ADDIS ABABA UNIVERSITY

Doctoral Dissertation

Alternative General Method of Moment Estimators in Dynamic Panel Data Models

Author:
Tegodie HIBSTU

Supervisors:
Dr. Emmanuel GABREYOHANNES,
Prof. Eshetu WENCHEKO

A dissertation submitted in fulfillment of the requirements for the degree of Doctor of Philosophy in Statistics

ADDIS ABABA UNIVERSITY, AAU

Addis Ababa University
Graduate Programs

Alternative General Method of Moment Estimators in Dynamic Panel Data Models

By

Tegodie Hibstu Mihretie

A dissertation submitted in fulfillment of the requirements for the degree of Doctor of Philosophy in Statistics

Approved by Examining Board:

Dr. Emmanuel GABREYOHANNES		June 14, 2024
Supervisor	Signature	Date
Prof. Eshetu WENCHEKO		June 14, 2024
Supervisor	Signature	Date
Prof. Sourafel GIRMA		June 14, 2024
Examiner	Signature	Date
Dr. Butte GOTTU		June 14, 2024
Examiner	Signature	Date
Dr. Merga BELINA		June 14, 2024
Chairman	Signature	Date

Declaration of Authorship

I, Tegodie Hibstu Mihretie, declare that this thesis titled: Alternative General Method of Moment Estimators in Dynamic Panel Data Models, and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- I have acknowledged all main sources of help.

Signed: ____

Date: May 10, 2024

ADDIS ABABA UNIVERSITY

Abstract

College of Natural and Computational Science
Department of Statistics, AAU

Doctor of Philosophy

Alternative General Method of Moment Estimators in Dynamic Panel Data Models

By Tegodie Hibstu Mihretie

Dynamic panel data models make it possible to address dynamic economic relationships through the inclusion of a lagged dependent variable among the explanatory variables. In the presence of lagged dependent variable as a regressor, however, the least squares-based estimators may not be consistent. The generalized method of moments (GMM) estimators are the most popular in dynamic panel data estimation. Two crucial issues in GMM estimation of dynamic panel models are the choice of the initial weighting matrix and the problem of instruments proliferation. In this study we propose alternative one-step and two-step system GMM estimators that utilize sub-optimal initial weighting matrices together with reduced instruments set (specifically lag-limited and partially collapsed instruments). Comparison of the performance of the proposed estimators against the Blundell-Bond system GMM estimator; the conventional system GMM estimator; and the sub-optimal system GMM estimator based on all available instruments was undertaken in terms of mean absolute bias, root mean squared error/standard deviation and coverage probabilities through Monte Carlo simulations. The small sample performance of the proposed estimators was also assessed using panel data of employment for manufacturing firms in Ethiopia.

The simulation studies show that sub-optimally weighted one-step GMM estimator based on collapsed instruments is the least biased for moderate and large values of the variance ratio of the individual-specific effects to that of random shocks. In terms of precision, sub-optimally weighted GMM estimator using all available instruments was found to be the most efficient, followed by that utilizing collapsed instruments set for moderate and large values of the variance ratio. As the value of T increases, the bias reduction from the latter outweighs the efficiency gain from the former in relative terms. The results also revealed that sub-optimally weighted two-step system GMM estimator adopting partially collapsed instruments outperforms the standard GMM estimator in terms of both bias and RMSE for large T and large variance ratios as the coefficient of the lagged dependent variable gets close to zero. Under these scenarios, the system GMM estimators based on reduced instruments were also found to perform well in terms of coverage probabilities. Moreover, for relatively small time dimensions and large variance ratios, the sub-optimal system GMM estimator based on all available instruments performs best in terms of precision. However, there is no considerable gain from the use of sub-optimal initial weights matrix in combination with reduced instruments as the process approaches a random walk for both one-step and two-step system GMM estimation. Our general conclusion is that the performance of sub-optimally weighted system GMM estimators which utilize collapsed as well as untransformed instruments seems promising.

Acknowledgments

I would like to thank the following people, without whom I would not have been able to complete my study. I am deeply grateful and indebted to my supervisor Dr. Emmanuel Gabreyohannes for his insightful comments, suggestions, guidance and overall assistance. His critical support was really influential in shaping my dissertation. I would like to extend my sincere thanks to Prof. Eshetu Wencheko for his advice, continuous support, and patience during my PhD study. His plentiful experience has encouraged me throughout my PhD dissertation project.

I would like to express my sincere gratitude to Mr. Mekonnen Taddese (Assoc. prof.), Dr. Bedlu Alamire, Dr. Dejen Tesfaw, Dr. Merga Belina, Dr.Shibru Temesgen and Dr. Taddesse Kassahun for their assistance at every stage of the current study.

I want to acknowledge my colleagues Mr. Mekonnen Elifnew (Asst. prof.), Shumetie Abeje, and Mr. Genanew Timerga (Asst. prof.) for their collaboration and support throughout this project

Last in this list but first in my heart, I would like to offer my special thanks to my wife Alem. Without her tremendous understanding and encouragement, it would be impossible for me to complete my study.

Tegodie Hibstu Mihretie

Addis Ababa, May 2024

Abbreviations

2SLS: Two Stages Least Squares

DPD: Dynamic Panel Data

ESS: Ethiopian Statistical Service

FD: First Difference

FE: Fixed effects

GIV: Generalized Instrumental Variables

GMM: Generalized Method of Moments

IV: Instrumental Variables

KI: Kantorovich Inequality

LMSMI: Large & Medium Scale Manufacturing Industries

MAB: Mean absolute bias

OLS: Ordinary Least Squares

RMSE: Root Mean square Error

Contents

Chapter 1

Introduction

In econometrics, panel data are multi-dimensional data involving repeated observations on the same cross-sectional units (individuals, households, firms, countries, etc.) over time. Studies based on dynamic panel data record changing information about the same cross section units over time. The sample may relate to individuals, households, firms, or collections of individuals in the form of cohorts or industries. The data may involve economic characteristics such as income, expenditure, and employment or indicators of health, well-being, and socioeconomic status.Panel data can be arranged in a matrix where each row tracks an individual or unit (i) at different points in time (t), thereby constituting a panel $\{y_{it} : i = 1, \cdots, n; t = 1, \cdots, T\}$ of observations of n individuals at T time periods leading to nT total observations of a particular variable (or vector of variables).

One of the major attractions of analyzing panel data rather than single indexed variables is that they allow us to cope with the empirically very relevant situation of unobserved heterogeneity correlated with included regressors. In other words, panel data provides a means of resolving econometric problems that often arise in empirical studies, specifically the presence of omitted (unobserved) variables that are correlated with explanatory variables (Arellano and Bond, 1991, Ahn and Schmidt, 1995 Blundell and Bond, 1998, Kiviet et al., 2017). Panel data estimation can be applied either in static panel model or dynamic panel model approaches. For static panel data models, one of the important issues is whether the individual specific term is correlated with the regressors or not. If the individual-specific effects are not correlated with the observed regressors, commonly we adopt random effects estimators. In this case, the unobservable individual effects are assumed to be random variables that are distributed independently of the regressors. Since the error components of the random effects model may be correlated over time for a given individual and/or may be heteroskedastic, we employ the feasible generalized least-square estimation procedures. On the other hand, if the individual-specific effects are correlated with the observed regressors, we use fixed effects estimator.

If lagged dependent variable terms are included in the usual individual-specific effects panel data model, the model is said to be a dynamic panel data model. Dynamic panel data models make it possible to address dynamic economic relationships through the inclusion of a lagged dependent variable among the explanatory variables. In this case, strict exogeneity of the regressors no longer holds. Since the lagged dependent variable is correlated with the individual specific effects (α_i), least squares-based estimators are not consistent. This inconsistency is referred to as dynamic panel bias (Nickel, 1981). In response to this problem of endogeneity, Anderson and Hsiao (1982) proposed instrumental variable (IV) estimators. In the 1990s, difference generalized method of moments (GMM) estimator was introduced by Arellano and Bond (1991) and then expanded to system GMM by Arellano & Bover (1995) and Blundell & Bond (1998). The GMM estimator is the most popular estimator due to its flexibility and to the few assumptions about the data generating process. In other words, even if we could have a known distribution for the error components of the model, the full description of the data generation process for the variables interest may not be known (see Harris and Matyas, 1999, Lai et al., 2008).

In dynamic panel data estimation, the number of GMM estimators is growing exponentially in the literature. However, it is hard to make a reasoned choice between many different possible implementations of these estimators. For example, using many instruments may improve the efficiency of GMM estimators, though, the asymptotically efficient GMM estimators may suffer from substantial bias when all available moment conditions are employed for small cross-sectional dimension sample sizes, n (Hall, 2005, Windmeijer, 2005, Bun and Kiviet, 2006; Doran and Schmidt, 2006, Roodman, 2009, Bun and Windmeijer, 2010; Kiviet et al., 2017, Croissant and Millo, 2019). This estimation problem calls for a GMM estimator with a reduced set of moment conditions. Second, the optimal initial weighting matrix for the system GMM estimator is only available when the variance of individual specific effects is zero. However, in practice, the variance of individual specific effects is often expected to be quite high even relative to the variance of idiosyncratic error terms (Bun and Windmeijer, 2010 and Jung et el., 2015). This necessitates a sub-optimal initial weighting matrix which makes use of the estimated variance ratio of the individual effects to that of the idiosyncratic error term.

To circumvent these bottlenecks, a number of alternative GMM estimators have been proposed. Bun and Kiviet (2006), Roodman (2009) and Fajeau (2020) tried to address the estimation problems due to large number of instruments. However, these studies did not address the estimation problems which may arise from the choice of an initial weight matrix in system GMM estimation. On the other hand, Kiviet (2007); Jung and kwon (2007); Jung, et al.(2015) and Youssef and Abonaze (2017) tried to address the estimation problems which may arise from the choice of an initial weight matrix in system GMM estimation, but did not consider the problems posed by instruments proliferation. In this dissertation, we attempted to address issues in connection with instruments proliferation along with the choice of an initial weight matrix in system GMM estimation procedures. In other words, we propose alternative system GMM estimators that utilize sub-optimal initial weighting matrices together with reduced instruments set.

The main objectives of this dissertation are:

1. To derive alternative system GMM estimators which utilize sub-optimal initial weighting matrices together with reduced instruments set (specifically, collapsed and lag limited instruments set),
2. To compare the performance of the proposed GMM estimators against the existing ones using Monte Carlo experiments, and
3. To compare the performance of the proposed estimators against the existing dynamic panel data models using empirical example.

According to Hansen, Heaten and Yaron (1996) the finite sample performance of the GMM estimator is sensitive to both the number of moments and the cross-sectional dimension sample size (n). A common choice of the instruments is a lower order polynomial of order $\mathrm{O}(T)$. In other words, the increment on the number of instruments should be linear in T.Therefore, first we reduced the instruments set using reduction techniques. Second, we adopt a sub-optimal initial weighting matrix which makes use of the estimated variance ratio of the individual effects to that of the idiosyncratic errors term. Finally, we propose alternative system GMM estimators that utilize suboptimal initial weighting matrices together with reduced instruments set (specifically lag-limited and partially collapsed instruments). Comparison of the performance of the proposed estimators against the existing estimators was undertaken through Monte Carlo experiments and real data analysis.

The rest of the dissertation is organized as follows. Chapter 2 introduces the AR (1) dynamic panel data model, least squares estimators and their limitations on dynamic panel data estimation, basic ideas of GMM estimation procedures, and the difference and system GMM estimators. Chapter 3 discusses the bottlenecks in GMM estimation in connection with instruments proliferation and moments' dimension reduction mechanisms, the asymptotic variance of GMM estimators, construction (derivation) details of our proposed estimators, and efficiency comparison of alternative estimators using kantorovich inequality efficiency bound. Simulation design, the results of experiments and empirical illustration are presented in chapter 4. Chapter 5 is devoted to conclusions and future research areas.

Chapter 2

Dynamic Panel Data Models

This chapter has four sections. The first section presents dynamic panel data model, and section 2 discusses about least squares estimation of dynamic panel data models. Section 3 presents the Anderson and Hsiao (1981, 1982) instrumental variable estimator. Section 4 is devoted to the discussion of generalized method of moments (GMM) estimators.

2.1 The model

The dynamic panel data model is given by:

$$y_{it} = \gamma y_{i,t-1} + x'_{it}\beta + \alpha_i + \epsilon_{it}; |\gamma| < 1;\ i = 1, \cdots, n, t = 1, \cdots, T. \tag{2.1}$$

where y_{it} is an observation on some series for individual i in period t, x_{it}, is $n \times (K+1)$ full rank matrix of K additional regressors, α_i is (unobserved) random individual specific effect, ϵ_{it} is time varying idiosyncratic error term. The model is linear in parameter (γ), and the error components (α_i, ϵ_{it}) and it can handle heteroscedasticity with in individual units. Further assumptions about dynamic micro panel data model are:

i) The idiosyncratic error terms and individual specific effects are identically and independently distributed: $\epsilon_{it} \sim iid(0, \sigma_\epsilon^2), \alpha_i \sim iid(0, \sigma_\alpha^2)$,
ii) ϵ_{it} and α_i are uncorrelated: E($\alpha_i \epsilon_{it}$)= 0,
iii) The initial observation satisfies: $y_{i1} = \frac{\alpha_i}{1-\gamma} + \omega_{i1}, i = 1, \cdots, n$, where $\omega_{i1} = \sum_{j=0}^{\infty} \gamma^j \omega_{i,1-j}$ is uncorrelated with both α_i and ϵ_{it} and ω_{i1} is covariance stationary.

We assume that we have observations on a large number of individuals (n) over a short time (T) and the model of interest is a regression model in which the lagged value of the dependent variable is one of the explanatory variables. In other words, we consider the commonly assumed and empirically relevant case of a large number of individuals (n) and a small number of time series observations per individual (i), so that we study the asymptotic properties of our estimators as $n \to \infty$ for fixed T (Ahn and Schmidt ,1995).

It is well known that exogeneity requires the explanatory variables to be uncorrelated with past, present, and future shocks. The presence of lagged dependent variable, however, violates strict exogeneity assumptions, that is, endogeneity may occur. The inclusion of lagged dependent variables as regressors has become growingly popular in modern econometrics in the last few decades and is now commonly adopted in empirical analyses. First, it offers the opportunity to control for unobserved individual heterogeneity, thus reducing the risk of biased estimates. Being safer against the omission of time-invariant explanatory variables, the need for good and relevant instruments is less compelling; hence we have less identification problems (Mammi and Calzolari, 2011). Secondly, the techniques help in estimating more complex models that also capture the dynamics of the variables of interest. More precisely, we can study the dynamics of the cross-section of interest over time, the transition probabilities among different states and the time-adjustment patterns.

2.2 Least squares estimation

For the panel model in (2.1), the individual-specific effect α_i is thought to be correlated with x_{it}. Furthermore, by construction, the lagged dependent variable is correlated with the individual specific effect, i.e. $E(\alpha_i|y_{i,t-1}) \neq 0$. Additionally, the covariate may also exhibit a non-zero correlation with the contemporaneous or lagged idiosyncratic errors such that E($\epsilon_{it}|x_{is}$)$\neq$ 0, for $t \leq s$. All these endogeneity issues imply that least squares-based estimators may not be consistent(Bun and Sarafidis, 2013). Specifically, the correlation between the lagged dependent variable and the individual effect makes the former endogenous, and consequently the estimators are not consistent. Nickel (1981) also shows that the fixed effects estimator has a non-vanishing bias for small T and large n. Consider the simple dynamic model:

$$y_{it} = \gamma y_{i,t-1} + \alpha_i + \epsilon_{it} \; ; i = 1, \cdots, n, \; t = 1, \cdots, T. \tag{2.2}$$

The model in deviations from the mean is given by :

$$y_{it} - \overline{y}_i = \gamma(y_{i,t-1} - \overline{y}_{i,-1}) + (\epsilon_{it} - \overline{\epsilon}_i) \tag{2.3}$$

where, $\overline{y}_{i,-1} = \frac{1}{T}\sum_{t=1}^{T} y_{i,t-1}$, $\overline{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$, $\overline{\epsilon}_i = \frac{1}{T}\sum_{t=1}^{T} \epsilon_{it}$

The fixed effects (FE) estimator of γ is then given by :

$$\begin{aligned}\hat{\gamma}_{FE} &= \left[\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{i,t-1}-\overline{y}_{i,-1})^2\right]^{-1}\left[\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{i,t-1}-\overline{y}_{i,-1})(y_{it}-\overline{y}_i)\right] \\ &= \gamma + \frac{(nT)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{i,t-1}-\overline{y}_{i,-1})(\epsilon_{it}-\overline{\epsilon}_i)}{(nT)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{i,t-1}-\overline{y}_{i,-1})^2} \qquad (2.4))\end{aligned}$$

The above transformation eliminated the unobserved heterogeneity α_i from the model. However,$(\epsilon_{it} - \overline{\epsilon}_i)$ is negatively correlated with $(y_{i,t-1} - \overline{y}_{i,-1})$.

We can show that :

$$\begin{aligned} &plim_{n\to\infty}\left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{i,t-1} - \overline{y}_{i,-1})(\epsilon_{it} - \overline{\epsilon}_i)\right] \\ &\qquad = -plim_{\to\infty}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\overline{y}_{i,-1}\overline{\epsilon}_i = \frac{-\sigma_\epsilon^2}{T^2}\frac{(T-1)-T\gamma+\gamma^T}{(1-\gamma)^2}\end{aligned}$$

Similarly, we can show that the denominator of $\hat{\gamma}_{FE}$ converges to :

$$plim_{n\to\infty}\left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{i,t-1}-\overline{y}_{i,-1})^2\right] = \frac{\sigma_\epsilon^2}{1-\gamma^2}\left[1-\frac{1}{T}-\frac{2\gamma}{(1-\gamma)^2}\times\frac{(T-1)-T\gamma+\gamma^T}{T^2}\right]$$

Thus, we have :

$$plim_{n\to\infty}(\hat{\gamma}_{FE}-\gamma) = -\frac{(1+\gamma)}{T-1}\left(1-\frac{1}{T}\frac{1-\gamma^T}{1-\gamma}\right)\times\left[\frac{1-2\gamma}{(1-\gamma)(T-1)}\left(1-\frac{1-\gamma^T}{T(1-\gamma)}\right)\right]^{-1}$$

(Nickel,1981).

This implies that the fixed effect estimator $\hat{\gamma}_{FE}$ is not consistent. Therefore, instrumental variable (IV) estimators have been proposed in response to the problem of endogeneity. Anderson and Hsiao (1982) were the first to propose an unbiased dynamic panel data (DPD) estimator with the notable trade-off between lag depth and sample size. In other words, there is a trade-off between the lag distance used to generate internal instruments and the depth of the sample (cross-sectional dimension of data, n) for estimation. The breakthrough came with difference Generalized Method of Moments (GMM) estimator by Arellano and Bond (1991) and System GMM by Arellano and Bover (1995) and Blundell and Bond (1998). The standard instrument set for difference GMM avoids the trade-off between instrument lag depth and sample depth by including separate instruments for each time period and zeroing out missing observations of lags (Holtz-Eakin, Newey and Rosen, 1988 and Arellano and Bond, 1991). Hence, the GMM estimator has become the most widely used estimator due to its flexibility and to the few assumptions about the data generating process. Its most appealing advantage is the availability of internal instruments, that is, the endogenous regressors are in fact instrumented by their previous realizations properly chosen according to meaningful moment conditions.

2.3 The Anderson-Hsiao (1981) estimator

The standard approach in cases where right-hand side variables are correlated with the residuals (when endogeneity occurs) is to estimate the equation using instrumental variables regression. The idea behind instrumental variables is to find a set of variables, termed instruments, which are correlated with the endogenous variables in the equation, but uncorrelated with the disturbances. Consider a dynamic panel as in (2.1). Taking the first difference to get rid of α_i we obtain:

$$\triangle y_{it} = \gamma \triangle y_{i,t-1} + \triangle x'_{it}\beta + \triangle \epsilon_{it}. \tag{2.5}$$

The OLS estimator for (γ, β)' obtained from (2.5) may not be consistent since $\triangle y_{i,t-1}$ and $\triangle \epsilon_{it}$ or more precisely $y_{i,t-1}$ and $\epsilon_{i,t-1}$ are by definition correlated. Anderson and Hsiao (1981) suggested using $y_{i,t-2}$ and/or $\triangle y_{i,t-2} = y_{i,t-2} - y_{i,t-3}$ as instruments for $\triangle y_{i,t-1}$. The lag difference of $X'_{it}s$ can be used as instruments for $\triangle x_{it}$ in (2.5) so that the instruments set for equation (2.5) becomes, $Z_i = (y_{i,t-2},\ \triangle x_{i,t-1})$. Hence, applying the generalized two-stages least square (2SLS) instrumental variable (IV) estimator, we obtain:

$$\begin{pmatrix} \hat{\gamma}_{IV} \\ \hat{\beta}_{IV} \end{pmatrix} = \left[\sum_{i=1}^{n} \sum_{t=2}^{T} \begin{pmatrix} (y_{i,t-1} - y_{i,t-2}) y_{i,t-2} & y_{i,t-2}(x_{it} - x_{i,t-1})' \\ (x_{it} - x_{i,t-1}) y_{i,t-2} & (x_{it} - x_{i,t-1})(x_{it} - x_{i,t-1})' \end{pmatrix} \right]^{-1} \times$$
$$\left[\sum_{i=1}^{n} \sum_{t=2}^{T} \begin{pmatrix} y_{i,t-2} \\ x_{it} - x_{i,t-1} \end{pmatrix} (y_{it} - y_{i,t-1}) \right]$$

Or using $\triangle y_{i,t-2} = y_{i,t-2} - y_{i,t-3}$ and $\triangle X_{it}$ as alternative instruments we have:

$$\begin{pmatrix} \hat{\gamma}_{IV} \\ \hat{\beta}_{IV} \end{pmatrix} = \left[\sum_{i=1}^{n}\sum_{t=3}^{T}\begin{pmatrix} (y_{i,t-1}-y_{i,t-2})(y_{i,t-2}-y_{i,t-3}) & (y_{i,t-2}-y_{i,t-3})(x_{it}-x_{i,t-1})' \\ (x_{it}-x_{i,t-1})(y_{i,t-2}-y_{i,t-3}) & (x_{it}-x_{i,t-1})(x_{it}-x_{i,t-1})' \end{pmatrix}\right]^{-1}$$
$$\left[\sum_{i=1}^{n}\sum_{t=3}^{T}\begin{pmatrix} y_{i,t-2}-y_{i,t-3} \\ x_{it}-x_{i,t-1} \end{pmatrix}(y_{it}-y_{i,t-1})\right]$$

The first estimation method has an advantage over the second one in that the minimum number of time periods required is two, whereas the second estimation requires $T \geq 3$. In practice, if $T \geq 3$ the choice between the two depends on the correlations between $\triangle y_{i,t-1}$ and $y_{i,t-2}$ or $\triangle y_{i,t-2}$. The two approach of estimations yield two consistent estimators, but does not exploit all relevant moment conditions that may not give efficient estimators.

2.4 Generalized Method of Moments (GMM) estimators

2.4.1 Basic ideas of GMM estimators

The general method of moments is an estimation procedure that allows economic models to be specified while avoiding often unnecessary assumptions. The GMM estimation is based on moment functions that depend on observable random variables and unknown parameters. These moment conditions are functions of the model parameters and the data, such that their expectation is zero at the true values of the parameters. In other words, it requires that the sample correlations between error terms and instruments is as close to zero as possible(depends on the optimal of weight matrices employed for estimation) and estimates are chosen to minimize the weighted distance between the theoretical and actual population values. The main peculiarity of the GMM estimator for dynamic panel data (DPD) is that it exploits internal instruments, namely the lags of the endogenous variables, and also allows the use of external instrumental variables. GMM estimators are also valid under heteroskedasticity of error terms and autocorrelation within individual unit errors (but not between themselves), that is, $E(\epsilon_{it}\epsilon_{is}) = 0$ for all $t \neq s$. Adopting the general notation of Wooldridge (2010), this is formalized as:

$$E\left[g(w_i, \theta)\right] = 0, \tag{2.6}$$

where $g(.)$ is l dimensional vector of moments, w_i is a vector of observable random variables that contains model variables in (2.1) and θ is a $(p \times 1)$ vector of unknown parameters. If the model is exactly identified $(l = p)$, then there are l moment conditions and they all hold exactly. This gives the standard IV estimator. If the model is over-identified $(l > p)$, however, there is no unique solution since not all sample moments corresponding to (2.6) will hold exactly (Hansen, 1982). In over-identified case, provided we have a random sample and equations provided by the sample moment conditions, we have an estimator of θ that brings the sample moments as close to zero as possible (Hansen, 1982). This is achieved by minimizing the quadratic form:

$$Q_n(\hat{\theta}) = \min_\theta \left[\textstyle\sum_{i=1}^n g_n(w_i, \theta) \right]' W_n \left[\textstyle\sum_{i=1}^n g_n(w_i, \theta) \right] \tag{2.7}$$

with respect to the parameters θ, where W_n is a positive definite $l \times l$ weighting matrix that yields a consistent estimator of θ. Differentiation can be used to find the solution, $\hat{\theta}$ which solves:

$$\left[\textstyle\sum_{i=1}^n G_n(w_i, \theta) \right]' W_n \left[\textstyle\sum_{i=1}^n g_n(w_i, \theta) \right] = 0,$$

where

$$G_n(w, \theta) = E \left[\tfrac{dg_n(w,\theta)}{d\theta} \right] = E \begin{bmatrix} \frac{dg_n(w_i,\theta)}{d\theta_1} \cdots \frac{dg_n(w_i,\theta)}{d\theta_p} \\ \vdots \\ \frac{dg_n(w_i,\theta)}{d\theta_1} \cdots \frac{dg_n)(w_i,\theta)}{d\theta_p} \end{bmatrix}$$

It is a matrix of derivatives with l rows and p columns where each row contains the derivative of one of the moment conditions with respect to all p parameters and each column contains the derivative of the l moment conditions with respect to a single parameter. Defining the orthogonality condition for the linear model, we have

$$g_n(w_i, \theta) = E\left[Z_i'(y_i - X_i'\theta)\right].$$

Hence,

$$G_n(w, \theta) = [-Z_i'x_{1i}, -Z_i'x_{2i}, \cdots, -Z_i'x_{pi}] = -Z_i'X_i \sim (L \times P).$$

The first-order condition becomes :

$$\left[\sum_{i=1}^{n} Z_i' X_i\right]' W_n \left[\sum_{i=1}^{n} Z_i'(y_i - X_i'\theta)\right].$$

Then, it follows that,

$$\left[\sum_{i=1}^{n} Z_i' X_i\right]' W_n \left[\sum_{i=1}^{n} Z_i' y_i\right] = \left[\sum_{i=1}^{n} Z_i' X_i\right]' W_n \left[\sum_{i=1}^{n} Z_i' X_i \theta\right].$$

Hence, the solution:

$$\hat{\theta} = \left(\left[\sum_{i=1}^{n} Z_i' X_i\right]' W_n \left[\sum_{i=1}^{n} Z_i' X_i\right]\right)^{-1} \left[\sum_{i=1}^{n} Z_i' X_i\right]' W_n \left[\sum_{i=1}^{n} Z_i' y_i\right] \quad (2.8)$$

defines the GMM estimator for the linear model, and W_n has not been specified other than requiring that this weighting matrix is positive definite. The best choice of weighting matrix is the inverse of the covariance of the population moment condition $S^{-1} = \lim_{n\to\infty} \left[\frac{1}{\sqrt{n}} g_n(w_i, \theta)\right]^{-1}$ Windmeijer, (2005). This weight matrix produces the GMM estimator with the smallest variance asymptotically (Wooldridge, 2010). The problem is that the inverse of the covariance matrix for population moments is unknown. So this estimator is infeasible since the covariance of the moment conditions generally depends on the unknown parameter vector. In practice, θ is obtained by solving the minimization problem in (2.7) with an inefficient choice of W_n that does not depend on θ (for example, the identity matrix of dimension, l). The two step estimation uses the residuals of the first-step estimate in estimating asymptotically efficient optimal weight matrix. Assume that a consistent estimation method is being used so that

$$\hat{S} \xrightarrow{p} S$$

$$W_n = \hat{S}^{-1} \xrightarrow{p} S^{-1}.$$

Even if the asymptotic distribution of the two-step procedure that delivers the efficient GMM is equivalent to the infeasible optimal GMM estimator that uses S^{-1} as a weighting matrix, the finite sample properties are affected by the first step estimation (for example see Altonji and Segal, 1996; Kiviet et al., 2017).

