ESTIMATING WIND SPEED IN BLACK SEA REGION WITH PANEL DATA ANALYSIS

Zeynep ATLI[1], Alper SİNAN[1]

[1]Sinop University Science and Agriculture Faculty, Department of Statistics, Sinop-Turkey, zatli@sinop.edu.tr, alpsin@sinop.edu.tr

ABSTRACT

Panel data analysis is a common method used especially in economic research and in other research in which time/unit relationship is significant. This study aims at predicting average wind speed during the years 2009-2010 in the cities of Sinop, Samsun, Ordu, Kastamonu, Bartın, Zonguldak and Karabük of Blacksea Region. In order to analyze the obtained data, of all panel data analysis methods, Random Effect Model and Fixed Effect Model have been used. Compatibility of these data sets have been compared via the use of graphics. The results have shown that Fixed Effect Model has turned out to be more effective in analyzing data according to units.

Key Words: *Panel data, Random Effects Model, Wind speed, Fixed Effect Model*

1. INTRODUCTION

Three types of data that can be used in econometrics research are time series, cross section and pooled data (Gujarati 2004). Time series is a set of observations that a variable takes a value in different times. Cross section refers to a data set that belongs to one or more variables at a single time point. Pooled data is an element of both time series and cross section data. Panel data is a special type of pooled data and panel data set refers to a data set has a data on cross section units over the time.

Panel data has some advantages over cross sectional and time series data. Panel data controls for individual heterogeneity. Time series and cross sectional studies don't control the heterogeneity. Time series models and cross sectional models which are not controlled the heterogeneity, run the risk of obtaining biased result (Baltagi 2005). Panel data has large number of observation over time series and cross sectional data. Increasing the number of observation increase degrees of freedom. Since explanatory variable vary over both time dimension and cross sectional dimension, multicollinearity reduce among explonarity variables in panel data. Increasing the degrees of freedom and reducing the multicollinearity, enhance the efficient of estimation (Tüzüntürk 2000).

2. REGRESSION MODELS FOR PANEL DATA

Panel data models have two dimensions on its variables differently from cross sectional and time series. Panel regression model can be written as

$$y_{it} = \alpha_{it} + x_{kit}\beta_{kit} + u_{it} \qquad i = 1,\ldots,N \quad t = 1,\ldots,T \tag{1}$$

where i donates units, t donates the time dimension, y_{it} is the observation on the dependent variable for i-th individual and t-th time period, α_{it} is the intercept varying over time and units. β_{it} is the slope coefficient , x_{kit} is the observation on the independent variable for i-th units and t-th time period. u_{it} is the error term. In this model all coefficient vary over units and time. The number of parameter to be estimated exceed the observation hence this model cannot estimate. Panel data models can be classified further, depending on whether the coefficients are assumed random or fixed (Hsiao 2003).

2.1.Fixed Effect Models

One way to incorporate for individual or time differences behavior is assumed that some of the regression coefficient or all of them are allowed to vary across individual or over time. (Hsiao 2003). The coefficients are allowed to vary time or / and individual, these models referred to fixed effects model.

In individual effect models we assumed that slope coefficient is constant over time and individuals however intercept is allowed to vary from individual to individual (Màtyàs and Sèvèstrè 2008). This case can be modeled as

$$y_{it} = \alpha_i + x_{kit}\beta_k + u_{it} \qquad i = 1,\dots,N \quad t = 1,\dots,T \tag{2}$$

i, donates units , t donates time period, x_{kit} is the it -th observation ok k explanatory variables. y_{it} is the value of the dependent variable for i -th units and t -th time. u_{it} is error term that is assumed to be identically and independently distributed with zero mean and constant variance ($u_{it} \sim i.i.d.(0,\sigma^2)$). In order to make difference among individuals, N dummy variables are added the model. Model can be written matrix notation as

