IMT Institute for Advanced Studies, Lucca

Lucca, Italy

Essays in GMM estimation of dynamic panel data models

PhD Program in Economics, Markets, Institutions

XXIII Cycle

By

Irene Mammi

2011

The dissertation of Irene Mammi is approved.

Program Coordinator: Prof. Fabio Pammolli, IMT Lucca

Supervisor: Prof. Giorgio Calzolari, Università di Firenze

Tutor: Dr. Mark Dincecco, IMT Lucca

The dissertation of Irene Mammi has been reviewed by:

Dr. Joaquim Oliveira Martins, OECD

Dr. Giammario Impullitti, Cambridge University

IMT Institute for Advanced Studies, Lucca

2011

A Roberto, il mio mentore
Ad Andrea, il mio “sempre”
Alla mia famiglia, il mio baluardo

Contents

List of Tables

Acknowledgements

I am almost at the end of this challenging adventure and I have the opportunity to thank the many people who have contributed to the success of my experience as a PhD student.

First and foremost I want to thank my husband Andrea, who has chosen to be with me in such a long journey and throughout our lives. His love and his patience have enabled me to stand the strain, the huge distance, the moments of discouragement of the last years, which were not easy at all. Moreover, his technical and mathematical support has been irreplaceable. We are crossing this finishing line together and together we will also continue pursuing our dreams. Thanks to him, I still believe dreams can come true. Hopefully, our life together will be lighter from now on.

Special thanks go to Roberto Golinelli. He has been my tireless guide in these years, and most of all a friend. He has always supported me and has encouraged me in the toughest moments. Without his unselfish support and the insightful discussions with him, the climbing would have been much harder.

My parents, my brothers and my sister are my strength. My family shaped me as the person I am and has led me where I am. I am particularly devoted to my mother and my sister for their support and understanding in the last months. Thousands of kilometers have not moved each away from the other an inch.

I wish to express my warm and sincere thanks to Professor Giorgio Calzolari, my PhD thesis supervisor, for the confidence he placed in me and for having introduced me to the amazing world of Monte Carlo simulations.

I thank all the colleagues and the academic staffs I have met in Lucca, in Colchester and in Bologna: they have made my study and research experience really enjoyable.

In particular, I would like to say the warmest thanks to Luisa, who has shared with me breakfasts, sleepless nights, anguish and satisfaction: she has become almost a sister. Hopefully, we will soon write a paper together, at last!

Thanks to Alessandro, who has made the days in Bologna so funny.

I am most grateful to Dr. Jens Mehrhoff for providing me with his Stata code file, to David Roodman for having made Stata codes and datasets available on the web and to Professor Roberto Golinelli for having given me his original dataset.

I have had the exceptional opportunity to attend Prof. Jeffrey Wooldridge's lectures in Bertinoro. I want to express my sincere gratitude to him for having been so friendly and having given me very useful suggestions about the second chapter of my dissertation.

The responsibility for any error or imprecision that may occur in this work is entirely my own.

Vita

October 27, 1983	Born, Scandiano, Italy
2002	Diploma di Maturità Classica Final mark: 100/100 cum laude Liceo Classico "L. Ariosto" Reggio Emilia, Italy
2005	B.A. Degree in Political Sciences Final mark: 110/110 cum laude University of Bologna Bologna, Italy
2007	M.A. Degree in Political Sciences Final mark: 110/110 cum laude University of Bologna Bologna, Italy
2007	Licenza del Collegio Superiore Collegio Superiore University of Bologna Bologna, Italy
2008-2011	PhD in Economics, Markets and Institutions IMT Institute for Advanced Studies Lucca Lucca, Italy
2010	M. Sc. Degree in Statistics and Econometrics Final mark: Distinction University of Essex Colchester, United Kingdom
2010-11	Visiting Graduate Student Department of Economics University of Bologna Bologna, Italy

Abstract

The aim of the work is twofold. First, we investigate the properties of the dynamic panel data (DPD) GMM estimator when the instrument count is high. We introduce the extraction of principal components from the instrument matrix as an effective strategy to reduce the number of instruments. Through Monte Carlo experiments, we want to compare the performances of the GMM estimators when the instrument set is factorized, collapsed or limited. Second, we estimate fiscal response functions on simulated panels and on real data to identify the best-performing estimator in this context, where endogeneity and instrument proliferation issues are unavoidable.