2.4.2 The Arellano-Bond (1991) estimator

The Arellano-Bond estimator is similar to that suggested by Anderson and Hsiao (1981, 1982) but exploits additional moment restrictions aimed at improving the efficiency of estimators. Arellano and Bond (1991) show that the list of instruments can be extended by exploiting additional moment conditions and letting their number vary with time. Here, we first applying differencing to eliminate the individual effects, and then estimate the parameters by the IV method using the values of the dependent variable lagged two or more periods as instruments. Consider the simple autoregressive panel in (2.2) and its first difference (FD) version:

$$\triangle y_{it} = \gamma \triangle y_{i,t-1} + \triangle \epsilon_{it} \ ; \ i = 1, \cdots, n, t = 1, \cdots, T. \tag{2.9}$$

Then for $T = 3$, we have y_{i1} as a valid instrument for $\triangle y_{i2}$ since it is highly correlated with $\triangle y_{i2}$ but not correlated with $\triangle \epsilon_{i3}$. When $t = 4$, y_{i2} as well as y_{i1} are valid instruments for $\triangle y_{i3}$ since both are not correlated with $\triangle \epsilon_{i4}$. Continuing in this fashion, the set of valid instruments at a time T becomes $(y_{i1}, y_{i2}, \cdots, y_{i,T-2})$. Define the $(T-2) \times (T-2)(T-1)/2$ instruments matrix for each individual i as:

$$Z_i^D = \begin{bmatrix} y_{i1} & 0 & 0 & \cdots & 0 \\ 0 & y_{i1} & y_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & y_{i1}, y_{i2}, \cdots, y_{i,T-2} \end{bmatrix} \tag{2.10}$$

Here each row contains the instruments that are valid for a given period. The set of all moment conditions can be written as $E\left[Z_i^{D\prime} \triangle \epsilon_i\right] = 0$, where $\triangle \epsilon_i = (\triangle \epsilon_{i3}, \triangle \epsilon_{i4}, \cdots, \triangle \epsilon_{iT})'$, equivalently $E\left[Z_i^D (\triangle y_i - \gamma \triangle y_{i,-1})\right] = 0$, where $\triangle y_i = (\triangle y_{i3}, \cdots, \triangle y_{iT})'$ and $\triangle y_{i,-1} = (\triangle y_{i2}, \cdots, \triangle y_{i,T-1})'$ are $(T-2) \times 1$ vectors. Define the $n(T-2) \times (T-2)(T-1)/2$ matrix of instruments for n cross sectional units as $(Z_1^{D\prime}, Z_2^{D\prime}, \cdots, Z_n^{D\prime})$ and rewrite the differenced equation in (2.9) as:

$$\triangle y = \gamma \triangle y_{-1} + \triangle \epsilon, \tag{2.11}$$

where

$$\triangle y = \begin{bmatrix} \triangle y_1 \\ \vdots \\ \triangle y_n \end{bmatrix}_{n(T-2)\times 1}, \triangle y_{-1} = \begin{bmatrix} \triangle y_{1,-1} \\ \vdots \\ \triangle y_{n,-1} \end{bmatrix}_{n(T-2)\times 1}, \triangle \epsilon = \begin{bmatrix} \triangle \epsilon_1 \\ \vdots \\ \triangle \epsilon_n \end{bmatrix}_{n(T-2)\times 1}$$

Arellano and Bond (1991) propose a two-step feasible GLS estimation procedure. The one step estimator is under the assumption that $\epsilon_{it}^{\prime}s$ are identically and independently distributed (iid) over both i and t. The residuals of the first step estimates are then used for the second step estimator. The one-step Arellano and Bond estimator is given by:

$$\hat{\gamma}^{D,1} = \left[(\triangle y_{-1})' Z^D (Z^{D\prime}(I_n \otimes D) Z^D)^{-1} Z^{D\prime}(\triangle y_{-1})\right]^{-1} \times$$
$$\left[(\triangle y_{-1})' Z^D (Z^{D\prime}(I_n \otimes D) Z^D)^{-1} Z^{D\prime}(\triangle y\right]. \qquad (2.12)$$

where $E\left[\triangle\epsilon\triangle\epsilon'\right] = \sigma_\epsilon^2 D$, and D is the matrix given by:

$$D = \begin{bmatrix} 2 & -1 & 0 & ,\cdots, & 0 & 0 \\ -1 & 2 & -1 & ,\cdots, & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & ,\cdots, & 2 & -1 \\ 0 & 0 & 0 & ,\cdots, & -1 & 2 \end{bmatrix}_{(T-2)\times(T-2)}$$

The one-step optimal differences GMM estimator can be obtained only if $\epsilon_{it}^{\prime}s$ are assumed to be homoscedastic and exhibit no serial autocorrelation. However, the GMM approach does not impose that the $\epsilon_{it}^{\prime}s$ are homoscedastic, and the weight matrix (W) can be estimated without imposing this restriction. To do so, we need to replace

$Z^{D\prime}(I_n \otimes D) Z^D = n^{-1}\sum_{i=1}^n Z_i^{D\prime} D Z_i^D$ by

$\hat{W}_{n(2)}^D = \left(n^{-1}\sum_{i=1}^n Z_i^{D\prime}\triangle\hat{\epsilon}\triangle\hat{\epsilon}' Z_i^D\right)^{-1}.$

where $\triangle\hat{\epsilon}_i = \triangle y_i - \hat{\gamma}^{D,1}\triangle y_{-1}$. The resulting two-steps Arellano and Bond GMM estimator is given by:

$$\hat{\gamma}^{D,2} = \left[(\triangle y_{-1})' Z^D \hat{W}_{n(2)}^D Z^D (\triangle\ y_{-1})^{-1}\left[(\triangle y_{-1})' Z^D \hat{W}_{n(2)}^D Z^D \triangle y\right]\right. \qquad (2.13)$$

The asymptotic variance of two-steps Arellano and Bond GMM estimator given as

$$var(\hat{\gamma}^{D,2}) = \left[(\triangle y_{-1})' Z^D \hat{W}^D_{n(2)} Z^{D\prime} (\triangle y_{-1})\right]^{-1}.$$

Now consider a general dynamic panel model:

$$y_{it} = \gamma y_{i,t-1} + \beta' x_{it} + \alpha_i + \epsilon_{it} \tag{2.14}$$

where x_{it} is the $k \times 1$ vector of regressors. Its first-difference version becomes:

$$\triangle y_{it} = \gamma \triangle y_{i,t-1} + \beta' \triangle x_{it} + \triangle \epsilon_{it}$$

There are two cases we will look into.

Case(1): Suppose that the k-dimensional regressors x_{it} are strictly exogenous such that $E(x_{it}\epsilon_{is}) = 0, \quad \forall, t, s$. Then all x_{it} are valid instruments for (2.14). If we use only one additional regressor x_{it} defined by $X_i = (x_{i1}, \cdots, x_{iT})'$ to be added to each diagonal element of Z_i^D in (2.10), we have the $(T-2) \times [(T-2)(T-1)/2 + T(T-2)]$ instruments matrix for each individual i:

$$Z_i^D = \begin{pmatrix} y_{i1} & x_{i1} & \cdots & x_{iT} & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & y_{i1} & y_{i2} & x_{i1} & \cdots & x_{iT} & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \cdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & \cdots & y_{i1} & \cdots & y_{i,T-2} & x_{i1} & \cdots, & x_{iT} \end{pmatrix}$$

Writing (2.14) in matrix form and pre-multiplying it by $Z^{D\prime}$ we obtain:

$$Z^{D\prime} \triangle y_i = Z^{D\prime} \gamma \triangle y_{i,-1} + Z^{D\prime} \beta \triangle x_i + Z^{D\prime} \triangle \epsilon_i \tag{2.15}$$

where $\triangle y_i, \triangle y_{i,-1}, \triangle \epsilon_i$, are as defined in (2.10), and $\triangle x_i$ is additional regressor in (2.15) instrumented by $(T-2) \times T(T-2)/2$ number of instruments.The two-step GLS estimator can then be obtained by:

$$\begin{pmatrix} \hat{\gamma}^{D,2} \\ \hat{\beta}^{D,2} \end{pmatrix} = \left[(\triangle k)' Z^D \hat{W}^D_n Z^{D\prime} (\triangle K)\right]^{-1} \left[(\triangle k)' Z^D \hat{W}^D_n Z^{D\prime} \triangle y\right] \tag{2.16}$$

where $\triangle k = (\triangle y_{-1}, \triangle x)$ are model explanatory variables.

Case (2): $E(x_{it}\epsilon_{is}) = 0,\ s \leq t.$

In this case, $x'_{it}s$ are predetermined rather than strictly exogenous (we still assume that all α_i are not correlated with $\triangle x_{it}$). Then only $(x_{i1}, x_{i2}, \cdots, x_{i,s-1})'$ are valid instruments for the differenced equation at period T. The instruments matrix in (2.14) becomes:

$$Z_i^D = \begin{bmatrix} y_{i1}, x_{i1}, x_{i2} & 0 & \cdots & 0 \\ 0 & y_{i1}, y_{i2}, x_{i1}, x_{i2}, x_{i3)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & y_{i1}, \cdots, y_{i,T-2}, x_{i1}, \cdots x_{i,T-1} \end{bmatrix}$$

Thus, the additional explanatory variable, $\triangle x_i$ is instrumented by $(T-2)\times(T-2)(T-1)/2$ number instruments. The two-step GLS estimator of $(\gamma, \beta)'$ can be obtained by using similar procedures as in (2.16).

2.4.3 The Arellano-Bover (1995) estimator

Arellano and Bover (1995) suggested the method which eliminates the individual effect from instrumental variables. In this approach, instruments for levels are lagged differences of the dependent variable (y_{it}) which are uncorrelated with the components of level equation. Specifically, Arellano and Bover (1995) considered the level model:

$$y_{it} = \gamma y_{i,t-1} + u_{it},\ u_{it} = \alpha_i + \epsilon_{it},\ i = 1, \cdots, n;\ t = 1, \cdots, T \tag{2.17}$$

and then they used the following matrix as instrumental variables:

$$Z_i^L = \begin{bmatrix} \triangle y_{i2} & 0 & \cdots & 0 \\ 0 & \triangle y_{i3} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \triangle y_{i,T-1} \end{bmatrix}$$

which is not contains individual effect and satisfies the orthogonal conditions

$$E(Z_i^{L\prime} u_i) = 0 \tag{2.17}$$

Using (2.18), Arellano and Bover (1995) one-step level GMM estimator is computed as follows:

$$\hat{\gamma}^L = \left(y'_{-1} Z^L W^L_{n1} Z^{L\prime} y_{-1}\right)^{-1} y'_{-1} Z^L W^L_{n1} Z^{L\prime} y \tag{2.18}$$

$y = (y_1, \cdots, y_n)'$; $y_{-1} = (y_{1,-1}, \cdots, y_{n,-1})'$; $Z^L = (Z^L_1, \cdots, Z^L_n)'$ and an initial weight matrix for level GMM estimator is given by:

$$W^L_{n1} = \left(\frac{1}{n}\sum_{i=1}^n (Z_i^{L\prime} Z_i^L\right)^{-1}$$

Youssef, et al. (2014) present an optimal weighting matrix (based on Liu and Neudecker (1997) Kantorovich inequality efficiency upper bound criteria) for Level GMM estimator and it is computed as:

$$W^{OL} = \left(\frac{1}{n}\sum_{i=1}^n (Z_i^{L\prime} J_{T-2} Z_i^L\right)^{-1}; \; J_{T-2} = I_{T-2} + \rho 1_{T-2} 1'_{T-2}; \; \rho = \frac{\sigma^2_\alpha}{\sigma^2_\epsilon}$$

where 1_{T-2} is a column matrix of ones with dimension $(T-2)\times 1$. The level estimator is more informative compared with the difference estimator when the series is persistent. However, it is less efficient as compared to system GMM estimator (since it has more instruments).

2.4.4 The Blundell and Bond (1998) estimator

The difference GMM estimator has a serious drawback when the auto regressive parameter is moderately large ($\gamma \to 1$) since the lagged values of the levels of the original regressors are frequently weak instruments for the differenced values of the regressors in this case (Blundell & Bond, 1998). Arellano-Bover (1995) develops GMM framework for looking at efficient instrumental variable estimators by removing the individual specific term from instrumental variables for DPD models which is more informative in cases where level instruments for first-differenced equation become weak. Arellano & Bover (1995) and Blundell & Bond (1998) proposed a system GMM estimator in which the moment conditions of differences and levels are jointly used to circumvent weak instruments problem and improve the efficiency of the estimator. Here, lagged levels are used as instruments for difference equations and lagged differences are used as instruments for level equations. Both moment conditions assume the same linear functional form and specification of the data generating process, and the whole system can be estimated by one single operation that applies to both systems of equations. Blundell and Bond (1998) have shown that the proposed estimator performs much better than the difference GMM estimator especially as γ approaches to one or $\sigma_\alpha^2/\sigma_\epsilon^2$ becomes large. The system GMM estimator requires an additional assumption that the first differences of instrument variables are uncorrelated with the individual specific unobserved effects. The moment conditions for the system GMM estimator include lagged levels as well as lagged differences, and are given by:

$$E\left[y_{i,t-s}\triangle\epsilon_{it}\right] = 0, \text{for } t = 3, \cdots, T; , 2 \leq s \leq t-1; i = 1, \cdots, n \qquad (2.19)$$

$$E\left[\triangle y_{t-1}u_{it}\right] = 0, \text{ for, } t = 3, \cdots, T; i = 1, \cdots, n. \qquad (2.20$$

where $u_{it} = \alpha_i + \epsilon_{it}$. The instruments matrix for system equations is given by:

$$Z_i^S = \begin{bmatrix} Z_i^D & 0 \\ 0 & Z_i^L \end{bmatrix}$$

where Z_i^D is as defined in (2.10) and Z_i^L is defined as:

$$Z_i^L = \begin{bmatrix} \triangle y_{i2} & 0 & \cdots & 0 \\ 0 & \triangle y_{i3} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \triangle y_{i,T-1} \end{bmatrix}_{(T-2)\times(T-2)} \qquad (2.21)$$

Using the moment conditions in (2.19) and (2.20) we have:

$$E\left(Z_i^S u_i^S\right) = 0 \qquad (2.22)$$

where $u_i^S = (\triangle u_i', u_i')$. The one-step system GMM estimator is computed as:

$$\hat{\gamma}^{S1} = \left(\tilde{y}_{-1}' Z^S W_{n(1)}^S Z^{S\prime} \tilde{y}_{-1}\right) \tilde{y}_{-1}' Z^S W_1^S Z^{S\prime} \tilde{y} \qquad (2.23)$$

where $\tilde{y}_{-1} = \left[(\triangle y_{1,-1}', y_{1,-1}'), \cdots, (\triangle y_{n,-1}', y_{n,-1}')\right]'$,

$\tilde{y}_1 = [(\triangle y_1', y_1'), \cdots, (\triangle y_n', y_n')]'$, $Z^S = \left(Z_1^{S\prime}, \cdots, Z_n^S\right)$, and

$W_{n(1)}^{S} = \lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i^{S\prime} H_I Z_i^{S}\right)^{-1}$ with

$$H_I = \begin{pmatrix} D & 0 \\ 0 & I_{T-2} \end{pmatrix}$$

where D is as defined in (2.12) and I_{T-2} is the identity matrix of dimension $T-2$. The two-step system GMM estimator can be obtained by applying similar procedures as in (2.13).

Moreover, consider multivariate dynamic model in (2.14) on system GMM estimation procedures, as suggested by Arellano and Bover (1995) suitably dated first differences $\triangle x_{is}$ can be used as instruments for level equations.

Case(1): If x_{it} is strictly exogenous or predetermined with respect to ϵ_{it}, we apply $(\triangle x_{i2}, \cdots, \triangle x_{iT})$ as instruments for the period T equation in levels i.e. we have $T-1$ moment conditions: $[\triangle x_{is}(\alpha_i + \epsilon_{it})] = 0$, $s = 2, \cdots, T$.

Case(2): If x_{it} is correlated with ϵ_{it}, we can only use $(\triangle x_{i2}, \cdots, \triangle x_{i,T-1})$ as instruments for the period T equation in levels. i.e., we have $T-2$ moment conditions of the form $[\triangle x_{is}(\alpha_i + \epsilon_{it})] = 0$, $s = 2, \cdots, T-1$.

Assumption on initial observation

Consider a unique additional explanatory variable x_{it} for simplicity, and allow the presence of endogeneity as follows:

$$x_{it} = \delta x_{i,t-1} + \tau\alpha_i + \lambda\epsilon_{it} + e_{it},\ i = 1, \cdots, n,\ t = 1, \cdots, T \qquad (2.24)$$

where δ captures the persistence of x_{it} , τ & λ determine the correlation of x_{it} with the individual effects & the idiosyncratic errors, respectively (see for example, Blundell & Bond, 1998; Ashley and Sun, 2016). We also assume that α_i, ϵ_{it} and e_{it} have the following properties:

$$E(e_{it}) = E(\alpha_i e_{it}) = 0,\ i = 1, \cdots, n;\ 1, \cdots, T,$$

$$E(e_{it}e_{is}) = 0\ ;\ E(\epsilon_{it}e_{is}) = 0 \text{ and } \forall_{t\neq s}.$$

We impose mean-stationarity restrictions on the initial conditions:

$$x_{i1} = \tfrac{\tau}{1-\delta}\alpha_i + v_{i1},\ i = 1, \cdots, n, \qquad (2.25)$$

$$y_{i1} = \tfrac{1}{1-\gamma}\left(1 + \tfrac{\beta\tau}{1-\delta}\right)\alpha_i + \omega_{it},\ i = 1, \cdots, n \tag{2.26}$$

$$E(v_{i1}) = E(\alpha_i v_{i1}) = E(\omega_{it}) = E(\alpha_i \omega_{it}) = 0;\ i = 1, \cdots, n$$

$$E(v_{i1}\epsilon_{it}) = E(\omega_{it}\epsilon_{it}) = 0,\ i = 1, \cdots, n;\ 1, \cdots, T.$$

In econometrics literature, instruments proliferation, the choice of an initial weight matrix and the validity of mean stationary assumption are the potential drawbacks in system GMM estimation (see details, Roodman, 2009).

Chapter 3

Alternative General Method of Moment Estimators in Dynamic Panel Data Models

This chapter has five sections. The first two sections discuss the bottlenecks in GMM estimation in connection with instruments proliferation and moments' dimension reduction mechanisms. Section 3 presents the asymptotic variance of GMM estimators. Section 4 presents derivation details of our proposed system GMM estimators. The efficiency comparison of alternative estimators based kantorovich inequality efficiency bound is presented in section 5.

3.1 Instruments proliferation and estimation problems

In dynamic panel data models, the number of moment restrictions increases rapidly with T. This dramatic increase of the instrument count is often referred to as instruments proliferation. Theoretically, the efficiency of the estimator improves as the number of instruments increases. However, this is unreliable guideline in small samples, as the inclusion of an excessive number of moment conditions results in more pronounced bias and severely underestimates the variance (standard errors) of the GMM estimator (Andersen et Sorensen, 1996; Bowsher, 2002; Windmeijer 2005; Doran & Schmidt, 2006; Park, 2007; Roodman, 2009 Okui, 2009, Fajeau, 2020). Cesar Alonso-Borrego and Manuel Arellano (1999) conclude that, as the number of instruments increases with T, the advantage of such extra instruments is often very small because they tend to be only weakly correlated with first-differenced endogenous variables. Ziliak (1997) also show that the bias in GMM is quite severe as the number of moment conditions expands, outweighing the gains in efficiency. The burden of instrument proliferation and weak instruments lies in the laps of the over-identifying restrictions generated by the difference part of the system GMM estimator. The main consequences of instruments proliferation are: over-fitting endogenous variables, imprecise estimates of the optimal weighting matrix, downward-bias in two-step standard errors and weak Hansen test of instrument validity (see, for example, Roodman, 2009).

3.1.1 Over-fitting endogenous variables

Over fitting occurs when the number of instruments, hence the number of elements in the weight matrix to be estimated becomes large compared with cross-sectional dimension of panel. In this case, the instruments failing to expunge the the endogenous components of instrumented variables (endogenous explanatory variables) and biasing estimates towards those from non-instrumented ones (see Roodman, 2009; Bun and Windmeijer, 2010).

3.1.2 Imprecise estimates of the optimal weighting matrix

The two step GMM estimation utilizes the weighting matrix, which is the inverse of the covariance of moments in (2.22), is of dimension of $(l \times l)$ and contains $l \times (l \times l)/2$ unique elements. Since every element of this matrix is estimated through an empirical average calculated over the individuals in the sample, obtaining precise estimates of the matrix elements is problematic unless n is large relative to l (Ziliak, 1997; Alonso-Borrego & Arellano, 1999; Windmeijer, 2005; Kiviet, 2007; Roodman, 2009; Croissant and Millo, 2019; Fajeau, 2020).

3.1.3 Downward-bias in two-step standard errors

The weight matrix used in the calculation of the efficient for two-step GMM estimator is based on the initial consistent parameter estimates. However, in an environment prone to instrument proliferation, the standard errors of the coefficient tend to be severely downward biased. The bias is caused by the correlation between poorly estimated weighting matrix and the sample moment conditions (Arellano and Bond, 1991; Blundell and Bond, 1998; Windmeijer, 2005; Doran and schmidt, 2006; Roodman, 2009; Kiviet et al. 2017).

3.1.4 Weakened Hansen test of instruments validity

Hansen test of instruments validity is implemented as a test of whether the sample moments corresponding to these restrictions are sufficiently close to zero. Accordingly, H_0: the restrictions are sufficiently close to zero; in which case, the instruments are valid. Hansen (1982) J -test is computed as:

$$J = \frac{1}{n} Z^{S\prime} \hat{u}(\hat{\theta}) \hat{W}_n(\hat{\theta}) Z^{S\prime} \hat{u}(\hat{\theta})$$

where $\hat{\theta} = (\hat{\gamma}, \hat{\beta})$ are estimators of the parameters, $\hat{W}_n(\hat{\theta}) = (\hat{W}_n(\hat{\theta}_1), \hat{W}_n(\hat{\theta}_2))$ are the estimated weight matrix utilizing residuals of one-step and two-step estimators, respectively, and $\hat{u}(\hat{\theta}) = (\hat{u}(\hat{\theta}_1), \hat{u}(\hat{\theta}_2))$ are the residuals of one-step and two-step estimators. The J-test is usually and reasonably thought of as a test of instrument validity. However, when the number of the instruments increases, the conventional J-test of over-identifying restrictions performs poorly. In other words, the instruments proliferation weakens Hansen test of instrument validity or makes it unable to reject the null hypothesis of instrument validity (for example, see Lai et al. 2008; Carrasco and doukali, 2022). As a solution, limiting the number of instruments to some degree usually reduces bias and therefore improves small sample properties of GMM estimators (Bun and Kiviet, 2006 ; Roodman; 2009; Kiviet et al. 2017). A rule of thumbs is that the instrument count should not exceed the number of cross-sectional dimension (n). However, this threshold has no particular theoretical basis due to the absence of formal procedure to determine how far one should go reducing the number of moment conditions (Okui, 2009; Roodman, 2009). A common choice to reduce the instrument count which is linear in T. To do so, Roodman (2009) proposed the use of a subset (only the most recent lags) of instruments and collapsing of instruments set.

3.2 Instruments reduction techniques

3.2.1 Lag-depth truncation

In the standard instruments set for system GMM estimator, moment conditions corresponding to differenced equation as in (2.19) contribute significantly to instrument proliferation. The most distant instruments often have the weakest correlation with the lagged dependent explanatory variable and are therefore less relevant. One of the solutions is to reduce the dimensionality of the instrument matrix by using only a subset of available instruments (Bun and Kiviet, 2006; Roodman, 2009; Wooldridge, 2010). In this case, we only exploit the moment conditions for a certain lag depth such that we can rewrite the moment condition in (2.19) as follows:

$$E\left[y_{i,t-s}\triangle\epsilon_{it}\right]=0, \quad \text{for } 2\leq s\leq\tau, t=3,...,T \tag{3.1}$$

where τ is the maximum lag depth to be considered. For difference GMM estimator, for example when $\tau = 2$, lag-depth truncation technique reduces the instrument count to $[(T-2)(T-1)/2]-[(T-4)(T-2)/2]$ for $T\geq 4$ and the instruments matrix in (2.10) is modified into:

$$\left((Z_i^D\right)^L=\begin{bmatrix} y_{i1} & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & y_{i2} & y_{i1} & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & y_{i3} & y_{i2} & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & y_{i4} & y_{i3} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & y_{i,T-2} & y_{i,T-3} \end{bmatrix}$$

For example, let $T = 6$ and we use the instruments set in (2.10), the instruments count is $[(6-2)(6-1)/2] = 10$. But, if we limit the maximum lag depth to two ($\tau = 2$), the instruments count is reduced into $10-[(6-4)(6-2)/2] = 7$. The numbers of instruments in (2.21) is not affected, since we use only the most recent lagged difference as instrument for level explanatory variable in each time t ($t = 3,\cdots,T$).

3.2.2 Collapsing of the instruments set

Collapsed instruments set is achieved by squeezing the number of columns in (2.10) based on only their lag distance for all periods but not for each period. In other words, we impose the following restrictions on the moment conditions in (2.19):

$$E\left[\sum_{s=1}^{t-1}y_{is}\triangle\epsilon_{it}\right]=0,\ t=3,\cdots,T, \tag{3.2}$$

The matrix of instruments for each individual i now becomes:

$$\left(Z_i^D\right)^C = \begin{bmatrix} y_{i1} & 0 & 0 & \cdots & 0 \\ y_{i2} & y_{i1} & 0 & \cdots & 0 \\ y_{i3} & y_{i2} & y_{i1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{i,T-2} & y_{i,T-3} & y_{i,T-4} & \cdots & y_{i1} \end{bmatrix}$$

Moreover, the matrix Z_i^L in (2.21) now reduces to $(Z_i^L)^C = (\triangle y_{i2}, \triangle y_{i3}, \cdots, \triangle y_{iT})'$. Consequently, the number of instruments in (2.10) reduced into $(T-2)$ and the number of instruments in (2.21) is reduced into only one. Here, we did not apply Roodman (2009) approach in collapsing the instruments set for the level part of moment conditions. This is due to the reason that our study approach requires a square matrix of dimension $(T-2)$ for applying sub-optimal weight matrix corresponds to moment conditions level part of system GMM. Moreover, Bun and Kiviet (2006) show that reducing the number of instruments further which inhibiting the block diagonal structure of the Z_i^L matrix no longer reduces the bias of estimator. They also show that the order of bias is not with respect to the reduction in number of instruments by removing the longest lags (using lag limited) instead it is the order of magnitude of their total number which determines the order of the bias (as in case of collapsed instruments reduced in to $T-2$ for difference part of system GMM). On the other hand, Mehrhoff (2009); Mammi and Calzolari (2011), and Bontempi and Mammi (2012, 2014) proposed factorization of the full instrument sets as a valid transformation for ensuring consistency of GMM estimator.

3.3 The asymptotic variance of GMM estimators

The GMM estimator is consistent and asymptotically normal under fairly weak, albeit technical assumptions. The intuition in developing sample moment conditions $E\left[(g_n(w,\theta)\right]$ on defining GMM estimation procedure is that the sample averages should have zero mean when the population moment condition is true. In order for the estimates to be reasonable $g_n(w_i,\theta)$ need to behave well. Specifically, W_n must be positive definite and the system must be identified. Positive definiteness of W_n is required to ensure that $Q_n(\theta)$ can only be minimized at unique value of θ_0. Consider the simple dynamic panel model (2.2) and define:

$$Z_i^S = \begin{bmatrix} y_{i1} & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & y_{i1} & y_{i2} & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & y_{i1} & \cdots & y_{i,T-2} & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & \triangle y_{i2} & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \triangle y_{i3} & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & \triangle y_{i,T-1} \end{bmatrix},$$

$$u_i = u_i(\gamma_0) = \begin{bmatrix} \triangle y_{i3} - \gamma_0 \triangle y_{i2} \\ \vdots \\ \triangle y_{iT} - \gamma_0 \triangle y_{i,T-1} \\ y_{i3} - \gamma_0 y_{i2} \\ \vdots \\ y_{iT} - \gamma_0 y_{i,T-1} \end{bmatrix}$$

and $g_i(\gamma_0) = Z_i^s u_i$. The moment conditions in (2.19) and (2.20) imply that $E\left[g_i(\gamma_0)\right] = 0$. Generally, using the moment conditions, the GMM estimator $\hat{\gamma}$ for γ_0 minimizes:

$$\left[\tfrac{1}{n}\textstyle\sum_{i=1}^{n} g_i(\gamma)\right) W_n \left(\tfrac{1}{n}\textstyle\sum_{i=1}^{n} g_i(\gamma)\right]$$

with respect to γ, where W_n is a positive semi-definite weight matrix which satisfies $lim_{n\to\infty} W_n = W$, with W a positive definite matrix. Under certain regularity conditions in place, it can be shown that

$$\lim_{n\to\infty} \mathrm{n}^{-1} \textstyle\sum_{i=1}^{n} g_i(\gamma_0) \quad \xrightarrow{p} \quad E\left[g_i(\gamma_0)\right]$$

and

$$lim_{n\to\infty} \tfrac{1}{\sqrt{n}} \textstyle\sum_{i=1}^{n} g_i(\gamma_0) \quad \xrightarrow{d} \quad N(0, S),$$

where S is as defined in (8). Furthermore, let

$$G = \lim_{n\to\infty} n^{-1} \textstyle\sum_{i=1}^{n} \tfrac{dg_i(\gamma_0)}{d\gamma} \quad \xrightarrow{p} \quad E\left[\tfrac{dg_i(\gamma_0)}{d\gamma}\right],$$

,

By the law of large numbers on the moment conditions, $\sqrt{n}(\hat{\gamma} - \gamma_0)$ has a limiting normal distribution, that is,

$$\sqrt{n}(\hat{\gamma} - \gamma_0) \quad \xrightarrow{d} \quad N(0, V_W),$$

where

$$V_W = (G'WG)^{-1}G'WSWG(G'WG)^{-1}. \qquad (3.3)$$

From the expression of the asymptotic variance matrix V_W in (3.3), it is clear that the efficiency of the GMM estimator is affected by the choice of the weighted matrix W. As we mentioned in (2.8), an optimal choice is a weight matrix for which $W = S^{-1}$. The asymptotic variance covariance matrix is then given by $(G'S^{-}G)^{-1}$. For any other W, the GMM estimator is less efficient as

$$(G'WG)^{-1-1} \leq (G'WG)^{-1}G'WSWG(G'WG)^{-1}.$$

To assess the potential loss in efficiency from the initial weighting matrix, the following expression for the upper bound of the efficiency loss has been derived by Liu and Neudecker (1997) on the basis of the Kantorovich inequality (KI):

$$(G'WG)^{-1} \leq (G'WG)^{-1}G'WSWG(G'WG)^{-1} \leq \tfrac{(\lambda_1+\lambda_p)^2}{4\lambda_1\lambda_p}(G'S^{-1}G)^{-1}$$

and the efficiency upper bound is:

$$UB_{KI} = \tfrac{(\lambda_1+\lambda_p)^2}{4\lambda_1\lambda_p}$$

where, $\lambda_i > 0 (i = 1, 2, \cdots, p)$ are the eigenvalues of the $p \times p$ matrix SW. For the most efficient estimator, $UB_{KI} = 1$. If UB_{KI} increased led to a further loss of efficiency. In system GMM estimation procedures, the weight matrix is optimal when $\sigma_\alpha^2 = 0$. In practice, it is common to use the inverse of the moment matrix of the instruments as an initial weight matrix. This procedure leads to the loss of efficiency. Therefore, in the next sections we will present varying initial weight matrices in system GMM estimation and we use the efficiency upper bounds as the efficiency comparison of the alternative system GMM estimators.

3.4 Sub-optimal weight matrix

The choice of weighting matrix is a crucial factor for efficiency of a GMM estimator as it defines which moments are given more attention in the minimization of (2.7). For the difference GMM estimator, an optimal initial weighting matrix is given by: $W_n^D = \left[\sum_{i=1}^n Z_i^{D\prime} D Z_i^D\right]^{-1}$. Since estimation of an optimal weighting matrix heavily depends on asymptotic characteristics, however, the estimator may perform poorly when small sample sizes are used. In particular, an optimal (asymptotically) initial weighting matrix for system GMM estimator depends on individual specific effects, and consequently, asymptotically efficient one-step GMM estimator is not easily driven. Blundell and Bond (1998) use the inverse of the sample moment matrix of the instruments as an initial weighting matrix which is far from optimal (Kiviet, 2007). As a solution, Kieviet (2007) suggest the use of a sub-optimal weight matrix given by:

$$W_n^J = \left(n^{-1}\sum_{i=1}^n Z_i^{S\prime} H_J Z_i^S\right)^{-1} \tag{3.4}$$

where

$$H_J = \begin{pmatrix} D & 0 \\ 0 & J_{T-2} \end{pmatrix}, \text{ D is as defined in (2.10) and:}$$

$$J_{T-2} = \begin{bmatrix} 1+\rho & \rho & \rho & \cdots & \rho \\ \rho & 1+\rho & \rho & \cdots & \rho \\ \rho & \rho & 1+\rho & \cdots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \cdots & 1+\rho \end{bmatrix}, \rho = \frac{\sigma_\alpha^2}{\sigma_\epsilon^2}$$

Since the variance ratio (ρ) is unknown in practice, Jung and Kwon (2007) suggest an estimate of the optimal weight matrix:

$$\hat{H}_J = \begin{pmatrix} D & 0 \\ 0 & \hat{J}_{T-2} \end{pmatrix} \tag{3.5}$$

where ρ in (3.4) is replaced by the estimated variance ratio $\hat{\rho} = \hat{\sigma}_\alpha^2/\hat{\sigma}_\epsilon^2$ in (3.5). The estimated variances of individual specific effects and idiosyncratic error terms are given by:

$$\hat{\sigma}_\epsilon^2 = \frac{\sum_{i=1}^n \triangle\hat{u}_i' \triangle\hat{u}_i}{2n(T-2)} \text{ and } \hat{\sigma}_\alpha^2 = \frac{\sum_{i=1}^n [\tilde{u}_i'\tilde{u}_i - (\triangle\tilde{u}_i'\triangle\tilde{u}_i/2)]}{n(T-2)}$$

where $\triangle\hat{u}_i$ are the residuals from one-step conventional difference estimator, while $\tilde{u}_i$ and $\triangle\tilde{u}_i$ are residuals from the first difference and the level equation in one-step conventional system GMM estimator, respectively.

