Analyzing the Consumption of Various Pharmaceuticals in Turkey: A Panel Approach

Türkiye'deki Çeşitli İlaçların Tüketim Analizi: Bir Panel Yaklaşımı

Barış AŞIKGİL[a]

[a]Department of Statistics, Mimar Sinan Fine Arts University Faculty of Science and Letters, İstanbul

Geliş Tarihi/*Received:* 30.10.2012
Kabul Tarihi/*Accepted:* 16.01.2013

Yazışma Adresi/*Correspondence:*
Barış AŞIKGİL
Mimar Sinan Fine Arts University Faculty of Science and Letters, Department of Statistics, İstanbul, TÜRKİYE/TURKEY
basikgil@msgsu.edu.tr

ABSTRACT Objective: The aim of this study is to analyze the pharmaceutical consumption in Turkey and investigate different types of models explaining this consumption by using the panel approach. **Material and Methods:** Different panel data models were constructed by using the panel data regression in order to explain the relationship between the sales and the total production costs of various pharmaceutical forms supplied by one of the leading pharmaceutical manufacturing companies in Turkey. In this study, 5 different forms of pharmaceuticals concerning general or beta-lactam pharmaceuticals were examined over 72 periods. **Results:** It can be seen that pharmaceutical consumption including general and beta-lactam pharmaceuticals was generally affected by time. Unit root analysis was performed for the variables used in this study and it was determined that the variables were stationary. Moreover, log-linear forms of the models were used in order to stabilize the variations in the variables. The findings implied that the dynamic panel two-way fixed effect model explained the relationship between the sales and the total production costs better than the others because of R^2 = 0,952, RMSE = 0,242, J-stat = 0,175 and the Hausman test result. **Conclusion:** The sales of various pharmaceutical forms over the long period were modeled by using the panel data regression. This analysis allows modeling of different approaches. It can be concluded that dynamic panel data regression models can also give good results in the studies related to health sciences depending on time.

Key Words: Regression analysis; linear models; beta-lactams

ÖZET Amaç: Bu çalışmanın amacı, Türkiye'deki ilaç tüketiminin analiz edilmesi ve bu tüketimi açıklayan farklı biçimdeki modellerin panel yaklaşımı kullanılarak araştırılmasıdır. **Gereç ve Yöntemler:** Türkiye'nin önde gelen ilaç imalat firmalarından biri tarafından sağlanan, çeşitli farmasötik formlardaki ilaçların satış ve toplam üretim maliyetleri arasındaki ilişkiyi açıklamak için panel veri regresyonu kullanılarak farklı panel veri modelleri oluşturulmuştur. Bu çalışmada, genel ya da beta-lactam grubuna ilişkin 5 farklı farmasötik formdaki ilaç 72 dönem üzerinden incelenmiştir. **Bulgular:** Genel ve beta-lactam grubu farmasötikleri içeren ilaç tüketiminin, ana hatlarıyla zaman tarafından etkilendiği görülebilir. Bu çalışmada kullanılan değişkenler için birim kök sınaması yapılmış ve değişkenlerin durağan oldukları saptanmıştır. Bunun yanında, değişkenlerdeki değişimleri dengelemek amacıyla incelenecek modellerin log-doğrusal biçimleri kullanılmıştır. Bulgular, dinamik panel iki-yönlü sabit etkiler modelinin satışlar ve toplam üretim maliyetleri arasındaki ilişkiyi, R^2 = 0.952, RMSE = 0.242, J-stat = 0.175 ve Hausman test sonucu nedenleriyle diğer modellerden daha iyi açıkladığını göstermiştir. **Sonuç:** Çeşitli farmasötik formlardaki ilaçların uzun dönemdeki satışları panel veri regresyonu kullanılarak modellenmiştir. Bu analiz, farklı yaklaşımların modellenmesine imkan sağlar. Sonuç olarak, dinamik panel veri regresyon modellerinin zamana bağlı sağlık bilimleri ile ilişkili çalışmalarda da iyi sonuçlar verebildiği kararına varılabilir.

