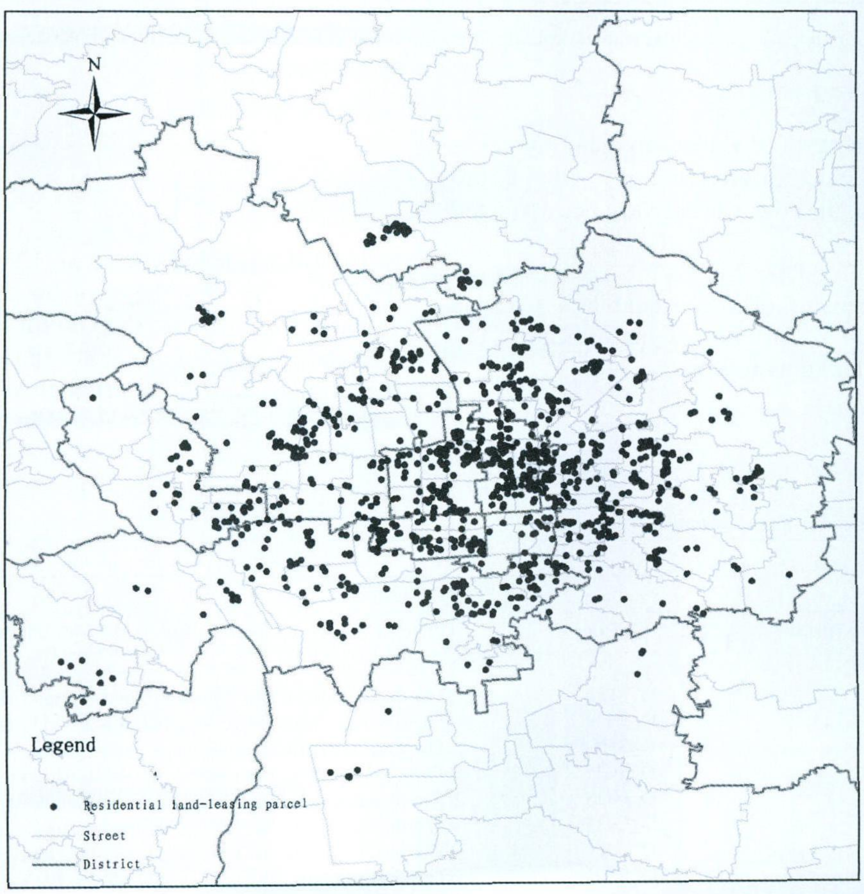

Fig. 2. Spatial distribution of the residential land parcels from 2004 to 2008 in Beijing.