3.4.1 Derivation of alternative initial weighting matrices

In this study we propose GMM estimators based on sub-optimal weighting matrix setup in combination with reduced instruments set (partially collapsed and lagged limited). As we mentioned in (3.2.2), we do not apply Roodman (2009) approach of collapsing instruments set for the level part of moment conditions.

Consider the FD version of the simple auto regressive panel in (2.2):

$$\triangle y_{it} = \gamma \triangle y_{i,t-1} + \triangle \epsilon_{it} \tag{3.6}$$

Assuming that the series is covariance stationary, we have:

$$E(y_{it}\alpha_i) = E\left[(\gamma y_{i,t-1} + \alpha_i + \epsilon_{it})\alpha_i\right]$$

$$= \gamma E(\alpha_i y_{i,t-1}) + E(\alpha_i^2) + E(\alpha_i \epsilon_{it})$$

$$= \gamma \sigma_{y\alpha} + \sigma_\alpha^2 \Rightarrow E(y_{it}\alpha_i) = \gamma\sigma_{y\alpha} + \sigma_\alpha^2$$

$$\Rightarrow \sigma_{y\alpha} = \tfrac{\sigma_\alpha^2}{1-\gamma} \tag{3.7}$$

Similarly,

$$var(y_{it}) = E(y_{it}y_{it}') = \sigma_y^2 = \left[\gamma y_{i,t-1} + \alpha_i + \epsilon_{it}\right]\left[\gamma y_{i,t-1} + \alpha_i + \epsilon_{it}\right]$$

$$= \gamma^2\sigma_y^2 + 2\gamma\sigma_{y\alpha} + \sigma_\alpha^2 + \sigma_\epsilon^2 = \gamma^2\sigma_y^2 + 2\gamma\left(\tfrac{\sigma^2}{1-\gamma}\right) + \sigma_\alpha^2 + \sigma_\epsilon^2.$$

$$\Rightarrow (1-\gamma^2)\sigma_y^2 = \tfrac{2\gamma\sigma_\alpha^2 + (1-\gamma)\sigma_\alpha^2}{(1-\gamma)} + \sigma_\epsilon^2 = \tfrac{(1+\gamma)\sigma_\alpha^2}{(1-\gamma)} + \sigma_\epsilon^2$$

$$\Rightarrow \sigma_y^2 = \tfrac{\sigma_\alpha^2}{(1-\gamma)^2} + \tfrac{\sigma_\epsilon^2}{1-\gamma^2} \tag{3.8}$$

and,

$$E(y_{i,t-1}y_{it}) = \left[y_{i,t-1}(\gamma y_{i,t-1} + \alpha_i + \epsilon_{it})\right] = \gamma\sigma_y^2 + \sigma_{y\alpha}$$

$$= \gamma\left(\tfrac{\sigma_\alpha^2}{(1-\gamma)^2} + \tfrac{\sigma_\epsilon^2}{1-\gamma^2}\right) + \tfrac{\sigma_\alpha^2}{1-\gamma} = \left(\tfrac{\gamma\sigma_\epsilon^2}{1-\gamma^2}\right) + \tfrac{\gamma\sigma_\epsilon^2}{1-\gamma^2} + \tfrac{\sigma_\alpha^2}{1-\gamma}$$

$$= \left(\tfrac{\gamma\sigma_\epsilon^2}{1-\gamma^2}\right) + \tfrac{\gamma\sigma_\alpha^2 + \sigma_\alpha^2(1-\gamma)}{(1-\gamma)^2} = \sigma_y^2 - \tfrac{\sigma_\epsilon^2}{1+\gamma} \tag{3.9}$$

The GMM estimation procedures heavily depend on the asymptotic theory (formulae) that it may not perform well in finite samples (Doran and Schmidt, 2006). Specifically, in system GMM estimation, an initial weighting matrix which is constructed for asymptomatically optimal weight matrix depends on individual specific effects, so asymptotically efficient one-step system GMM estimator is not easily driven (Altonji and Segal, 1996; Windmeijer, 2000; Kiviet, 2007; Croissant and Giovanni Millo, 2019). Kiviet et al. (2017) also show that the small sample performance of the GMM estimator depends on the chosen degree of robustness of the adopted initial weighting matrix. In practice, we commonly use less efficient initial weight matrix which is required to converge in probability to a positive definite optimal weight matrix. In general, an appropriate weighting matrix is not readily available on system GMM estimation.

The three commonly used initial weight matrices in system GMM estimation are: the Blundell and Bond (1998) initial weight matrix which yields the simple generalized instrumental variable estimator, the conventional initial weight matrix for system GMM estimator (see for example, Doornik et al. 2002) and the sub-optimal weight matrix.

a) The Blundell and Bond (1998) initial weight matrix is given by

$$W_n^{BB} = \left[n^{-1}\sum_{i=1}^{n} Z_i^{S\prime} H_I^{BB} Z_i^S\right]^{-1}, \tag{3.10}$$

where $H_I^{BB} = \begin{pmatrix} I_{T-2} & 0 \\ 0 & I_{T-2} \end{pmatrix}$

b) The initial weight matrix for the conventional system GMM estimator is given as:

$$W_n^{C} = \left[n^{-1}\sum_{i=1}^{n} Z_i^{S\prime} H_I Z_i^S\right]^{-1}, \tag{3.11}$$

where $H_I = \begin{pmatrix} D & 0 \\ 0 & I_{T-2} \end{pmatrix}$, and D is as defined in (2.12)

c) The sub-optimal weight matrix is given by

$$W_n^{J} = \left(n^{-1}\sum_{i=1}^{n} Z_i^{S\prime} \hat{H}_J Z_i^S\right)^{-1}, \tag{3.12}$$

where $H_J = \begin{pmatrix} D & 0 \\ 0 & J_{T-2} \end{pmatrix}$,

$$J_{T-2} = \begin{bmatrix} 1+\rho & \rho & \rho & \cdots & \rho \\ \rho & 1+\rho & \rho & \cdots & \rho \\ \rho & \rho & 1+\rho & \cdots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \cdots & 1+\rho \end{bmatrix}, \rho = \frac{\sigma_\alpha^2}{\sigma_\epsilon^2}$$

As stated above, our study approach is on the basis of applying sub-optimal initial weight matrix together with reduced instruments set for system GMM estimation. This is not viable at $T = 3$ since we have only two instruments y_{i1} and $\triangle y_{i2}$ which correspond to difference and level parts of the equations, respectively. Therefore, we have to consider more suitable environment in terms of the number of instruments utilized in system GMM estimation. Here, we consider $T = 4$ for our derivations and assume that conditions (i) to (iii) in section (2.1) hold. Under the orthogonal conditions in (2.19) and (2.20), there are four over-identifying moment conditions:

$$E\left[y_{i1}(\triangle y_{i3} - \gamma \triangle y_{i2})\right] = 0, \quad E\left[y_{i2}(\triangle y_{i3} - \gamma \triangle y_{i2})\right] = 0,$$

$$E\left[\triangle y_{i2}(y_{i3} - \gamma y_{i2})\right] = 0, \quad E\left[\triangle y_{i3}(y_{i3} - \gamma y_{i2})\right] = 0,$$

We consider the following cases:

1. Using untransformed instruments set in combination with initial weight matrix setups in (3.10), (3.11) and (3.12) yields (5×5) initial weight matrices for Blundell and Bond (1998) estimator, conventional system GMM estimator and sub optimal system GMM estimator (denoted by W_n^{BB}, W_n^{C} and W_n^{Ju}, respectively).

2. Using collapsed instruments set in combination with initial weight matrix setup in (3.12) results in (4×4) sub-optimal initial weight matrix (denoted by W_n^{Jc}).

3. Using lag-limited instruments set in combination with initial weight matrix setup in (3.12) gives (4×4) sub-optimal initial weight matrix (denoted by W_n^{Jl}).

In order to compare the efficiency estimators using UB_{KI}, the product matrices (SW) are also computed.

1) Using untransformed (all available) instruments, the asymptotic covariance matrix for sample moments is given as:

$$S^u = \lim_{n\to\infty} \left(\tfrac{1}{n}\textstyle\sum_{i=1}^{n} Z_i^{S\prime} u_i u_i' Z_i^S\right)$$

$$= E\left[\begin{pmatrix} y_{i1} & 0 & 0 & 0 \\ 0 & y_{i1} & 0 & 0 \\ 0 & y_{i2} & 0 & 0 \\ 0 & 0 & \triangle y_{i2} & 0 \\ 0 & 0 & 0 & \triangle y_{i3} \end{pmatrix} u_i u_i' \begin{pmatrix} y_{i1} & 0 & 0 & 0 & 0 \\ 0 & y_{i1} & y_{i2} & 0 & 0 \\ 0 & 0 & 0 & \triangle y_{i2} & 0 \\ 0 & 0 & 0 & 0 & \triangle y_{i3} \end{pmatrix}\right]$$

$$= \sigma_\epsilon^2 \begin{bmatrix} 2\sigma_y^2 & -\sigma_y^2 & -\delta & -(1-\gamma)\sigma_y^2 & 0 \\ -\sigma_y^2 & 2\sigma_y^2 & 2\delta & (1-\gamma)\sigma_y^2 & -\frac{\gamma\sigma_\epsilon^2}{1+\gamma} \\ -\delta & 2\delta & 2\sigma_y^2 & -\frac{\sigma_\epsilon^2}{1+\gamma} & -\frac{\sigma_\epsilon^2}{1+\gamma} \\ -(1-\gamma)\sigma_y^2 & (1-\gamma)\sigma_y^2 & -\frac{\sigma_\epsilon^2}{1+\gamma} & \frac{2(1+\rho)\sigma_\epsilon^2}{1+\gamma} & \delta - \frac{\gamma\sigma_\epsilon^2}{1+\gamma} \\ 0 & -\frac{\gamma\sigma_\epsilon^2}{1+\gamma} & -\frac{\sigma_\epsilon^2}{1+\gamma} & \delta - \frac{\gamma\sigma_\epsilon^2}{1+\gamma} & \frac{2(1+\rho)\sigma_\epsilon^2}{1+\gamma} \end{bmatrix}$$

where $u_i = (\triangle\epsilon_{i3},\ \triangle\epsilon_{i4},\ \alpha_i + \epsilon_{i3},\ \alpha_i + \epsilon_{i4})'$ is the vector of error terms in system GMM estimation procedures.

a) The initial weighting matrix as in Blundell and Bond (1998) is computed as:

$$W_n^{BB} = \left(plim\tfrac{1}{n}\textstyle\sum_{i=1}^{n} Z_i^{S\prime} Z_i^S\right)^{-1}$$

$$= \left(E\left[\begin{pmatrix} y_{i1} & 0 & 0 & 0 \\ 0 & y_{i1} & 0 & 0 \\ 0 & y_{i2} & 0 & 0 \\ 0 & 0 & \triangle y_{i2} & 0 \\ 0 & 0 & 0 & \triangle y_{i3} \end{pmatrix} \begin{pmatrix} y_{i1} & 0 & 0 & 0 & 0 \\ 0 & y_{i1} & y_{i2} & 0 & 0 \\ 0 & 0 & 0 & \triangle y_{i2} & 0 \\ 0 & 0 & 0 & 0 & \triangle y_{i3} \end{pmatrix}\right]\right)^{-1}$$

$$= \left[\begin{pmatrix} \sigma_y^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_y^2 & \delta & 0 & 0 \\ 0 & \delta & \sigma_y^2 & 0 & 0 \\ 0 & 0 & 0 & \frac{2\sigma_\epsilon^2}{1+\gamma} & 0 \\ 0 & 0 & 0 & 0 & \frac{2\sigma_\epsilon^2}{1+\gamma} \end{pmatrix}\right]^{-1} = \begin{bmatrix} \frac{1}{\sigma_y^2} & 0 & 0 & 0 & 0 \\ 0 & \frac{-\sigma_y^2}{\delta^2-\sigma_y^4} & \frac{\delta}{\delta^2-\sigma_y^4} & 0 & 0 \\ 0 & \frac{\delta}{\delta^2-\sigma_y^4} & \frac{-\sigma_y^2}{\delta^2-\sigma_y^4} & 0 & 0 \\ 0 & 0 & 0 & \frac{1+\gamma}{2\sigma_\epsilon^2} & 0 \\ 0 & 0 & 0 & 0 & \frac{1+\gamma}{2\sigma_\epsilon^2} \end{bmatrix}$$

Consequently,

$$S^u W_n^{BB} = \sigma_\epsilon^2 \begin{bmatrix} 2\sigma_y^2 & -\sigma_y^2 & -\delta & -(1-\gamma)\sigma_y^2 & 0 \\ -\sigma_y^2 & 2\sigma_y^2 & 2\delta & (1-\gamma)\sigma_y^2 & -\frac{\gamma\sigma_\epsilon^2}{1+\gamma} \\ -\delta & 2\delta & 2\sigma_y^2 & -\frac{\sigma_\epsilon^2}{1+\gamma} & -\frac{\sigma_\epsilon^2}{1+\gamma} \\ -(1-\gamma)\sigma_y^2 & (1-\gamma)\sigma_y^2 & -\frac{\sigma_\epsilon^2}{1+\gamma} & \frac{2(1+\rho)\sigma_\epsilon^2}{1+\gamma} & \delta - \frac{\gamma\sigma_\epsilon^2}{1+\gamma} \\ 0 & -\frac{\gamma\sigma_\epsilon^2}{1+\gamma} & -\frac{\sigma_\epsilon^2}{1+\gamma} & \delta - \frac{\gamma\sigma_\epsilon^2}{1+\gamma} & \frac{2(1+\rho)\sigma_\epsilon^2}{1+\gamma} \end{bmatrix} \times$$

$$\begin{bmatrix} \frac{1}{\sigma_y^2} & 0 & 0 & 0 & 0 \\ 0 & \frac{-\sigma_y^2}{\delta^2-\sigma_y^4} & \frac{\delta}{\delta^2-\sigma_y^4} & 0 & 0 \\ 0 & \frac{\delta}{\delta^2-\sigma_y^4} & \frac{-\sigma_y^2}{\delta^2-\sigma_y^4} & 0 & 0 \\ 0 & 0 & 0 & \frac{1+\gamma}{2\sigma_\epsilon^2} & 0 \\ 0 & 0 & 0 & 0 & \frac{1+\gamma}{2\sigma_\epsilon^2} \end{bmatrix}$$

$$= \sigma_\epsilon^2 \begin{bmatrix} 2 & -1 & 0 & -\frac{(1-\gamma)\sigma_y^2}{3\sigma_\epsilon^2} & 0 \\ -1 & 2 & 0 & \frac{(1-\gamma)\sigma_y^2}{3\sigma_\epsilon^2} & -\frac{\gamma}{2} \\ \frac{-\delta}{\sigma_y^2} & 0 & 2 & -\frac{1}{2} & -\frac{1}{2} \\ -(1-\gamma) & \frac{\sigma_y^2\left[\sigma_y^2(\gamma^2-1)-\delta\right]}{(1+\gamma)(\delta^2-\sigma_y^4)} & \frac{\sigma_y^2\left[(1-\gamma^2)\delta+\sigma_\epsilon^2\right]}{2(1+\gamma)(\delta^2-\sigma_y^4)} & (\rho+1) & \frac{\rho\sigma_\epsilon^2(2\gamma-\gamma^2-1)}{(1-\gamma^2)} \\ 0 & \frac{\sigma_\epsilon^2\left[\gamma\sigma_y^2-\delta\right]}{2(1+\gamma)(\delta^2-\sigma_y^4)} & \frac{\sigma_\epsilon^2\left[\sigma_y^2-\gamma\delta\right]}{2(1+\gamma)(\delta^2-\sigma_y^4)} & \frac{\rho\sigma_\epsilon^2(\gamma-\gamma^2-1)}{(1-\gamma^2)} & (\rho+1) \end{bmatrix}$$

b) The initial weight matrix for conventional system GMM estimation is given by

$$W_n^C = \left(plim \tfrac{1}{n} \textstyle\sum_{i=1}^n Z_i^{S\prime} H_I Z_i^S\right)^{-1}$$

$$= \left[\begin{pmatrix} 2\sigma_y^2 & \sigma_y^2 & -\delta & 0 & 0 \\ -\sigma_y^2 & 2\sigma_y^2 & 2\delta & 0 & 0 \\ -\delta & 2\delta & 2\sigma_y^2 & 0 & 0 \\ 0 & 0 & 0 & \frac{2\sigma_\epsilon^2}{1+\gamma} & 0 \\ 0 & 0 & 0 & 0 & \frac{2\sigma_\epsilon^2}{1+\gamma} \end{pmatrix}\right]^{-1}$$

$$= \begin{bmatrix} \frac{2}{3\sigma_y^2} & \frac{1}{3\sigma_y^2} & 0 & 0 & 0 \\ \frac{1}{3\sigma_y^2} & \frac{2\sigma_y^4-\delta^2}{2\sigma_y^2(\sigma_y^4-\delta^2)} & \frac{-\delta}{4(\sigma_y^4-\delta^2)} & 0 & 0 \\ 0 & \frac{-\delta}{\sigma_y^4-\delta^2} & \frac{\sigma_y^2}{2(\sigma_y^4-\delta^2)} & 0 & 0 \\ 0 & 0 & 0 & \frac{1+\gamma}{2\sigma_\epsilon^2} & 0 \\ 0 & 0 & 0 & 0 & \frac{1+\gamma}{2\sigma_\epsilon^2} \end{bmatrix}$$

then,

$$S^u W_n^C = \sigma_\epsilon^2 \begin{bmatrix} 1 & 0 & \frac{\sigma_y^2(\delta-1-\gamma)\sigma_y^2)}{\sigma_y^4-\delta^2} & -\frac{(1-\gamma^2)\sigma_y^2}{3\sigma_\epsilon^2} & 0 \\ 0 & \frac{2(\sigma_y^4-\delta^2)}{2\sigma_y^4-\delta^2} & \frac{\delta\sigma_y^2}{(\sigma_y^4-\delta^2)} & \frac{(1-\gamma^2)\sigma_y^2}{3\sigma_\epsilon^2} & -\frac{\gamma}{2} \\ 1 & 0 & \frac{2\sigma_y^4-\delta^2}{(\sigma_y^4-\delta^2)} & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{(1-\gamma)}{3} & \frac{(1-\gamma^2)(2\sigma_y^4+\delta^2)+3\delta^2)}{3(1+\gamma)(2\sigma_y^4-\delta^2)} & -\frac{\sigma_y^2\left[2(1-\gamma^2)\delta+3\sigma_\epsilon^2\right]}{6(1+\gamma)(2\sigma_y^4-\delta^2)} & (\rho+1) & \frac{\rho(2\gamma-\gamma^2-1)}{2(1-\gamma)} \\ -\frac{\gamma\sigma_\epsilon^2}{3(1+\gamma)\sigma_y^2} & \frac{\sigma_\epsilon^2\left[3\delta\sigma_y^2-\gamma(4\sigma_y^4-\delta^2)\right]}{3(1+\gamma)(2\sigma_y^4-\delta^2)} & \frac{\sigma_\epsilon^2\left[\gamma\delta-3\sigma_y^2\right]}{6(1+\gamma)(2\sigma_y^4-\delta^2)} & \frac{\rho(2\gamma-\gamma^2-1)}{2(1-\gamma)} & (\rho+1) \end{bmatrix}$$

c) The sub-optimal weight matrix which is defined in (3.12) is given by

$$W_n^{Ju} = \left(plim\frac{1}{n}\sum_{i=1}^n Z_i^{S\prime} H_J Z_i^S\right)^{-1}$$

$$= \left[\begin{pmatrix} 2\sigma_y^2 & \sigma_y^2 & -\delta & 0 & 0 \\ -\sigma_y^2 & 2\sigma_y^2 & 2\delta & 0 & 0 \\ -\delta & 2\delta & 2\sigma_y^2 & 0 & 0 \\ 0 & 0 & 0 & 2(1-\gamma)(\rho+1)\sigma_y^2 & 2\rho\gamma(1-\gamma)\sigma_y^2 \\ 0 & 0 & 0 & 2\rho\gamma(1-\gamma)\sigma_y^2 & 2(1-\gamma)(\rho+1)\sigma_y^2 \end{pmatrix}\right]^{-1}$$

$$= \begin{bmatrix} \frac{2}{3\sigma_y^2} & \frac{1}{3\sigma_y^2} & 0 & 0 & 0 \\ \frac{1}{3\sigma_y^2} & \frac{2\sigma_y^4-\delta^2}{2\sigma_y^2(\sigma_y^4-\delta^2)} & \frac{-\delta}{4(\sigma_y^4-\delta^2)} & 0 & 0 \\ 0 & \frac{-\delta}{\sigma_y^4-\delta^2} & \frac{\sigma_y^2}{2(\sigma_y^4-\delta^2)} & 0 & 0 \\ 0 & 0 & 0 & \frac{1+\rho}{2(1-\gamma)[(1+\rho)^2-(\gamma\rho)^2]\sigma_y^2} & \frac{-\gamma\rho}{2(1-\gamma)[(1+\rho)^2-(\gamma\rho)^2]\sigma_y^2} \\ 0 & 0 & 0 & \frac{-\gamma\rho}{2(1-\gamma)[(1+\rho)^2-(\gamma\rho)^2]\sigma_y^2} & \frac{1+\rho}{2(1-\gamma)[(1+\rho)^2-(\gamma\rho)^2]\sigma_y^2} \end{bmatrix}$$

and,

$$S^u W_n^{Ju} = \sigma_\epsilon^2 \begin{bmatrix} 1 & 0 & \frac{\sigma_y^2(\delta-1-\gamma)\sigma_y^2)}{\sigma_y^4-\delta^2} & -\frac{2(1-\gamma)\sigma_y^2}{C} & \frac{AB\sigma_y^2}{C} \\ 0 & \frac{2(\sigma_y^4-\delta^2)}{2\sigma_y^4-\delta^2} & \frac{\delta\sigma_y^2}{(\sigma_y^4-\delta^2)} & \frac{2(1-\gamma)+\gamma B}{C} & -\frac{(\rho+1)\left[(1-\gamma)B+\gamma\sigma_\epsilon^2\right]}{C} \\ 1 & 0 & \frac{2\sigma_y^4-\delta^2}{(\sigma_y^4-\delta^2)} & -\frac{(B-2A)\sigma_\epsilon^2}{(1-\gamma)^2(\rho+1)C} & \frac{(\rho+1)(B-1)\sigma_\epsilon^2}{C} \\ -\frac{(1-\gamma)}{3} & \frac{(1-\gamma^2)(2\sigma_y^4+\delta^2)+3\delta^2)}{3(1+\gamma)(2\sigma_y^4-\delta^2)} & -\frac{\sigma_y^2\left[2(1-\gamma^2)\delta+3\sigma_\epsilon^2\right]}{6(1+\gamma)(2\sigma_y^4-\delta^2)} & \frac{(4A^2-B)\sigma_\epsilon^2}{(1+\gamma)^2AC} & \frac{B(\rho+1)[1-2(\rho+1)]\sigma_\epsilon^2}{(1-\gamma^2)C} \\ -\frac{\gamma\sigma_\epsilon^2}{3(1+\gamma)\sigma_y^2} & \frac{\sigma_\epsilon^2\left[3\delta\sigma_y^2-\gamma(4\sigma_y^4-\delta^2)\right]}{3(1+\gamma)(2\sigma_y^4-\delta^2)} & \frac{\sigma_\epsilon^2\left[\gamma\delta-3\sigma_y^2\right]}{6(1+\gamma)(2\sigma_y^4-\delta^2)} & \frac{2B(\rho+1)\left[(1-\rho^2)-A\right]\sigma_\epsilon^2}{(1+\gamma)^2AC} & \frac{(\rho+1)\left[2A-B^2\right]\sigma_\epsilon^2}{(1-\gamma^2)C} \end{bmatrix}$$

where $A = (1-\gamma)(\rho+1)$, $B = \rho(2\gamma-\gamma^2-1)$, and $C = (2A-B^2)(1-\gamma)\sigma_\epsilon^2$

2) Using collapsed instruments, the asymptotic covariance matrix for sample moments is computed as:

$$S^c = \left(plim \tfrac{1}{n}\textstyle\sum_{i=1}^n (Z_i^S)^{C\prime} u_i u_i' (Z_i^S)^C\right)$$

$$= E\left[\begin{pmatrix} y_{i1} & y_{i2} & 0 & 0 \\ 0 & y_{i1} & 0 & 0 \\ 0 & 0 & \triangle y_{i2} & 0 \\ 0 & 0 & 0 & \triangle y_{i3} \end{pmatrix} u_i u_i' \begin{pmatrix} y_{i1} & 0 & 0 & 0 \\ y_{i1} & y_{i2} & 0 & 0 \\ 0 & 0 & \triangle y_{i2} & 0 \\ 0 & 0 & 0 & \triangle y_{i3} \end{pmatrix}\right]$$

$$= \sigma_\epsilon^2 \begin{bmatrix} 4\sigma_y^2 - 2\delta & 2\delta - \sigma_y^2 & -\frac{2\sigma_\epsilon^2}{(1+\gamma)} & -\frac{\sigma_\epsilon^2}{(1+\gamma)} \\ 2\delta - \sigma_y^2 & 2\sigma_y^2 & -(1-\gamma)\sigma_y^2 & -\frac{\gamma\sigma_\epsilon^2}{(1+\gamma)} \\ -\frac{2\sigma_\epsilon^2}{(1+\gamma)} & -(1-\gamma)\sigma_y^2 & \frac{2(\rho+1)\sigma_\epsilon^2}{(1+\gamma)} & \frac{\rho(2\gamma-\gamma^2-1)\sigma_\epsilon^2}{(1-\gamma^2)} \\ -\frac{\sigma_\epsilon^2}{(1+\gamma)} & -\frac{\gamma\sigma_\epsilon^2}{(1+\gamma)} & \frac{\rho(2\gamma-\gamma^2-1)\sigma_\epsilon^2}{(1-\gamma^2)} & \frac{2(\rho+1)\sigma_\epsilon^2}{(1+\gamma)} \end{bmatrix}$$

The corresponding sub-optimal initial weighting matrix is computed as:

$$W_n^{Jc} = \left(plim \tfrac{1}{n}\textstyle\sum_{i=1}^n (Z_i^S)^{C\prime} H_J (Z_i^S)^C\right)^{-1}$$

$$= \begin{bmatrix} \frac{2\sigma_y^2}{7\sigma_y^4-4\delta^2} & \frac{\sigma_y^2-2\delta}{7\sigma_y^4-4\delta^2} & 0 & 0 \\ \frac{\sigma_y^2-2\delta}{7\sigma_y^4-4\delta^2} & \frac{4\sigma_y^4-2\delta}{7\sigma_y^4-4\delta^2} & 0 & 0 \\ 0 & 0 & \frac{2(1-\gamma)}{(1+\gamma)C} & -\frac{(\rho+1)B}{C} \\ 0 & 0 & -\frac{B}{(1+\gamma)(\rho+1)C} & \frac{(\rho+1)}{C} \end{bmatrix}$$

and,

$$S^c W_n^{Jc} = \sigma_\epsilon^2 \begin{bmatrix} 1 & 0 & \frac{(B-4A)\sigma_\epsilon^2}{(1+\gamma)^2)(\rho+1)C} & \frac{(\rho+1)(2B-1)\sigma_\epsilon^2}{(1+\gamma)C} \\ 0 & 1 & \frac{\left[\gamma B\sigma_\epsilon^2-(1+\gamma)^2)A\sigma_y^2\right]}{(1+\gamma)^2)(\rho+1)C} & \frac{(\rho+1)\left[(1-\gamma)^2)B\rho+1)\sigma_y^2-\gamma\sigma_\epsilon^2\right]}{2(\rho+1)C} \\ -\frac{\sigma_y^2\left[2\sigma_\epsilon^2+(1-\gamma^2)(\sigma_y^2+-2\delta)\right]}{(1+\gamma)(7\sigma_y^4-4\delta^2)} & \frac{4\delta\sigma_\epsilon^2-\sigma_y^2\left[2\sigma_\epsilon^2+(1-\gamma^2)(4\sigma_y^4-2\delta)\right]}{2(1+\gamma)(7\sigma_y^4-4\delta^2)} & \frac{\left[A^2-B^2\right]\sigma_\epsilon^2}{(1+\gamma)^2)AC} & \frac{(\rho+1)B[(1+\gamma)-2(\rho+1)]\sigma_\epsilon^2}{(1+\gamma)C} \\ \frac{\left[2\gamma\delta-\sigma_y^2(1+\gamma)\right]\sigma_\epsilon^2}{(1+\gamma)(7\sigma_y^4-4\delta^2)} & \frac{\left[2\delta(1+2\gamma)-\sigma_y^2(1+4\gamma)\right]\sigma_\epsilon^2}{2(1+\gamma)(7\sigma_y^4-4\delta^2)} & 0 & \frac{(\rho+1)\left[2(1+\gamma)A-B^2\right]\sigma_\epsilon^2}{(1+\gamma)(1-\gamma^2)C} \end{bmatrix}$$

3) Using lag-limited instruments set ($\tau = 1$), the asymptotic covariance matrix for sample moments is computed as:

$$S^L = plim\left(\tfrac{1}{n}\textstyle\sum_{i=1}^n (Z_i^S)^{L\prime} u_i u_i' (Z_i^S)^L\right)$$

$$= E\left[\begin{pmatrix} y_{i1} & 0 & 0 & 0 \\ 0 & y_{i2} & 0 & 0 \\ 0 & 0 & \triangle y_{i2} & 0 \\ 0 & 0 & 0 & \triangle y_{i3} \end{pmatrix} u_i u_i' \begin{pmatrix} y_{i1} & 0 & 0 & 0 \\ 0 & y_{i2} & 0 & 0 \\ 0 & 0 & \triangle y_{i2} & 0 \\ 0 & 0 & 0 & \triangle y_{i3} \end{pmatrix}\right]$$

$$= \sigma_\epsilon^2 \begin{bmatrix} 2\sigma_y^2 & -\delta & -(1-\gamma)\sigma_y^2 & 0 \\ -\delta & 2\sigma_y^2 & -\frac{\sigma_\epsilon^2}{(1+\gamma)} & -\frac{\sigma_\epsilon^2}{(1+\gamma)} \\ -(1-\gamma)\sigma_y^2 & -\frac{\sigma_\epsilon^2}{(1+\gamma)} & \frac{2(\rho+1)\sigma_\epsilon^2}{(1+\gamma)} & \frac{\rho(2\gamma-\gamma^2-1)\sigma_\epsilon^2}{(1-\gamma^2)} \\ 0 & -\frac{\sigma_\epsilon^2}{(1+\gamma)} & \frac{\rho(2\gamma-\gamma^2-1)\sigma_\epsilon^2}{(1-\gamma^2)} & \frac{2(\rho+1)\sigma_\epsilon^2}{(1+\gamma)} \end{bmatrix}$$

The corresponding sub optimal initial weighting matrix is computed as:

$$\mathrm{W}_n^{Jl} = \left(plim\tfrac{1}{n}\textstyle\sum_{i=1}^n (Z_i^S)^{L\prime} H_J (Z_i^S)^L\right)^{-1}$$

$$= \begin{bmatrix} \frac{2\sigma_y^2}{7\sigma_y^4-4\delta^2} & \frac{\sigma_y^2-2\delta}{7\sigma_y^4-4\delta^2} & 0 & 0 \\ \frac{\sigma_y^2-2\delta}{7\sigma_y^4-4\delta^2} & \frac{4\sigma_y^4-2\delta}{7\sigma_y^4-4\delta^2} & 0 & 0 \\ 0 & 0 & \frac{1+\rho}{2(1-\gamma)[(1+\rho)^2-(\gamma\rho)^2]\sigma_y^2} & \frac{-\gamma\rho}{2(1-\gamma)[(1+\rho)^2-(\gamma\rho)^2]\sigma_y^2} \\ 0 & 0 & \frac{-\gamma\rho}{2(1-\gamma)[(1+\rho)^2-(\gamma\rho)^2]\sigma_y^2} & \frac{1+\rho}{2(1-\gamma)[(1+\rho)^2-(\gamma\rho)^2]\sigma_y^2} \end{bmatrix}$$

Finally,

$$S^l W_n^{Jl} = \sigma_\epsilon^2 \begin{bmatrix} 1 & 0 & -\frac{2(1-\gamma)\sigma_y^2}{(1+\gamma)C} & -\frac{AB\sigma_y^2}{C} \\ 0 & 1 & \frac{(B-2A)\sigma_\epsilon^2}{(1+\gamma^2)(\rho+1)C} & \frac{(\rho+1)(B-1)\sigma_\epsilon^2}{(1+\gamma)C} \\ -\frac{\left[2(1-\gamma^2)\sigma_y^4+\delta\sigma_\epsilon^2\right]}{(1+\gamma)(4\sigma_y^4-\delta^2)} & -\frac{2\sigma_y^2\left[(1-\gamma^2)\delta+2\sigma_\epsilon^2\right]}{(1+\gamma)(4\sigma_y^4-\delta^2)} & \frac{(2A^2-B^2)\sigma_\epsilon^2}{(1+\gamma)^2 AC} & \frac{(\rho+1)(B-A)\sigma_\epsilon^2}{(1-\gamma)^2 C} \\ -\frac{2\delta\sigma_\epsilon^2}{(1+\gamma)(4\sigma_y^4-\delta^2)} & -\frac{4\sigma_y^2\sigma_\epsilon^2}{(1+\gamma)(4\sigma_y^4-\delta^2)} & 0 & \frac{(\rho+1)(2A-B^2)}{(1-\gamma)^2 C} \end{bmatrix}$$

3.4.2 The proposed alternative system GMM estimators

The proposed alternative system GMM estimators are given as follows:

1. One step and two step sub-optimal system GMM estimators which utilize the sub-optimal weighting matrix and collapsed instruments set. The one step system GMM estimator is given by:

$$\hat{\gamma}^{Sc1} = \left(\tilde{y}'_{-1}(Z^S)^C W^{Jc}_{n(1)}(Z^S)^{C\prime}\tilde{y}_{-1}\right)^{-1}\left(\tilde{y}'_{-1}(Z^S)^C W^{Jc}_{n(1)}(Z^S)^{C\prime}\tilde{y}\right) \quad (3.13)$$

   where $(Z^S)^C = \begin{bmatrix} (Z^D)^C & 0 \\ 0 & Z^L \end{bmatrix}$, $(Z^D)^C$ is as defined in (3.2),

   Z^L is the instruments set in (2.21) and $W^{Jc}_{n(1)}$ is the initial sub optimal weight matrix based on collapsed instruments.