Anahtar Kelimeler: Regresyon çözümlemesi; doğrusal modeller; beta-laktamlar

Turkiye Klinikleri J Biostat 2013;5(2):55-60

Drugs are chemical substances providing health care or performance enhancement for human life. These substances having medicinal properties are called pharmaceuticals. It is well known with the help of extensive surveys that different drugs can be recommended for each different disease in medical sciences. For this reason, the terms "pharmacology" and "medication" have great importance nowadays.

Pharmaceuticals can be defined into two forms which are general pharmaceuticals and beta-lactam pharmaceuticals. General pharmaceuticals contain mainly tablets, gels, ampoules, syrups etc. However, beta-lactam pharmaceuticals are known as two types of antibiotics which are cephalosporins and penicillins. The production of these substances has to be reliable and high quality for health care. Therefore, pharmaceutical production takes place in suitable laboratory conditions by using high technology.

Tablets, which are pharmaceutical solid dosage forms, contain mixtures of active and inactive substances usually in powder form.[1] Powders are prepared by using two types of granulation given as dry and wet. Moreover, tablets have different types of coating. The most important ones are sugar and film. Sugar coating is regarded as the oldest method for tablet coating. It relies mainly on the use of sucrose. However, film coating, which is the deposition of a thin polymeric film onto the tablet, is a popular alternative to the former. Film coating can be preferred for some reasons. It provides preventing unpleasant taste of the tablet and swallowing it easily for the patient.[2] Some other general pharmaceuticals, gels (semi-solid form), ampoules (liquid form, protection from air) and syrups (liquid form, aqueous solutions containing sugar), also have different properties.

Beta-lactam antibiotics, which are named for the beta-lactam ring in their chemical structure, are used for the treatment of bacterial infections.[3] The cephalosporins are one class of these antibiotics whose purpose is to prevent the public health. They are used in two forms, injectable or oral. The penicillins are the other class of these antibiotics and they inhibit the division of bacteria like the other antibiotics. An important thing about the beta-lactam antibiotics is that these antibiotics can have adverse affects on some patients such as severe allergic reactions.[4] Therefore, the use of them should be very carefully.

Time series data, growths or changes over time are seen not only in econometrics but also in biological and medical sciences. Therefore, the methods related to time series data analysis can be used in these areas.[5] In empirical researches, data sets can have both cross-sectional and time series dimensions instead of pure time-series. These data sets are called panel (longitudinal) data sets which consist of independently sampled observations.[6] Panel data is used for controlling the individual heterogeneity, obtaining more degrees of freedom, more information, more efficiency and less collinearity.[7]

The study presents the results of different approaches obtained by using the panel data regression in one area of the health sciences. The rest of the paper is organized as follows. In section 2, the panel data regression, the sample and the statistical analysis are given. In section 3, the results of the several analyses are presented. Section 4 is the discussion part.

MATERIAL AND METHODS

PANEL DATA REGRESSION

In panel approach, different cross-sectional units are surveyed over time. If each unit has the same number of observations, the approach is called a balanced panel. Otherwise, it will be an unbalanced panel. Also, a panel is defined long when the number of time periods is greater than the number of cross-sectional units.[8]

The panel data regression is considered in order to analyze panel data sets which can be modeled as

$$y_{it} = \alpha + X_{it}'\beta + u_{it} \quad i = 1, 2, ..., N; \;\; t = 1, 2, ..., T \qquad (1)$$

where y_{it} denotes itth observation in the dependent variable, X_{it} denotes itth observation in K explanatory variables and u_{it} denotes the error term. α is a scalar and β is the $K \times 1$ vector of the unknown pa-

rameters. The cross-section dimension is given as *i* and the time series dimension is given as *t*. The sample consists of $N \times T$ observations.[7]

