ROBUST INFERENCE IN PANEL DATA MODEL

NURUL SIMA BINTI MOHAMAD SHARIFF

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

INSTITUTE OF MATHEMATICAL SCIENCES
FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2012

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **Nurul Sima binti Mohamad Shariff**
I.C./Passport No.: **810730-10-5408**
Registration / Matrix No.: **SHB070002**
Name of Degrees: **Doctor of Philosophy (PhD)**
Title of Project Paper / Research Report / Dissertation / Thesis ("this Work"):
Robust Inference In Panel Data Model

Field of Study: **Applied Statistics**

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:
Name:

Designation:

ABSTRAK

Data panel adalah satu kumpulan yang terdiri dari beberapa individu, dan cerapan bagi sesuatu pembolehubah dibuat untuk tempoh masa tertentu. Secara umum, penyelidik mungkin cenderung untuk mengabungkan unit tersebut di mana setiap unit adalah dikira sebagai tak bersandaran diantaranya. Andaian tersebut adalah tidak sah kerana kebanyakan data ekonomi adalah saling bersandaran di antaranya, yang mungkin wujud dari pengaruh yang sama dan seterusnya yang memberi kesan kepada semua unit. Ini dikenali sebagai bersandaran keratan rentas (CD). Kehadiran data terpencil boleh menyebabkan penolakan hipotesis nol, yang menyokong ketidakbersandaran keratan rentas. Untuk mengatasi masalah itu, kaedah alternatif yang tidak sensitif terhadap kehadiran data terpencil adalah diperlukan.

Untuk menangani masalah kesilapan statistik ujian dan anggaran parameter dengan kehadiran CD dan data terpencil, kajian ini akan memberi tumpuan kepada beberapa bahagian. Pertama, ujian CD versi teguh dicadangkan untuk menyiasat kehadiran CD dan data terpencil dalam kedua-dua model panel statik dan dinamik. Ciri-ciri anggaran parameter dan kajian simulasi bagi sampel terhingga dipertimbangkan berdasarkan kajian simulasi Monte Carlo. Kami mendapati bukti bahawa, kewujudan CD sederhana (rendah) dan data terpencil dalam panel, ujian kami mengatasi ujian CD yang kerap digunakan dalam kajian iaitu ujian LM dan PCD.

Kedua, kami mencadangkan penganggar *Common Correlated Mean Group (CMG)* versi teguh iaitu penganggar RCMG, untuk menganggar parameter dalam model statik. Beberapa sifat-sifat dan inferens statistik untuk parameter juga dipertimbangkan. Untuk lebih memahami kelakuan anggaran parameter dalam sampel terhingga, kami

menjalankan kajian simulasi Monte Carlo. Penganggar yang dicadangkan menghasilkan anggaran saksama dengan MSE yang kecil walaupun dengan kehadiran data terpencil pada arah X dan Y. Ujian hipotesis menunjukkan penganggar RCMG mempunyai saiz dan kuasa yang munasabah dan mengatasi penganggar CMG dalam panel, sama ada wujud atau tidak, data terpencil. Di samping ukuran kesaksamaan dan MSE, ketepatan juga dibuat dengan ukuran panjang selang keyakinan bagi menyokong keputusan ini.

Ketiga, kami meneroka ujian punca unit dalam rangka panel dinamik. Ujian sedia ada seperti ADF dan CIPS adalah mudah terjejas oleh kehadiran data terpencil yang kemudiannya menyebabkan keputusan hipotesis yang salah dengan seterusnya memihak kepada hipotesis nol (iaitu punca unit). Satu alternatif kepada CIPS iaitu RCIPS diperkenalkan berdasarkan prosedur RCMG. Prestasi dan kekukuhan RCIPS dibincangkan dan perbandingan dibuat untuk beberapa pendekatan yang sedia ada dalam kesusasteraan. Keputusan simulasi kami menunjukkan bahawa walaupun CIPS memberi keputusan yang baik untuk T besar, RCIPS cenderung untuk memberi keputusan saiz dan kuasa yang lebih baik walaupun untuk N dan T yang lebih kecil, sama ada wujud atau tidak, data terpencil.

Akhir sekali, kami mengkaji semula dua set data sebenar yang berkaitan dengan data panel; 1) data gasoline untu model yang statik dan 2) panel PPP Negara ASIA dan CEEC untuk model yang dinamik. Di sini, kami menggunakan kaedah yang dibincangkan di atas dan menganalisa data sewajarnya.