2. One step and two step sub-optimal system GMM estimators which employ sub optimal weighting matrix and using lag-limited instruments set. The one step estimator is given by

$$\hat{\gamma}^{Sl1} = \left(\tilde{y}'_{-1}(Z^S)^L W^{Jl}_{n(1)}(Z^S)^{L\prime}\tilde{y}_{-1}\right)^{-1}\left(\tilde{y}'_{-1}(Z^S)^L W^{Jl}_{n(1)}(Z^S)^{L\prime}\tilde{y}\right) \quad (3.14)$$

   where $(Z^S)^L = \begin{bmatrix} (Z^D)^L & 0 \\ 0 & Z^L \end{bmatrix}$, $(Z^D)^L$ is as defined in (3.1) and

   $W^{Jl}_{n(1)}$ is the sub optimal initial weight matrix using lag limited instruments.

The two step estimation is implemented following the usual procedures in system GMM estimations as in (2.13).

3.5 Efficiency comparison of alternative system GMM estimators

Figures 3.1 up to 3.4 display the efficiency upper bounds of different system GMM estimators constructed based on varying initial weighting matrices. Specifically, we considered the efficiency upper bounds based on: an initial weight matrix as in Blundell and Bond (1998) (SYS_BB); an initial weight matrix for conventional system GMM estimator (SYS_Conv); and sub-optimal initial weight matrix utilizing all available instruments (Sub_full), collapsed instruments (Sub_coll) and lag-limited instruments (Sub_lag). The graphs were constructed by fixing the values of γ at $0, 0.4, 0.9$ and values of ρ at $0.5, 1, 5$ and 10.

We can see from the figures that system GMM estimators which utilize sub-optimal weight matrices have smaller efficiency bounds regardless of the type of instruments adopted. In other words, applying initial sub-optimal weighting matrix on system GMM estimation is advantageous regardless of whether we use untransformed, partially collapsed or lag-limited instruments. Specifically, the system GMM estimator based on sub-optimal weight matrix and using untransformed (or all available) instruments has the smallest efficiency bounds in all cases, followed by the estimators employing sub-optimal weight matrix in combination with reduced instruments set. In contrast, the estimator employing an initial weight matrix as in Blundell and Bond is the least efficient for all values of ρ and/or γ. The efficiency upper bound based on an initial weight matrix for conventional system GMM estimation procedure is in between the two.

In general, the Kantorovich inequality efficiency bounds show that the efficiency loss is large when we adopt the conventional and Blundell and Bond (1998) initial weight matrices compared to those based on initial sub-optimal weight matrices. This is more reflected when the value of γ is small and/or the value of ρ gets large. On the other hand, the advantage of sub-optimal weighting matrix decreases when γ is close to one in relative terms.

Efficiency comparisons of system GMM estimators using efficiency bounds, by fixing values of γ at 0, 0.4, 0.9 & ρ at 0.5, 1, 5 & 10, respectively for $T = 4$

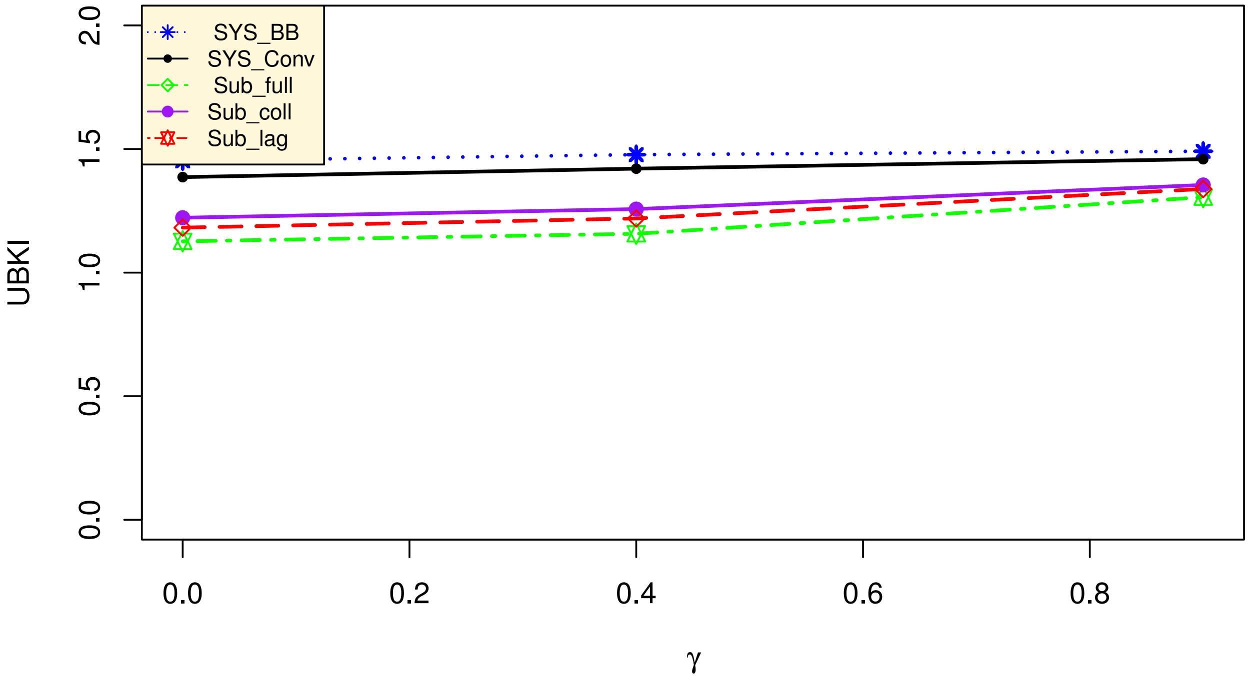

Figure 3.1: UBKI for $T = 4$, and $\rho = 0.5$

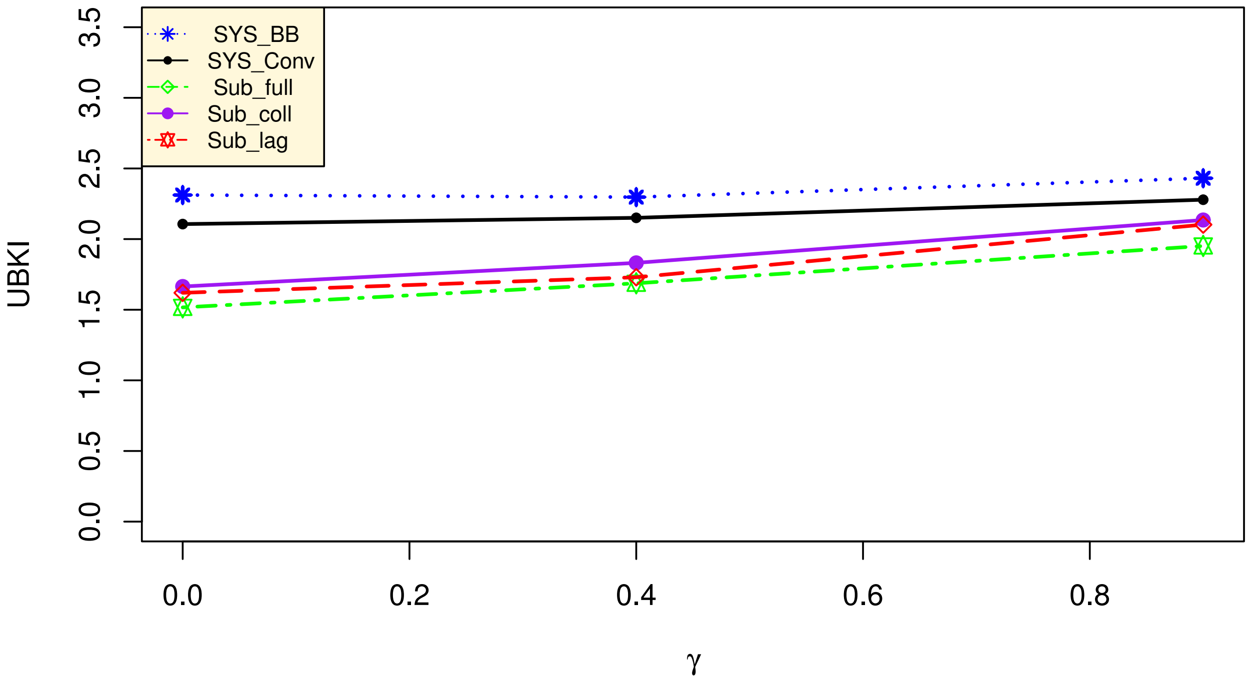

Figure 3.2: UBKI for $T = 4$, and $\rho = 1$

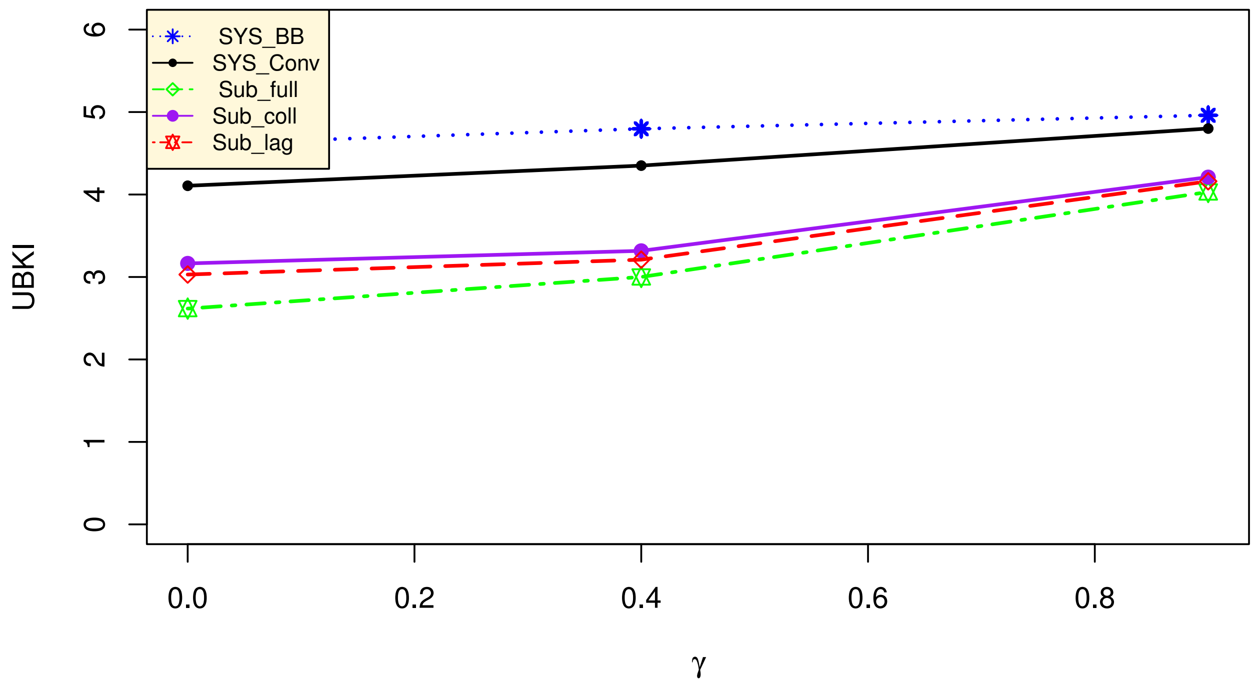

Figure 3.3: UBKI for $T = 4$, and $\rho = 5$

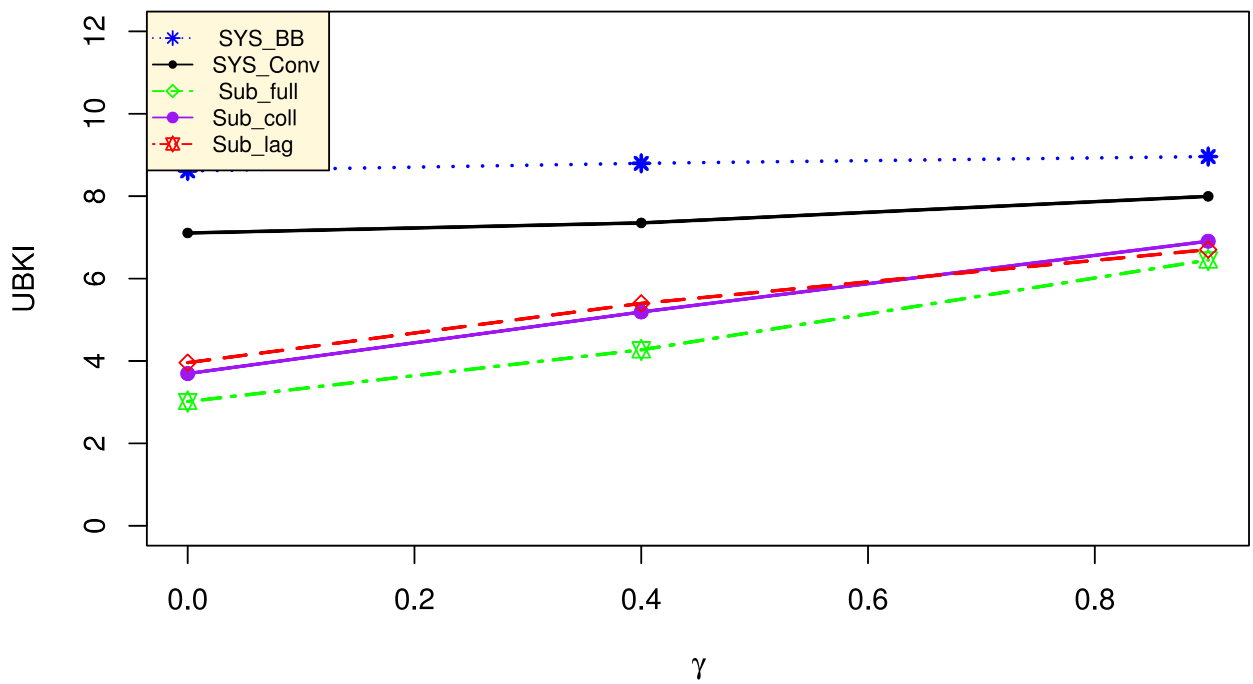

Figure 3.4: UBKI for $T = 4$, and $\rho = 10$

Chapter 4

Simulation Design and the results of experiments

This chapter has four sections. The first section presents the simulation design. Section 2 presents comparison of one-step system GMM estimators. Comparison of two-step system GMM estimators are presented in section 3. Section 4 is devoted to empirical illustration.

4.1 Simulation Design

In this section, we assess the performance of system GMM estimators which utilize various initial weight matrices in combination with different instruments set (untransformed, partially collapsed and lagged limited). A number of studies (Blundell and Bond, 1998; Bun and Kiviet, 2006; Youssef and Abonazel, 2017) have shown that the performance of GMM estimators improves as the cross-sectional dimension (n) increases. In this dissertation, we chose to compare the performance of various GMM estimators by varying the time dimension for fixed cross-sectional dimension since it is also well documented in the literature that GMM estimation problems arise in connection with large time dimension (T) of panel data models relative to the cross-sectional dimension (Windmeijer, 2005; Roodman, 2009; Fajeau, 2021). Monte Carlo experiments are carried out using R4.0.5 software based on the following data generating process for model in (2.2).

1. The individual effects α_i, were generated as independent normally distributed random variables with mean zero and variance σ^2_α and is independent across i.
2. The disturbances (ϵ_{it}) were generated as independent normally distributed random variables with mean zero and variance σ^2_ϵ, and inde-

pendent across i and t.

3. We generate the initial conditions (y_{i1}) as: $y_{i1} = \frac{\alpha_i}{1-\gamma} + \omega_{i1}, i = 1, \cdots, n$, where $\omega_{i1} = \sum_{j=0}^{\infty} \gamma^j \omega_{i,1-j}$ is independent of both α_i and ϵ_{it} and is covariance stationary.

4. The values of the variance ratio $\rho = \frac{\sigma_\alpha^2}{\sigma_\epsilon^2}$ considered are $0.5, 1, 2, 5$ and 10.

5. n is at 100 & 300; T is $4, 5, 10$ and 20; and $\gamma = \{0.2, 0.4, 0.9\}$.

6. For all experiments, we ran 5000 replications and the results of all separate experiments are obtained by precisely the same series of random numbers.

In order to compare the small sample performance of system GMM estimators, different estimation procedures are considered based on the initial weighting matrices that are utilized. These are system GMM estimator (SYS_BB) as in Blundell and Bond (1998), the conventional system GMM estimator (SYS_Conv), sub-optimal system GMM estimator (Sub_full) utilizing untransformed (all available) instruments, sub-optimal system GMM estimator (Sub_col) employing collapsed instruments, and sub-optimal system GMM estimator (Sub_lag) based on lag-limited instruments. The performance of estimators was compared in terms of mean absolute bias (MAB), standard deviation (std.dev.), root mean square error (RMSE) and coverage probability. The MAB and RMSE for a Monte Carlo experiment are computed as:

$$\text{MAB} = \tfrac{1}{5000} \textstyle\sum_{i=1}^{50000} |\hat{\gamma}_i - \gamma|, \tag{4.1}$$

$$\text{RMSE} = \sqrt{\tfrac{1}{5000} \textstyle\sum_{i=1}^{50000} (\hat{\gamma}_i - \gamma)^2}, \tag{4.2}$$

where γ is the true value for lagged dependent coefficient in (2), and $\hat{\gamma}$ is the estimated value for γ. Moreover, in order to assess the uncertainty associated with the point estimates, empirical coverage probabilities at 95% nominal levels are also calculated and it is given by

$$\text{coverage}(\gamma) = p\left[\hat{\gamma} - Z_{\alpha/2} \times s.e(\hat{\gamma}) < \gamma < \hat{\gamma} + Z_{\alpha/2} \times s.e(\hat{\gamma}) : \gamma\right],$$

where $\hat{\gamma} = \frac{1}{5000} \sum_{i=1}^{5000} \hat{\gamma}_i$, $s.e(\hat{\gamma})$ is the standard deviation of $\hat{\gamma}$ from 5000 bootstraps replicates, since standard deviation of bootstraps can serve as the standard error of coefficients (see, Fox & Weisberg, 2008). $Z_{\alpha/2}$ is the upper $\alpha/2$ part of the standard normal distribution.

4.2 Comparison of one-step system GMM estimators

Figures 4.1 to 4.3 pertain to the MAB for one stage system GMM estimators (based on the results given in Tables 5.1 and 5.2, APPENDIX I). These are the one-step system GMM estimator based on an initial weight matrix as in Blundell-Bond (1998) (SYS_BB); the conventional one-step system GMM estimator (SYS_Conv); and the system GMM estimators based on sub-optimal initial weight matrix setup in (3.12) in combination with untransformed, collapsed and lag limited instruments (denoted by Sub_full, Sub_col, and Sub_lag, respectively) at $T = 4, 10$; $n = 100$ & $\gamma = \{0.4, 0.9\}$.

For $T = 4$ and $\gamma = 0.4$ (Figure 4.1), there is no marked difference in the MAB among all estimators for small values of ρ ($\rho = 0.5$, 1). For moderate and large values of ρ, however, sub-optimally weighted GMM estimator based on collapsed instruments is the least biased. In contrast, the conventional one-step system GMM estimator and that based on an initial weight matrix as in Blundell and Bond (1998) are the most biased. Moreover, the MAB of all estimators keeps on increasing as the variance ratio increases. For $T = 4$, and $\gamma = 0.9$ (Figure 4.2) there is no marked difference in terms of bias across the board regardless of the values of the variance ratios. For $T = 10$, and $\gamma = 0.4$ (Figure 4.3), sub-optimally weighted GMM estimator that employs collapsed instruments is the least biased, and the difference against competing estimators is more pronounced for moderate and large values of ρ ($\rho = 5$, 10). The estimator based on sub optimal initial weights matrix and untransformed instruments is also doing good for $\rho = 5$. The reverse is true for the conventional one-step system GMM estimator and that based on an initial weight matrix as in Blundell and Bond (1998) for all values of the variance ratios.

Figure 4.4 (based on 5.2, APPENDIX I) presents the standard deviation of one stage system GMM estimators at For $T = 10$, and $\gamma = 0.4$. We can observe that, for small value of the variance ratio ($\rho = 0.5$), there is no marked difference in the performance of system GMM estimators (same for the sub-optimally weighted estimator based on lag limited instruments). For moderate and large values of ρ ($\rho = 5$ and $\rho = 10$), however, the sub-optimally weighted GMM estimator using all available (or untransformed) instruments is more efficient, followed by those with reduced instruments set (collapsed and lag limited). In contrast, the conventional one-step system GMM estimator and that based on an initial weight matrix as in Blundell and Bond (1998) have relatively large standard deviations, especially when the value of ρ gets large. For $T = 10$, $\gamma = 0.9$, there is no notable difference across all estimators in terms of precision. We also observe that the dispersion of all estimators decreases as the values of ρ gets larger. This might be explained by the underestimation of the true variance of the parameters as the value of γ approaches to one (Blundell and Bond, 1998; Doran and Schmidt, 2006; Bun and Windmeijer, 2010).

In general terms, the comparative analysis of estimators using mean absolute bias and standard deviation shows that sub-optimally weighted GMM estimators utilizing collapsed instruments and untransformed instruments fared well, in particular for moderate and large values of the variance ratio (ρ). For $T = 10$, $\gamma = 0.4$, for instance, the former estimator registered 8.4% and 23.5% reduction in MAB over the latter for $\rho = 5$ and $\rho = 10$, respectively. On the other hand, the sub-optimally weighted GMM estimator utilizing all available instruments exhibited 4.2% and 6.3% reduction in standard deviation over the estimator utilizing collapsed instruments for the same pair of variance ratios. Thus, we can conclude that, in relative terms, the substantial reduction in MAB of sub-optimally weighted GMM estimator based on collapsed instruments more than compensated for the loss in precision (or imprecision).

One step system GMM estimators based on the initial weight matrices as in Blundell-Bond, 1998 (SYS_BB), conventional one-step system GMM estimators and the system GMM estimators based sub-optimal initial weights matrix in combination with untransformed, lag limited and collapsed instruments (denoted by, SYS_Subf, SYS_Subl, and SYS_Subcol, respectively) at $\gamma = \{0.4, 0.9\}$; $T = \{4, 10\}$ & $n = 100$.