There are mainly three approaches to estimate the panel data regression model. These are pooled ordinary least squares (POLS), fixed effects (FE) including least squares dummy variables (LSDV) and random effects (RE).[8]

Ordinary panel data regression models are specialized cases of linear regression models for panel data. The most important ones are known as unobserved effects panel data regression models. The unobserved effects are examined in the error term.[7,8] Therefore, the error term is called composite error and given by

$$u_{it} = \mu_i + \nu_{it} \quad i = 1,2,...,N; \quad t = 1,2,...,T \tag{2}$$

where μ_i denotes the unobservable time-invariant effect and ν_{it} denotes the remainder error with independent and identically distributed IID$(0,\sigma^2_\nu)$. If μ_i is fixed, the model is called FE. However, the model is called RE when $\mu_i \sim$ IID$(0,\sigma^2_\mu)$ and μ_i is independent of ν_{it}. In addition, there can be unobservable time effects (λ_t) in the composite error.[7]

All panel data regression models used in this study are given in Table 1. Log-linear forms of the models are considered in order to stabilize the variations in the variables of the panel data regression. If one or more lagged (past) values of the dependent variable are evaluated as regressors with the other explanatory variables, then the models are called dynamic panel data regression models.[8] In these models, there is a basic problem because of the correlation between the error term (u_{it}) and the lagged dependent variable ($y_{i,t-1}$). This problem is known as a violation of one of the assumptions given in regression models. Therefore, the parameter estimates obtained by OLS become biased and inconsistent in this situation. Generalized method of moments (GMM) is proposed to overcome the problem.[7]

The use of GMM is preferable for dynamic panel data regression models since it can provide efficient results in the presence of heteroscedasticity and autocorrelation problems. However, the performance of GMM can be influenced by instrumental variables (IV).[9] By considering the matrix form of equation (1),

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{u} \tag{3}$$

and the matrix of IV ($\mathbf{Z}$) including *L* instruments, the GMM estimator is given by

TABLE 1: Dynamic panel data regression models.

Name	Model	Error term
Pooled	$\log y_{it} = \alpha + \beta_1 \log x_{it} + \beta_2 \log y_{i,t-1} + \gamma t + u_{it}$	$u_{it} \sim \text{IID}\left(0, \sigma^2_u\right)$
One-way FE	$\log y_{it} = \alpha + \beta_1 \log x_{it} + \beta_2 \log y_{i,t-1} + \gamma t + u_{it}$; $u_{it} = \mu_i + \nu_{it}$	μ_i is fixed $\nu_{it} \sim \text{IID}\left(0, \sigma^2_\nu\right)$
Two-way FE	$\log y_{it} = \alpha + \beta_1 \log x_{it} + \beta_2 \log y_{i,t-1} + u_{it}$; $u_{it} = \mu_i + \lambda_t + \nu_{it}$	μ_i and λ_t are fixed $\nu_{it} \sim \text{IID}\left(0, \sigma^2_\nu\right)$
One-way RE	$\log y_{it} = \alpha + \beta_1 \log x_{it} + \beta_2 \log y_{i,t-1} + \gamma t + u_{it}$; $u_{it} = \mu_i + \nu_{it}$	$\mu_i \sim \text{IID}\left(0, \sigma^2_\mu\right)$ $\nu_{it} \sim \text{IID}\left(0, \sigma^2_\nu\right)$
Two-way RE	$\log y_{it} = \alpha + \beta_1 \log x_{it} + \beta_2 \log y_{i,t-1} + \gamma t + u_{it}$; $u_{it} = \mu_i + \lambda_t + \nu_{it}$	$\mu_i \sim \text{IID}\left(0, \sigma^2_\mu\right)$ $\lambda_t \sim \text{IID}\left(0, \sigma^2_\lambda\right)$ $\nu_{it} \sim \text{IID}\left(0, \sigma^2_\nu\right)$

$$\hat{\boldsymbol{\beta}}_{GMM} = (\mathbf{X}'\mathbf{Z}\mathbf{W}\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\mathbf{W}\mathbf{Z}'\mathbf{y}, \quad (4)$$

where $\mathbf{W}$ is an $L \times L$ weight matrix. In equation (4), the choice of the optimal weight matrix is important.[10] The IV estimator, which is a special case of the GMM estimator, is defined by

$$\hat{\boldsymbol{\beta}}_{IV} = \left(\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}. \quad (5)$$

