Cairo University
Institute of Statistical Studies and Research

Some Properties of Random Coefficients Regression Estimators

By

Mohamed Reda Sobhi Mohamed

Supervised by

Prof. Amany Mousa Mohamed
Professor of Applied Statistics
Dept. of Applied Statistics and Econometrics

Prof. Ahmed Hassen Youssef
Professor of Applied Statistics
Dept. of Applied Statisiics and Econometrics

A Thesis Submitted to the
Department of Applied Statistics and Econometrics
In Partial Fulfillment of the Requirements for the Degree of
MASTER OF SCIENCE
In
Applied Statistics

2009

Cairo University
Institute of Statistical Studies and Research

Approval Sheet

Some Properties of Random Coefficients Regression Estimators

By

Mohamed Reda Sobhi Mohamed

A Thesis Submitted to the
Department of Applied Statistics and Econometrics
In Partial Fulfillment of the Requirements for the Degree of
MASTER OF SCIENCE
In
Applied Statistics

Approved by the Examining Committee

Name	Signature
Prof. **Amany Mousa Mohamed**	
Prof. **Sohair Fahmy Higazi**	
Prof. **Sayed Mesheal Elsayed**	
Prof. **Ahmed Hassen Youssef**	

Date: / / 2009

Contents

Glossary of Notation

Notation	Definition
ANCOVA	Analysis of Covariance
AR	Autoregressive
ARMA	Autoregressive Moving Average
BLUE	Best Linear Unbiased Estimator
CP	Classical Pooling
EM	Expectation Maximization
FGLS	Feasible Generalized Least Squares
GLM	General Linear Model
GLS	Generalized Least Squares
MA	Moving Average
MG	Mean Group
Mixed RCR	Mixed Random Coefficient Regression
MLE	Maximum Likelihood Estimator
MSE	Mean Square Error
MVUE	Minimum Variance Unbiased Estimator
OLS	Ordinary Least Squares
RCR	Random Coefficient Regression
REML	Restricted Maximum Likelihood
SSE	Sum Square Error
SUR	Seemingly Unrelated Regression
TSCS	Time-Series-Cross-Section

Acknowledgments

I'm greatly indebted to **Prof. Amany Mousa**, Professor of Applied Statistics, Dept. of Applied Statistics and Econometrics, Institute of Statistical Studies and Research for her valuable and generous assistance. My sincere thanks are also dedicated to her for this constructive guidance and warm encouragement throughout the preparation of this thesis.

Prof. Ahmed Hassen, Professor of Applied Statistics, Dept. of Applied Statistics and Econometrics, Institute of Statistical Studies and Research, deserves my deepest gratitude and appreciation for his kind supervision, continuous help and active discussions during the preparation of this thesis.

I would like to express my thanks to **Prof. Sohair Higazi**, Professor of Applied Statistics, Dept. of Applied Statistics, Tanta University for her generous acceptance of discussion of this thesis, and to **Prof. Sayed Mesheal**, Professor of Applied Statistics, Head of the Department of Applied Statistics and Econometrics, Institute of Statistical Studies and Research for his continuous help and his generous acceptance of discussion of this thesis.

Summary

An important assumption of the General Linear Model (GLM) is that the vector of regression coefficients is fixed vector, so the model will be called "Fixed Model". But when we assumed that the regression coefficients are random variables, so the model will be called "Random Coefficient Regression (RCR) Model" examined by Swamy in several publications (Swamy 1970, 1971, 1973, and 1974). And if the regression coefficients in model contain both random and fixed coefficients, so the model will be called "Mixed Random Coefficient Regression (Mixed RCR) model".

In this thesis, we studied the properties of RCR and Mixed RCR models. And also we studied the Swamy's estimator (RCR estimator) for RCR model in panel data, and we proposed the alternative estimators for RCR model, such as unit by unit OLS, Mean Group (MG), Classical Pooling (CP), and Stein-rule estimators.

In this thesis, we used the Monte Carlo simulation to study the behavior of the Swamy's estimator in small, medium and large samples in panel data. The parameters were set at several values, to allow the study of estimators under several situations, to know when the RCR model will be properly and improperly. This simulation provides some insight into how well the RCR estimator performs in different samples size. Also, we used the Mote Carlo simulation again for comparison between the behavior of RCR, CP, and MG estimators in three models (RCR, fixed, and Mixed RCR models). And we used the R language to conduct the Monte Carlo simulation study.

The thesis includes five chapters:

Chapter 1: Definitions and Notations

This chapter involved some definitions which be used in this thesis.

Chapter 2: Introduction to Random Coefficient Regression Models

This chapter presented an introduction to the general linear model estimators under the classical assumptions in section (2.1). While section (2.2) discussed RCR model for panel data, and we estimated the random coefficients by the generalized least square method. The literature review for RCR models in section (2.3). Finally, section (2.4) introduced the applications for RCR models.