The dissertation consists of three chapters. The first reviews the literature of DPD estimation and presents the issue of instrument proliferation in DPD GMM estimation. The second introduces the principal component analysis (PCA) to reduce the dimension of the instrument matrix and compares the performances of the factorized, limited and collapsed GMM estimators, finding them similar. Though the simulated models are extremely simplified, the PCA seems to be promising. The third chapter simulates fiscal response functions and investigates the properties of DPD estimators in fiscal rules estimation; the fiscal rules are then estimated on real data for EMU Countries. The system GMM estimator is the best-performing here. Instrument proliferation does not bias the estimates; collapsing and lag truncation of the instrument matrix can lead to misleading results, while the factorized estimator performs well. Discretionary policies within the EMU are systematically found a-cyclical.

Introduction

Panel data techniques have become growingly popular in modern econometrics in the last few decades and are now commonly adopted in empirical analyses. The increasing availability of reliable data for cross-sections of individuals at different points in time has greatly enhanced the development of panel data estimators. The implementation of proper procedures to estimate panel data models in many econometric softwares has contributed further to the spread of these techniques.

In particular, the estimation of dynamic panel data (DPD) models, that give the opportunity to study individual dynamics over time, has gained a leading role in panel data econometrics and is now common practice in many microeconomic and macroeconomic empirical applications. Unfortunately, the inclusion of a lagged dependent variable in a model where individual effects are present gives rise to the well known dynamic panel bias: the correlation between the lagged dependent variable and the individual effect makes the former endogenous, so that the estimates are inconsistent.

The classical DPD estimators[1] do not succeed in solving the bias. Instrumental variable (IV) estimators have therefore been proposed to tackle the endogeneity of the lagged dependent variable.

The Generalized Method of Moments (GMM) estimator has recently become the most widely used DPD IV estimator, thanks to its flexibility and to the few assumptions about the data generating process it requires. Its most appealing advantage is the availability of "internal" in-

[1]We refer here to the pooled OLS, first-difference and within-group estimators.

struments: the endogenous regressors are in fact instrumented by their previous realizations, properly chosen according to meaningful moment conditions. GMM estimates are now also easily implementable in many econometric softwares. As a consequence, the GMM estimator is often blindly applied in empirical works and is often seen as the panacea for all the drawbacks of the classical DPD estimators.

However, the GMM estimator itself is not free of severe drawbacks[2]. A relevant issue in this context is the instrument proliferation problem. In GMM estimation, the instrument count gets very high as the number of periods in the sample becomes moderately large. This comes along with potentially serious problems that affect the estimates, such as the overfitting of the endogenous variables, the weakening of useful specification tests and imprecise estimates of the covariance matrix of the moments. The problem of instrument proliferation is rarely investigated and detected in empirical analysis and the related risks are generally not properly considered in GMM estimation. Sometimes, the lag depth of the instrument is truncated or the instrument set is collapsed to reduce the instrument count. However, there is not a clear indication in the literature about which is a safe number of instruments. Most of all, the robustness of the estimates to alternative instrument reduction strategies has not been investigated properly. It is unclear whether the truncation in the lag depth or the collapsing are always safe or can lead to misleading results in some specifications of the GMM estimator.

The GMM estimation of DPD models is the thread of our work. In particular, we address the issue of instrument proliferation in this context and compare alternative strategies to reduce the instrument count.

The aim of the work is twofold.

Firstly, we want to assess the performance of the GMM estimator when the instrument count becomes worryingly high. Through extensive Monte Carlo experiments, we compare the behavior of the difference and system GMM estimators when some instrument reduction strategy is adopted. We introduce an alternative technique to reduce the instru-

[2]It frequently suffers, among the others, weak identification problems, small sample biases and instrument proliferation issues.

ment count: the extraction of principal components from the instrument matrix.

Secondly, we want to investigate which is the best-performing DPD estimator in the estimation of fiscal rules. Through unprecedented simulation experiments, we are interested in assessing the properties of the estimators in a framework in which many endogenous regressors are included in the model, with the unpleasant consequence of a proliferation of instruments.

The dissertation consists of three chapters.