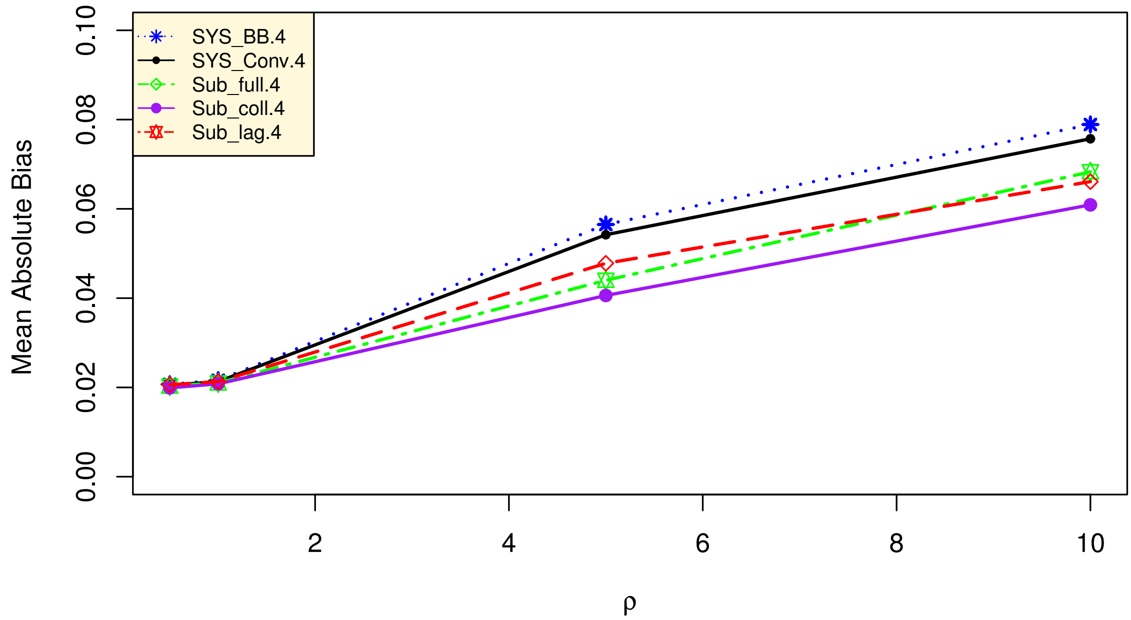

Figure 4.1: Mean absolute Bias for $T = 4$, and $\gamma = 0.4$

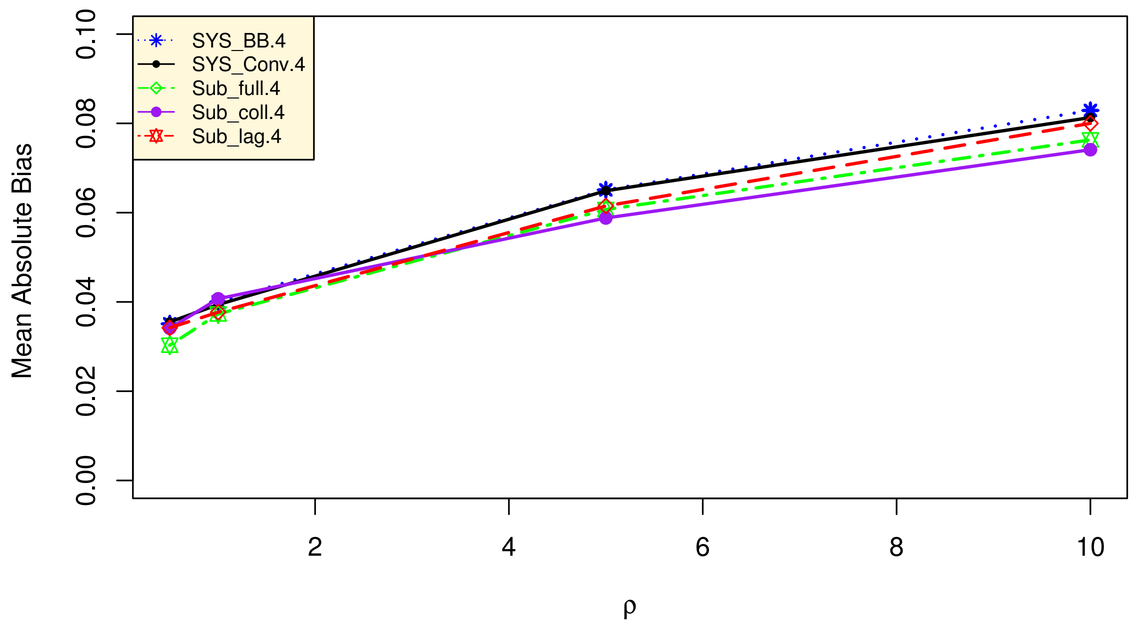

Figure 4.2: Mean absolute Bias for $T = 4$, and $\gamma = 0.9$

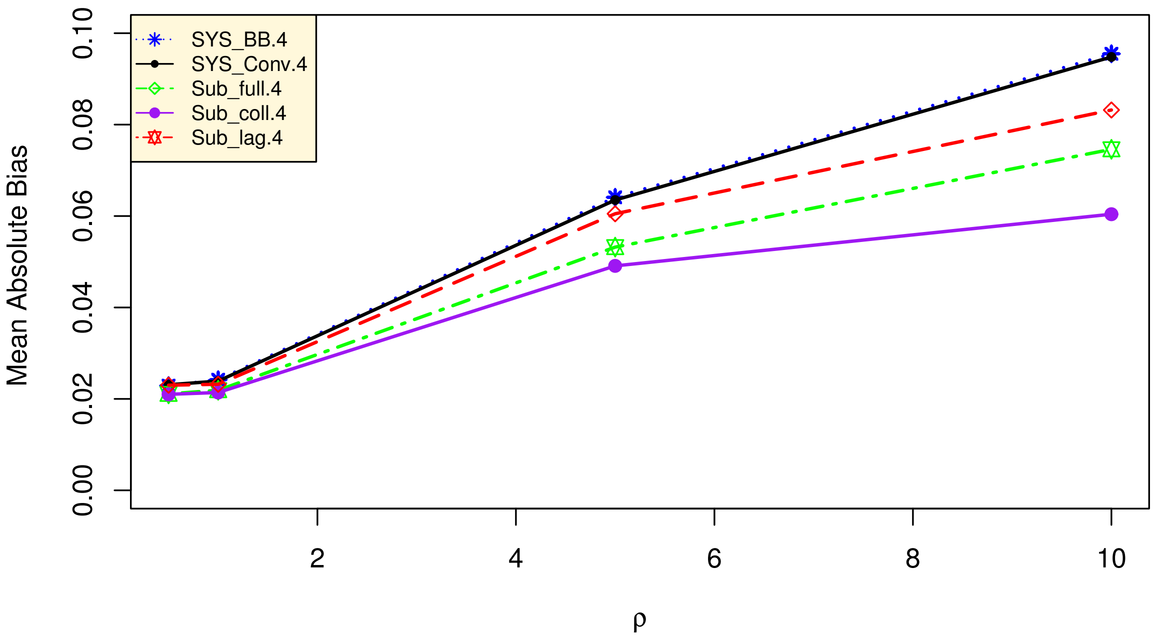

Figure 4.3: Mean absolute Bias for $T = 10$, and $\gamma = 0.4$

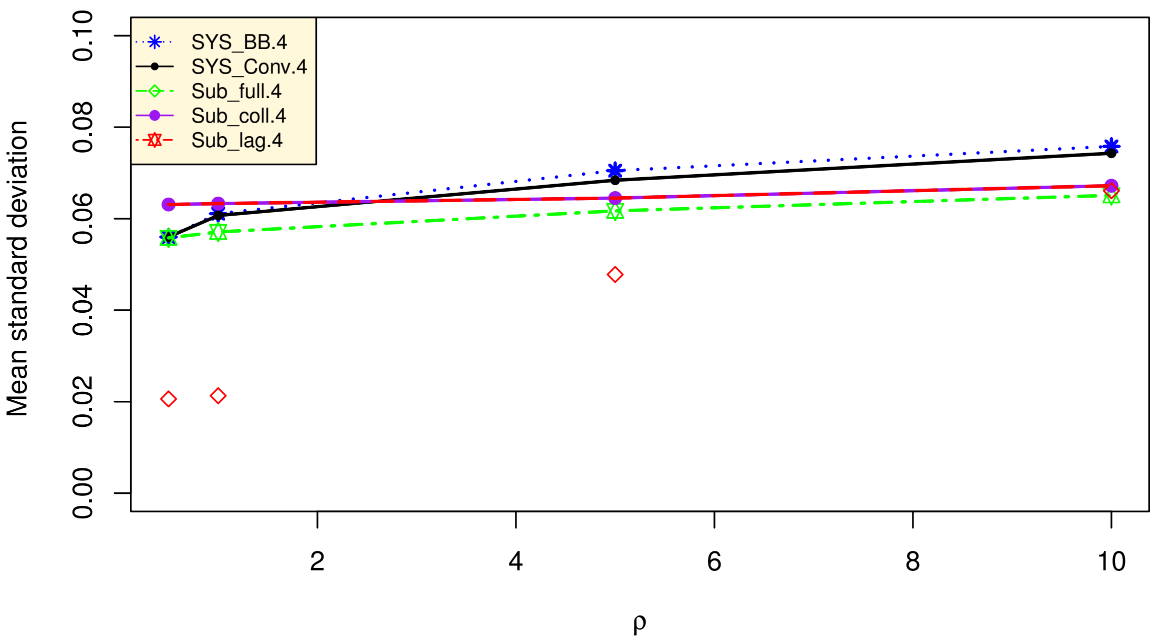

Figure 4.4: Standard deviation for $T = 10$, and $\gamma = 0.4$

4.3 Comparison of two-step system GMM estimators

a) Comparison of estimators based on mean absolute bias

Figures 4.5 to 4.7 pertain to the MAB for two stage system GMM estimators at $n = 100$ & $T = 5, 10$ & 20, respectively. The properties of the estimators for selected values of γ and various levels of variance ratio (ρ) are discussed below.

For $\gamma = 0.2$**:** At $T = 5$ in Table 5.3 (Figure 4.5), the conventional GMM estimator has the smallest MAB for small to moderate variance ratios ($\rho = 0.5, 1, 2$) as compared to sub-optimally weighted estimators based on partially collapsed as well as lag limited instruments. For $T = 10$ (Table 5.4) and small variance ratios ($\rho = 0.5, 1$), there is no marked difference in the MAB between the conventional and sub-optimally weighted GMM estimators. On the other hand, the conventional estimator has relatively higher MAB as compared to sub-optimally weighted GMM estimators for $T = 20$ with few exceptions. Moreover, the results indicate that the estimator which utilizes sub-optimal initial weight matrix and partially collapsed instruments is effective in controlling the bias of estimates for large T and large variance ratios. For small values of T and small variance ratios, however, the conventional GMM estimator seems to do better in terms of bias.

For $\gamma = 0.9$**:** For relatively smaller time dimensions ($T = 5$ and 10), there is no marked difference in the MAB of the conventional as well as the sub-optimally weighted GMM estimators. For $T = 20$ (Table 5.5) and small to moderately large variance ratios ($\rho = 0.5, 1, 2, 5$), however, sub-optimally weighted GMM estimator based on partially collapsed instruments has the smallest MAB. Moreover, for all time dimensions considered ($T = 5, 10$ and 20), the MAB of all estimators keeps on increasing as the variance ratio increases.

In general, for $T = 5$ and smaller variance ratios ($\rho = 0.5, 1$), the estimators at $\gamma = 0.9$ have smaller MAB as compared to those at $\gamma = 0.2$. The opposite is true for moderate to larger variance ratios($\rho = 2, 5, 10$). For $T = 10$, the sub-optimal estimator based on partially collapsed instruments at $\gamma = 0.2$ has smaller MAB as compared to the conventional estimator and sub-optimally weighted estimator based on lag-limited instruments at $\gamma = 0.2$ as well $\gamma = 0.9$ regardless of the values of variance ratios. For $T = 20$, all estimators at $\gamma = 0.9$ have smaller MAB as compared to those at $T = 20$, in general. Even though the findings are mixed, the estimators seem to do well in terms of MAB at $\gamma = 0.2$ as compared to $\gamma = 0.9$ for relatively smaller time dimensions ($T = 5, 10$) and moderate to large variance ratios. The opposite is true when the process approaches a random walk ($\gamma = 0.9$) and $T = 20$.

b) Comparison of estimators based on RMSE

Figures 4.8 to 4.10 display the RMSE of system GMM estimators for selected values of γ and various levels of variance ratios (ρ) at $n = 100$ and for time dimensions $T = 5, 10$ and 20, respectively.

For $\gamma = 0.2$: At $T = 5$, the conventional GMM estimator has the smallest RMSE for all variance ratios considered as compared to sub-optimally weighted estimators based on partially collapsed as well as lag-limited instruments. The estimator based on partially collapsed instruments in particular is the least efficient. For $T = 10$, the conventional estimator still has the smallest RMSE for moderate variance ratio ($\rho = 2$), while there is no marked difference in the RMSE between the conventional and sub-optimally weighted GMM estimators when $\rho = 1$. For large variance ratios ($\rho = 5, 10$) however, the estimator based on partially collapsed instruments outperforms the other two. For $T = 20$, sub-optimally weighted GMM estimator based on partially collapsed instruments is the most efficient across all variance ratios considered, while the conventional estimator based on untransformed instruments is the least efficient. The estimator based on lag-limited instruments is also doing well for extreme variance ratio ratios ($\rho = 5, 10$).

The results indicate that the conventional GMM estimator has the least RMSE regardless of the values of variance ratios for small value of T. However, as the value of T and ρ increase, the GMM estimator adopting sub-optimal initial weight matrix and partially collapsed instruments outperforms all other estimators. For both $T = 5$ and $T = 10$, the RMSE of all estimators keeps on increasing as the variance ratio increases. However, there does not seem to be similar upward trend in the RMSE for the estimator based on partially collapsed instruments for $T = 20$.

For $\gamma = 0.2$: At $T = 5$ and $T = 10$, there is no marked difference in the RMSE of the conventional as well as the sub-optimally weighted estimators regardless of the size of the variance ratio. For $T = 20$ and small to moderately large variance ratios ratios ($\rho = 0.5, 1, 2, 5$), sub-optimally weighted GMM estimator based on lag-limited instruments has relatively higher RMSE, while there is no marked difference in the efficiency of the conventional estimator and the estimator utilizing partially collapsed instruments. Moreover, the RMSE of the estimators keeps on increasing as the variance ratio increases in general. Our conclusion is that, when γ gets closer to one, there is no difference in the performance all estimators in terms of RMSE. However, sub-optimally weighted GMM estimator based on lag-limited instruments has larger RMSE, especially as T gets large ($T = 20$) and for small and moderate values of ρ.

Generally speaking, for all time dimensions, the estimators tend to have lower RMSE for $\gamma = 0.9$ as compared to that of $\gamma = 0.2$, and this is more so for higher variance ratios. According to Doran and Schmidt (2006), this usually comes from severe underestimation of the actual variance for two-step system estimator as γ gets close to one. This implies that our suggested estimator, sub-optimally weighted GMM estimator based partially collapsed instruments, is also vulnerable to weak instruments problems when γ is close to one.

c) Comparison of two-step sub-optimally weighted system GMM estimators

Figures 4.11 to 4.12 compared the MAB and RMSE of the conventional system GMM estimator (SYS_Conv) and sub-optimally weighted GMM estimators utilizing the full set of (or untransformed) instruments (Sub_full), collapsed instruments (Sub_col) and lag-limited instruments (Sub_lag) at $T = 10$ and $\gamma = 0.2$. In terms of MAB, sub-optimally weighted GMM estimator with collapsed instruments by far outperforms all other estimators regardless of the size of the variance ratio. In contrast, the sub-optimally weighted GMM estimator based on full instruments set has relatively smaller RMSE, followed by that constructed using collapsed instruments. Judging by the considerable reduction in MAB with relatively smaller loss of efficiency, we can conclude that the latter (proposed) estimator is preferred to the former. The other two estimators are the worst

in terms of both performance measures.
Tables 5.6 – 5.9 present simulation results of one-step and two-step alternative system GMM estimators at $T = \{5, 10\}$, $n = 30$ and $\gamma = 0.2$. The results revealed that all estimators are less biased and they have smaller RMSE compared with the estimates at $n = 100$ in most cases. Regards to Hansen J tests, estimators based on all available instruments or lag limited never reject H_0 at $T = 10$ and 20. In this case, system GMM estimators based on collapsed instruments seems performing more favorably (or J test has better power). More over, J test shows better size properties for estimators at $n = 300$ compared to the estimates at $n = 100$.

Two stage system GMM estimates based on the initial weight matrix for conventional system GMM at $\gamma = 0.2$ (SYS_Conv.2) & $\gamma = 0.9$ (SYS_Conv.9) against to two stage sub optimal weighted and using reduced instruments: namely partially collapsed at $\gamma = 0.2$ (Sub_coll.2) & $\gamma = 0.9$ (Sub_coll.9), lag limited (the most recent 2 lags in each T) at $\gamma = 0.2$ (Sub_lag.2) & $\gamma = 0.9$ (Sub_lag.9) a $n = 100$ & for $T = \{5, 10 \text{ and } 20\}$, respectively.

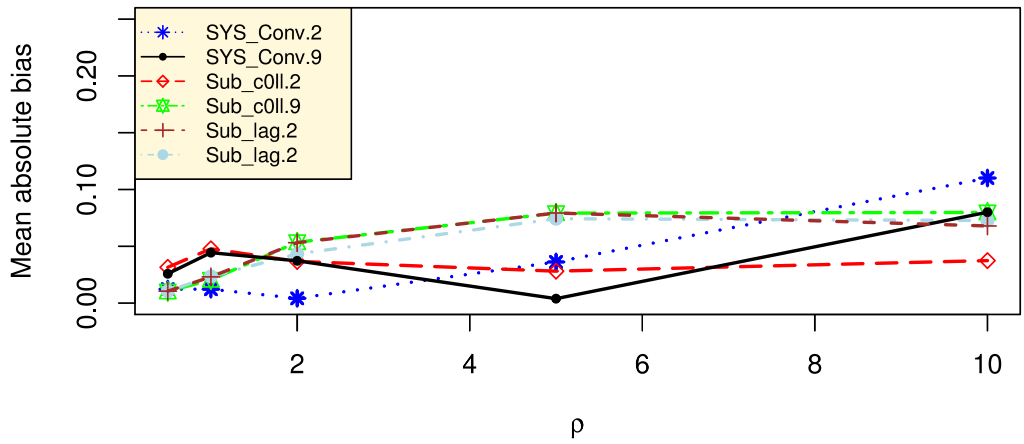

Figure 4.5: Mean absolute bias for $T = 5$ at $\gamma = 0.2$ & $\gamma = 0.9$

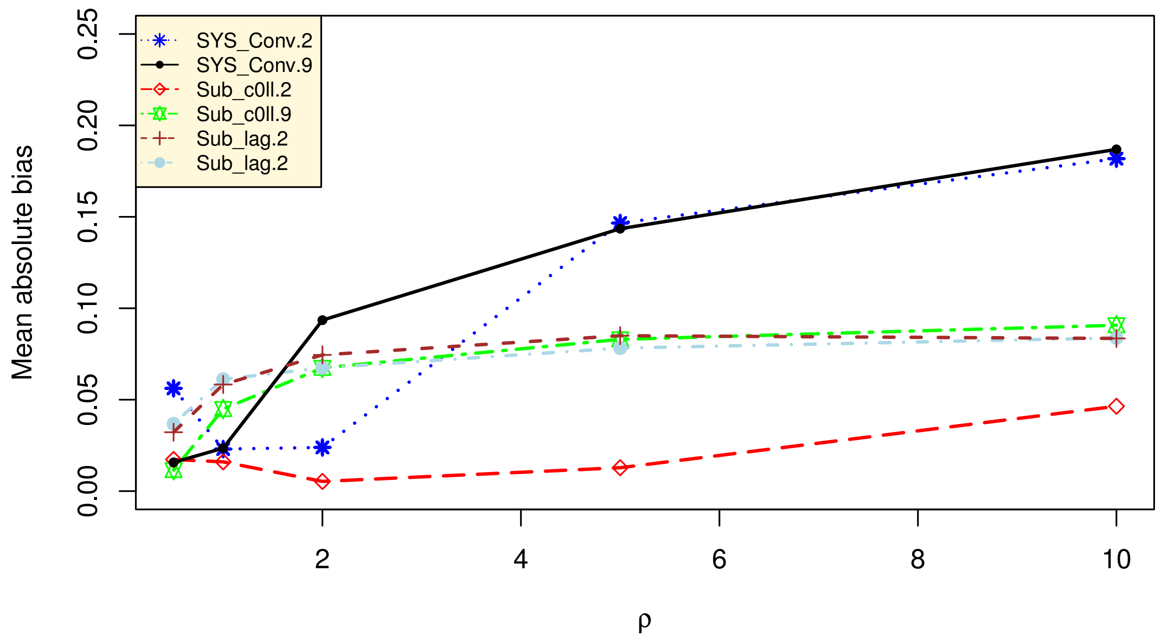

Figure 4.6: Mean absolute bias for $T = 10$ at $\gamma = 0.2$ & $\gamma = 0.9$

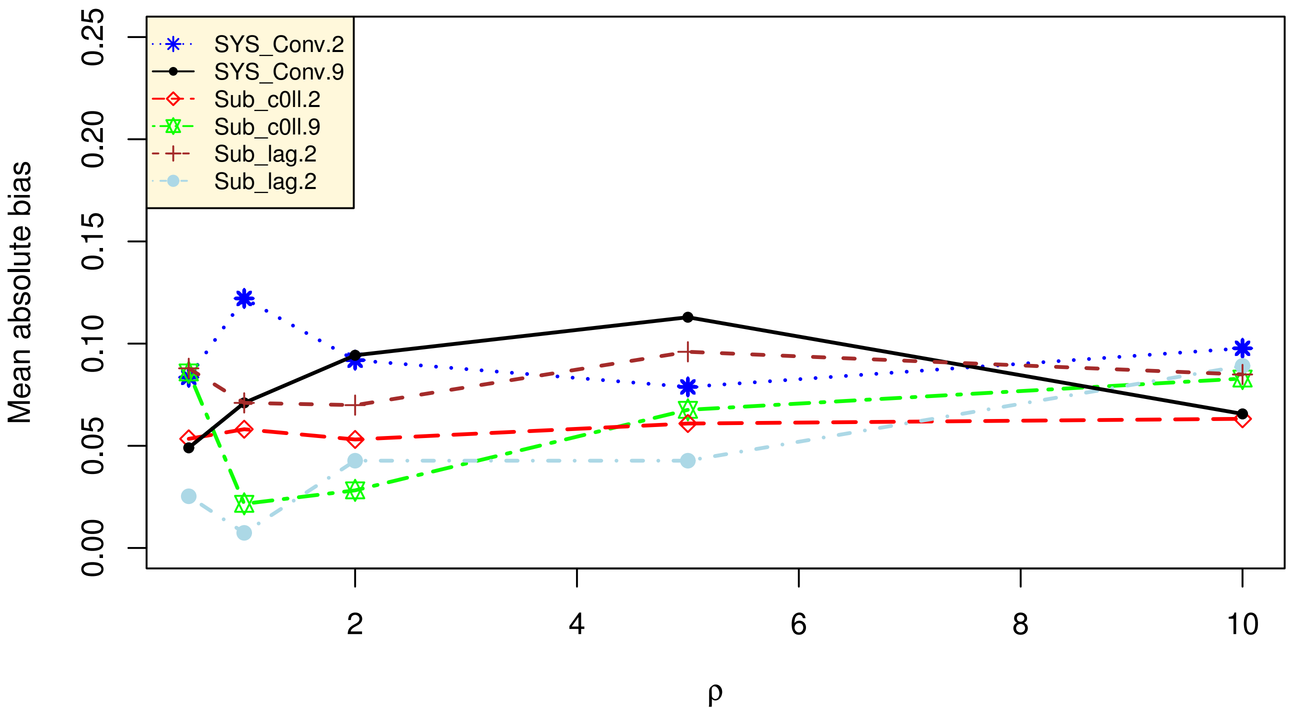

Figure 4.7: Mean absolute bias for $T = 20$ at $\gamma = 0.2$ & $\gamma = 0.9$

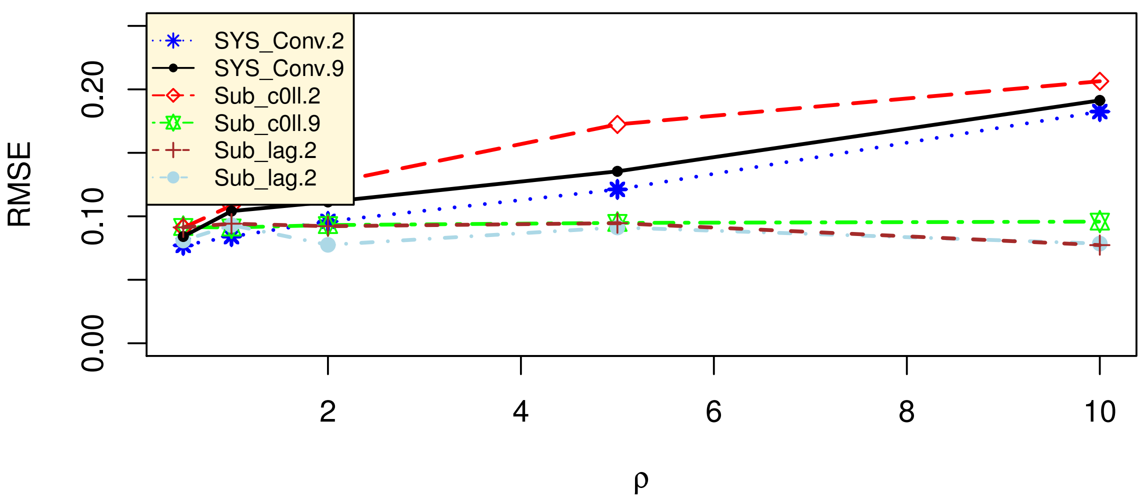

Figure 4.8: RMSE for $T = 5$ at $\gamma = 0.2$ & $\gamma = 0.9$

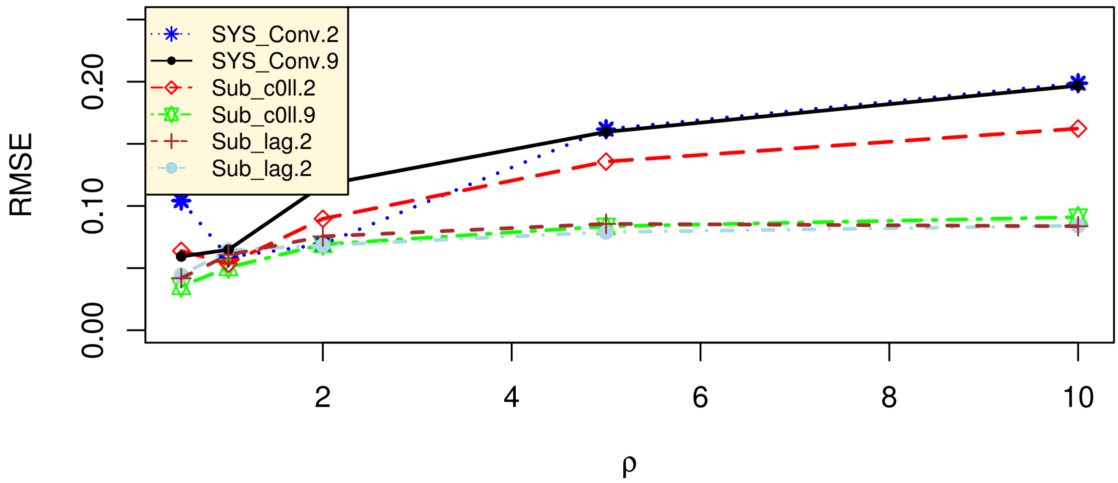

Figure 4.9: RMSE for $T = 10$ at $\gamma = 0.2$ & $\gamma = 0.9$

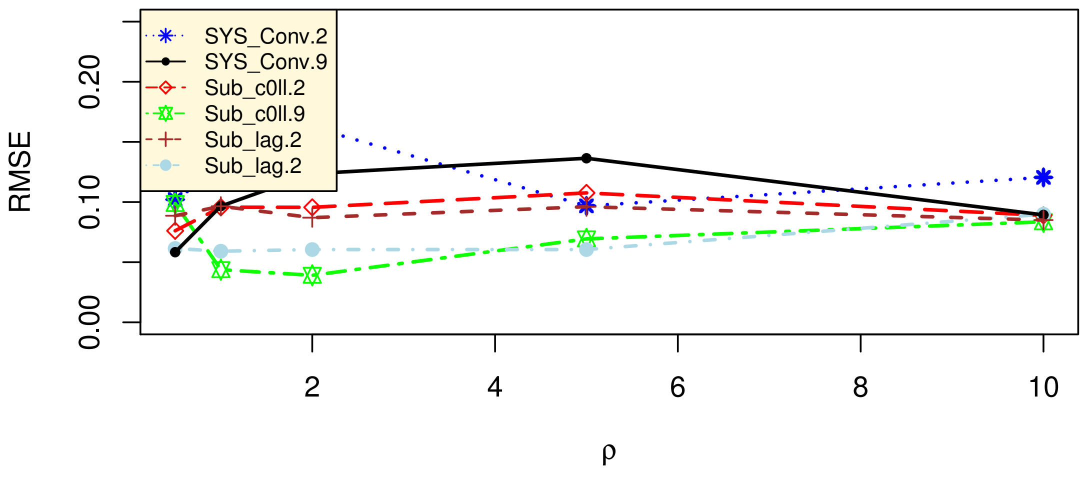

Figure 4.10: RMSE for $T = 20$ at $\gamma = 0.2$ & $\gamma = 0.9$

Two stage system GMM estimates for conventional system GMM estimator (SYS_Conv.2), sub optimal weighted & based on all available instruments (SYS_full.2), sub optimal weighted and using partially collapsed instruments (Sub_col.2) and sub optimal weighted using lag-limited instruments (Sub_lag.2) at $\gamma = 0.2$; $T = 10$ & $n = 100$.

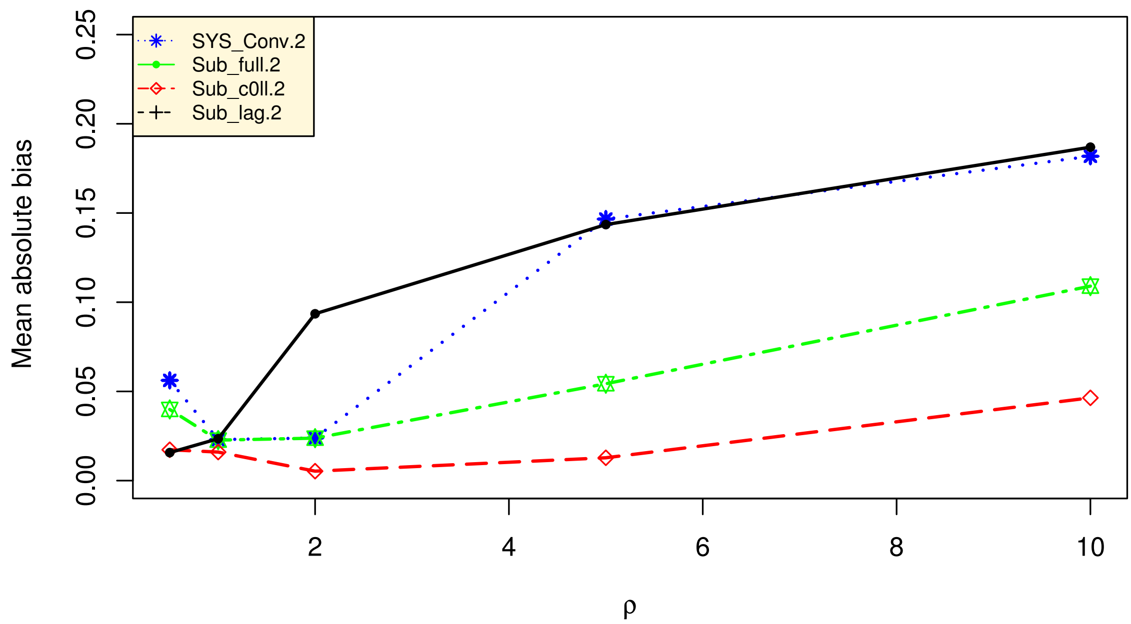

Figure 4.11: Mean absolute bias for $T = 10$ & $\gamma = 0.2$

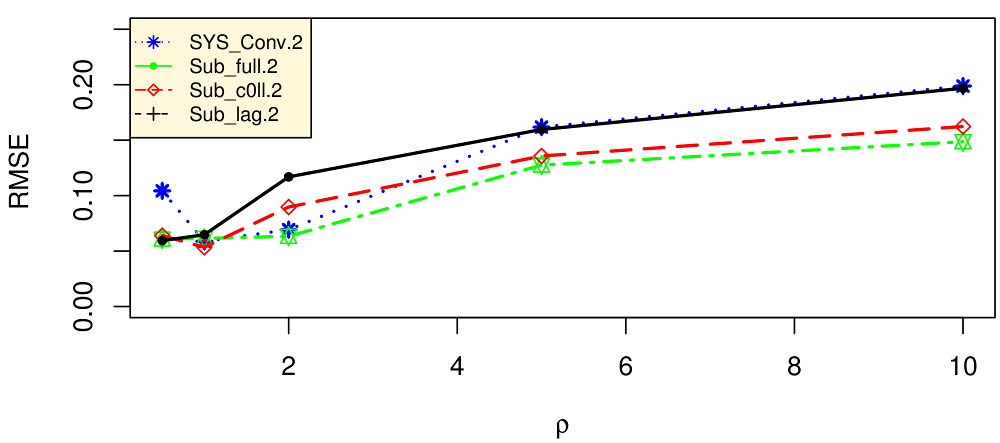

Figure 4.12: RMSE for $T = 10$ & $\gamma = 0.2$

d) Comparison of estimators based on coverage probability

To obtain the empirical coverage probabilities for the system GMM estimators at the nominal 95% confidence level, we ran 5000 replications, estimate the 95% intervals for γ and then computed the proportion of intervals containing the true value. The results for the two-step system GMM estimators are presented in Tables 5.10-5.12 (see APPENDIX I).

For $\gamma = 0.2$**:** At $T = \{5, 10\}$; $n = 100$ and small variance ratios($\rho = 0.5, 1$), the coverage probabilities for all estimators are considerably lower than the nominal 95% level with no marked difference. The same is true for $T = 20$ and at $\rho = 0.5$. For small time dimension ($T = 5$) and higher values of ρ, however, the empirical coverage probabilities are all in line with the nominal level. The benefit of instruments reduction is clearly demonstrated for higher time dimensions ($T = 10$ and 20) and higher values of ρ ($\rho = 5, 10$). Under these scenarios, the empirical coverage probabilities of sub-optimally weighted GMM estimators based on lag-limited and collapsed instruments are closer to the nominal level than that based on untransformed instruments (and this is more so for $T = 20$). In particular, the conventional estimator performed poorly for $T = 20$ where the coverage probabilities are way off from the nominal confidence level regardless of the size of the variance ratio. In general, the coverage probabilities of all estimators keep on increasing as the variance ratio increases.

For $\gamma = 0.9$**:** For all time dimensions considered ($T = 5, 10$ and 20) , the coverage probabilities for all estimators are considerably lower than the nominal level regardless of the size of the variance ratio, and there is no marked difference in the coverage probabilities between the conventional and sub-optimally weighted GMM estimators. It is well documented that the two-step standard errors of GMM estimators systematically underestimate the real standard deviation of the estimates when the process approaches a random walk (Doran and Schmidt, 2006). This results in narrower confidence intervals, and hence lower coverage probabilities.