ABSTRACT

Panel data is a group of many individual units observed for a specific time period. In general, researchers may tend to pool the units together where each observation is treated as independently among the others. Such restriction is invalid because most of the economic data are cross correlated between cross sectional units which may arise from a common influence which affects all units. This is known as cross sectional dependence (CD). The presence of outliers may result in rejection of null hypothesis, that is, in support of cross sectional independence. To overcome such problem, alternative methods which are insensitive to the presence of outliers are needed.

To address the problems of incorrect test statistics and parameter estimates in the presence of CD and outliers, this study will focus on several parts. Firstly, robust versions of CD tests are proposed to investigate the presence of CD and outliers in both the pure static and dynamic models. The asymptotic behaviours and simulation study of power for the finite sample behaviour based on Monte Carlo simulation study are considered. We find evidence that, in the presence of mild (low) CD and outliers in panels, our tests outperform the commonly used CD tests that are the LM and CD tests.

Secondly, we propose a robust version of Common Correlated Mean Group (CMG), namely RCMG, for estimating parameters in pure static model. Some properties and statistical inference for the parameter are also considered. To better understand the finite sample behavior of these approaches, we run a Monte Carlo simulation study. Our proposed estimator yields unbiased estimates with small MSE in the presence of outliers occur in X and Y directions. The hypothesis test for the robust estimator indicates that RCMG estimator has reasonable size and power with and without the presence of

outliers and outperform CMG estimator in contaminated panel. In addition to the measure of bias and MSE, its accuracy is also measured by the length of confidence interval for RCMG estimator to supports these findings.

Thirdly, we further explore the unit root tests for the dynamic framework. The currently available tests such as ADF and CIPS are very much affected by the presence of outliers which subsequently result in wrong decision making by favoring to the null hypothesis of a unit root. An alternative of CIPS denoted by RCIPS is introduced based on the RCMG procedure. The performance and robustness of the RCIPS is discussed and comparisons are made to ADF ad CIPS. Our simulation results show that while the CIPS performs well for large T, the RCIPS tends to provide a good size and power even for smaller N and T, as well as with and without the presence of outliers.

Finally, we revisit two real datasets that are related to panel data; 1) the gasoline data for the pure static case; and 2) the PPP panel of ASIAN and CEEC countries for the dynamic model. Here, we employ the methods discussed above and reanalyze the data accordingly.

ACKNOWLEDGEMENT

Firstly, I wish to express my sincere and deepest appreciation to Professor Dr. Nor Aishah Hamzah for the guidance, advice and idea throughout her supervision in this PhD project. Special thanks to Professor Dr. Ahmad Zubaidi Baharumshah, as my second supervisor for his guidance and lesson on economic approach towards completing my PhD.

I would like to forward my special thanks to Kementerian Pengajian Tinggi dan Universiti Sains Islam Malaysia (USIM) for this precious and valuable scholarship for me to pursue my PhD and Professor Dr. Jalani Sukaimi as a dean of Faculty Science and Technology (USIM) for his continuous support, advice and idea especially during the my hardest time of the study period.

Thank you so much to my husband, Muazam Shaari, my son, Muhammad Adam Mu'min, my families and my families in law, for their continuous support and understanding. Last but not least, I would like to thank all my friends for their moral support and assistance especially when I have problem in completing this thesis. Thanks to all of you.

Sincerely from,

Nurul Sima Mohamad Shariff
Institute of Mathematical Sciences
Faculty of Science
University of Malaya

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

CD	Cross sectional dependence
iid	Identically and independently distributed
OLS	Ordinary Least Squares
PPP	Purchasing Power Parity
RER	Real Exchange Rates
APT	Arbitrage Pricing Theory
$\hat{\rho}_{ij}$	Pair-wise correlation coefficient of the residuals
$\lvert\hat{\rho}\rvert$	Absolute average of pair-wise correlation coefficient of the residuals
AO	Additive outliers
IO	Innovation outlier
LS	Level shift
TC	Temporary change
LP	Leverage point
MO	Multiple outliers
GMM	Generalized Method of Moments
LM	Breusch and Pagan (1980) Lagrange Multiplier Test
PCD	Pesaran's Cross Sectional Test (2004)
OECD	Organization for Economic Co-operation and Development
CMG	Common Correlated Effect Mean Group of Pesaran (2006)
ADF	Augmented Dickey-Fuller Test
CADF	Common Correlated ADF
N	Number of cross sectional units
T	Length of time periods
R_{ave}	Spearman's rank correlation coefficient
$f(x_{it}, y_{it}; \alpha_i, \beta_i)$	function of x_{it} and y_{it}
x_{it}	observed regressor (independent variable) on the i th cross section unit at time t
y_{it}	dependent variable (response variable) on the i th cross section unit at time t
α_i, β_i	parameters which are allowed to vary across i