Chapter 3: Estimation of Random Coefficient Regression Model in Panel Data

This chapter presented the RCR model in panel data when regression coefficients are viewed as invariant over time, but varying from one unit to another in section (3.1). While section (3.2) discussed the alternatives estimators for RCR model. Finally, in section (3.3) we proposed the Mixed RCR model as special case of the RCR model.

Chapter 4: Estimation of Time and Cross-Sectionally Varying Parameter Models

In this chapter, an introduction for the time and cross-sectionally varying parameter models and we proposed the Hsiao model in section (4.1). While in section (4.2), we discussed the estimation and tests of hypotheses for the random coefficients of Swamy and Mehta model. In section (4.3), we proposed other models with time varying coefficients. Finally, section (4.4) introduced the applications for time and cross-sectionally varying parameter models.

Chapter 5: Monte Carlo Simulation for Efficiency of Random Coefficient Regression Estimators

In this chapter, we used the Monte Carlo simulation to study the behavior of the Swamy's estimator in small, medium and large samples in panel data in section (5.1). While in section (5.2), we explained the results of simulation study. We used the Mote Carlo simulation again for comparison between the behavior of RCR, CP, and MG estimators in section (5.3). While in section (5.4), we explained the simulation results for the three estimators. Ending with section (5.5), we focused on displaying the concluding remarks of the Monte Carlo simulation studies.

Finally, the Monte Carlo simulation results suggest that the RCR estimators perform well in small samples if the coefficients are random but it does not in fixed or Mixed RCR models. But if the samples sizes are medium or large, the RCR estimators performs well for the three models. While CP estimators perform well in the fixed model only. But the MG estimators perform well if the coefficients are random or fixed. So, we can say that the MG method is the general estimation method for fixed, RCR, and Mixed RCR models.

Chapter One

Definitions and Notations

Chapter 1

Definitions and Notations

This chapter involves some definitions which will be used in the following chapters.

Methods of Point Estimation

In general, if X is a random variable with probability density function $f(x,\theta)$, characterized by the unknown parameter θ, and if $x_1, x_2, \ldots, x_n$ is a random sample of size n from X, then the statistic $U = u(x_1, x_2, \ldots, x_n)$ is called a point estimator of θ. There are several methods are used to obtain point estimators for the parameters of a given probability distribution and there are important statistical properties of point estimators.

Properties of Point Estimators

Various statistical properties of estimators can be used to decide which estimator is most appropriate in a given situation, which will expose us to the smallest risk, which will give us the most information at the lowest cost, and so forth. Here the properties of point estimation which are unbiasedness, efficiency, consistency and sufficiency will be reviewed.

(i) Unbiasedness

A statistic $\hat{\theta}$ is said to be an unbiased estimator of a parameter θ if the average or expected value of the statistic is equal to the parameter. The corresponding observed value is called an unbiased estimate.

Suppose that we know a random variable X has a given probability function (or density function) depending upon an unknown parameter θ. Let $\hat{\theta}$ a statistic that is a function of one or more random variables having the given probability function. If $E(\hat{\theta}) = \theta$ then $\hat{\theta}$ is an unbiased estimator of θ, Walpole *et al* (1998) gave the following definitions:

Definition (1)

A statistic $\hat{\theta}$ is an unbiased estimator for the parameter θ if and only if $E(\hat{\theta}) = \theta$.

Definition (2)

Let $x_1, x_2, \ldots, x_n$ be a random sample of X, whose probability density function depends on an unknown parameter θ, and let $\hat{\theta}$ be any statistic. Then the Mean Square Error (MSE) of $\hat{\theta}$ is

$$MSE(\hat{\theta}) = E[\hat{\theta} - \theta]^2$$
$$= Var(\hat{\theta}) + [bias(\hat{\theta})]^2$$

(ii) Efficiency

If an estimator $\hat{\theta}_1$ has a MSE smaller than the MSE for another estimator $\hat{\theta}_2$ when estimating the unknown parameter θ from a given random sample. The estimate $\hat{\theta}_1$ is thought of as more efficient use of the observations. Thus an estimator $\hat{\theta}_1$ is more efficient than $\hat{\theta}_2$ if

$$E(\hat{\theta}_1 - \theta)^2 \leq E(\hat{\theta}_2 - \theta)^2$$

i.e. $MSE(\hat{\theta}_1) \leq MSE(\hat{\theta}_2)$,

with strict inequality for some θ. The relative efficiency of $\hat{\theta}_2$ with respect to $\hat{\theta}_1$ is the ratio

$$e(\hat{\theta}_1, \hat{\theta}_2) = \frac{E(\hat{\theta}_1 - \theta)^2}{E(\hat{\theta}_2 - \theta)^2} = \frac{MSE(\hat{\theta}_1)}{MSE(\hat{\theta}_2)}.$$

If $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased estimators of the parameter θ, the relative efficiency is just the ratio of their variances; and one with minimum variances would be the most efficient.