In the first chapter, we extensively review the literature about the estimation of DPD models, we present the alternative DPD estimators that have been proposed and discuss relevant econometric issues in this framework. We mainly focus on the GMM estimator for DPD. Our aim is to cast light on the complexity of the GMM estimators, despite their huge flexibility, and on the potential sensitiveness of the estimates to alternative specifications of the estimators. We critically argue that the blind application of a GMM approach, made particularly easy by the availability of "buttons to push" in most econometric softwares, can be dangerous. The GMM estimator has many facets[3] and the choices of the researcher about the setting of the estimator are far from being irrelevant. We extensively treat the issue of instrument proliferation in DPD GMM estimation, discuss the dangers deriving from it and present the strategies that have been proposed to face it.

In the second chapter, we introduce the application of the principal component analysis (PCA) as an effective strategy to reduce the number of instruments in GMM estimation. We extensively discuss the rationale of the factorization of the instrument matrix, its advantages and its novelty in DPD estimation. Through extensive Monte Carlo simulations, we compare the performance of the difference and system estimators when an instrument reduction strategy[4] is adopted or when the techniques are

[3]It is enough to think that we can use difference or system GMM, one-step or two-step GMM, alternative weighting matrices, we can reduce the number of lags included or collapse the instruments, and so on.

[4]We aim at assessing the properties of the GMM estimators when the instrument set is limited, collapsed or factorized.

combined. We simulate and estimate the standard panel AR(1) model first, as this is generally the baseline model in Monte Carlo simulations. We then move to a multivariate DPD model that also includes an additional endogenous regressor. Though we estimate the mainstream AR(1) model and we also add an additional explanatory variable, we argue that these models are probably too bare to appreciate in depth the risks of instrument proliferation and to exploit the potentialities or to identify the drawbacks of the different instrument reduction techniques. The analysis of a more complex model could help to better understand the importance of the instrument proliferation issue and to assess more safely whether a drastic reduction in the instrument count is dangerous in more informative models.

The third chapter moves exactly in the direction suggested in conclusion of the second chapter. We simulate here fiscal response functions for advanced economies and compare the performance of alternative DPD GMM estimators to estimate the parameters of the model. There are many advantages in the choice of the fiscal policy as our field of interest. From an econometric perspective, fiscal rules are always dynamic models that include endogenous variables, such as the output gap, the primary balance and the public debt, so that they are a proper field of application for the estimators presented in Chapters 1 and 2. The underlying relationships among the variables in a fiscal rule come from standard national accountancy rules, so that they are perfectly known. No behavioral assumption is needed. When we simulate fiscal rules, we are thus able to generate a complex realistic scenario having though a strict control on the underlying processes of the variables and on the relations among them, on the sources of endogeneity and on the relevance of fixed effects and error terms. Most of all, differently from a pure AR(1) process, the variables have here an economic meaning as we have a precise model in mind. The analysis becomes both more realistic and less an end in itself. From the economic point of view, the estimation of fiscal response functions is of undoubted interest as it allows to investigate whether discretionary fiscal actions are pro-cyclical or countercyclical. An answer to this question is particularly important for the European Monetary Union, where the constraints imposed by the Stability Growth Pact may affect the response of discretionary fiscal policy to the economic cycle. We simulate here the fiscal rule and check the robustness of the estimates to different specifications of the GMM estimators and to alternative settings of the parameters. We compare the instrument reduction strategies for the estimation of a model which is more exposed to endogeneity and to instrument proliferation than the models previously considered. We then estimate the fiscal rule on real data for the European Monetary Union. We expect to obtain clearer indications about the advisability of applying the instrument reduction techniques introduced in the second chapter.

The work is conceived to be read as a whole, provided with internal consistency. Chapter after chapter we aim at casting more light on the issues of interest and at providing more and more practical indications for the estimation of DPD models on real data. We first draw the attention to serious econometric issues in DPD estimation; we then propose strategies to handle the instrument proliferation problem; finally we test the solidity of the remedies we suggest in a context that is more tricky, but also more interesting from an economic perspective and closer to the type of models that are commonly estimated in applied macroeconomics.