In summary, when the process approaches a random walk, there is no added value in utilizing sub-optimally weighted GMM estimators in terms of coverage probabilities. For smaller values of the coefficient of the lagged dependent variable, on the other hand, estimators based on reduced instruments seem superior for higher time dimensions and higher values of the variance ratio.

micro-panel data: Data covered a large number of individuals (n) over a short time (T). It can provide insights into specific trends such as consumer behavior, individual labor markets, and the theory of firms.

macro-panel data: Data characterized by having a relatively large time dimension (T) and a relatively small number of individuals (n). Macro data can provide insights into broad trends such as aggregate demand, national output and inflation.

4.4 Empirical illustration

In this section we illustrate the performance of alternative GMM estimators using real data set obtained from a series of Large and Medium Scale Manufacturing Industries (LMSMI) surveys conducted in Ethiopia by the Ethiopian Statistical Service (ESS) from 2010 to 2019. The ESS defines large and medium scale manufacturing industries as all manufacturing establishments which employ ten or more people and use power-driven machines. This definition is based on the International Standard Industrial Classification (ISIC Revision-3.1). The panel dataset for our comparative analysis comprises 284 establishments and 2840 observations (balanced panel). Real wage and real working capital were computed by deflating the nominal values using the price deflator series obtained from the Ministry of Planning and Development.

We consider the labour demand (employment) equation of the form:

$$y_{it} = \beta_w w_{it} + \beta_k k_{it} + \delta_t + \alpha_i + \epsilon_{it} + m_{it}; i = 1, \cdots, n; t = 2, \cdots, T, \quad (4.3)$$

$$\epsilon_{it} = \phi\epsilon_{i,t-1} + e_{it},\ |\phi| < 1,\ e_{it},\ m_{it} \sim MA(0)$$

where y_{it} is log employment of firm i in year t, w_{it} is log real wage and k_{it} is log real working capital, and δ_t is a year-specific intercept, α_i is firm specific time-invariant unobserved effect, ϵ_{it} is a random shock and m_{it} refers serially uncorrelated measurement errors. The explanatory variables we considered (w_{it}, k_{it}) are potentially correlated with the individual-specific effects, the random shocks and measurement errors, so that the instrument set for each variable of interest will be constructed accordingly. The model has dynamic (common factor) representation defined as:

$$\begin{aligned} y_{it} &= \phi y_{i,t-1} + \beta_w w_{it} - \phi\beta_w w_{i,t-1} + \beta_k k_{it} - \phi\beta_k k_{i,t-1} + \\ &\quad (\delta_t - \phi\delta_{t-1}) + \alpha_i(1-\phi) + e_{it} + m_{it} - \phi m_{i,t-1} \qquad (4.4) \end{aligned}$$

Equation (4.4) can be expressed as:

$$y_{it} = \pi_1 y_{i,t-1} + \pi_2 w_{it} + \pi_3 w_{i,t-1} + \pi_4 k_{it} + \pi_5 k_{i,t-1} + \tilde{\delta}_t + \tilde{\alpha}_i + e_{it} + \eta_{it} \quad (4.5)$$

subject to two non-linear (common factor) restrictions, $\pi_3 = -\pi_1\pi_2$ and $\pi_5 = -\pi_1\pi_4$, where $\tilde{\delta}_t = (\delta_t - \phi\delta_{t-1})$, $\tilde{\alpha_i} = \alpha_i(1-\phi)$ and $\eta_{it} = e_{it} + m_{it} - \phi m_{i,t-1}$. Consequently, both the unrestricted parameter vector $\pi = (\pi_1, \cdots, \pi_5)$ and and the restricted ones (β_w, β_k, ϕ) can be estimated. Notice that η_{it} is serially uncorrelated if there are no measurement errors or $\eta_{it} \sim MA(1)$ otherwise.

During the implementation of various GMM estimation procedures, the validity of the moment restrictions was tested. We begin by using lagged levels dated $(t-2)$ and earlier as instruments for the differenced equation and differenced lags dated $(t-1)$ as instruments for levels equation. However, Sargan test of over identifying restrictions rejects the validity of instruments. This may arise from the presence of measurement errors. Hence, we exclude lagged levels dated $(t-2)$ and differenced lags dated $(t-1)$ from the instruments set and apply $(t-3)$ and earlier level lags as instruments for difference equations and differenced lags dated $(t-2)$ for level equations. The validity of instruments under this specification was confirmed by the Sargan test. The common factor restrictions are accepted for various implementations of system GMM estimators, while they are rejected by OLS, FE and first difference GMM estimators.

Table 5.13 presents the result. The coefficient on the lagged dependent variable is higher for sub-optimally weighted system GMM estimators. Specifically, sub optimally weighted system GMM estimators utilizing untransformed instruments (denoted by Sub_full) and collapsed instruments (denoted by Sub_col) have larger coefficient for lagged dependent variable. The standard deviation of these estimates is smaller as compared to those from the conventional system GMM estimator and sub-optimally weighted system GMM estimator employing lag limited instruments. In addition, the coefficient of the lagged dependent variable is higher than the within groups estimate and below the OLS levels estimate as expected. One unexpected result is that the elasticity of employment with respect to real wage is positive and significant across all estimators. This probably is attributed to the decline in the real wage over time for LMSMI workers. Our finding is consistent with those of Berhe et al.(2020) who argued that this is likely a reflection of fixed nominal public wage in the presence of very high inflation rate. As a consequence, very low real wage rate can be considered as one of the reasons for unemployment (especially, voluntary unemployment) in the manufacturing sector of Ethiopia. The result may suggest the importance revising wage rate (at least minimum wage rate) in the sector. The elasticity of employment with respect to working capital is positive. However the coefficients for all estimators are very small, with relatively larger sizes for conventional and proposed estimators. In general, sub-optimal weight system GMM estimators utilizing either untransformed or collapsed instruments may yield reasonable estimates.

To check the robustness of the estimators, we re-fit the models by omitting the first 4 observations for each firm, i.e, we consider only data from 2014 to 2019 and the results are presented in Table 5.14. The coefficient estimates are closer to those obtained from the full sample period (2010-2019), while their standard errors are relatively larger.The results of tests of instruments validity and serial correlation and common factor restrictions are consistent with those obtained for the full sample.

Chapter 5

Conclusions and Future Research

5.1 Conclusions

The generalized method of moments (GMM) estimators are the most popular in dynamic panel data estimation. Two crucial issues in GMM estimation of dynamic panel models are the choice of an initial weighting matrix and the problem of instruments proliferation. In this study, we compared the performance of one-step and two-step system GMM estimators utilizing various instrument sets and initial weight matrices.

For one stage estimation, comparison was made between the Blundell and Bond (1998) estimator, the conventional system GMM estimator (for example, Doornik, 2001), the sub-optimal system GMM estimator using all available instruments (for example, Jung et. al., 2015) and the proposed sub-optimal system GMM estimators with reduced instruments (i.e., collapsed instruments and lag-limited instruments). The simulation results indicated that sub-optimally weighted GMM estimator based on collapsed instruments is the least biased for moderate and large values of the variance ratio, and this is more so as the time dimension increases and for relatively smaller autoregressive coefficients (for example, at $T = 10$ and $\gamma = 0.4$). The reverse is true for the conventional one-step system GMM estimator and that based on an initial weight matrix as in Blundell and Bond (1998). When the autoregressive coefficient gets close to one, however, there is no marked difference in the mean absolute bias across the board regardless of the values of the variance ratio. Moreover, the bias of all estimators keeps on increasing as the variance ratio (ρ) increases. In terms of precision, there is no notable difference across all estimators for small values of the variance ratio as well as when γ gets close to one. For moderate and large values of the variance ratio, however, sub-optimally weighted GMM estimator using all available instruments is the most efficient, followed by that of utilizing collapsed instruments set.

For two stage estimation, we compared the performance of the conventional system GMM estimator, the sub-optimal optimal system GMM estimator using all available instruments, sub-optimal system GMM estimator employing collapsed instruments and sub-optimal system GMM estimator utilizing lag-limited instruments. Our simulation results revealed that sub-optimally weighted GMM estimator based on partially collapsed instruments outperforms all other estimators in most cases. This is more visible when the coefficient of the lagged dependent variable is small ($\gamma = 0.2$) and the values of T and/or ρ are increased. Under these scenarios, the system GMM estimators based on reduced instruments were also found to perform well in terms of coverage probabilities. However, as γ gets close to one, there is no considerable gain from the use of the proposed estimators. Regarding precision, for relatively small time dimensions and large variance ratios, the sub-optimal system GMM estimator based on all available instruments performs best. The simulation results also indicated that the RMSE of all estimators keeps on increasing as the variance ratio increases in general.

In general terms, the performance of sub-optimally weighted system GMM estimators which utilize collapsed and untransformed instruments seems promising. When we compare these two estimators, the bias reduction from the former estimator outweighs the efficiency gain from the latter estimator in relative terms as the value of T increases. In this regard, utilizing these estimators (instead of the conventional system GMM estimators) might provide more insights in panel data analysis for practitioners.

5.2 Future Research

We consider fixed value of n by assuming that the performance of estimators usually improves as the cross sectional dimension of panel data increases (based on literature). However, when the time dimension of panel becomes large relative to the value of the cross-sectional dimension ($n = 100$), the suggested estimator may not effectively control the bias. In our future study, we will assess the performance of estimators using different values of n in combination with varying time dimension (T) of panel. In addition, the empirical illustration of this study considers balanced panel data. However, exit (not surviving over survey periods) and entry of new firms are common phenomena in empirical studies. In the future, we will assess the performance of estimators using unbalanced panel data.

Moreover, some authors in the area (for example, Hansen, 2001; Bun and Carree, 2005) suggest bias corrected fixed effects estimation techniques. These authors show that bias corrected fixed effects estimators are more efficient compared to conventional GMM estimators in their simulation studies. Studies by Chigira and Yamamoto (2006) and Breitung et al. (2022) further suggest bias corrected GMM estimators. So, our future research will assess the performance of our estimator compared to that of bias corrected least squares as well as bias corrected GMM estimators.

Bibliography

[1] Ahn S.C. & Schmidt. P.(1995). Efficient estimation of models for dynamic panel data, Journal of Econometrics 68, 5-27.

[2] Andersen, T.G, & Sorensen, B.E.(1996). GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study. Journal of Business and Economic Statistics 14(3):328–352

[3] Antoine.B . & Renault.E .(2010). Efficient Inference with Poor Instruments: a General Framework, Simon Fraser university Department of Economics Working Papers series, 12-04.

[4] Anderson T. W. & Hsiao. C.(1981). Estimation of Dynamic Models with Error Components, Journal of the American Statistical Association, Volume 76, Issue 375,598-606.

[5] Arellano, M. & Bond, S.R.(1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. The Review of Economic Studies 58, 2, 277-297.

[6] Arellano, M. & Bover, O.(1995). Another look at the instrumental variable estimation of error-component Models. Journal of Econometrics 68, 29-51.

[7] Breitung et al. (2002). Bias-corrected method of moments estimators for dynamic panel data models, Econometrics and Statistics, vol. 24, 116-132

[8] Berhe et al., (2020).Ethiopia Productivity Report, National Graduate Institute for Policy Studies.

[9] Blundell, R. & Bond, S.(1998). Initial conditions and moment restrictions in dynamic panel data models, Journal of Econometrics 87, 115 -143.

[10] Bontempi, E.B & Mammi, I.(2012). A strategy to reduce the count of moment conditions in panel data GMM. Munich Personal RePEc Archive (MPRA) Paper No. 40720.

[11] Bontempi, E. B. & Mammi, I.(2014). Implementing a strategy to reduce the instrument count in panel GMM. The Stata Journal, 15, No.4, 1075-1097.

[12] Bowsher.C.G. (2002). On testing over identifying restrictions in dynamic panel data models, Economics Letters 77 (2002) 211–220.

[13] Bun, M.J.G. & Carree , M.A. (2001).Bias-Corrected Estimation in Dynamic Panel Data Models, Journal of Business & Economic Statistics, 23(2), 200-210

[14] Bun, M.J.G.& Kiviet, J.F.(2006). The effects of dynamic feedbacks on LS and MM estimator accuracy in panel data models. Journal of Econometrics 132, 409-444.

[15] Bun.J.G., & Poldermans, R.W. (2015). Weak identification robust inference in dynamic panel data Models. Tinbergen Institute and Amsterdam School of Economics, University of Amsterdam.

[16] Bun, M.,& Sarafidis.V.(2013). Dynamic Panel Data Models, University of Amsterdam. Econometrics discussion Paper: 2013/01.

[17] M. Carrasco, M. & Doukali, M.(2022). Testing over identifying restrictions with many instruments and heteroscedasticity using regularised jack-knife instruments; Econometrics Journal, volume 25, 71-97.

[18] Alonso-Borrego, C. & Arellano, M. (1999). Symmetrically Normalized Instrumental-Variable Estimation Using Panel Data, Journal of Business and Economic Statistics, Vol. 17, No. 1

[19] Chigira, H. & Yamamoto,T.(2006). A Bias-Corrected Estimation for Dynamic Panel Models in Small Samples, Hi-Stat Discussion Paper Series d06-177, Institute of Economic Research, Hitotsubashi University.

[20] Croissant.Y & Millo, G.(2019). Panel Data Econometrics with R, John Wiley and Sons Ltd, The Atrium, Southern Gate, Chichester,West Sussex, PO19 8SQ, UK.

[21] Doornik, et al., (2002). Panel Data estimation using DPD for Ox, DPD package, Version 1.25

[22] Doran, H.E. & Schmidt, P. (2006). GMM estimators with improved finite sample properties using principal components of the weighting

matrix, with an application to the dynamic panel data model, Journal of Econometrics 133, 387-409,

[23] Eakin, et al., (1988). Estimating Vector Autoregressions with Panel Data, Econometrica, Vol. 56, No. 6

[24] Fajeau, M.(2020). Too Much Finance or Too Many Weak Instruments? International Economics, S2110-7017(20)30271-7.

[25] Fox,J.]& Weisberg, S. (2008). An R Companion to Applied Regression, third edition, Sage Publications.

[26] Hall, A. R.(2005). Generalized Method of Moments , Oxford University press.

[27] Hansen, G.(2001). A Bias-Corrected Least Squares Estimator of Dynamic Panel Models, Allgemeines Statistisches Archiv, 85, 127-140

[28] Hansen, L. P.(1982). Large Sample Properties of Generalized Method of Moments Estimators, Econometrica, 50 (4) 1029-1054.

[29] Hansen, et al., (1996). Finite-Sample Properties of Some Alternative GMM Estimators, Journal of Business & Economic Statistics, 14(3), 262-280.

[30] Harris, D. & Matyas, L. (1999). Introduction to the generalized method of moments estimation. In Generalized Method of Moments Estimation (L. Matyas, ed.). Cambridge University Press, New York.

[31] Hayakawa, K.(2007) Small sample bias properties of the system GMM estimator in dynamic panel data models. Economics Letters 95(1), 32–38

[32] Jung, H. & Kwon, H.U.(2007). An Alternative System GMM Estimation in Dynamic Panel Models. Hitotsubashi University Repository. Discussion Paper Series No.217

[33] Jung, et al., (2015). An Alternative System GMM Estimation in Dynamic Panel Models. Journal of Economic Theory and Econometrics, 26 (2)

[34] Kiviet, J.F.(2007a). Judging contending estimators by simulation: tournaments in dynamic panel data models. Tinbergen Institute. Discussion Paper, No. 05-112/4.

[35] Kiviet, J.F.(2007b). On the optimal weighting matrix for the GMM system estimator in dynamic panel data models. Tinbergen Institute, Discussion Paper, No. 2007/02.

[36] Kiviet, J., (2017). Accuracy and Efficiency of Various GMM Inference Techniques in Dynamic Micro Panel Data Models," Econometrics, MDPI, 5(1),1-54. Liu. S. & Neudecker. H.(1997). Kantorovich inequalities and efficiency comparison for several classes of estimators in linear models. Statistica Neerlandica,51, 345-355.

[37] Lai et al., (2008). Journal of Statistical Planning and Inference. Elsevier, 138, Issue 9, 2763-2776

[38] Mammi, I. & Calzolari, G.(2011). Essays in GMM estimation of dynamic panel data models. IMT (Institutions, Markets, Technologies), School for Advanced Studies Lucca, Italy.

[39] Mehrhoff, J.(2009). A solution to the problem of too many instruments in dynamic panel data GMM. European Central Bank discussion paper series 1: Economic studies No 31/2009.

[40] Nickell, S.(1981). Biases in Dynamic Models with Fixed Effects, Econometrica, 49 (6), 1417-1426.

[41] Phillips, C.B. & Han,C.(2019). Dynamic panel GMM using R, Handbook of Statistics, Vol. 41.

[42] Okui, R.(2009). The optimal choice of moments in dynamic panel data model, Journal of Econometrics, 151, 1-16

[43] Park, et al., (2007). Semi-parametric efficient estimation of dynamic panel data models, Journal of Econometrics 136, 281–301

[44] Roodman, D.(2009a). A note on the theme of too many instruments. Oxford bulletin of Economics and Statistics, 71, 1.

[45] Roodman, D.(2009b). An introduction to difference and system GMM in Stata. The Stata Journal.No.1, 86 -136.

[46] Stock. J.H. & Wright. J.H.(2000). General method of moments estimator with weak identification,Econometrica, 68, No.5, 1055-1096.

[47] Staiger, D. & Stock, J.H.(1997). Instrumental Variables Regression with Weak Instruments, Econometrica, 65 (3), pp. 557-586.

[48] Tauchen, G.(1986). Statistical Properties of Generalized Method-of-Moments Estimators of Structural Parameters Obtained From Financial Market Data, American Statistical Association, Journal of Business & Economic Statistics, 4 (4).

[49] Windmeijer, F.(2000). Efficiency comparisons for a system GMM estimator in dynamic panel data Models Institute for Fiscal Studies. Working Papers, No. W98/01.

[50] Windmeijer, F.(2005). A finite sample correction for the variance of linear efficient two-step GMM, Journal of Econometrics, 126, 25–51.

[51] Wooldridge, J.M.(2010). Econometric analysis of cross section and panel data (Second edition). The Massachusetts Institute of Technology, (MIT) Press

[52] Youssef, et al., (2014). Improving the Efficiency of GMM Estimators for Dynamic Panel Models. Vol.47, No.2, Munich Personal RePEc Archive (MPRA) paper No.68675.

[53] Youssef, A. H. & Abonazel, M. R.(2017). Alternative GMM Estimators for First-order Autoregressive Panel Model: An Improving Efficiency Approach. Communications in Statistics - Simulation and Computation,ISSN: 0361-0918.

[54] Ziliak, J.P.(1997). Efficient estimation with panel data when instruments are predetermined: an Empirical comparison of moment-condition estimators, Journal of Business and Economic Statistics, 15(4), 419-431.

APPENDIX I: Tables

Table 5.1: **Mean absolute bias** and **Standard deviation (std.dev.)** of alternative one step GMM estimators based on original (**untransformed**), **lag limited** and **collapsed** instruments from **top** to **bottom**, respectively, ($T = 4$) & ($n = 100$)

		ρ	SYS_BB	SYS_Conv	SYS_Sub
Untransformed IVs	$\gamma = 0.4$	0.5	0.0207 (*0.0806*)	0.0205(*0.0806*)	0.0203 (*0.0791*)
		1	0.0216 (*0.0814*)	0.0213 (*0.0811*)	0.0210 (*0.0803*)
		5	0.0565(*0.1082*)	0.0542 (*0.1060*)	0.0440 (*0.0922*)
		10	0.0789(*0.1190*)	0.0757(*0.1032*)	0.0612 (*0.1001*)
		J-test	0.0761	0.0783	0.0855
	$\gamma = 0.9$	0.5	0.0351 (*0.1009*)	0.0355(*0.1006*)	0.0303 (*0.1003*)
		1	0.0400 (*0.1003*)	0.0394 (*0.0942*)	0.0373 (*0.0950*)
		5	0.0651(*0.0871*)	0.0649 (*0.0875*)	0.0607 (*0.0883*)
		10	0.0829 (*0.0723*)	0.0813(*0.0826*)	0.0763 (*0.0839*)
		J-test	0.0980	0.0952	0.0911
Lag limited IVs	$\gamma = 0.4$	0.5	0.0210 (*0.0863*)	0.0209 (*0.0861*)	0.0206 (*0.0859*)
		1	0.0217 (*0.0870*)	0.0217(*0.0867*)	0.0213 (*0.0863*)
		5	0.0568 (*0.0882*)	0.0533 (*0.1100*)	0.0478 (*0.0939*)
		10	0.0785 (*0.1038*)	0.0741 (*0.1031*)	0.0628(*0.1198*)
		J-test	0.0723	0.0721	0.0714
	$\gamma = 0.9$	0.5	0.0342 (*0.0995*)	0.0358 (*0.0993*)	0.0342 (*0.1080*)
		1	0.0414 (*0.0981*)	0.0415 (*0.0983*)	0.0377(*0.0979*)
		5	0.0666 (*0.0848*)	0.0663 (*0.0849*)	0.0615 (*0.0886*)
		10	0.0885 (*0.0825*)	0.0885 (*0.0824*)	0.0800 (*0.0865*)
		J-test	0.0756	0.0732	0.0733
Collapsed IVs	$\gamma = 0.4$	0.5	0.0205(*0.0874*)	0.0203(*0.0862*)	0.0199 (*0.0867*)
		1	0.0211 (*0.0876*)	0.0210 (*0.0871*)	0.0208 (*0.0969*)
		5	0.0555 (*0.0978*)	0.0537 (*0.0953*)	0.0406 (*0.1095*)
		10	0.0745 (*0.1210*)	0.0708 (*0.1140*)	0.0609 (*0.1183*)
		J-test	0.0704	0.0699	0.0701
	$\gamma = 0.9$	0.5	0.0368 (*0.0952*)	0.0371 (*0.0956*)	0.0341 (*0.1069*)
		1	0.0414 (*0.0937*)	0.0415(*0.0937*)	0.0407 (*0.0980*)
		5	0.0660 (*0.0902*)	0.0661 (*0.0908*)	0.0588 (*0.0869*)
		10	0.0883 (*0.0881*)	0.0866 (*0.0888*)	0.0741(*0.0854*)
		J-test	0.0729	0.0715	0.0725

Note that: In each tables $(5.1 - 5.2)$ the standard deviation is presented with in parentheses.

J-test: represents Hansen J test (mean p-value) for each specification of Monte Carlo experiments (tables $5.1 - 5.9$)

We adopt bootstrap replication in estimation, so that the standard deviation is an estimate of the standard error of the coefficients.

Table 5.2: **Mean absolute bias** and **Standard deviation (std.dev.)** of alternative one step GMM estimators based on original (**untransformed**), **lag limited** and **collapsed** instruments from **top** to **bottom**, respectively, ($T = 10$) & ($n = 100$)

		ρ	SYS_BB	SYS_Conv.	SYS_Sub
Untransformed IVs	$\gamma = 0.4$	0.5	0.0229 (*0.0560*)	0.0231 (*0.0560*)	0.0211 (*0.0558*)
		1	0.0242 (*0.0611*)	0.0497 (*0.0607*)	0.0219 (*0.0571*)
		5	0.0641 (*0.0705*)	0.0635 (*0.0684*)	0.0532 (*0.0617*)
		10	0.0955 (*0.0758*)	0.0948 (*0.0743*)	0.0746 (*0.0651*)
		J-test	0.4913	0.4908	0.4411
	$\gamma = 0.9$	0.5	0.0351 (*0.0516*)	0.0355 (*0.0513*)	0.0303 (*0.0522*)
		1	0.0400 (*0.0516*)	0.0394 (*0.0512*)	0.0373 (*0.0520*)
		5	0.0651 (*0.0456*)	0.0649 (*0.0452*)	0.0607 (*0.0465*)
		10	0.0829 (*0.0429*)	0.0813 (*0.0438*)	0.0763 (*0.0441*)
		J-test	0.4501	0.4426	0.4458
Lag limited IVs	$\gamma = 0.4$	0.5	0.0224 (*0.0508*)	0.0224 (*0.0630*)	0.0221 (*0.0631*)
		1	0.0227 (*0.0667*)	0.0226 (*0.0665*)	0.0222 (*0.0633*)
		5	0.0673 (*0.0694*)	0.0661 (*0.0690*)	0.0605 (*0.0645*)
		10	0.0903 (*0.0738*)	0.0895 (*0.0731*)	0.0832 (*0.0672*)
		J-test	0.3615	0.3627	0.3210
	$\gamma = 0.9$	0.5	0.0331 (*0.0556*)	0.0320 (*0.0553*)	0.0342 (*0.0547*)
		1	0.0597 (*0.0552*)	0.0581 (*0.0551*)	0.0377 (*0.0547*)
		5	0.0914 (*0.0469*)	0.0879 (*0.0475*)	0.0615 (*0.0488*)
		10	0.0924 (*0.0449*)	0.0934 (*0.0448*)	0.0800 (*0.0473*)
		J-test	0.2796	0.2829	0.2850
Collapsed IVs	$\gamma = 0.4$	0.5	0.0217 (*0.0544*)	0.0215 (*0.0537*)	0.0209 (*0.0561*)
		1	0.0220 (*0.0684*)	0.0219 (*0.0576*)	0.0214 (*0.0629*)
		5	0.0523 (*0.0770*)	0.0518 (*0.0733*)	0.0491 (*0.0643*)
		10	0.0632 (*0.0894*)	0.0631 (*0.0783*)	0.0604 (*0.0692*)
		J-test	0.2130	0.2073	0.1604
	$\gamma = 0.9$	0.5	0.0382 (*0.0504*)	0.0387 (*0.0504*)	0.0341 (*0.0510*)
		1	0.0619 (*0.0503*)	0.0612 (*0.0501*)	0.0407 (*0.0508*)
		5	0.0882 (*0.0466*)	0.0882 (*0.0468*)	0.0588 (*0.0471*)
		10	0.0930 (*0.0460*)	0.0930 (*0.0479*)	0.0741 (*0.0463*)
		J-test	0.1699	0.1619	0.1586

Table 5.3: **Mean absolute bias** and **RMSE** of various GMM estimators based on original (**untransformed**), **lag limited** (the first two lags) and **collapsed** instruments from **top** to **bottom**, respectively, $(T = 5)$ & $(n = 100)$

.

Note that: In each cell of table 5.3 up to table 5.5 **bold italic** within brackets **refers** to **RMSE**.

			SYS_Conv		SYS_Sub	
		ρ	SYS_Conv1	SYS_Conv2	SYS_Sub1	SYS_Sub2
Untransformed IVs	$\gamma = 0.2$	0.5	-0.0124 (***0.0773***)	-0.0123(***0.0771***)	-0.0121 (***0.0770***)	-0.0123 (***0.0763***)
		1	-0.0124 (***0.0843***)	-0.0122 (***0.0840***)	-0.0121 (***0.0823***)	-0.0122 (***0.0823***)
		2	-0.0045(***0.0961***)	-0.0042 (***0.0960***)	-0.0042 (***0.0953***)	-0.0243 (***0.0933***)
		5	0.0360 (***0.1292***)	0.0362 (***0.1212***)	0.0352 (***0.1213***)	0.0351(***0.1199***)
		10	-0.1103 (***0.1909***)	0.1101 (***0.1825***)	0.1102 (***0.1838***)	0.0502 (***0.1806***)
		J-test	0.0634	0.0615	0.0614	0.0616
	$\gamma = 0.9$	0.5	-0.0104 (***0.0919***)	-0.0102(***0.0916***)	-0.0101 (***0.0833***)	-0.0101 (***0.0815***)
		1	0.0234 (***0.0921***)	0.0205 (***0.0912***)	0.0233 (***0.0896***)	0.0233 (***0.0897***)
		2	0.0534 (***0.0938***)	0.0536 (***0.0931***)	0.0519 (***0.0918***)	0.0506 (***0.0910***)
		5	0.0793(***0.0951***)	0.0794 (***0.0948***)	0.0791 (***0.0950***)	0.0792 (***0.0941***)
		10	0.0898 (***0.0967***)	0.0798(***0.0958***)	0.0799 (***0.0868***)	0.0698 (***0.0866***)
		J-test	0.0894	0.0789	0.0851	0.0788
Two lags only IVs	$\gamma = 0.2$	0.5	-0.0318(***0.0864***)	-0.0298(***0.0843***)	-0.0267 (***0.0842***)	-0.0258 (***0.0841***)
		1	-0.0445 (***0.1052***)	-0.0445 (***0.1051***)	-0.0444 (***0.1043***)	-0.0444 (***0.1041***)
		2	-0.0392 (***0.1211***)	-0.0391 (***0.1194***)	-0.0372 (***0.1204***)	0.0372 (***0.1113***)
		5	0.0042 (***0.1540***)	0.0340 (***0.1353***)	0.0440 (***0.1449***)	0.0039 (***0.1354***)
		10	0.0907 (***0.2127***)	0.0901 (***0.2040***)	0.0802 (***0.2005***)	0.0800 (***0.1913***)
		J-test	0.0580	0.0559	0.0618	0.0559
	$\gamma = 0.9$	0.5	-0.0106 (***0.0929***)	-0.0105 (***0.0913***)	-0.0107 (***0.0918***)	-0.0105 (***0.0913***)
		1	0.0232 (***0.0956***)	0.0231 (***0.0951***)	0.0229(***0.0953***)	0.0231 (***0.0942***)
		2	0.0534 (***0.0939***)	0.0535 (***0.0936***)	0.0533 (***0.0926***)	0.0532(***0.0922***)
		5	0.07941 (***0.0956***)	0.0794 (***0.0952***)	0.0792 (***0.0947***)	0.0793 (***0.0946***)
		10	0.0899 (***0.0966***)	0.0739 (***0.0944***)	0.0699 (***0.0767***)	0.0679 (***0.0773***)
		J-test	0.0679	0.0682	0.0761	0.0686
Collapsed IVs	$\gamma = 0.2$	0.5	-0.0379(***0.0937***)	-0.0377 (***0.0915***)	-0.0381 (***.0916***)	-0.0315 (***0.0910***)
		1	-0.0508 (***0.1152***)	-0.0506 (***0.1148***)	-0.0446(***0.1128***)	-0.0478(***0.1084***)
		2	-0.0410 (***0.1359***)	-0.0408 (***0.1357***)	-0.0394 (***0.1282***)	0.0367 (***0.1259***)
		5	0.0148 (***0.1833***)	0.0246 (***0.1810***)	-0.0104 (***0.1794***)	-0.0281 (***0.1724***)
		10	0.1204 (***0.2317***)	0.1195 (***0.2329***)	0.0477 (***0.2217***)	0.0374(***0.2064***)
		J-test	0.0583	0.0565	0.0564	0.0562
	$\gamma = 0.9$	0.5	0.0067 (***0.0824***)	0.0066 (***0.0818***)	0.0065 (***0.0814***)	0.0105 (***0.0809***)
		1	0.0363 (***0.0950***)	0.0363(***0.0950***)	0.0104 (***0.0955***)	0.0235(***0.0931***)
		2	0.0624 (***0.0959***)	0.0624 (***0.0948***)	0.0435 (***0.0857***)	0.0433 (***0.0775***)
		5	0.0838 (***0.0972***)	0.0838 (***0.0971***)	0.0741 (***0.0920***)	0.0743 (***0.0912***)
		10	0.0925 (***0.0983***)	0.0824 (***0.0873***)	0.0768 (***0.0834***)	0.0727 (***0.0785***)
		J-test	0.0680	0.0559	0.0628	0.0577

Table 5.4: **Mean absolute bias** and **RMSE** of various GMM estimators based on original (**untransformed**), **lag limited** (the first two lags) and **collapsed** instruments from **top** to **bottom**, respectively, ($T = 10$) & ($n = 100$)

.