This estimator is also known as two-stage least squares (2SLS) estimator and it must be $L \geq K$ for the identification of equation (4) and equation (5).[10]

PHARMACEUTICAL SAMPLE

In this study, a long and balanced panel data is examined. The real data set supplied by one of the leading pharmaceutical manufacturing companies in Turkey consists of two dimensions. The first dimension is cross-section which contains 5 different and randomly selected forms of pharmaceuticals:

- General pharmaceuticals

1. Tablets (dry and wet granulation, direct compression)
2. Film coated tablets
3. Sugar coated tablets
4. Syrups

- Beta-lactam pharmaceuticals

5. Cephalosporins (injectable and oral)

The second dimension is time which contains 72 periods from November 2004 to October 2010. The two-dimensional structure of the balanced panel data is shown in Table 2. The variables used in this panel are sales (y) and total production costs (x). Sales are considered as the consumption of the pharmaceuticals in Turkey. Total production costs consist mainly of laboratory analysis (microbiological examination), quality control analysis, machine run time etc.

RESULTS

In this study, the sample has 360 observations. Each cross-section (pharmaceutical) includes 72 observations. EViews 7 is used to analyze this balanced panel data set. Firstly, area graphs related to the sales of tablets, film coated tablets, sugar coated tablets, syrups and cephalosporins respectively are given in Figure 1 in order to observe the trends of sales. The descriptive statistics including minimum, maximum and mean±standard deviation sales of each pharmaceutical are given in Table 3. It can be said in view of Table 3 that tablets are consumed at most between the years 2004 and 2010 with the mean sales 2635576±1103411. Then, trend and seasonality analyses for this data set are performed and the results are not given here. It can be concluded that the data set has a trend, but it doesn't include month specific effect. Therefore, the trend term "*t*" is used in the models given in Table 1. However, only the two-way FE model doesn't have this term because of the matrix singularity problem for the estimation method.

Before several panel data regression models are examined for explaining the pharmaceutical consumption in Turkey, the stationarity of each variable should be checked. Therefore, several unit root tests are performed for each variable and the results are given in Table 4. Both the variables "sales" and "costs" are stationary because of $p < 0.05$ for all unit root tests which means the rejection of nonstationarity.

TABLE 2: The structure of the balanced panel data.

	Forms of pharmaceuticals									
	1		2		3		4		5	
Time	y	x	y	x	y	x	y	x	y	x
Nov04	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
Oct10	.	.	.	.	.	.	.	.	.	.

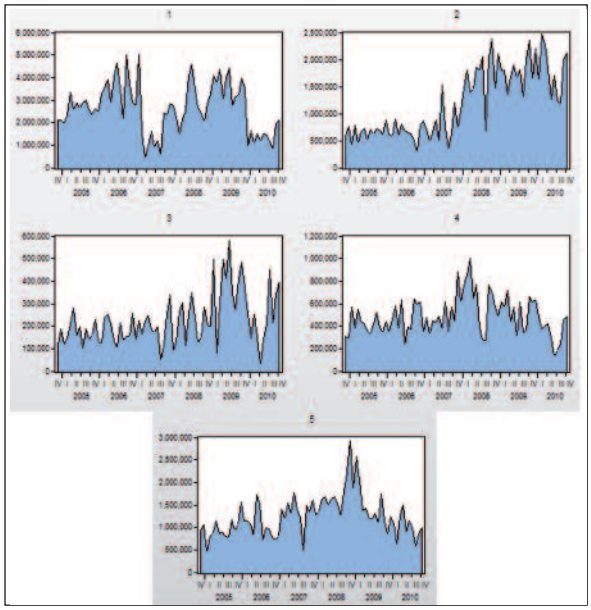

FIGURE 1: The sales of each pharmaceutical.