Chapter 1

Opening the "black box": a review on GMM estimation of dynamic panel data models

1.1 Introduction

"We estimate a panel data model with fixed effects..." or "We consider a dynamic model for panel data..." or again "We estimate the model by GMM...". How many times we have read similar sentences in economic papers in recent years! And how may times we have been left with doubts about the precise estimation approach adopted by the authors!

We will try here to cast some light on the estimation of panel data models, in particular on the Generalized Method of Moments (GMM) estimation of dynamic models. We aim at enlightening what often seems an impenetrable "black box": to do so, we present the alternative methodologies that have been proposed to estimate panel data models and review the main contributions in the massive literature on the topic. We would also like to launch a warning to the researcher against the blind use of the GMM estimator in the estimation of dynamic panel data (DPD) models. We move from the consideration that simply saying "GMM estimator" could mean almost nothing, as GMM estimation has an incredible number of facets that lead to different estimation strategies. One-step or two-step, a weighting matrix rather than another one, untransformed or collapsed or limited instruments... are not simply options in a command for GMM estimation in an econometric software: different choices in this regard can bring to very different results.

In the remainder of the chapter we proceed as follows: we first discuss the recent flourishing of panel data models; we then present the standard DPD model and the classical estimators for it: we focus here in particular on the endogeneity issues intrinsic in this context and on instrumental variable estimators that address the problem; we then introduce the general linear GMM estimator and discuss its adoption in the estimation of DPD models; we then focus on some controversial econometric issues in GMM estimation in this field, with a particular attention to the problem of instrument proliferation and to strategies to detect and face it.

1.2 A flourishing interest in panel data models

Panel data techniques have become increasingly popular in the last few decades and are now one of the most exploited tools in modern econometrics, both in microeconomic and macroeconomic applications. The increasing availability of reliable data in recent times has made the estimation of panel data models easier and has given a decisive boost to the development of appropriate methodologies for panel datasets.

We refer to *panel data* when we deal with repeated observations over time for the same cross-section of individuals: we therefore have a time dimension T, a cross-section dimension N and observations y_{it}.

Panel data are very appealing, thanks to their advantages over both a purely cross-section or time-series approach[1].

[1] Hsiao [2003, 2007], Baltagi [2008], Wooldridge [2010] and Arellano [2003], among the others, extensively discuss the advantages of panels over time series and cross-sections.

First, they offer the opportunity to control for unobserved individual heterogeneity, thus reducing the risk of biased estimates. They allow to handle time-invariant unobserved variables and to alleviate the risk of omitted variable bias problems. Being safer against the omission of time-invariant explanatory variables, the need of good and relevant instruments is less compelling, hence we have less identification problems.

Secondly, thanks to the availability of repeated observations for the same individual, we can estimate more complex models that also capture the dynamics of the variables of interest[2]. 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.

Moreover, with panel data, we can exploit two different sources of variability, as we have information on both inter-temporal dynamics, that are captured by the variation of the variables of interest over time for the same individual (*within variability*), and on the variability across individuals in a given moment of time (*between variability*). Repeated observations also weaken the effect of dangerous measurement errors in the variables, thus mitigating well-known under-identification problems. More degrees of freedom can usually be exploited, so that the inference on the parameters is likely to be more accurate. The problem of multicollinearity among the variables is largely reduced in panel data. There is a huge gain in terms of efficiency of the estimates and the opportunity to construct and test more complex behavioral hypotheses[3].

1.3 "Panel": a very generic word

Referring simply to a "panel model", however, is probably too generic and not very informative. The specification of a panel data model depends on the type of data we have: in fact, we could deal with micro panels or macro panels, static or dynamic panels, short or long panels,

[2]It is obviously impossible to estimates dynamic processes when observations only for a single point in time are available.

[3]See Hsiao [2007] and Baltagi [2008] for further discussion on these advantages.

small or large panels, each of them having different relevant characteristics. In micro panels, some characteristics of a cross-section of a large number of individuals, households or firms are generally observed for a limited period of time: the size of the sample N is usually large while the time dimension T is generally quite small. The asymptotic properties of the estimators are therefore evaluated for large N and fixed T. In macro panels in general we have aggregated data at a national level for a limited or moderate number of Countries N over a sufficiently large number of years T: in this case the two dimensions of the panel are often similar in magnitude and the asymptotics is studied for both large N and large T.