$$y = D\alpha + X\beta + u = Z\delta + u \tag{3}$$

where y is an $NTx1$ vector, X is an $NTxq$ matrix of explanatory variable, u_{it} is an $NTx1$ vector of error terms and β is a $qx1$ vector and α is an $Nx1$ vector. $Z = [D\ X]$ and $\delta' = [\alpha\ \beta]'$. D is an NT x N matrix of dummy variable and has the following Kronocker product representation,

$$\mathrm{D}=\mathrm{I}_N \otimes \iota_T \tag{4}$$

where I_N denotes $N \times N$ identity matrix and ι_T is a vector of ones of dimension T. This model also referred to as the least squares dummy variable model (LSDV). This model is a classical regression model and if N is small we can estimate the model by ordinary least squares with q regressor in X and N columns in D (Greene 2003). But if N is large, inverting a matrix of $(N+q)\times(N+q)$ dimension is difficult and this might be a problem and to avoid this problem partitioned regression can be used (Roy 1997). Using the dummy variable matrix D, one can obtain transform matrix

where

$$M_N = I_{NT} - D(D'D)^{-1}D' \tag{5}$$

And by using the transform matrix M_N, we can obtain OLS estimator of β

$$\hat{\beta}_{SE} = (X'M_N X)^{-1}(X'M_N Y) \tag{6}$$

And intercept term for each i

$$\hat{\alpha}_i = \overline{Y}_{i.} - \sum_{k=1}^{q} \overline{x}_{ki.}\hat{\beta}_{kSE} \tag{7}$$

The estimated variance of $\hat{\beta}_{SE}$ and $\hat{\alpha}$ can be obtain,

$$Var(\hat{\beta}) = \sigma^2 (X'M_N X)^{-1} \tag{8}$$

$$Var(\hat{\alpha}) = \sigma^2 (\frac{1}{T} X_{i.} Var(\hat{\beta}) \overline{X}_{i.}) \tag{9}$$

One can test to see whether there is a difference between individual. In this case the null hypothesis is that there is no difference between individuals.

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_N$$
$$H_1 : \alpha_1 \neq \alpha_2 \neq \cdots \neq \alpha_N$$

To test this hypothesis, F test can be used

Let HKT_1 be the sum of squared residuals obtained from restricted model and HKT_2 be the sum of squared residuals obtained from unrestricted model. One can use the F test with (N-1) and (NT-N-K) degrees of freedom.

$$F = \frac{HKT_1 - HKT_2/(N-1)}{HKT_2/(NT-N-K)}$$

If F ratio is larger than the critical value of F, null hypothesis will be rejected. Thus one can decided to units are homogeneity.

In time effects model, assumed that intercept term vary over the time period. Time effects model written as,

$$y_{it} = \lambda_t + x_{kit}\beta_k + u_{it} \qquad i = 1,\ldots,N \quad t = 1,\ldots,T \tag{10}$$

λ_t is a intercept term which is constant over time. The model can be written as matrix notation,

$$y = D_T\lambda + X\beta + u \tag{11}$$

Where D_T is a NT x T matrix of time dummies and λ is T x 1 vector of varying intercept.

In time effect models obtaining the parameter estimation resemble by fixed effect models. One can be used partitioned regression to obtain the parameter estimator. Transform matrix in time effect model,

$$M_T = I_{NT} - D_T(D_T^{'}D_T)^{-1}D_T^{'} \tag{12}$$

Estimation of parameter is exactly the same with fixed effect.

