Available online at www.sciencedirect.com

ScienceDirect

Procedia Economics and Finance 35 (2016) 241 – 248

Procedia Economics and Finance

www.elsevier.com/locate/procedia

7th International Economics & Business Management Conference, 5th & 6th October 2015

Panel Data Analysis for Sabah Construction Industries: Choosing the Best Model

Anwar Fitrianto[a,b,c]*, Nur Farhanah Kahal Musakkal[a]

[a]*Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, 43400 UPM, Serdang, Malaysia*
[b]*Laboratory of Computational Statistics and Operations Research, Institute for Mathematical Research, Univeriti Putra Malaysia*
[c]*Department of Statistics, Faculty of Mathematics and Natural Resources, Bogor Agricultural University, Indonesia*

Abstract

Analysis of panel data by using statistical models is rapidly growing. It is sometime tough for the novice users of panel data to make an informed choice of what estimators best suit their research questions. This paper is meant to find best model among few types of models such as panel data models and ordinary least squares (OLS) regression for Sabah construction industries. The best model will be chosen based on lowest Root Mean Square Errors (RMSE). The purpose of comparing between models is to find the most efficient model which will be useful for prediction. After analyzing the data using SAS software, it was found that two-way fixed effect panel data model provide the lowest RMSE for the Sabah construction industries.

Peer-reviewed under responsibility of Universiti Tenaga Nasional

Keywords: Panel data, Regression, Time series, Random effects, Fixed effects, Sabah

1. Introduction

Econometric analysis of panel data has been started some time ago, such as what has been done by Balestra and Nerlove (1966) and Hoch (1962). Panel data analysis can benefit to industries because it provides information from dataset which behavior of cross sectional is observed across time. The pooling of cross section and time series data

* Corresponding author. Tel.: +6038976-7025
E-mail address: anwarstat@gmail.com

Peer-reviewed under responsibility of Universiti Tenaga Nasional
doi:10.1016/S2212-5671(16)00030-7

have been increasingly important and popular way to determine economic relationship. In panel data, each series yields information and results which others series do not have. Combination of both will highly lead to produce more accurate and reliable results compared to one type of series alone. Panel data analysis is highly recommended if the main purpose of research to estimate relationship at an individual or disaggregated level.

There are several types of panel data models including fixed effect model, random effect model, between estimators, within estimator, dummy variable estimator, first differencing estimator, Feasible Generalized Least Square (FGLS), Ordinary Least Square (OLS), Monte Carlo approaches and many others (Wooldridge, 2012). A lot of research about estimation of panel data has been done. However, most of them just focused on the estimation itself and lack of attention to efficiency and consistency of the estimations.

Estimator is used to infer the value of an unknown parameter in statistical models. This is because in real world the exact value of the population is not known. Reed and Ye (2009) in their research mentioned that the most common estimators in panel data are Generalized Least Square (GLS) and Feasible Generalized Least Square (FGLS). Since variance covariance is often unknown, FGLS is more frequently used rather than GLS. However, the poor performance of the FGLS estimator arises because the true value for variance covariance is unknown (Reed and Ye, 2009). Furthermore, Reed and Ye (2009) also mentioned that Ordinary Least Square (OLS) is also one of the preferable estimators in panel data analysis. Unfortunately, the OLS estimator is generally inconsistent when the independent variable and random error disturbance are correlated. To remedy this inconsistency, one method can be used which is method of Instrumental Variable (IV).

Based in above background, the purpose of the study is to determine the best panel data model to estimate parameters for Sabah construction industries among fixed effects, random effect model and pooled Ordinary Least Squares (OLS). It is important to choose appropriate model which leads a better consistency and efficiency by considering several important aspect such as RMSE and other test if needed in order to choose the best one. The study is limited on the data of Sabah construction Industry which have four cross section variables, which are residential construction, non-residential construction, civil engineering construction and special trade construction.

2. **Literature Reviews**

2.1. Panel data

Panel data refers to data sets consisting of multiple observations on each sampling unit. This could be generated by pooling time-series observations across a variety of cross-sectional units (Baltagi, 2013). An example of panel data are annual unemployment rates of each state over several years, quarterly sales of individual stores over several quarters and wages for the same worker, working at several different jobs. Hsiao (1986) proposed one of the benefits of panel data sets where it provides much larger data sets with more variability and less collinearity among variables compared to typical of cross-section or time series data alone. In addition, he also mentioned other benefits of panel data including panel data sets are more informative and able to control for individuals heterogeneity. Controlling for individual heterogeneity is necessary because it is can cause to bias estimate.

Kasprzyk et al. (1989) said that limitations in panel data sets include problems in the design, data collection and data management of panel surveys. These include the problems of coverage or also known as incomplete account of population of interest, non-response which might be due to the lack of cooperation among respondent or might because of interviewer errors, recall because some of respondent are not remembering correctly, frequency of interviewing, interview spacing, reference period, the use of bounding to prevent the shifting of events from outside the recall period into the recall period and time in sample bias. Another limitation of panel data sets are the distortions due to measurement errors. The measurement errors may arise because of faulty of response due to unclear questions, memory errors, deliberate distortion of responses (e.g., prestige bias), inappropriate informants, miss-recording of responses and interviewer effects.