TABLE 3: Descriptive statistics for the sales of each pharmaceutical.

Pharmaceutical	Minimum	Maximum	Mean ± Std. dev.
Tablets	441740	5030457	2635576 ± 1103411
Film coated tablets	301781	2470036	1194279 ± 618909.5
Sugar coated tablets	30000	577139	224290.1 ± 113470.4
Syrups	135256	1001205	481766.2 ± 170293.4
Cephalosporins	441122	2908850	1242595 ± 450944.5

The logarithmic transformation is considered for each variable and different types of dynamic panel data regression models are constructed. Panel GMM with 2SLS instrument weighting matrix and White diagonal standard errors and covariance is used to obtain efficient parameter estimates. Some results related to this analysis are presented in Table 5. In this table, coefficient of determination (R^2) and root mean squared error (RMSE) give some information about the goodness of fit of a dynamic panel data regression model. J-statistic states that a model fits the data well or not (a misspecified model). It comes from the chi-square distribution. Durbin-Watson (DW) statistic is used to determine the presence of autocorrelation in the residuals. It can be said in view of Table 5 that the two-way FE model is better than the others since $R^2 = 0.952$ is the greatest, RMSE = 0.242 and J-stat = 0.175 are the smallest for the model. It can be concluded that there is not an autocorrelation problem for the model since DW test statistic is close to 2. Moreover, the Hausman test is performed to confirm that the preferable model is RE or FE. Since the result of this test is $p = 0.000 < 0.05$, the FE model can be chosen in this study. Therefore, the parameter estimates for the one-way and two-way FE models are given in Table 6. All parameter estimates except the constant term are statistically significant ($p < 0.05$). This result points out that the total production costs affect the consumption positively like the sales of the past periods. Residuals, actual and fitted values related to the two-way FE model are displayed in Figure 2. The actual and fitted values are nearly same. Therefore, it can be said that the two-way FE model explains the relationship among the variables in the pharmaceutical data very well.

DISCUSSION AND CONCLUSION

Ordinary regression modeling approaches are not valid for the cross-sectional data including time.

TABLE 4: Results related to several unit root tests for each variable.

Variable	Unit root	Method	Stat	P value
Sales	Common	Levin, Lin and Chu	-10.632	0.000
		Breitung	-7.489	0.000
	Individual	Im, Pesaran and Shin	-8.128	0.000
		ADF-Fisher	76.601	0.000
		PP-Fisher	86.999	0.000
Costs	Common	Levin, Lin and Chu	-7.992	0.000
		Breitung	-2.365	0.009
	Individual	Im, Pesaran and Shin	-7.075	0.000
		ADF-Fisher	66.370	0.000
		PP-Fisher	94.225	0.000

TABLE 5: Results related to the dynamic panel data regression models.

Model	R^2	RMSE	J-stat	DW-stat
Pooled	0.887	0.330	4.540	2.186
One-way FE	0.917	0.285	2.309	1.976
Two-way FE	0.952	0.242	0.175	1.754
One-way RE	0.887	0.330	4.540	2.186
Two-way RE	0.905	0.293	3.768	2.139

TABLE 6: Parameter estimates of the FE models.