2.2.Random effects models

In fixed effects models the individual-varying variables represented by α_i and it is possible to indicate over individual-varying in random effect models. If one considered that error term u_{it} represents to unobservable effects and omitted variable that vary over individuals and over time, α_i could be treated random as error term u_{it} and α_i is a component of the error term. New error term is assumed that consist of two components.

$$v_{it} = \alpha_i + u_{it}$$

and we assumed that α_i and u_{it} ,

3. α_i is independent and identically distributed with zero mean and σ^2. $\alpha_i \sqcup iid(0,\sigma^2)$.
4. u_{it} is independent and identically distributed with zero mean and σ^2. $u_{it} \sqcup iid(0,\sigma^2)$.
5. $E(\alpha_i u_{it}) = 0$
6. $E(\alpha_i X_{it}) = 0$ and $E(u_{it} X_{it}) = 0$

Random effect model can be written as,

$$y_{it} = \mu + x_{kit}\beta_k + v_{it} \qquad i = 1,\dots,N \quad t = 1,\dots,T \tag{13}$$

And the model rewriting in matrix form,

$$y = \mu + X\beta + v \tag{14}$$

Where β is q x 1 vector, X is $(NT \times k)$ design matrix, v is $(NT \times 1)$ error term. μ represents mean intercept.

The parameter of model can be estimated by using generalized least square in random effects model since GLS estimators are the best linear unbiased (BLUE) estimators.

The variance covariance matrix of error term v_i is

$$V = \sigma_u^2 I_T + \sigma_\alpha^2 \iota_T \iota_T^{'} \tag{15}$$

The inverse of variance covariance matrix V can be written as

$$V^{-1} = \left[I_T - \frac{\sigma_\alpha^2}{\sigma_u^2 - T\sigma_\alpha^2} \iota_T \iota_T^{'} \right]$$

By using equations (18), $\hat{\beta}$ estimator can be obtained,

$$\hat{\beta}_{GEKK} = (X'V^{-1}X)^{-1}(X'V^{-1}Y) \tag{16}$$

Mean intercept can be expressed,

$$\hat{\mu} = \overline{y}_{..} - \hat{\beta}_{RE}\overline{X}_{..} \tag{17}$$

In order to get inverse of variance covariance matrix of V, one should be compute the estimators of σ_α^2 and σ_u^2. Suggested estimator by Swamy and Arora (1972) for σ_α^2 and σ_u^2 as follows,

$$\hat{\sigma}_u^2 = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left[y_{it} - \overline{y}_{i.} - (x_{kit} - \overline{x}_{ki.})\hat{\beta}_{SE} \right]^2}{N(T-1) - q} \tag{18}$$

$$\hat{\sigma}_\alpha^2 = \frac{\sum_{i=1}^{N} (\overline{y}_{i.} - \hat{\mu}_G - \overline{x}_{i.} \hat{\beta}_G)^2}{N - (q+1)} - \frac{1}{T} \hat{\sigma}_u^2$$

where $\hat{\beta}_G$ refers to between groups estimators and is given as,

$$\hat{\beta}_G = \left[\sum_{t=1}^{T} (\overline{x}_{ki.} - \overline{x}_{..})'(\overline{x}_{ki.} - \overline{x}_{..}) \right]^{-1} \left[\sum_{i=1}^{N} (\overline{x}_{ki.} - \overline{x}_{..})'(\overline{y}_{i.} - \overline{y}_{..}) \right] \tag{19}$$

where $\hat{\mu}_G = \overline{y}_{..} - \overline{x}_{..} \hat{\beta}_G$.

Breusch and Pagan (1979) suggested using Lagrange Multiplier to test for random effects model. To check for absence of individual effects, the null hypothesis and alternative hypothesis are given as

$$H_0 : \sigma_\alpha^2 = 0$$
$$H_1 : \sigma_\alpha^2 > 0$$

The LM test statistics is written as,

$$LM = \frac{NT}{2(T-1)} \left[\frac{\sum_{i=1}^{N} (\sum_{t=1}^{T} \hat{v}_{it}^2)}{\sum_{i=1}^{N} \sum_{t=1}^{T} (\hat{v}_{it})^2} - 1 \right] \tag{20}$$

and the statistics is distributed as a χ^2 with 1 degrees of freedom. Eventually if the statistics computed from OLS regression residuals is larger than critical value of χ^2 the null hypothesis will be rejected.