The distinction we are interested in the most is that between static and dynamic models. In our analysis we will focus specifically on linear models for DPD, in which several lags of the dependent variable appear among the regressors.

1.4 The linear dynamic panel data model

Throughout our analysis, we will focus on a first-order autoregressive error-components model[4] with a cross-section dimension N larger than the time dimension T, so that that $N >> T$. The baseline full DPD model has the form:

$$y_{it} = \alpha y_{i,t-1} + x'_{it}\beta + \varepsilon_{it} \tag{1.4.1}$$

where $i = 1, 2, ..., N$, $t = 1, 2, ..., T$, $|\alpha| < 1$ is a scalar and is the parameter of interest, x_{it} is a $1 \times K$ vector of regressors and β is a $K \times 1$ vector of coefficients to be estimated.

The unobservable error term ε_{it} is assumed to follow a one-way error component model consisting of two components:

$$\varepsilon_{it} = \mu_i + u_{it} \tag{1.4.2}$$

[4] For the sake of simplicity and without loss of generality, we include here only the first lag of the dependent variable among the regressors. Higher-order AR process could also be considered.

where μ_i is an unobservable individual time-invariant effect, that accounts for individual heterogeneity, and u_{it} is an idiosyncratic term. We assume here that $\mu_i \sim$ i.i.d.$(0, \sigma_\mu^2)$ and $u_{it} \sim$ i.i.d.$(0, \sigma_u^2)$, that the μ_i and the u_{it} are independent each of the other and across individuals and that the idiosyncratic shocks are serially uncorrelated. We thus have

$$\mathrm{E}[\mu_i] = \mathrm{E}[u_{it}] = \mathrm{E}[\mu_i u_{it}] = 0. \tag{1.4.3}$$

For the sake of simplicity, we are here assuming homoskedasticity of the residuals and lack of autocorrelation.

We further assume that $|\alpha| < 1$, as we desire the y_{it} series to be stationary. However, heterogeneity in the means of the y_{it} series across individuals is allowed through the inclusion in the model of the μ_i.

The vector x_{it} may include both predetermined and strictly exogenous regressors and possibly also lagged values of these variables and higher-order lags of the dependent variable. The individual effects may be correlated with the variables included in the vector x_{it}. The dynamics is obviously introduced in the model through the inclusion of $y_{i,t-1}$ among the regressors.

The inclusion of a lagged dependent variable in a model with individual effects gives rise to the well-known *dynamic panel bias*, due to the correlation between the lagged variable y_{t-1} and the individual effect μ_i.

According to the model in equation (1.4.1) y_{it} is a function of μ_i: the same relationship lagged one period implies that also $y_{i,t-1}$ is a function of μ_i, as the μ_i are time-invariant. It follows that the lagged dependent variable, hence one of the regressors, is correlated with one component of the error term, thus causing a severe endogeneity problem.

The standard OLS estimator is neither consistent nor unbiased when applied to the model in (1.4.1), as the assumption of exogeneity of the regressors is not satisfied. The estimates remain biased and inconsistent even when we assume that the μ_i are random effects and even though the u_{it} are serially uncorrelated. We neither recover consistency as $T \to \infty$ nor as the sample size increases. The OLS estimators can be shown to be upward biased, as the correlation between $y_{i,t-1}$ and the composite error term is positive, i.e. $\mathrm{E}(y_{i,t-1}, \varepsilon_{it}) > 0$: α will be thus overestimated.

The endogeneity problem needs therefore to be tackled. We can adopt two alternative strategies or even apply them together for this purpose. We can both transform the data in order to remove the individual fixed effects μ_i, so that we eliminate the correlation with the regressors, or we can face the endogeneity problem by looking for relevant and valid instruments for $y_{i,t-1}$.