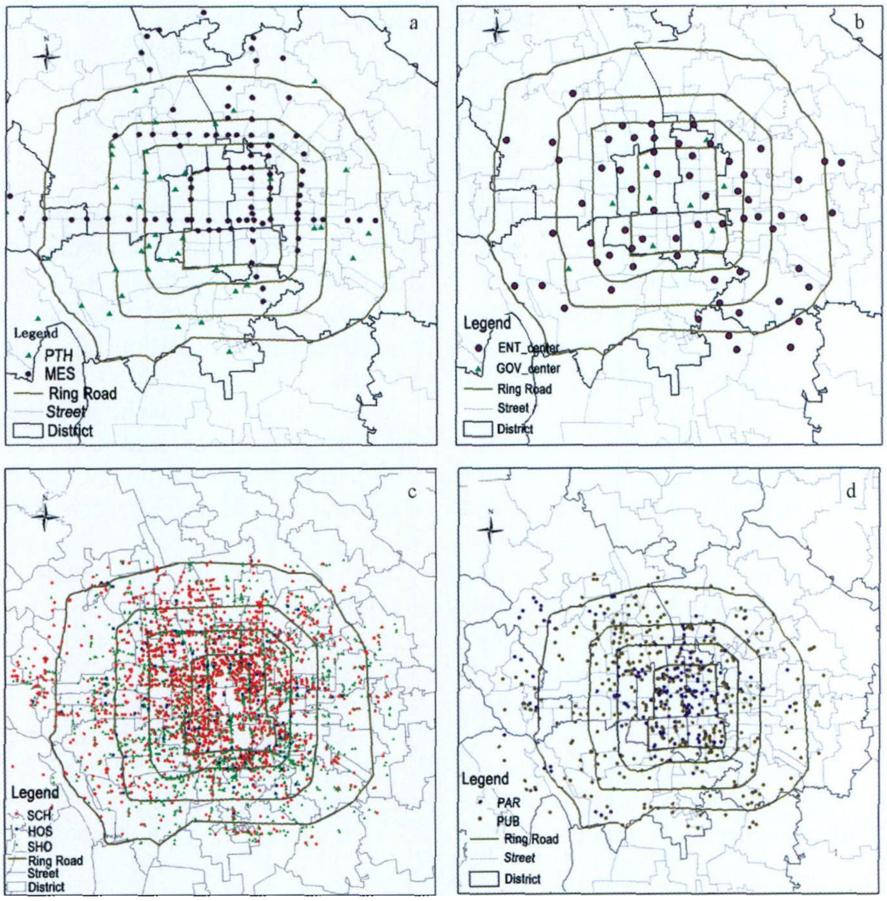

Fig. 3. Spatial distribution of public service facilities and workplace centers in Beijing. (a) Traffic facilities; (b) workplace centers; (c) livability facilities; (d) environment facilities.

Then, the reasonableness of the measurement model is evaluated by testing the composite reliability (CR), Cronbach's coefficient alpha, and the average variance extracted (AVE).

(i) CR is a measurement of scale reliability. CR can assess the internal consistency of the indicator [18]. The formula is shown as follows:

$$\rho_{\xi i} = \frac{\left(\sum \lambda_{ij}\right)^2}{\left[\sum \lambda_{ij}^2 + \sum \Theta\right]} \qquad (4)$$

In Eq. (4), $\rho_{\xi i}$ is the CR of an LV; λ_{ij} is the standardized factor loading; Θ is the error for each MV.

Table 2 shows that the CR for all the LEXVs is above 0.7. Therefore, it is reasonable to use these MVs to measure the LEXVs, and the internal consistency of the measurement model should be reasonably good.

Table 1
Description of the index system of model variables.

LV	MV	
LENV	Name	Definition
Residential land-leasing price (price)	Sig_Price	Price of residential land-leasing parcel (ø10,000/m^2)
LEXV		
Traffic	D_MES	The distance from the residential land-leasing parcel to the closest MES (km)
	D_PTH	The distance from the residential land-leasing parcel to the closest PTH (km)
Livability	D_SHC	The distance from the residential land-leasing parcel to the closest SHC (km)
	D_SCH	The distance from the residential land-leasing parcel to the closest SCH (km)
	D_HOS	The distance from the residential land-leasing parcel to the closest HOS (km)
Workplace center	D_GOV	The distance from the residential land-leasing parcel to the closest GOV (km)
	D_ENT	The distance from the residential land-leasing parcel to the closest ENT (km)
Environment	D_PAR	The distance from the residential land-leasing parcel to the closest PAR (km)
	D_PUB	The distance from the residential land-leasing parcel to the closest PUB (km)