3. APPLICATION

This study aims at predicting average wind speed during the years 2009-2010 in the cities of Sinop, Samsun, Ordu, Kastamonu, Bartın, Zonguldak and Karabük of Blacksea Region. Wind speed is a dependent variable, the average minimum temperature; the average maximum temperature and the average humidity are independent variables.

First of all, individual effects are analyzed for fixed effects model. Only individual effects model is written as,

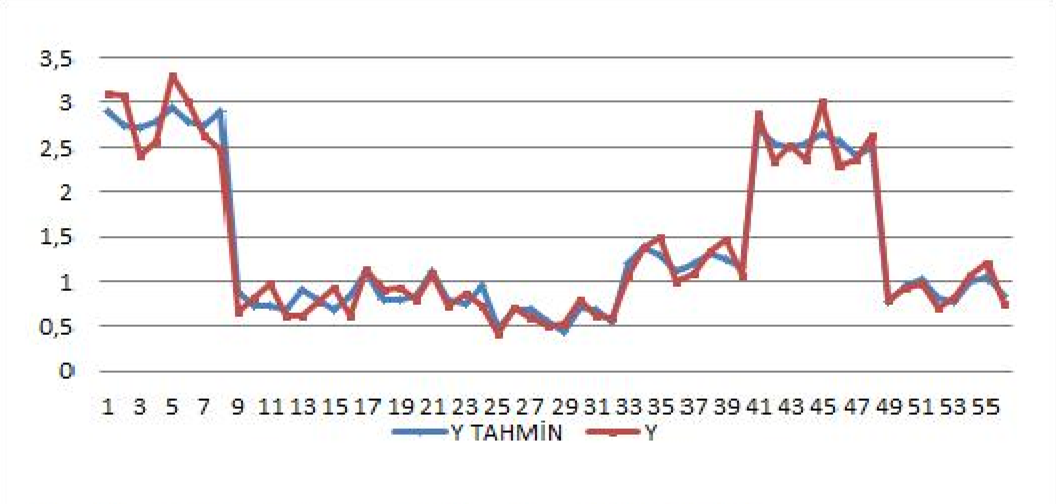
$$Y_{it} = \alpha_i + x_{1it}\beta_1 + x_{2it}\beta_2 + x_{3it}\beta_3 \quad i = 1,\ldots,7 \;\; t = 1,\ldots,8 \tag{21}$$

Computed value of α_i for each city and the slope coefficient $\hat{\beta}_k$ which is the same over units, given as in Table 1.

Table 1. Parameter estimation in only individual effects model.

i	$\hat{\alpha}_i$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
1	4,7137	-0,0014	0,07406	-
			0,02177	
2	2,5165			
3	2,6439			
4	2,3899			
5	3,1809			
6	4,2853			
7	2,7096			

In Table 1. $\hat{\alpha}$ and $\hat{\beta}$ estimation is shown for individual effects model. Predicted of $\hat{\alpha}$ is obtained as a large value on the ground that $\hat{\beta}$ has rather lower value.



Model $Y_{it} = \alpha_i + \beta_1 x_{1it} + \beta_2 x_{2it} + \beta_3 x_{3it} + u_{it}$	**HKO =** 0.0286