$$e(\hat{\theta}_1,\hat{\theta}_2)=\frac{E(\hat{\theta}_1-\theta)^2}{E(\hat{\theta}_2-\theta)^2}=\frac{Var(\hat{\theta}_1)}{Var(\hat{\theta}_2)}.$$

To check whether a given unbiased estimator has the smallest possible variance, that is, it is a Minimum Variance Unbiased Estimator (MVUE), we use the fact that if $\hat{\theta}$ is an unbiased estimator of θ, it can be shown under very general conditions that the variance of $\hat{\theta}$ must satisfy the Cramér-Rao inequality

$$Var(\hat{\theta})\geq\frac{1}{n\times E\left(\frac{\partial\ln f(x\backslash\theta)}{\partial\theta}\right)^2},$$

where $f(x,\theta)$ is the value of the population density at x and n is the size of the random sample. This inequality, leads to the following result.

Theorem (1)

If $\hat{\theta}$ is unbiased estimator of θ and $Var(\hat{\theta})=1/n\times E(\partial\ln f(x\backslash\theta)/\partial\theta)^2$. Then $\hat{\theta}$ is minimum variance unbiased estimator of θ (see Miller and Miller (1999)). Here, the quantity in the denominator is referred to as the information about θ that is supplied by the sample. Thus the smaller the variance is the greater the information.

(iii) Consistency

suppose that $\hat{\theta}_n$ is an estimator of parameter θ based on random variables $x_1, x_2, ..., x_n$. As n increases, it seems reasonable to expect that the sampling distribution of $\hat{\theta}_n$ should becomes increasingly concentrated around the true parameter value θ. This property of the sequence of estimators $\{\hat{\theta}_n\}$ is known as consistency, Mendenhall *et al* (1981) gave the following definition:

Definition (3)

The estimator $\hat{\theta}_n$ is said to be a consistent estimator of θ if, for any positive number ε,

$$\lim_{n\to\infty} p\left[\left|\hat{\theta}_n - \theta\right| \le \varepsilon\right] = 1$$

or, equivalently,

$$\lim_{n\to\infty} p\left[\left|\hat{\theta}_n - \theta\right| > \varepsilon\right] = 0.$$

In the fact the consistency is an asymptotic property, that is, a limiting property of an estimator. Informally, the pervious definition says that when n is sufficiently large, we can be practically certain that the error made with a consistent estimator will be less than any small preassigned positive constant. In practice, we can often judge whether an estimator is consistent by using the following sufficient condition which, in fact, is an immediate consequence of Chebyshev's theorem.

Theorem (2)

If $\hat{\theta}_n$ is an unbiased estimator of the parameter θ and $\mathrm{Var}(\hat{\theta}_n) \to 0$ as, $n \to \infty$, then $\hat{\theta}_n$ is a consistent estimator of θ, (see Mendenhall *et al* (1981)).

(iv) Sufficiency

The statistic which summarizes all the information about an unknown parameter is called a sufficient statistic. This is clear in the following definition:

Definition (4)

The statistic U is a sufficient estimator of the parameter θ if and only if for each value of U the condition probability distribution or density of the random sample $x_1, x_2, \ldots, x_n$, given $U=u$, is independent of θ (see Miller and Miller (1999)).

Since we can be very tedious to check whether a statistic is a sufficient estimator of a given parameter based directly on the pervious definition, it is usually easier to base it instead on the following factorization theorem which mentioned by Mendenhall *et al* (1981).

Theorem (3)

The statistic U is a sufficient estimator of the parameter θ if and only if the joint probability distribution of the random sample can be factored into two non-negative functions,

$$f(x_1, x_2, \ldots, x_n; \theta) = g(u, \theta).h(x_1, x_2, \ldots, x_n),$$

where $g(u, \theta)$ depends only on u and θ, and $h(x_1, x_2, \ldots, x_n)$ does not depend on θ. Mendenhall *et al* (1981) concluded that if U is a sufficient estimator of θ, then any single valued function $Y = w(U)$, not involving θ, is also a sufficient estimator of θ, and therefore of $w(\theta)$.

Least Squares Method

A method used for estimating parameters, particularly in regression models, by minimizing the sum of squared differences between the observed response and the value predicted by the model (see Darnell (1994)).

Definition (5)

Gauss-Markov theorem is a theorem that proves that if the error terms in a multiple regression have the same variance and are uncorrelated, then the estimators of the parameters in the model produced by least squares estimation are better (in the sense of having lower dispersion about the mean) than any other unbiased linear estimator (Everitt (2006)).

Maximum Likelihood Method

The Maximum Likelihood (ML) method is one of the most important methods in the theory of estimation. Engelhardet and Bain (1987) gave the following definition:

Definition (6)

The likelihood function of n random variables $X_1, X_2, \ldots, X_n$ is the joint density of the n random variables and is often denoted by $L(\theta)$. If $x_1, x_2, \ldots, x_n$ represents a random sample from $f(x \setminus \theta)$, then the likelihood function is given by