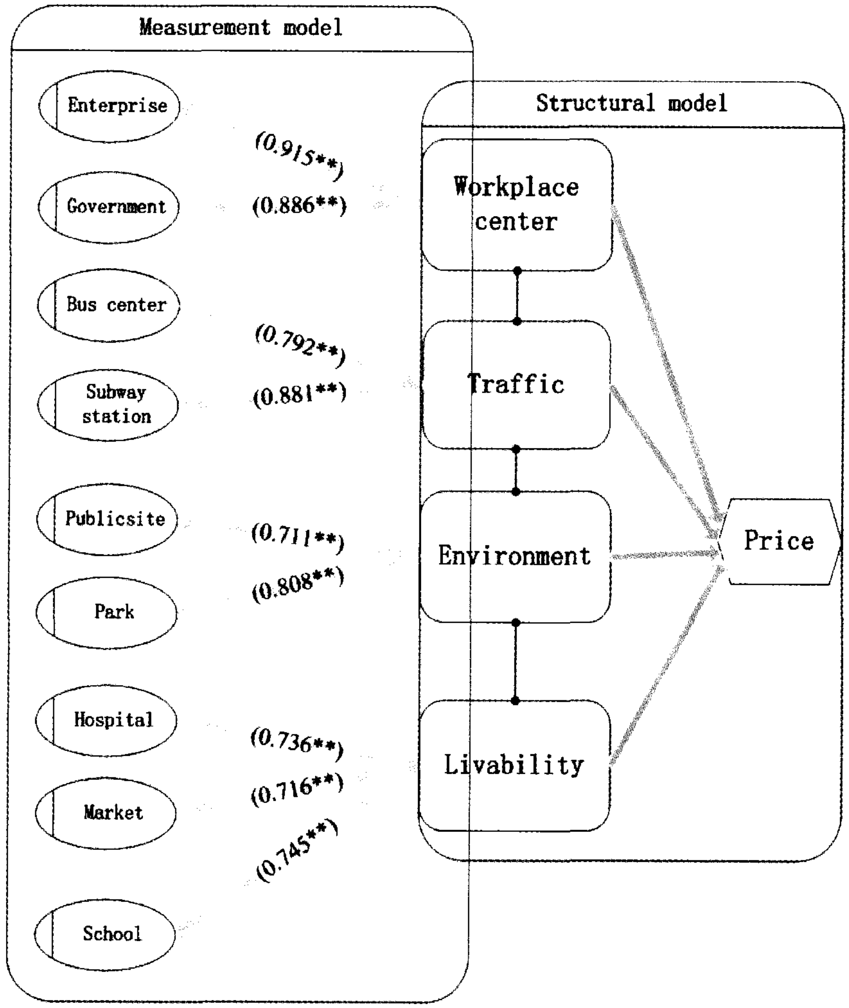

Fig. 4. PLS path model of residential land price determinants. The numbers in the () represent the standardized factor loading; ** represent significance at the 5% level.

(i) Cronbach's coefficient alpha, used to assess the extent to which the MVs can explain the constructed LVs [19], is also a reliability measurement tool. It is generally accepted that if the value of the alpha coefficient is greater than 0.70, then the constructed measurement model will have both satisfactory reliability and stability [20]. Therefore, the measurement model constructed in this PLS path model can meet the reliability requirements (Table 2).

(ii) The average variance extracted (AVE) can measure the discriminate validity between the LVs [18]. The formula is shown as follows:

Table 2
Measure model's results.

	AVE	Composite reliability (CR)	Cronbach's coefficient alpha
Workplace center	0.846	0.851	0.782
Traffic	0.766	0.782	0.736
Environment	0.631	0.699	0.701
Livability	0.736	0.708	0.726

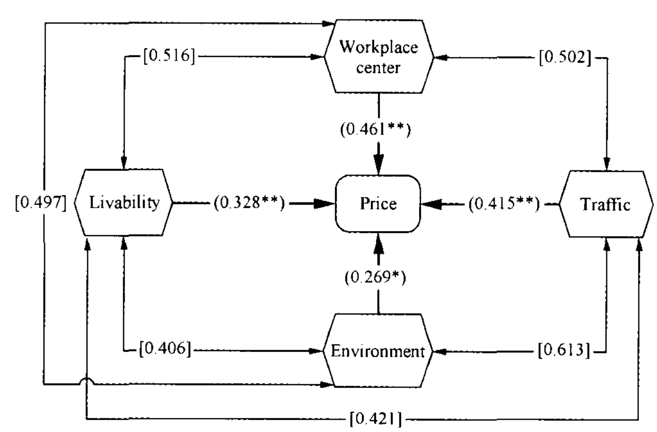

Fig. 5. Structural model results overview. The numbers in the [ ] represent the correlation coefficient among the LVs; the numbers in the () represent the path coefficients.

$$\mathrm{AVE} = \left(\sum \lambda_i^2\right) \Big/ \left(\left(\sum \lambda_i^2\right) + \left(\sum 1 - \lambda_i^2\right)\right) \quad (5)$$

In Eq. (5), λ_i^2 is the square of standardized factor loading. It is generally believed that the AVE should be at least

Table 3
Structure model's path coefficients (t-test).

LENV	LEXV				R-Square
	Workplace center	Livability	Environment	Traffic	
Price	0.461**(16.121)	0.328**(7.636)	0.269*(6.812)	0.415**(10.185)	0.791

* Significance at the 10% level.
** Significance at the 5% level.

greater than 0.5. It can be seen from Table 2 that the AVE values for the LEXVs are all greater than 0.6, which confirmed that the measurement model had relatively good discriminate validity.

3.4.2. Analysis of the structural model

The structural model can be demonstrated by the path coefficient (Γ) and the correlation coefficient among the LVs (Fig. 5).

The results show as follows:

(i) Although certain correlation relationships exist among LEXVs, their correlation coefficients are smaller than their corresponding AVE (Fig. 5). Thus, the impact of the LEXVs on the LENV is certain (Table 3).
(ii) The degree of proximity to the workplace center has the greatest impact on the residential land price. According to the structural model, the residential land price will increase by 0.461 per unit if accessibility to the workplace center rises by one unit. Similarly, the degree of proximity of the other three LEXVs also has a significant impact on the LENV, and the residential land price will increase 0.415 per unit, 0.328 per unit, and 0.269 per unit, respectively, if accessibility of traffic, environment and livability rise by one unit (Fig. 5).