			SYS_Conv		SYS_Sub	
		ρ	SYS_Conv1	SYS_Conv2	SYS_Sub1	SYS_Sub2
	$\gamma = 0.2$	0.5	-0.0502 (***0.0729***)	-0.0562 (***0.1543***)	-0.0500 (***0.0720***)	0.0399 (***0.0608***)
		1	-0.0261 (***0.0632***)	-0.0237 (***0.0588***)	-0.0219 (***0.0611***)	-0.0227 (***0.0614***)
		2	0.0235 (***0.0692***)	0.0239 (***0.0693***)	0.0236 (***0.0629***)	0.0238 (***0.0633***)
		5	0.1426 (***0.1664***)	0.1466 (***0.1618***)	0.1411 (***0.1590***)	0.0543 (***0.1275***)
		10	0.1820 (***0.2012***)	0.1818 (***0.1988***)	0.1810 (***0.1982***)	0.1090 (***0.1486***)
		J-test	0.4869	0.4829	0.4321	0.4301
Untransformed IVs						
	$\gamma = 0.9$	0.5	0.0080 (***0.0368***)	0.01150 (***0.0353***)	0.0109 (***.0356***)	0.0111 (***0.0360***)
		1	0.0447 (***0.0509***)	0.0449 (***0.0505***)	0.0443 (***0.0508***)	0.0449(***0.0493***)
		2	0.0673(***0.0694***)	0.0674(***0.0691***)	0.0663 (***0.0682***)	0.0664 (***0.0680***)
		5	0.0840 (***0.0846***)	0.0830 (***0.0836***)	0.0740 (***0.0749***)	0.0751 (***0.0768***)
		10	0.0907 (***0.0910***)	0.0907 (***0.0910***)	0.0904 (***0.0908***)	0.0813(***0.0906***)
		J-test	0.4527	0.4506	0.4458	0.4447
	$\gamma = 0.2$	0.5	-0.0163 (***0.0630***)	-0.0159 (***0.0606***)	-0.0157 (***0.0603***)	0.0157 (***0.0594***)
		1	0.0232 (***0.0712***)	0.0236 (***0.0674***)	0.0221 (***0.0653***)	0.0235(***0.0648***)
		2	0.0932 (***0.1193***)	0.0931 (***0.1165***)	0.0934 (***0.1157***)	0.0935 (***0.1169***)
		5	0.1440 (***0.1652***)	0.1421 (***0.1598***)	0.1433 (***0.1601***)	0.1435 (***0.1596***)
		10	0.2880 (***0.2986***)	0.2890 (***0.2988***)	0.2872 (***0.2962***)	0.1869 (***0.2968***)
		J-test	0.3754	0.3658	0.3076	0.3050
Two lags only IVs						
	$\gamma = 0.9$	0.5	0.0320 (***0.0428***)	0.0322 (***0.0426***)	0.0310 (***0.0431***)	0.0322(***0.0419***)
		1	0.0581 (***0.0619***)	0.0583 (***0.0618***)	0.0562 (***0.0622***)	0.0583 (***0.0607***)
		2	0.0753 (***0.0769***)	0.0754 (***0.0771***)	0.0728 (***0.0742***)	0.0745 (***0.0755***)
		5	0.0879 (***0.0885***)	0.0880 (***0.0886***)	0.0779 (***0.0787***)	0.0850 (***0.0856***)
		10	0.0934 (***0.0937***)	0.0935 (***0.0937***)	0.0834 (***0.0838***)	0.0835 (***0.0837***)
		J-test	0.2839	0.2830	0.2850	0.2840
	$\gamma = 0.2$	0.5	-0.0176 (***0.0649***)	-0.0174 (***0.0634***)	-0.0175 (***0.0628***)	-0.0173 (***0.0639***)
		1	0.0349 (***0.0778***)	0.0352 (***0.0636***)	0.0330 (***0.0402***)	-0.016 (***0.0534***)
		2	0.1180 (***0.1406***)	0.1182 (***0.1408***)	0.1177 (***0.1362***)	-0.0053 (***0.1097***)
		5	0.1762 (***0.1941***)	0.1761 (***0.1942***)	0.1735 (***0.1914***)	0.0128 (***0.1357***)
		10	0.2152 (***0.2260***	0.2150 (***0.2264***)	0.2142 (***0.2205***)	0.0464 (***0.1623***)
		J-test	0.2064	0.1839	0.1514	0.1380
Collapsed IVs						
	$\gamma = 0.9$	0.5	0.0387(***0.0466***)	0.0363 (***0.0449***)	0.0365 (***0.0445***)	0.0368 (***0.0443***)
		1	0.0612 (***0.0644***)	0.0614 (***0.0643***)	0.0611 (***.0637***)	0.0612 (***0.0645***)
		2	0.0766 (***0.0781***)	0.0767 (***0.0782***)	0.0625 (***0.0645***)	0.0674 (***0.0683***)
		5	0.0882 (***0.0884***)	0.0873 (***0.0879***)	0.0872 (***0.0878***)	0.0782 (***0.0789***)
		10	0.0930 (***0.0933***)	0.0931 (***0.0932***)	0.0833 (***0.0833***)	0.0838 (***0.0841***)
		J-test	0.1519	0.1513	0.2018	0.1954

Table 5.5: **Mean absolute bias** and **RMSE** of various GMM estimators based on original (**untransformed**), **lag limited** (the first two lags) and **collapsed** instruments from **top** to **bottom**, respectively, ($T = 20$) & ($n = 100$)

.

		SYS_Conv		SYS_Sub	
	ρ	SYS_Conv1	SYS_Conv2	SYS_Sub1	SYS_Sub2
Untransformed IVs, $\gamma = 0.2$	0.5	-0.0960 (***0.1016***)	-0.0835 (***0.1021***)	-0.0970 (***0.1020***)	-0.0799 (***0.0966***)
	1	-0.1028 (***0.1067***)	-0.1221 (***0.1274***)	-0.1021 (***0.1068***)	-0.1121 (***0.1178***)
	2	-0.0872 (***0.0929***)	-0.0919 (***0.1622***)	-0.0864 (***0.1033***)	-0.0920 (***0.1220***)
	5	-0.0794 (***0.0923***)	-0.0788 (***0.0868***)	-0.0782 (***0.0927***)	-0.0758 (***0.1281***)
	10	0.0876 (***0.0872***)	-0.0977(***0.1205***)	0.0378 (***0.0870***)	-0.0889 (***0.1295***)
	J-test	0.8364	0.7950	0.8002	0.7641
Untransformed IVs, $\gamma = 0.9$	0.5	-0.0871 (***0.0991***)	-0.0859 (***0.0988***)	-0.0859 (***0.0987***)	-0.0858 (***0.0389***)
	1	-0.0225 (***0.0429***)	-0.0216 (***0.0438***)	-0.0217 (***0.0439***)	-0.0216 (***0.0446***)
	2	0.0277 (***0.0376***)	0.0282 (***0.0391***)	0.0281 (***0.0391***)	0.0271 (***0.0393***)
	5	0.0335 (***0.0824***)	0.0676 (***0.0694***)	0.0193 (***0.0823***)	0.0675 (***0.0697***)
	10	0.0828 (***0.0834***)	0.0829 (***0.0836***)	0.0829 (***0.0836***)	0.0875 (***0.0830***)
	J-test	0.8251	0.8103	0.7933	0.7821
Two lags only IVs, $\gamma = 0.2$	0.5	-0.0554 (***0.0570***)	-0.0051 (***0.0591***)	-0.0552 (***0.0591***)	-0.0049 (***0.0583***)
	1	0.0711 (***0.0967***)	0.0712 (***0.1001***)	0.0712 (***0.0991***)	0.0710 (***0.0966***)
	2	0.1051 (***0.1113***)	0.0950 (***0.1221***)	0.1011 (***0.1513***)	0.0943 (***0.1236***)
	5	0.1239 (***0.1291***)	0.1134 (***0.1333***)	-0.0652 (***0.1284***)	0.1129 (***0.1364***)
	10	0.0650 (***0.1020***)	0.0663 (***0.0847***)	0.0661 (***0.0990***)	0.0656 (***0.0893***)
	J-test	0.4873	0.4508	0.3298	0.3207
Two lags only IVs, $\gamma = 0.9$	0.5	0.0712 (***0.1114***)	0.0912 (***0.0914***)	0.0612 (***0.1114***)	0.0879 (***0.0886***)
	1	0.0858 (***0.1259***)	0.0712 (***0.0996***)	0.0814 (***0.1215***)	0.0710 (***0.0965***)
	2	0.0859 (***0.0813***)	0.0788 (***0.0936***	0.0958 (***0.1259***)	0.0699 (***0.0870***)
	5	0.0968 (***0.1378***)	0.0958 (***0.0995***)	0.0948 (***0.1263***)	0.0960 (***0.0961***)
	10	0.1160 (***0.1059***)	0.1217 (***0.1219***)	0.0959 (***0.1260***)	0.0849 (***0.0850***)
	J-test	0.46209	0.4533	0.4110	0.3856
Collapsed IVs, $\gamma = 0.2$	0.5	-0.0556 (***0.0712***)	-0.0554 (***0.0712***)	-0.0552 (***0.0718***)	-0.0534 (***0.0761***)
	1	-0.0392 (***0.0646***)	0.0689 (***0.1061***)	-0.0388 (***0.0660***)	0.0581 (***0.0957***)
	2	-0.0547 (***0.0614***)	0.0689 (***0.0988***)	-0.0443 (***0.0642***)	0.0531 (***0.0956***)
	5	0.0885 (***0.1162***)	0.0692 (***0.0988***)	0.0819 (***0.1192***)	0.0609 (***0.1076***)
	10	0.1074 (***0.1231***)	0.0775 (***0.1195***)	0.1074 (***0.1080***)	-0.0632 (***0.0889***)
	J-test	0.2395	0.2201	0.2108	0.2067
Collapsed IVs, $\gamma = 0.9$	0.5	0.0238 (***0.0380***)	0.0236 (***0.0381***)	0.0235 (***0.0384***)	-0.0253(***0.0614***)
	1	0.0522 (***0.0574***)	0.0521 (***0.0575***)	0.0520 (***0.0576***)	0.0074(***0.0591***)
	2	0.0724 (***0.0747***)	0.0723 (***0.0747***)	0.0723 (***0.0747***)	0.0427 (***0.0605***)
	5	0.0874 (***0.0883***)	0.0723 (***0.0747***)	0.0874 (***0.0883***)	0.0427 (***0.0605***)
	10	0.0932 (***0.0937***)	0.0932 (***0.0936***)	0.0818 (***0.1057***)	0.0891 (***0.0899***)
	J-test	0.2305	0.2281	0.2155	0.2098

Table 5.6: **Mean absolute bias** and **RMSE** of alternative **one-step** GMM estimators based on original (**untransformed**), **lag limited** and **collapsed** instruments from **left** to **right**, respectively, ($T = 5$) & ($n = 300$)

	ρ	SYS_Conv1	SYS_subU1	SYS_lag1	SYS_Subcol1
$\gamma = 0.2$	0.5	0.0030 (*0.0467*)	0.0030 (*0.0466*)	0.0028 (*0.0519*)	0.0028 (*0.0560*)
	1	0.0111 (*0.0523*)	0.0110(*0.0518*)	0.0108 (*0.0530*)	0.0102 (*0.0672*))
	5	0.0513 (*0.0935*)	0.0511 (*.0932*)	0.0499 (*0.0963*)	0.0407 (*0.0929*)
	10	0.1061 (*0.1350*)	0.1058 (*0.1343*)	0.1039 (*0.15213*)	0.0638 (*0.1267*)
	J-test	0.0849	0.07503	0.0743	0.0705

Table 5.7: **Mean absolute bias** and **RMSE** of alternative **two-step** GMM estimators based on original (**untransformed**), **lag limited** and **collapsed** instruments from **left** to **right**, respectively, ($T = 5$) & ($n = 300$)

	ρ	SYS_Conv2	SYS_subU2	SYS_lag2	SYS_Subcol2
$\gamma = 0.2$	0.5	0.0030 (*0.0467*)	0.0029 (*0.0432*)	.0031 (*0.0571*)	0.0026 (*0.0650*)
	1	0.0111 (*0.0522*)	0.0108 (*0.0503*)	0.0096 (*0.0577*)	0.0063 (*0.0750*))
	5	0.0511 (*0.0836*)	0.0506 (*0.0820*)	0.0509 (*0.0854*)	0.0376 (*0.0865*)
	10	0.1057 (*0.1351*)	0.1026 (*0.1309*)	0.1078 (*0.1400*)	0.0539 (*0.1103*)
	J-test	0.0847	0.0742	0.0738	0.0.0702

Table 5.8: **Mean absolute bias** and **RMSE** of alternative **one-step** GMM estimators based on original (**untransformed**), **lag limited** and **collapsed** instruments from **left** to **right**, respectively, ($T = 10$) & ($n = 300$)

	ρ	SYS_Conv1	SYS_subU1	SYS_lag1	SYS_Subcol1
$\gamma = 0.2$	0.5	0.0300 (*0.0465*)	0.0300 (*0.0560*)	0.0211 (*0.0564*)	0.0210 (*0.0567*)
	1	0.0308 (*0.0473*)	0.0316 (*0.0560*)	0.0310 (*0.0516*)	0.0214 (*0.0527*))
	5	0.0347 (*0.0489*)	0.0339 (*0.0560*)	0.0335 (*0.0639*)	0.0221 (*0.0739*)
	10	0.0453 (*0.0511*)	0.0365 (*0.0560*)	0.044 (*0.0864*)	0.0251 (*0.9031*)
	J-test	0.1429	0.1416	0.1208	0.1073

Table 5.9: **Mean absolute bias** and **RMSE** of alternative **two-step** GMM estimators based on original (**untransformed**), **lag limited** and **collapsed** instruments from **left** to **right**, respectively, ($T = 10$) & ($n = 300$)

	ρ	SYS_Conv2	SYS_subU2	SYS_lag2	SYS_Subcol2
$\gamma = 0.2$	0.5	0.0302 (*0.0465*)	0.0246 (*0.0445*)	0.0211 (*0.0558*)	0.0209 (*0.0551*)
	1	0.0371 (*0.0483*)	0.0296 (*0.0447*)	0.0211 (*0.0558*)	0.0216(*0.0632*))
	5	0.0345 (*0.0597*)	0.0336(*0.0514*)	0.0211 (*0.0558*)	0.0226(*0.0820*)
	10	0.0395 (*0.0628*)	0.0345 (*0.0536*)	0.0211 (*0.0558*)	0.0244(*0.0881*)
	J-test	1418	0.1414	0.1072	0.0911

Table 5.10: **Percentage (%)** confidence intervals that covered γ for various system GMM estimator, namely, conventional system GMM **(SYS_Conv)**, sub optimal and based on lag limited instruments system GMM **(Sub_lag)**, and sub-optimal and using collapsed instruments system GMM **(Sub_col)** from left to right, respectively, ($T = 5$) & ($n = 100$) .

	ρ	SYS_Conv	Sub_lag	Sub_col
$\gamma = 0.2$	0.5	0.8572	0.8530	0.8567
	1	0.8646	0.8592	0.8623
	5	0.9404	0.9488	0. 9328
	10	0.9510	0.9502	0.9510
$\gamma = 0.9$	0.5	0.8420	0.8422	0.8494
	1	0.8562	0..8564	0.8604
	5	0.8752	0.8754	0.8768
	10	0.8782	0.8780	0.8788

Table 5.11: **Percentage** (%) confidence intervals that covered γ for various system GMM estimator, namely, conventional system GMM **(SYS_Conv)**, sub optimal and based on lag limited instruments system GMM **(Sub_lag)**, and sub-optimal and using collapsed instruments system GMM **(Sub_col)** from left to right, respectively, ($T = 10$) & ($n = 100$)
.

	ρ	SYS_Conv	Sub_lag	Sub_col
$\gamma = 0.2$	0.5	0.8316	0.8330	0.8354
	1	0.8896	0.8996	0.8946
	5	0.9074	0.9318	0. 9393
	10	0.9112	0.9394	0.9356
$\gamma = 0.9$	0.5	0.8468	0.8572	0.8596
	1	0.8616	0.8680	0.8680
	5	0.8750	0.8762	0.8768
	10	0.8770	0.8783	0.8782

Table 5.12: **Percentage** (%) confidence intervals that covered γ for various system GMM estimator, namely, conventional system GMM **(SYS_Conv)**, sub optimal and based on lag limited instruments system GMM **(Sub_lag)**, and sub-optimal and using collapsed instruments system GMM **(Sub_col)** from left to right, respectively, ($T = 20$) & ($n = 100$)
.

	ρ	SYS_Conv	Sub_lag	Sub_col
$\gamma = 0.2$	0.5	0.7831	0.8550	0.8504
	1	0.8020	0.9464	0.9374
	5	0.8444	0.9428	0. 9426
	10	0.8556	0.9446	0.9503
$\gamma = 0.9$	0.5	0.7980	0.8366	0.8328
	1	0.8339	0.8856	0.8878
	5	0.8704	0.8870	0.8734
	10	0.8750	0.8868	0.8855

Table 5.13: **Estimates for employment equation, (2010 − 2019)**

	OLS	FE	DIF	SYS_Conv	Sub_full	Sub_lag	Sub_col
$y_{i,t-1}$	0.852	0.0432	0.0864	0.7802	0.7882	0.7648	0.7865
	(0.0128)	(0.0238)	(0.0719)	(0.0637)	(0.0516)	(0.0761)	(0.0608)
w_{it}	0.5092	0.2568	0.4310	0.5340	0.5291	0.4809	0.5062
	(0.0107)	(0.0294)	(0.0720)	(0.0530)	(0.0508)	(0.0666)	(0.0529)
$w_{i,t-1}$	-0.0014	-0.2004	-0.1910	-0.1202	-0.1042	-0.1130	-0.1008
	(0.0114)	(0.0194)	(0.0706)	(0.0531)	(0.0499)	(0.0682)	(0.0524)
k_{it}	0.0518	0.0108	0.0189	0.0308	0.0317	0.0445	0.0480
	(0.0082)	(0.0073)	(0.0689)	(0.0441)	(0.0415)	(0.0584)	(0.0551)
$k_{i,t-1}$	-0.0046	-0.0167	0.0109	-0.0102	-0.0100	-0.0114	-0.0098
	(0.0081)	(0.0071)	(0.0610)	(0.0469)	(0.0428)	(0.0480)	(0.0461)
m_1	-	-	-9.3207	-6.5911	-6.6201	-7.8140	-10.0087
m_2	-	-	-1.5674	-2.3671	-2.3085	-2.2760	-1.8944
J-test	-	-	0.2475	0.3187	0.2920	0.2733	0.2219
β_w	0.5589	0.4529	0.4647	0.5030	0.5151	0.5048	0.5219
	(0.0358)	(0.0388)	(0.0623)	(0.0595)	(0.0583)	(0.0611)	(0.0603)
β_k	0.0532	0.0110	0.0493	0.0754	0.0791	0.0637	0.0770
	(0.0326)	(0.0469)	(0.0508)	(0.0381)	(0.0375)	(0.0432)	(0.0418)
ϕ	0.5144	-0.0442	-0.3868	0.3048	0.2933	0.3001	0.3200
	(0.0342)	(0.0399)	(0.0488)	(0.0332)	(0.0317)	(0.0391)	(.0346)
Comfac.	0.0012	0.0092	0.0187	0.1098	0.1164	0.2153	0.1676

Comfac.: represents the p-values of common factor restrictions the under the null common factor restrictions are valid.

Table 5.14: **Estimates for employment equation,** **(**$2014 - 2019$**)**

	OLS	FE	DIF	SYS_Conv	Sub_full	Sub_lag	Sub_col
$y_{i,t-1}$	0.8619	0.0289	0.0728	0.7713	0.7806	0.7620	0.7812
	(0.0164)	(0.0296)	(0.0795)	(0.0766)	(0.0749)	(0.1262)	(0.0784)
w_{it}	0.4988	0.2390	0.3684	0.5227	0.5106	0.4611	0.5022
	(0.0139)	(0.0132)	(0.0745)	(0.0782)	(0.0760)	(0.1136)	(0.0799)
$w_{i,t-1}$	-0.0023	-0.2061	-0.1933	-0.1216	-0.1070	-0.1205	-0.1044
	(0.0148)	(0.0208)	(0.0760)	(0.0698)	(0.0894)	(0.1745)	(0.0962)
k_{it}	0.0486	0.0102	0.0163	0.0112	0.0415	0.0335	0.0401
	(0.0107)	(0.0102)	(0.0764)	(0.0623)	(0.0623)	(0.0917)	(0.0701)
$k_{i,t-1}$	-0.0050	-0.0235	-0.0218	-0.0137	-0.0129	-0.0132	-0.0125
	(0.0106)	(0.0104)	(0.0651)	(0.0632)	(0.0638)	(0.1216)	(0.0719)
m_1	-	-	-5.0822	-4.1720	-4.0663	-4.9001	-5.4330
m_2	-	-	-2.9520	-3.7764	-3.1069	-2.9982	-3.0556
J-test	-	-	0.2475	0.3187	0.2920	0.2733	0.2219
β_w	0.5413	0.4453	0.4609	0.4844	0.5070	0.4992	0.5166
	(0.03090)	(0.0386)	(0.0853)	(0.0765)	(0.0722)	(0.0826)	(0.0770)
β_k	0.0511	0.0095	0.0442	0.0726	0.0733	0.0618	0.0750
	(0.0483)	(0.0491)	(0.0722)	(0.0690)	(0.0668)	(0.0705)	(0.0728)
ϕ	0.5032	-0.0501	-0.3891	0.3005	0.2743	0.2889	0.3115
	(0.0519	(0.0433)	(0.0510)	(0.0418)	(0.0375	(0.0433)	(0.0410)
Comfac.	0.0060	0.0092	0.0187	0.0798	0.0997	0.1866	0.1305