Model	Parameter	Estimation	Std. error	P value
One-way FE	α	0.250	0.883	0.777
	β_1	0.638	0.050	0.000
	β_2	0.403	0.085	0.000
	γ	-0.008	0.001	0.000
Two-way FE	α	0.872	0.900	0.334
	β_1	0.626	0.050	0.000
	β_2	0.346	0.087	0.000

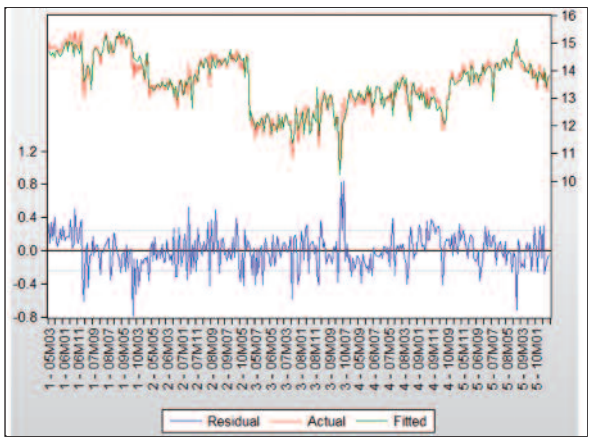

FIGURE 2: The view of residuals, actual and fitted values for the two-way FE model.

Therefore, panel data regression has been introduced for this type of data. The panel data regression provides evaluating different cross-sections on the same time interval together and estimating the best model. Panel data regression analysis can also be used in the clinical studies depending on time.

In this paper, 5 different and independent forms of pharmaceuticals have been analyzed by monthly between 2004 and 2010. Pharmaceuticals have an important role in medical sciences. The use of them is a need for the patient. For this reason, it is expected that the pharmaceutical consumption increases over time periods. It can be said in view of Figure 1 that the sales of the tablets have decreased when the sales of the film coated tablets have increased in recent years. This is an expected result because the latter provide convenience of use for the patient.

However, it is also expected that the total production costs increase because of high quality standards. The question is whether the pharmaceutical consumption still increases as the total production costs of the pharmaceuticals increase. The dynamic panel data regression models can be considered for the analysis of the question. The best model satisfying the regression assumptions should be determined in order to answer the question correctly. It has been seen in this paper that the answer of the question is affirmative. This is a sign that the pharmaceuticals are indispensable in human life.

In conclusion, it can be emphasized that dynamic panel data regression analysis is preferable for estimating the models in the studies related to medical sciences depending on time.

Acknowledgement

The author would like to thank a specialist, whose name is not given because of the privacy policy of the pharmaceutical company, for providing data and helpful information about pharmaceuticals.

REFERENCES

1. Kibbe AH. Introduction. Handbook of Pharmaceutical Excipients. 3rd ed. Washington: American Pharmaceutical Association; 2000. p.1-14.
2. Porter SC. Coating of pharmaceutical dosage forms. In: Troy DB, Hauber MJ, eds. The Science and Practice of Pharmacy. 21st ed. Philadelphia: Lippincott Williams and Wilkins; 2006. p.929-38.
3. Holten KB, Onusko EM. Appropriate prescribing of oral beta-lactam antibiotics. Am Fam Physician 2000;62(3):611-20.
4. Rossi S. Drugs and therapeutics. Australian Medicines Handbook. 2nd ed. Adelaide: Australian Medicines Handbook Pty Ltd; 2000. p.2-9.
5. Agung IGN. Continuous growth models. Time Series Data Analysis Using EViews. 1st ed. Singapore: John Wiley and Sons (Asia); 2009. p.25-6.
6. Wooldridge JM. Pooling cross sections across time: simple panel data methods. Introductory Econometrics: A Modern Approach. 3rd ed. New York: Thomson South-Western; 2005. p.444-80.
7. Baltagi BH. Introduction, The one-way and the two-way error component regression models. Econometric Analysis of Panel Data. 3rd ed. Chichester: John Wiley and Sons Ltd; 2005. p.4-52.
8. Gujarati DN, Porter DC. Panel data regression models. Basic econometrics. 5th ed. Singapore: The McGraw-Hill Companies; 2009. p.591-616.
9. Roodman D. A Note on the theme of too many instruments. Oxford Bulletin of Economics and Statistics 2009;71(1):135-58.
10. Baum CF, Schaffer ME, Stillman S. Instrumental variables and GMM: estimation and testing. The Stata Journal 2003;3(1):1-31.