4. Conclusions

The methodology of the PLS path model presented in this paper offers new opportunities to deal with the relationships between land price and its determinants. Based on the earlier analysis, the following conclusions can be drawn:

From a statistical analysis of the relationships between the residential land price and its determinants, the ability of the PLS path model to deal with the highly complex nature of variable systems and limited data availability has been confirmed. The model identified the relationship between the residential land price and its important determinants, such as accessibility of traffic, environment, and livability public facilities as well as the workplace centers. The results can provide a basis for developing quantitative insights into the residential land price decision-making process for government and real estate developers.

As the availability of land-leasing parcel data expands, the PLS path model provides an attractive method for handling multiple variables. Future research should investigate whether the indicator system is sufficiently comprehensive and whether larger datasets will improve the results for all types of land-leasing parcels in this region. Moreover, changing government policies [21] in relation to land use and its price may yield different evaluation results. Furthermore, the case pertaining to Beijing city cannot be taken as reflective of the picture for the entire country. Further studies are needed to analyze the inter-relationship between the determinants and the price of land in Chinese cities with different administrative, demographic, and geographic characteristics.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (Grant Nos. 40635026, 40571048).

References

[1] Wold H. Soft model by latent variables: the nonlinear iterative partial least squares approach. In: Perspectives in probability and statistics, papers in honour of M.S. Bartlett. London: Academic Press; 1975. p. 520–40.
[2] Wold S, Ruhe H, Wold H, et al. The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverse. SIAM J Sci Stat Comput 1984;5(3):735–43.
[3] Wold H. Partial least squares. In: Encyclopedia of statistical sciences. vol. 6. New York: Wiley; 1985. p. 581–91.
[4] Tenenhaus M. L'approche PLS. Rev Stat Appl 1999;47(2):5–40.
[5] Veldkamp A, Fresco LO. Reconstructing land use drivers and their spatial scale dependence for Costa Rica (1973 and 1984). Agric Syst 1997;55(1):19–43.
[6] Verburg PH, Chen YQ. Multiscale characterization of land-use patterns in China. Ecosystems 2000;3(4):369–85.
[7] Bowes DR, Ihlanfeldt KR. Identifying the impacts of rail transit stations on residential property value. J Econ 2001;50(1):1–25.
[8] Fehribach FA, Rutherford RC, Eakin ME. An analysis of the determinants of industrial property valuation. J Real Estate Res 1993;8(3):365–77.
[9] Sirpal R. Empirical model of the relative impacts of various sizes of shopping centers on the value of surrounding residential properties. J Real Estate Res 1994;9(4):487–505.
[10] Walden ML. Magnet schools and the differential impact of quality on residential property values. J Real Estate Res 1990;5(2):221–30.
[11] Zheng SQ, Kahn ME. Land and residential property markets in a booming economy: New evidence from Beijing. J Urban Econ 2008;63(2):743–57.
[12] Horst P. Relations among m sets of variables. Psychometrika 1961;26(2):129–49.
[13] Carroll JD. A generalization of canonical correlation analysis to three or more sets of variables. In: Proc 76th Conv Am Psych Assoc; 1968. p. 227–8.

[14] Gardner RH. Pattern, process, and the analysis of spatial scales. In: Ecological scale: theory and applications. New York: Columbia University Press; 1998. p. 7–34.
[15] Öztürk F, Akdeniz F. Ill-conditioning and multicollinearity. Linear Algebra Appl 2000;321(1–3):295–305.
[16] Beijing Municipal Bureau of Statistics. Beijing Statistical Yearbook 2007. Beijing: Chinese Statistics Press; 2008. p. 1–26 [in Chinese].
[17] Ringle CM, Wende S, Will A. Customer segmentation with FIMIX-PLS. In: Proceedings of the PLS'05 international symposium. SPAD Test&go, France; 2005. p. 507–14.
[18] Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Marketing Res 1981;18(3):39–50.
[19] Pedhazur EJ, Schmelkin LP. Measurement, design, and analysis: an integrated approach. Hillsdale: Lawrence Erlbaum; 1991. p. 104.
[20] Churchill GA. Marketing research: methodological foundation. 5th ed. Fort Worth: The Dryden Press; 1995. p. 539.
[21] Brueckner JK. Growth controls and land values in an open city. Land Econ 1990;66(3):237–48.