APPENDIX II: R Codes

```
R Codes example for computing alternative
System GMM estimators

Definition of basic R codes or abbrevations

yit_1: Dependent variable for dynamic panel data model
gamma: value of lagged coefficient
D.T_D.T: First differnce operator matrix for diffrence estimator
ZD: Instrument sets for the diffrence estimator
ZL: Instrument sets for level estimator
ZS: Instrument sets for system GMM estimator

Fc_T (T= 5,.., 10):Transformation matrix into
                   collapsed instruments
ZDC:  Collapsed instrument sets for differnce equation

WD1C: The initial weight matix for differnce estimator
         using collapsed instruments

ZSC: Collapsed instrument sets for system GMM estimator
WS1C:Initial weight matrix  for conventional system GMM
      using collapsed instruments
WS1_JC: Sub optimal initial weight matrix for system GMM
        using collapsed instruments

L2_T(T=5,...,10): Transformation matrix into
                  lag limited instruments

ZDL2: Lag limited instrument sets (the recent two lags)
         for differnced equation
```

```
ZSL2: Lag limited instrument sets for system GMM

WD1L2: Initial weight matix for diffrence estimator
           using lag limited instruments

WS1L2:  Initial weight matrix  for conventional
           system GMM using lag limited instruments

WS1_JL2: Initial weight matrix for sub optimal
          system GMM using lag limited instruments

alpha.i:Unoberved individual effects
sd_alpha: Standard deviation of unoberved individual effects
epsilon.it: Random shocks
sd_epsilon: Standard deviation of random shocks

required R libraries

library(simex);library(plm); library(MASS)

C.1: R codes to compute conventional and
     suboptimal system GMM estimators
    using collapsed instruments

          rm (list = ls())

          set.seed(111)

yit_1<- function(n, T, nT = n*T, T1=T+1, nT_2=n*(T-2),
         gamma, sd_alpha , sd_epsilon){

 # Generating  alpha , epsilon and uit

 alpha.i <- rnorm (n, 0, sd_alpha);
 alpha.it <- rep(alpha.i, each = T)
 epsilon.it <- rnorm (nT,0, sd_epsilon)
    u.it <- alpha.it+epsilon.it;

 ## calculate yit_1
```

```
 dim(u.it) <-c(T,n)
 u.it<-rbind(0, u.it); dim(u.it)
 yit <-matrix(0, nrow=T1,ncol=n)

 for(j in 1:n){
  yit[1,j]<- (alpha.i[j]/(1- gamma))+rnorm (1,0, sd_alpha)
  for (i in 2:T1){
   yit[i,j]<- gamma*yit[i-1,j]+u.it[i,j]
     }
     }

 y_full<-yit[-1,]; dim(y_full)<-c(nT,1)
 yit_1 <-yit[-1,] ;dim(yit_1)
 y_model <- yit[4:T1,]
 dim(y_model)<-c(nT_2,1)
 return(yit_1)
 }
   yit_1 <- yit_1()

##  calculate D for any  T

DF <- function (n , T ){
 D.T <-diag(-1, nrow=T-1,ncol=T)
 for (i in 1:T-1) D.T [i,(i+1)]<- 1
 D.T_correct <- D.T [1:(T-2),-T]
 D.T_D.T <- D.T_correct %*% t(D.T_correct)
 D<- kronecker (diag(1,n), D.T_D.T )
 Dlist<- list(D.T_D.T= D.T_D.T , D= D)
 return(D)}
D<- DF(n =100 , T =5)

## calculate D.T_D.T for any  T

DF <- function (n , T ){
 D.T<-diag(-1, nrow=T-1,ncol=T)
 for (i in 1:T-1) D.T [i,(i+1)]<- 1
 D.T_correct <- D.T [1:(T-2),-T]
 D.T_D.T <- D.T_correct %*% t(D.T_correct)
```

```
 D<- kronecker (diag(1,n), D.T_D.T)
 Dlist<- list(D.T_D.T= D.T_D.T, D= D)
 return(D.T_D.T)}
D.T_D.T<- DF(n =n, T = T)

##  Transformation Matrix for collapsed IVS
## for T = 5
(zi_5<- matrix(c(0,1,0,0,0,0,2,0,0,0,1,0,0,0,0,
3,0,0,0,2,0,0,0,1),ncol = 6))
(Fc_5 = matrix(c(1,1,0,1,0, 0,0,0,1,0,1,0,
0,0,0,0,0,1), ncol = 3))

##  For T = 6
(Fc_6 = matrix(c(1,1,0,1,0, 0,1,0,0,0,0,0,1,0,1,
0,1,0,0, 0,0,0,0,0,1,0,0,1,0,0,
0,0,0,0, 0,0,0,0,1), ncol = 4))

(zi_6 = matrix(c(0,1,0,0,0, 0,0,2,0,0,0,0,1,0,0,0,
0,0, 3,0, 0,0,0,2,0,0, 0, 0, 1,0,
0,0,0,0,4, 0, 0,0,0,3, 0,0,0,0,2,
0,0,0,0,1),ncol = 10))

## For T = 7
(Fc = matrix(c(1,1,0,1,0, 0,1,0,0,0,0,0,1,0,1, 0,0,1,
0,0, 0,0,0,0,0,1,0,0,1,0, 0,0,0, 0,0,
0,0,0,0,1,0,0,0,0,0,0,0,0,0,0), ncol = 5))

Id_5 <- diag(5);(Fc_7 <- rbind(Fc,Id_5));Fc_7

## For T= 8
(Fc = matrix(c(1,1,0,1,0, 0,1,0,0,0,0,0,1,0,1, 0,0,1,
0,0, 0,0,0,0,0,1,0,0,1,0, 0,0,0, 0,0,
0,0,0,0,1,0,0,0,0,0,0,0,0,0,0),ncol = 5))

(Fc_7 <- rbind(Fc,Id_5)) ;
(y7c<- matrix(c(0,0,0,0,0,0, 0,0,0,0,0,0,0,0,0),
ncol = 1))
(Fc_8i<- cbind(Fc_7,y7c));Id_6<- diag(6)
(Fc_8<- rbind(Fc_8i, Id_6))
```

```
## For T= 9
(y8c<- matrix(c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0), ncol = 1))

(Fc_9i<- cbind(Fc_8,y8c)) ; Id_7<- diag(7)
(Fc_9<- rbind(Fc_9i, Id_7))

## For T= 10
(y9c<- matrix(c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0, 0,0,0,0,0,0,0),
ncol = 1))
(Fc_10i<- cbind(Fc_9,y9c)) ; Id_8<- diag(8)
(Fc_10<- rbind(Fc_10i, Id_8))


##  ZD function = instruments for differenced equation

ZD.matrix <- function (yit_1){
 n<-ncol(yit_1);
 T<-nrow(yit_1)

 ## calculate delta y

 delta_y<- apply(yit_1[(2:T),],2,diff)
 dim(delta_y)<-c((T-2)*n,1)
 delta_y_1<- apply(yit_1[(1:T-1),],2,diff)
 dim(delta_y_1)<-c((T-2)*n,1)

 ## calculate ZD: calculate list_for_Z.1

 list_for_Z.1<-list()
 for (z in 1: (T-2)) {
  list_for_Z.1[[z]]<- c( yit_1[ (1:z) , 1])
 }
 Z.1<- diag.block (list_for_Z.1)

 ## calculate ZD
 Z<- Z.1
```

```
 for (i in 2:n){
  Z.i<-list(0)
  for (z in 1: (T-2)) {
   Z.i [[z]]<- c( yit_1 [ (1:z) , i])
  }
  Z<- cbind(Z,diag.block (Z.i))
 }
 ZD<- t(Z)
 return(ZD)
}

## ZD = untransformed instruments for differenced equation
ZD<-ZD.matrix(yit_1)
## ZDC = collapsed instruments for differenced  equation
ZDC <- ZD%*%Fc_T
## ZL function = Instruments for level equation

ZL.matrix <- function (yit_1){
 n <-ncol(yit_1); T<-nrow(yit_1)
 delta_y_1<- apply(yit_1[(1:T-1),],2,diff)

## calculate ZL
 ZL<-rbind(matrix(diag (delta_y_1[,1]),
 nrow=( T-2), ncol=(T-2)))
 for (j in 2: n)
 ZL<- rbind(ZL, diag(delta_y_1[,j]))
 return(ZL)}
ZL<-ZL.matrix(yit_1)

## ZSC = IVs for system based on collapsed_IVs

ZS.matrix_C <- function (yit_1, ZDC,ZL ){

 n<- ncol(yit_1)
 T<- nrow(yit_1)

 ##  calculate ZS: f factor for any Z_i
 f<- matrix (0,nrow= n+1, ncol=1)
 for (j in 1 : n) f[j+1,1]<- j* (T-2)
```

```
 zs_list<-list()
 zs_list[[1]]<- ZDC[(f[1,1]+1) : f[1+1,1],]
 zs_list[[2]]<- ZL[(f[1,1]+1) : f[1+1,1],]

 ZSC <-diag.block(zs_list)

 ## from 2 to n

 for (j in 2 : n){
  zs_list<-list()
  zs_list [[1]]<- ZDC[(f[j,1]+1) : f[j+1,1],]
  zs_list [[2]]<- ZL[(f[j,1]+1) : f[j+1,1],]
  ZSC <-rbind(ZSC, diag.block (zs_list)) }
 return(ZSC)}
ZSC<-  ZS.matrix_C(yit_1,ZDC,ZL);ZSC

## Arellano-Bond estimator: based on collapsed_IVs

Arellano.Bond_C<- function (yit_1, gamma,
D.T_D.T, ZDC) {

 n<-ncol(yit_1);T<-nrow(yit_1)
 ## calculate delta_y
 delta_y<- rbind(apply(yit_1[(2:T),],2,diff));
 dim(delta_y)<-c((T-2)*n,1)
 delta_y_1<- rbind(apply(yit_1[(1:T-1),],2,diff));
 dim(delta_y_1)<-c((T-2)*n,1)

 ## calculate D
 D<- kronecker(diag(1, n), D.T_D.T)
## calculate WD1C
 WD1C<- t(ZDC)%*%D%*%ZDC
 ## calculate  gamma.hat step (1)
 part1C <- solve(t(delta_y_1)%*% ZDC %*%
 solve(WD1C)%*%t(ZDC)%*% delta_y_1)
 part2C <- t(delta_y_1)%*%ZDC%*%solve(WD1C)
 %*% t(ZDC)%*% delta_y

 gamma.hat.step1C<- part1C%*% part2C
```

```
 dim(gamma.hat.step1C) <- NULL
 Bias.ABond.step1C<- gamma.hat.step1C-gamma
 SE.gamma.hat.step1C<-sqrt(diag(part1C))
## Two step estimates

 delta_y<- rbind(apply(yit_1[(2:T),],2,diff))
 delta_y_1<- rbind(apply(yit_1[(1:T-1),],2,diff))
 delta_yit.hat1C<- gamma.hat.step1C*delta_y_1

## calculate  delta_epsilon.hat

 delta_epsilon.hat.iC1<- delta_y- delta_yit.hat1C

## calculate sigma.2.epsilon.hat

 sigma.2.epsilon.hat.ABC_1<- (sum(unlist(lapply(
 delta_epsilon.hat.iC1, function(x)
 t(x)%*% x)))) /(2*n*(T-2))

 f<- matrix (0,nrow= n+1,ncol=1)
 for (j in 1 : n) f[j+1,1]<- j* (T-2)
 W_sumC<- 0
 for(j in 1 : n){
  W_sumC<- W_sumC+(t(ZDC[(f[j,1]+1): f[j+1,1],])
  %*%delta_epsilon.hat.iC1[,j]%*%
  t(delta_epsilon.hat.iC1[,j])%*%
  (ZDC[(f[j,1]+1) : f[j+1,1],]))
 }
   WD2C <- W_sumC

## Two step estimation

 dim(delta_y)<-c((T-2)*n,1)
 dim(delta_y_1)<-c((T-2)*n,1)

 part1_step2C<- solve(t(delta_y_1)%*%ZDC%*%
 solve(WD2C)%*%t(ZDC) %*% delta_y_1)
 part2_step2C<- t(delta_y_1)%*%ZDC%*%solve(WD2C)
 %*%t(ZDC)%*% delta_y
```

```
 gamma.hat.step2C<- part1_step2C%*%part2_step2C
 dim(gamma.hat.step2C)<- NULL
 Bias.ABond.step2C<- gamma.hat.step2C - gamma
 SE.gamma.hat.step2C<- sqrt(diag(part1_step2C))

## Conventional SYS GMM for "one sample"

  SYSC<- function (yit_1, gamma, HI, D.T_D.T,
   ZDC, ZL, SYSC){

  n<-ncol(yit_1);
  T<-nrow(yit_1)
  ## calculate y_s and y_s_1

  y <- yit_1[(3:T),]
  y_1<- yit_1 [(2:(T-1)),]
  delta_y_1 <- apply(yit_1 [(1:T-1),],2,diff)
  delta_y <- apply(yit_1 [(2:T),],2,diff)
  y_s <- rbind(delta_y,y)
  dim(y_s) <-c( 2*(T-2)*n,1)
  y_s_1 <- rbind(delta_y_1,y_1)
  dim(y_s_1) <-c(2*(T-2)*n,1)

## Calculate HI

  Z  S_HI<-list(D.T_D.T , diag(1,T-2));
  HI_T<- diag.block(ZS_HI)
  HI<- kronecker(diag(1, n), HI_T)

## calculate WS1

  WS1C<- t(ZSC) %*% HI%*% ZSC

# One step estimation

  part1_SYS_step1C<- solve(t(y_s_1)%*%ZSC %*%
  solve(WS1C)%*%t(ZSC)%*% y_s_1)
```

```
  part2_SYS_step1C<- t(y_s_1)%*%ZSC%*%
  solve(WS1C)%*% t(ZSC)%*% y_s

  gamma.hat.SYS.step1C<- part1_SYS_step1C%*%
  part2_SYS_step1C

  dim(gamma.hat.SYS.step1C) <- NULL
  Bias.SYS.step1C <- gamma.hat.SYS.step1C - gamma
  SE.gamma.hat.SYS.step1C<- sqrt(diag(part1_SYS_step1C))

  ## Two step estimation : calculate residual u.hat.i_S

  y_s <- rbind(delta_y,y) ;
  y_s_1<- rbind(delta_y_1,y_1)
  yit.hat_SYSC <- gamma.hat.SYS.step1C*y_s_1
  u.hat.i_SC<- y_s - yit.hat_SYSC

## calculate sigma.2.alpha.hat

  delta_u_i_tilde_C <- lapply(u.hat.i_SC,
  function(x) head(x, (T-2)))

  u_i_tilde_C<- lapply(u.hat.i_SC, function(x)
  tail(x, (T - 1)))

  sigma.2.alpha.hat.SYSC_1<- (sum(unlist(lapply(u_i_tilde_C,
  function(x) t(x)%*% x))-(unlist
  (lapply(delta_u_i_tilde_C, function(x)
  t(x) %*% x)) / 2))) / (n*(T-2))

  ##  Compute estimated variance ratio ("hat.rho")

  rc<- sigma.2.alpha.hat.SYSC_1/sigma.2.epsilon.hat.ABC_1

  f<-matrix (0,nrow= n+1,ncol=1)
  for (j in 1 : n) f[j+1,1]<- j*(T-2)

  ## Calculate WS2C
```

```
  zs_list<-list(); S2C<-0
  for (j in 1 : n){
   zs_list [[1]]<- ZDC[(f[j,1]+1) : f[j+1,1],]
   zs_list [[2]]<- ZL[(f[j,1]+1) : f[j+1,1],]
   ZSiC<- diag.block (zs_list)
   WS2C<- WS2C+(t(ZSiC) %*%u.hat.i_SC[,j] %*%
   t(u.hat.i_SC[,j])%*%ZSiC )
  }

  ## Two step  estimation

  dim(y_s_1)<-c(2*(T-2)*n,1); dim(y_s)<-c(2*(T-2)*n,1)
  part1_SYS_step2C<- solve(t(y_s_1)%*% ZSC%*%
  solve(WS2C)%*%t(ZSC)%*%y_s_1)
  part2_SYS_step2C<- t(y_s_1) %*%ZSC %*%
  solve(WS2C) %*% t(ZSC)%*%y_s
  gamma.hat.SYS.step2C<- part1_SYS_step2C%*%
  part2_SYS_step2C

  dim(gamma.hat.SYS.step2C)<- NULL
  Bias.SYS.step2C<-  gamma.hat.SYS.step2C - gamma
  SE.gamma.hat.SYS.step2C<- sqrt(diag(part1_SYS_step2C))

## One step estimation conventional: using R replicates

  R = R ; gamma.hat.SYS.step1C<-numeric(R)

  y_s <- Data_SYSC[,1, drop = FALSE];
  y_s_1<- Data_SYSC[,2, drop = FALSE]
  HSC <- Data_SYSC[, 3:2*(T-1)]
  n1<-nrow(y_s)

for(i in 1:R) {
 ind<-sample(n1, n1, replace = TRUE)
 boot<-  Data_SYSC[ind,];
 WS1C<- t(boot[,3:2*(T-1)])%*%HI%*%boot[,3:2*(T-1)]
  part1_SYS_step1C<- solve(t(boot[,2])%*%
 boot[,3:2*(T-1)] %*% solve(WS1C)%*%
 t(boot[,3:2*(T-1)])%*%boot[,2])
```

```
part2_SYS_step1C<- t(boot[,2]) %*%
boot[,3:2*(T-1)]%*%solve(WS1C)%*%
t(boot[,3:2*(T-1)])%*% boot[,1]

gamma.hat.SYS.step1C[i]<- part1_SYS_step1C%*%
part2_SYS_step1C
}
mean(gamma.hat.SYS.step1C)
(RMSE<- sqrt(mean((gamma.hat.SYS.step1C-gamma)^2)))
sd(gamma.hat.SYS.step1C)
(Bias =  mean(gamma.hat.SYS.step1C)- gamma)

## Suboptimal estimation based collapsed instruments

##  One step: sub-optimal estimates: one sample
n<-ncol(yit_1) ;T<-nrow(yit_1)

## calculate y_s and y_s_1

y <- yit_1 [(3:T),] ;
y_1<- yit_1 [(2:(T-1)),]
delta_y_1<- apply(yit_1 [(1:T-1),],2,diff)
delta_y<- apply(yit_1 [(2:T),],2,diff) ;
y_s<- rbind(delta_y,y) ;
dim(y_s)<-c( 2*(T-2)*n,1)
y_s_1 <- rbind(delta_y_1,y_1)
dim(y_s_1)<-c(2*(T-2)*n,1)

## Calculate H_JC = sub optimal weighting matrix

JC <- diag(1, T - 2)+rc*ones(T - 2)
ZS_H_JC <-list(D.T_D.T , JC);
H_JC_T<- diag.block( ZS_H_JC)
H_JC <- kronecker(diag(1,n), H_JC_T)

## calculate  WS1_JC
WS1_JC<- t(ZSC) %*%H_JC%*% ZSC
## calculate gamma.hat.step (1): sub optimal
```

```
part1_SYS_step1_JC<- solve(t(y_s_1)%*%
ZSC%*%solve(WS1_JC)%*% t(ZSC)%*%y_s_1)

part2_SYS_step1_JC<- t(y_s_1)%*% ZSC %*%
solve(WS1_JC)%*% t(ZSC)%*% y_s

gamma.hat.SYS.step1_JC<- part1_SYS_step1_JC%*%
part2_SYS_step1_JC

dim(gamma.hat.SYS.step1_JC)<- NULL
Bias.SYS.step1_JC<- gamma.hat.SYS.step1_JC - gamma
SE.gamma.hat.SYS.step1_JC<- sqrt(diag(part1_SYS_step1_JC))

## sub-optimal two step  estimates

y_s <- rbind(delta_y,y); y_s_1<- rbind(delta_y_1,y_1)
yit.hat_SYS_JC<- gamma.hat.SYS.step1_JC*y_s_1
u.hat.i_S_JC<- y_s - yit.hat_SYS_JC

## calculate sigma.2.alpha.hat

delta_u_i_tilde_JC <- lapply(u.hat.i_S_JC,
function(x) head(x, (T-2)))
u_i_tilde_JC <- lapply(u.hat.i_S_JC,
function(x) tail(x, (T - 1)))

sigma.2.alpha.hat.SYSC_J1<- (sum(unlist(lapply(u.hat.i_S_JC,
function(x) t(x) %*% x))-(unlist(
lapply(delta_u_i_tilde_JC, function(x)
t(x)%*%x))/2)))/(n*(T-2))

## calculate: WS2_J = weight for two step

f<-matrix (0,nrow= n+1,ncol=1)
for (j in 1 : n) f[j+1,1]<- j*(T-2)
zs_list <-list()
WS2_JC <- 0
for (j in 1 : n){
```

```
zs_list[[1]]<- ZDC[(f[j,1]+1) : f[j+1,1],]
zs_list[[2]]<- ZL[(f[j,1]+1) : f[j+1,1],]
ZSiC<- diag.block (zs_list)
WS2_JC<- AS2_JC+(t(ZSiC)%*%u.hat.i_S_JC[,j]%*%
t(u.hat.i_S_JC[,j])%*%ZSiC)
}

## sub-optimal two step

dim(y_s_1) <-c(2*(T-2)*n,1)
dim(y_s) <-c( 2*(T-2)*n,1)

part1_SYS_step2_JC<- solve(t(y_s_1)%*%ZSC %*%
solve(WS2_JC)%*% t(ZSC)%*% y_s_1)

part2_SYS_step2_JC<- t(y_s_1)%*%ZSC%*%
solve(WS2_JC)%*%t(ZSC)%*%y_s

gamma.hat.SYS.step2_JC<- part1_SYS_step2_JC%*%
part2_SYS_step2_JC

dim(gamma.hat.SYS.step2_JC)<- NULL
Bias.SYS.step2_JC<-  gamma.hat.SYS.step2_JC - gamma
SE.gamma.hat.SYS.step2_JC<-sqrt(diag(part1_SYS_step2_JC))

## sub-optimal estimates: R replications(bootstrap)

## 1) one step  estimation with  R replicates

R = R ; gamma.hat.SYS.step1_JC<-numeric(R)

y_s <- Data_SYSC[,1, drop = FALSE];
y_s_1<- Data_SYSC[,2, drop = FALSE]
ZSC <- Data_SYSC[, 3:2*(T-1)];
n1<-nrow(y_s)

##  sub-optimal weight matrix = WS1_JC

WS1_JC<- t(ZSC) %*% H_JC%*% ZSC
```

```
for(i in 1:R) {
ind<-sample(n1, n1, replace = TRUE)
boot<-  Data_SYSC[ind,]
part1_SYS_step1_JC<- solve(t(boot[,2])%*%
boot[,3:2*(T-1)]%*% solve(WS1_JC)%*%
t(boot[,3:2*(T-1)]) %*% boot[,2])
part2_SYS_step1_JC<- t(boot[,2])%*%boot[,3:2*(T-1)]%*%
solve(WS1_JC) %*%
t(boot[,3:2*(T-1)]) %*% boot[,1]
gamma.hat.SYS.step1_JC[i]<- part1_SYS_step1_JC%*%
part2_SYS_step1_JC
}
mean(gamma.hat.SYS.step1_JC)
(RMSE<- sqrt(mean((gamma.hat.SYS.step1_JC-gamma)^2)))
sd(gamma.hat.SYS.step1_JC)
Bias =  mean(gamma.hat.SYS.step1_JC)- gamma

#  Two step Sub-optimal

R = R;resc_JC <- tcrossprod(u.hat.i_S_JC)

Hres_JC <- kronecker(diag(1,n), resc_JC)
WS_JC2<- t(ZSC)%*%Hres_JC%*%ZSC

gamma.hat.SYS.step2_JC<-numeric(R)
y_s <- Data_SYSC[,1, drop = FALSE];
y_s_1<- Data_SYSC[,2, drop = FALSE]
ZSC <- Data_SYSC[, 3:2*(T-1)]
n1<-nrow(y_s)

for(i in 1:R) {
ind<-sample(n1, n1, replace = TRUE)
boot<-  Data_SYSC[ind,]
part1_SYS_step2_JC<- solve(t(boot[,2])%*%
boot[,3:2*(T-1)]%*%solve(WS_JC2)%*%
t(boot[,3:2*(T-1)]) %*% boot[,2])
part2_SYS_step2_JC<- t(boot[,2])%*%
boot[,3:2*(T-1)]%*% solve(WS_JC2)%*%
```

```
t(boot[,3:2*(T-1)]) %*% boot[,1]
gamma.hat.SYS.step2_JC[i]<- part1_SYS_step2_JC%*%
part2_SYS_step2_JC
}
mean(gamma.hat.SYS.step2_JC)
(Bias =  mean(gamma.hat.SYS.step2_JC)-gamma)
(RMSE<-sqrt(mean((gamma.hat.SYS.step2_JC-gamma)^2)))
sd(gamma.hat.SYS.step2_JC)

C.2 R codes to compute  conventional and
suboptimal system GMM estimators
using lag limited instruments

## Matrix transformation for lagged limit
(the recent 2lags each period)

##  Let for T = 5

I_5 <- diag(5)
y<- matrix(c(0,0,0,0,0), ncol = 5);
L2_5<- rbind(I_5, y)

##  for T=6
(L2_6 = matrix(c(1,0,0,0,0, 0,0,0,0,0, 0,1,0,0,0, 0,0,0,
0,0, 0,0,1,0,0,0,0,0,0,0, 0,0,0,1,0, 0,
0,0,0,0, 0,0,0,0,1, 0,0,0,0,0, 0,0,0,0,0,
0,1,0,0,0, 0,0,0,0,0, 0,0,1,0,0),ncol = 7))

##  For T = 7

(L2_7 = matrix(c(1,0,0,0,0, 0,0,0,0,0, 0,1,0,0,0, 0,0,0,0,0,
0,0,1,0,0,0,0,0,0,0, 0,0, 0,1,0, 0,0,0,0,0,
0,0,0,0,1, 0,0,0,0,0,0,0,0,0,0, 0,1,0,0,0,
0,0,0,0,0, 0,0,1,0,0, 0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0), ncol = 9))

(y<- matrix(c(0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,1, 0,0,0,0,
0,0, 0,0,0,0,0,0,0, 0,0,0, 0,0,0, 0,0,0,0,0,0,
0,0,0), ncol = 9, byrow = TRUE))
```

```
(L2_7<- rbind(L2_7, y))

## For T= 8

(y1<- matrix(c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0, 0,0,0,0,
0,0,0,0,0,0,0,0),ncol = 2, byrow = TRUE))

(l2_8i<- cbind(L2_7,y1))

(y<- matrix(c(0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0, 0,0,0,1,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0, 0,0,0,0,0,0,0,0),
ncol = 11, byrow = TRUE))

(L2_8<- rbind(l2_8i, y))


## For T = 9

(y8<- matrix(c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0, 0,0,0,0,
0, 0,0, 0,0,0,0,0,0,0,0,0, 0,0,0,0, 0,0,0,0,0,0),
ncol = 2, byrow = TRUE))

(L2_9i<- cbind(L2_8,y8))

(y9<- matrix(c(0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0, 0,0,0,0,0,
0,0, 1,0,0,0,0,0,0, 0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0, 0,0, 0,0,0,0,0,0,0, 0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0, 0,0,0,0,0,0,0,0,0,0, 0,0,0,0,0,0,0,0),
ncol = 13, byrow = TRUE))

(L2_9<- rbind(L2_9i, y9))


## For T= 10

(y9<- matrix(c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0, 0,0,0,
0,0,0,0, 0,0,0,0,0,0,0,0,0, 0,0,0,0, 0,0,0,0,
```

```
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0),
ncol = 2, byrow = TRUE))

L2_10i<- cbind(L2_9,y9)
(y10<- matrix(c(0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,
0,0,0, 0,0,0,0,0,1,0,0,0,0, 0,0,0,0,0,0,0,0,
0,0,0,0, 0,0, 0,0, 0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0, 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0, 0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0, 0,0,0,0,0,0,0,0,
0,0,0, 0,0,0,0, 0,0,0,0,0,0,0,0,0,0,0,0),
ncol = 15, byrow = TRUE))

(L2_10<- rbind(L2_10i, y10))

# ZDL2 = Two lagged IVs for differences equation

ZDL2<- ZD%*%L2_T
## ZL:  IVs for Lvel equation as it is
## ZSL2 = lagged IVs for System equation

ZS.matrixL2<- function (yit_1, ZDL2 ,ZL ){

n<-ncol(yit_1)
T<-nrow(yit_1)

## 1- calculate ZS: f factor for any Z_i

f<-matrix (0,nrow= n+1,ncol=1)
for (j in 1 : n) f[j+1,1]<- j*(T-2)
zs_list<-list()
zs_list [[1]]<- ZDL2[(f[1,1]+1) : f[1+1,1],]
zs_list [[2]]<- ZL[(f[1,1]+1) : f[1+1,1],]
ZSL2 <-diag.block(zs_list)

## from 2 to n
for (j in 2 : n){
 hs_list<-list()
 zs_list [[1]]<- ZDL2[(f[j,1]+1):f[j+1,1],]
 zs_list [[2]]<- ZL[(f[j,1]+1): f[j+1,1],]
```

```
 ZSL2 <-rbind(ZSL2,diag.block (zs_list)) }
return(ZSL2)}

ZSL2<- ZS.matrixL2(yit_1, ZDL2, ZL)

## One step conventiona system GMM using lagged IVs

Data_ZSL2 <- function (yit_1){
n <-ncol(yit_1)
T<-nrow(yit_1)

## Calculate y_s and y_s_1

y <- yit_1 [(3:T),] ;
y_1 <- yit_1 [(2:(T-1)),]
delta_y_1 <- apply(yit_1 [(1:T-1),],2,diff)
delta_y <- apply(yit_1 [(2:T),],2,diff)
y_s <- rbind(delta_y,y) ;
dim(y_s) <-c( 2*(T-2)*n,1)
y_s_1 <- rbind(delta_y_1,y_1);

dim(y_s_1)<-c(2*(T-2)*n,1)
ZSL2 == ZSL2
return(cbind(y_s,y_s_1, ZSL2))
}

Data_ZSL2 <- Data_ZSL2(yit_1)
y_s <- Data_ZSL2[,1, drop = FALSE];
y_s_1<- Data_ZSL2[,2, drop = FALSE]
ZSL2 <- Data_ZSL2[, 3:[2+(T-2)(T-1)/2]-[(T-4)(T-3)/2]

library(MASS)

ZSL2_HI<-list(D.T_D.T, diag(1,T-2));
ZSL2_HI_T<- diag.block(ZSL2_HI)
HIL2<- kronecker(diag(1, n), ZSL2_HI_T)

## calculate  weight matrix = WS1L2
```

```
WS1L2<- t(ZSL2)%*%HIL2%*%ZSL2
part1_SYS_step1L2<- solve(t(y_s_1) %*%
ZSL2 %*% solve(WS1L2)%*%
t(ZSL2)%*% y_s_1)

part2_SYS_step1L2<- t(y_s_1)%*%ZSL2%*%
solve(WS1L2)%*% t(ZSL2)%*%y_s
gamma.hat.SYS.step1L2<- part1_SYS_step1L2%*%
part2_SYS_step1L2

dim(gamma.hat.SYS.step1L2) <- NULL
Bias.SYS.step1L2 <- gamma.hat.SYS.step1L2-gamma
SE.gamma.hat.SYS.step1L2<-sqrt(diag(part1_SYS_step1L2))

## One step estimation conventional system GMM

R = R ; gamma.hat.SYS.step1_l2<-numeric(R)

delta_y <- Data_ZSL2[,1, drop = FALSE]
delta_y_1<- Data_ZSL2[,2, drop = FALSE]
ZS_l2v<- Data_ZSL2[, 3:[2+(T-2)(T-1)/2]-(T-4)(T-3)/2]
n1<-nrow(y_s)
WS1L2<- t(boot[,3:[2+(T-2)(T-1)/2]-[(T-4)(T-3)/2])%*%
HIL2%*%boot[,3:[2+(T-2)(T-1)/2]-[(T-4)(T-3)/2]

for (i in 1:R) {
ind<-sample(n1, n1, replace = TRUE)
boot<-  Data_SYSl2[ind,]
part1_SYS_step1L2<- solve(t(boot[,2])%*%
boot[,3:[2+(T-2)(T-1)/2]-[(T-4)(T-3)/2]%*%
solve(WS1L2)%*t([,3:[2+(T-2)(T-1)/2]-(T-4)(T-3)/2])
%*% boot[,2])

part2_SYS_step1L2<- t(boot[,2])%*%
boot[,3:[2+(T-2)(T-1)/2]-[(T-4)(T-3)/2] %*%
solve(WS1L2)%*%t(boot[,3:[2+(T-2)(T-1)/2]-(T-4)(T-3)/2])
%*% boot[,1]

gamma.hat.SYS.step1L2 [i]<- part1_SYS_step1L2%*%
```

```
part2_SYS_step1L2
}

mean(gamma.hat.SYS.step1L2 )
(Bias =  mean(gamma.hat.SYS.step1L2)-gamma)
(RMSE<- sqrt(mean((gamma.hat.SYS.step1L2-gamma)^2)))
sd(gamma.hat.SYS.step1L2 )

# Diffrence estimator using lagged IVs

Arellano.Bond<-function(yit_1, gamma,
D.T_D.T, ZDL2) {

n<-ncol(yit_1)
T<-nrow(yit_1)

## calculate delt_y

delta_y <- rbind(apply(yit_1[(2:T),],2,diff))
dim(delta_y) <-c((T-2)*N,1)
delta_y_1 <- rbind(apply(yit_1[(1:T-1),],2,diff))
dim(delta_y_1)<-c((T-2)*N,1)

# calculate D using my function

D<- kronecker(diag(1,N), D.T_D.T)
# calculate WDL2:
WDL2<- t(ZDL2)%*%D%*%ZDL2

part1_l2<- solve(t(delta_y_1)%*%ZDL2%*%
solve(WDL2)%*%t(ZDL2)%*%delta_y_1)

part2_l2<- t(delta_y_1)%*%ZDL2 %*%
solve(WDL2)%*%t(ZDL2)%*%delta_y

gamma.hat.step1_l2<- part1_l2%*%part2_l2

## Two step estimation
```

```
## calculate residual delta_u.hat

delta_y <- rbind(apply(yit_1[(2:T),],2,diff))
delta_y_1 <- rbind(apply(yit_1 [(1:T-1),],2,diff))
delta_yit.hatl2 <- gamma.hat.step1_l2*delta_y_1

## Calculate  delta_u.hat
delta_epsilon.hat.il2<- delta_y- delta_yit.hatl2
## calculate sigma.2.epsilon.hat

sigma.2.epsilon.hat.ABL2<- (sum(unlist(lapply
(delta_epsilon.hat.il2, function(x)
t(x)%*% x))))/(2*N*(T-2))

##  Conventional System  GMM  using lagged IVs

SYSL2<- function(yit_1, gamma, HI ,
D.T_D.T, ZDL2,HL,ZSL2){

 n <-ncol(yit_1)
 T<-nrow(yit_1)

 ## calculate y_s and y_s_1

 y <- yit_1 [(3:T),]
 y_1 <- yit_1 [(2:(T-1)),]
 delta_y_1 <- apply(yit_1 [(1:T-1),],2,diff)
 delta_y <- apply(yit_1 [(2:T),],2,diff) ;
 y_s <- rbind(delta_y,y) ;
 dim(y_s) <-c( 2*(T-2)*n,1)
 y_s_1 <- rbind(delta_y_1,y_1) ;
 dim(y_s_1) <-c(2*(T-2)*n,1)

 # calculate HIL2
 ZSL2_HI <- list(D.T_D.T , diag(1,T-2));
 HIL2_T <- diag.block(ZSL2_HI)
 HIL2 <- kronecker(diag(1,n), HI_T)

 ## calculate WS1L2
```

```
WS1L2<- t(ZSL2) %*%HIL2%*%ZSL2
# calculate gamma.hat.step (1)
part1_SYS_step1L2<- solve(t(y_s_1)%*% HSL2 %*%
solve(WS1L2)%*%t(ZSL2)%*%y_s_1)
part2_SYS_step1L2<- t(y_s_1)%*%ZSL2%*%solve(AS1_l2)
%*% t(ZSL2) %*%y_s
gamma.hat.SYS.step1L2<- part1_SYS_step1L2%*%
part2_SYS_step1L2

##  Two step  estimation: calculate residual u.hat.i_SL2

y_s <- rbind(delta_y,y)
y_s_1<- rbind(delta_y_1,y_1)
yit.hat_SYSL2<- gamma.hat.SYS.step1L2*y_s_1
u.hat.i_SL2<- y_s - yit.hatL2

## Calculate sigma.2.alpha.hat

delta_u_i_tilde_SL2<- lapply(u.hat.i_SL2,function(x)
head(x,(T-2)))
u_i_tilde_SL2<- lapply(u.hat.i_SL2,function(x)
tail(x,(- 1)))

sigma.2.alpha.hat.SYSL2<-(sum(unlist(lapply(u_i_tilde_SL2,
function(x) t(x)%*% x))-(unlist(lapply
(delta_u_i_tilde_SL2,function(x)
t(x)%*%x))/ 2)))/(n*(T-2))

## variance ratio "hat.rho" using lagged instruments

rL2<- sigma.2.alpha.hat.SYSL2/sigma.2.epsilon.hat.ABL2

## one step sub-optimal using lagged IVs

Data_SYSL2 <- function (yit_1){

 n <-ncol(yit_1)
 T <-nrow(yit_1)
```

```
 ## calculate y_s and y_s_1

 y <- yit_1 [(3:T),] ;
 y_1 <- yit_1 [(2:(T-1)),]
 delta_y_1 <- apply(yit_1 [(1:T-1),],2,diff)
 delta_y <- apply(yit_1 [(2:T),],2,diff)
 y_s <- rbind(delta_y,y)
 dim(y_s)<- c(2*(T-2)*n,1)
 y_s_1<- rbind(delta_y_1,y_1)
 dim(y_s_1)<- c(2*(T-2)*n,1)
 ZSL2 ==  ZSL2
 return(cbind(y_s,y_s_1, ZSL2))
}
Data_SYSL2<- Data_SYSL2(yit_1)
R = R ; n1<-nrow(y_s)
gamma.hat.SYS.step1_JL2<-numeric(R)

delta_y<- Data_SYSL2[,1, drop = FALSE]
delta_y_1<- Data_SYSl2[,2, drop = FALSE]
ZSL2<- Data_SYSL2[, 3:[2+(T-2)(T-1)/2]-(T-4)(T-3)/2]
WS1_JL2<- t(boot[,3:[2+(T-2)(T-1)/2]-(T-4)(T-3)/2])%*%
H_JL2%*%boot[, 3:[2+(T-2)(T-1)/2]-(T-4)(T-3)/2]

for(i in 1:R) {
 ind<-sample(n1, n1, replace = TRUE)
 boot<-  Data_SYSL2[ind,]
 part1_SYS_step1_JL2<- solve(t(boot[,2])%*%
 boot[,3:[2+(T-2)(T-1)/2]-(T-4)(T-3)/2]%*%
 solve(WS1_JL2)%*%
 t(boot[,3:[2+(T-2)(T-1)/2]-(T-4)(T-3)/2])%*%boot[,2])

 part2_SYS_step1_JL2<-t(boot[,2])%*%
 boot[, 3:[2+(T-2)(T-1)/2]-(T-4)(T-3)/2]%*%
 solve(WS1_JL2)%*%
 t(boot[,3:[2+(T-2)(T-1)/2]-(T-4)(T-3)/2])%*%boot[,1]

 gamma.hat.SYS.step1_JL2[i]<- part1_SYS_step1_JL2%*%
 part2_SYS_step1_JL2
```

```
}
mean(gamma.hat.SYS.step1_JL2)
Bias =  mean(gamma.hat.SYS.step1_JL2)- gamma)
(RMSE<- sqrt(mean((gamma.hat.SYS.step1_JL2-gamma)^2)))
sd(gamma.hat.SYS.step1_JL2)
```