There are two alternative ways of transforming the data in order to get rid of the problematic individual effects. For the sake of simplicity, we assume from here onwards that no other regressor is included, so that the model in (1.4.1) becomes:

$$y_{it} = \alpha y_{i,t-1} + \varepsilon_{it}. \tag{1.4.4}$$

First differencing[5] consists of subtracting from equation (1.4.4) the same equation lagged one period, namely

$$y_{i,t-1} = \alpha y_{i,t-2} + \varepsilon_{i,t-1}, \tag{1.4.5}$$

thus yelding

$$\begin{aligned}(y_{it} - y_{i,t-1}) &= \alpha(y_{i,t-1} - y_{i,t-2}) + (\varepsilon_{it} - \varepsilon_{i,t-1}) \\ &= \alpha(y_{i,t-1} - y_{i,t-2}) + (\mu_i - \mu_i) + (u_{it} - u_{i,t-1}) \\ &= \alpha(y_{i,t-1} - y_{i,t-2}) + (u_{it} - u_{i,t-1}). \end{aligned} \tag{1.4.6}$$

The transformation will result in a system of $T-1$ equations in first-differences, as the first-differencing drops T first-period observations. In unbalanced panels, the loss of observations due to first differencing can be very huge.

Defining $\Delta = (1-L)$ the first difference operator, with L being the lag operator, we can write the transformed model more compactly as:

$$\Delta y_{it} = \alpha \Delta y_{i,t-1} + \Delta u_{it} \quad \text{for} \quad i = 1, ..., N \quad \text{for} \quad t = 2, ..., T. \tag{1.4.7}$$

The individual effects μ_i cancel out. In other terms, the first-difference transformation is obtained by multiplying the model in (1.4.4) by $I_N \otimes$

[5]This transformation will be the basis of the Arellano-Bond [1991] GMM estimator.

M_Δ, where I_N is an identity matrix of dimension N and M_Δ is $(T-1) \times T$ matrix which has the form[6]:

$$M_\Delta = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 \end{pmatrix}. \tag{1.4.8}$$

The first-difference (FD) estimator is the OLS estimator of equation (1.4.7), that is

$$\hat{\alpha} = \sum_{i=1}^{N} \left[(\Delta y_{i,t-1})' \Delta y_{i,t-1} \right]^{-1} \sum_{i=1}^{N} (\Delta y_{i,t-1})' \Delta y_{it}. \tag{1.4.9}$$

As we have assumed that $u_{it} \sim$ i.i.d.$(0, \sigma_u^2)$, the transformation induces a MA(1) process for the Δu_{it}, so that the FD estimator is inconsistent. In fact, the variance of Δu_{it} will be given by the $(T-1) \times (T-1)$ matrix[7]

$$\text{var}(\Delta u_{it}) = \sigma_u^2 M_\Delta M_\Delta' = \sigma_u^2 \begin{pmatrix} 2 & 1 & 0 & 0 & \dots & 0 & 0 \\ 1 & 2 & 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & 2 \end{pmatrix}. \tag{1.4.10}$$

A generalized least square (GLS) estimator is therefore required. It can be shown[8] that the optimal GLS estimator here is the Within-Group (WG) estimator.

Following Arellano [2003], we can define Q as the within group operator that transforms the series y_{it} in the series of deviations from time means $(y_{it} - \bar{y}_i)$. In other words, we subtract from the model in (1.4.4) the time averaged model

$$\bar{y}_i = \alpha \bar{y}_{i,-1} + \mu_i + \bar{u}_i, \tag{1.4.11}$$

[6]See Roodman [2009a], Hsiao [2007] and Arellano [2003], among others, for a formal presentation of the estimator.

[7]See Arellano [2003] for a detailed presentation.

[8]See Arellano [2003] for a proof.

where $y_i = (y_{i2}, \ldots, y_{iT})'$ and $y_{i,-1} = (y_{i1}, \ldots, y_{i(T-1)})'$, so that the transformed model becomes:

$$(y_{it} - \bar{y}_i) = \alpha(y_{i,t-1} - \bar{y}_{i,-1}) + (\mu_i - \mu_i) + (u_{it} - \bar{u}_i). \qquad (1.4.12)$$

In Arellano [2003]'s notation, we can write the transformation as $\tilde{y}_i = Qy_i$ whose elements are $\tilde{y}_{it} = y_{it} - \bar{y}_i$. Q can be shown to be:

$$Q \equiv M'_\Delta (M_\Delta M'_\Delta)^{-1} M_\Delta. \qquad (1.4.13)$$

The WG estimator is the OLS applied to equation (1.4.12), so that $(y_{it} - \bar{y}_i)$ is regressed on $(y_{i,t-1} - \bar{y}_{i,-1})$.