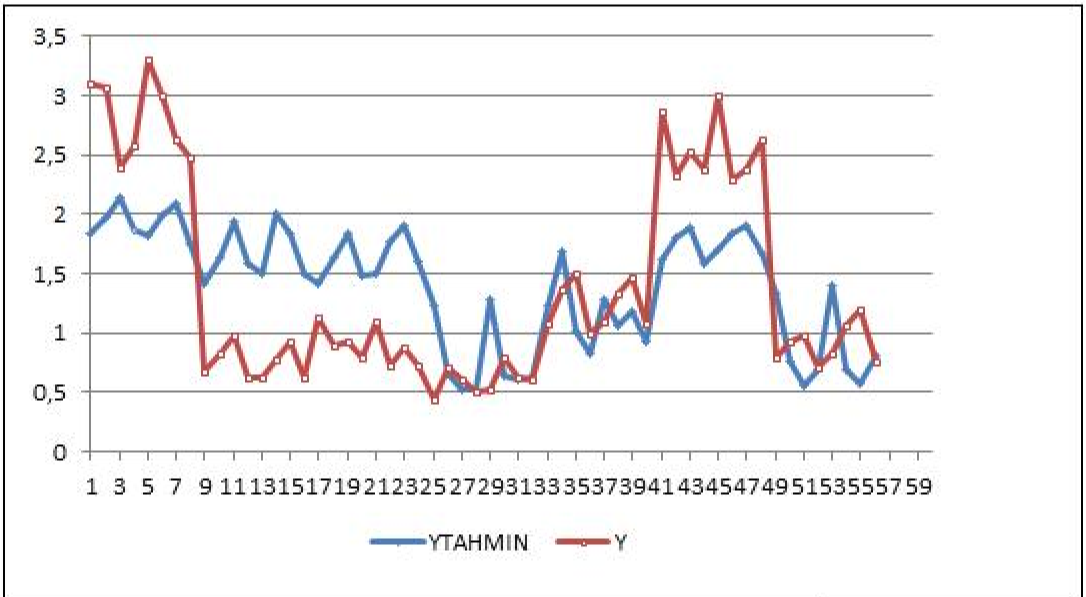
Figure 1. Graphs of individual effects model

The observed y and predicted values of y are shown in graph 1. Predicted value which is obtained from individual effects is close to observed y. Error mean square is 0.0286 for the individual effects model. F statistics equal to s 132.601 and it is greater than the critical value of $F_{(6,46)}$ =2,305. Consequently, null hypothesis is rejected so we decided to there is a difference between units fixed effect model is significant

Table 2. Parameter estimation in only Time effects model.

t	λ_t	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
2009Q1	1,2192	-0,1166	0,151 0,0112	
2009Q2	1,3447			
2009Q3	1,2879			
2009Q4	1,0306			
2010Q1	1,2154			
2010Q2	1,4057			
2010Q3	1,1273			
2010Q4	1,0175			

The estimate of λ_t and $\hat{\beta}_k$ estimators is shown in table 2. Critical level of F is greater than the F, null hypothesis is accepted and it means that there is no differences between time period over the units.



Model $Y_{it} = \lambda_t + \beta_1 x_{1it} + \beta_2 x_{2it} + \beta_3 x_{3it} + u_{it}$	HKO = 0.543

Figure 2. Time Effects Model Graph

In Fig.2 observations of Y and predicted values of y are given together. In time effect model, mean error square is 0.543 and it's greater than the individual effect model. Random effect model is written as ,

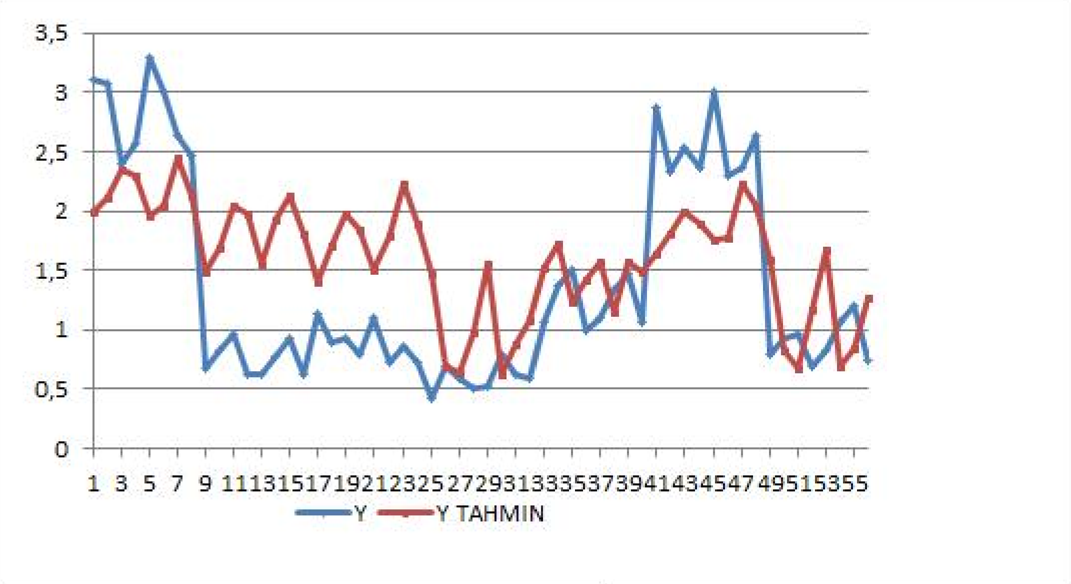
$$Y_{it} = x_{1it}\beta_1 + x_{2it}\beta_2 + x_{3it}\beta_3 \quad i = 1,\ldots,7 \;\; t = 1,\ldots,8 \tag{22}$$