i	cross sectional units
t	time
e_{it}	random errors component on the i th cross section unit at time t
y_{it-1}	first lagged value of y_{it}
$\Delta \boldsymbol{y_{it}}$	difference between y_{it} and y_{it-1} $(y_{it} - y_{it-1})$
f_t	latent (unobserved) factors
γ_i	factor loadings that is common for all i
ε_{it}	random errors of e_{it} on the i th cross section unit at time t
$d_i(x_{it})$	regressor of the outlyingness "X" from its mean value
$\hat{e}_{it}$	Fitted values of random errors component on the i th cross section unit at time t
$\hat{y}_{it}$	Fitted values of the dependent variables on the i th cross section unit at time t
$\bar{y}_t$	Cross section average of the dependent variables
$\bar{x}_t$	Cross section average of the independent variables
$\boldsymbol{\beta}$	the matrix of β_i
$\mathbf{x}_{it}$	$k \times 1$ vector of independent variables
H_0	Null hypothesis
H_1	Alternative hypothesis
rv	Random variables
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
RREG	Robust regression
RLM1	Robust version of the LM test based on Huber function
RLM2	Robust version of the LM test based on diagnostic tool
RPCD1	Robust version of the PCD test based on Huber function
RPCD2	Robust version of the PCD test based on diagnostic tool
LTS	Least Trimmed Squares
c	Tuning constant
LMS	Least Median Squares
k	number of regressor

MAD	Median Absolute Deviation
d	critical value for robust tool using diagnostic approach
CLT	Central Limit Theorem
$\gamma_i = 0$	Cross section independence
$\gamma_i \sim iidU(0.1,0.3)$	Mild (Low) CD
$\gamma_i \sim iidU(0.5,1.5)$	Strong CD
δ	Percentage of contaminations
DGP	Data Generating Process
MSE	Mean Squared Error
AIC	Akaike's information criterion
BIC	Bayesian information criterion
MLE	Maximum Likelihood Estimator
$\xrightarrow{d}$	convergence in distribution
$\xrightarrow{qm}$	convergence in quadratic mean (or mean square error)
$\xrightarrow{p}$	convergence in probability
R^2	Coefficient of determination
RR^2	Robust version of R^2
CV	Cross Validation
CV^2	CV based on squared residuals
RCV	Robust version of CV
RCV^2	Robust version of CV^2
SSR	regression sum of squares
SST	total sum of squares
SSE	residuals sum of squares
RCMG	Robust CMG
Type I error	Probability of rejecting the null when the null is true
Type II error	Probability of rejecting the alternatives when the alternative is true
I(0)	Integrated with order 0 (Stationary process)
I(1)	Integrated with order 1
MG	Mean Group Estimator
SUR	Seemingly unrelated mean groroup Estimat
PC	Principal Component
$nsimul$	Number or simulation

$\overline{\beta}$	Sample mean of parameter estimates
$SD(\beta)$	Sample standard deviation of parameter estimates
RMSE	Root Mean Square Error
CI	Confidence Interval
DF	Dickey-Fuller Unit Root Test
AR	autoregressive model
IPS	Im, Pesaran and Shin (1997) unit root test
CIPS	Cross-sectionally augmented IPS
CIPS*	truncated version of CIPS
RCIPS	Robust version of CIPS
GLS	Generalized Least Squares
CEEC	Central and Eastern Europe
US	United States
CNY	China
HK	Hong Kong
IND	India
INDO	Indonesia
KOR	Korea
MYS	Malaysia
MMR	Myanmar
PAK	Pakistan
PHL	Philippines
WSM	Samoa
SGP	Singapore
LKA	Sri Lanka
TWN	Taiwan
THA	Thailand
AL	Algeria
CR	Croatia
CZ	Czech Republic
EE	Estonia
HU	Hungary
LV	Latvia
LT	Lithuania
MK	Macedonia

PL	Poland
RO	Romania
SK	Slovakia
SL	Slovenia