Table 3. Random effect models estimation

$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
-0,099	0,1374	0,0248
Lagrange Multiple (LM)=145.744		

Table 3. contain LM test statistics and the estimate of β. Critical level of χ^2 with 1 degrees of freedom is 3,84 and LM test statistics is greater than the critical level.

Random effects model is significant. Observations of Y and graphs of predicted values of Y shown in fig.3. For random effects model predicted values are not as good as prediction values of Y in fixed effect.



Model $Y_{it} = \mu + \beta_1 x_{1it} + \beta_2 x_{2it} + \beta_3 x_{3it} + v_{it}$	**HKO** = 0.607

Figure 3. Random Effects Model Graphs

4. CONCLUSION

Panel data is common used method for economic data but it is rarely used for the meteorological data. This study aims at predicting average wind speed during the years 2009-2010 in the cities of Sinop, Samsun, Ordu, Kastamonu, Bartın, Zonguldak and Karabük of Blacksea Region by using wind speed, the average of maximum temperature, the average of minim temperature and the average of humidity. For individual effects model, MSE is 0,0286 and FE model is statistically significant. MSE is 0,543 in time effects model. In consequence of F test, one concludes that there are differences across the individual but not any differences over time period. The result shown that fixed effect model is significant and has turn out to be more effective in analyzing these data. Constant term which represented to individual is significant hence individuals are important to modelling for wind speed.

In this paper we showed that, panel data analysis can be used for meteorological data sets which include unit and time relation as an alternative method.

REFERENCES

Baltagi, H. Badi 2005. Econometric Analysis of Panel Data. John Wiley&Sons Ltd
Greene, W.H. 2002, Econometric Analysis, Prentice Hall
Gujarati,D. Basic Econometrics .2004. The McGraw−Hill Companies
Hsiao Cheng.2003. Analysis of Panel Data.Cambridge University Press
Matyas, L. , Sevestre . P., The Econometrics Of Panel Data : Advanced Studies In Theoretical And Applied Econometrics Volume 46, Fundamentals and Recent Developments in Theory and Practise, 2008. Springer
Sayyan H., 2000 Dinamik Panel Veri Modelleri Ve OECD Ülkeleri Para Talebi Uygulaması Doktora Tezi Marmara Üniversitesi Sosyal Bilimler Enstitüsü Ekonometri Anabilim Dalı
Tüzüntürk, S., 2005, İşlem Sıklığı Ve Hacmi İle Fiyat Volatilitesi İlişkisi : İMKB Örneği Yüksek Lisans Tezi Uludağ Ünver sitesi Sosyal Bilimler Enstitüsü Ekonometri Anabilim Dalı
Roy, N. 1997, Nonparametric and Semiparametric Analysis of Panel Data Models: An Application to Calorie-Income Relation for Rural South India. Doctor of Philosophy, Graduate Program in Economics, University of California.