The WG transformation cancels out the μ_i, but unfortunately it does not fix the dynamic panel bias.

To see this, we can write the transformed lagged dependent variable as

$$y_{i,t-1} - \frac{1}{T-1}\left(y_{i1} + \cdots + y_{iT-1}\right) \qquad (1.4.14)$$

and the transformed idiosyncratic term as

$$u_{it} - \frac{1}{T-1}\left(u_{i2} + \cdots + u_{iT}\right). \qquad (1.4.15)$$

Since y_{it} is correlated with u_{it} according to (1.4.4), also $y_{i,t-1}$ is correlated with $u_{i,t-1}$. As a consequence, $y_{i,t-1}$ is also correlated with $\bar{u}_i$ which contains $u_{i,t-1}$.

We therefore have correlation between the component $-\frac{y_{it}}{T-1}$ in the transformed lagged dependent variable and the term u_{it} in the transformed idiosyncratic term; similarly, the component $-\frac{u_{i,t-1}}{T-1}$ of the new error term is correlated with $y_{i,t-1}$ in the dependent variable.

Nickell [1981] shows that this correlation is negative and dominates over positive correlations that are induced by the WG transformation. Unfortunately, the inconsistency of the WG estimator persists and the estimates come out to be downward biased even as $N \to \infty$.

The WG estimator would instead be consistent as $T \to \infty$, as the correlated term would become insignificant and the correlation would tend to disappear[9].

[9] See Roodman [2009a] and Bond [2002] for a treatment of this issue.

It is therefore easier to handle with the endogeneity problem induced by the FD transformation rather than that implied by the WG transformation: in the former, only the previous realization of the error term is included in the model; in the latter, all the previous realizations are included.

The pooled OLS, the FD and the WG estimators therefore do not allow to solve the dynamic panel bias. It is thus necessary to adopt alternative strategies in order to handle the problem.

Hsiao [2003] presents the Maximum Likelihood (ML) estimators for the AR(1) panel data model that have been developed in the literature. He stresses the importance of the assumptions about the initial conditions that are required for the ML function to be specified; he also warns about the potential inconsistency of the ML estimator when the initial conditions are misspecified. Because of the delicacy of the assumptions on the initial conditions, the ML estimators have not had big luck and the Instrumental Variable estimators (IVE) have been largely preferred.

IVEs have been widely developed in the context of the AR(1) panel model from the works of Anderson and Hsiao [1981, 1982] onwards and have become the most exploited tool to deal with the endogeneity problem intrinsic in DPD estimation.

The FD model in (1.4.7) needs to be estimated with an IV estimator, as the FD transformation is not enough to recover consistency when the OLS estimator is applied to the transformed data.

Anderson and Hsiao [1981, 1982] suggest on using the FD transformation to purge the data of the fixed effects; they then present two formulations of the 2-stage Least Squares (2SLS) estimator for the first-differenced AR(1) that both use lags of the dependent variable to instrument the transformed lagged dependent variable.

In the model in (1.4.7) the $y_{i,t-1}$ term included in $\Delta y_{i,t-1}$ is correlated with the $u_{i,t-1}$ in $\Delta u_{i,t}$. However, deeper lags of the regressors can be shown to be uncorrelated to the error term and can therefore be used as instruments.

Anderson and Hsiao [1981, 1982] suggest on using $y_{i,t-2}$ as an instrument for $\Delta y_{i,t-1}$, as it is correlated with $\Delta y_{i,t-1} = y_{i,t-1} - y_{i,t-2}$ but it is also orthogonal to the error term Δu_{it} once we assume that the error terms are serially uncorrelated[10].

In addition to serial uncorrelation of the u_{it}, stationarity is not necessary and only a weak assumption on the initial conditions is required in order to have $y_{i,t-2}$ not correlated with Δu_{it}. In fact, the initial conditions y_{i1} are required to be predetermined, i.e. uncorrelated with u_{it} for $t = 2, 3, \ldots, T$.

This Anderson-Hsiao 2SLS estimator exploits the following linear IV moment restrictions:

$$\mathrm{E}\left[\sum_{t=2}^{T} y_{i,t-2}(\Delta y_{i,t} - \alpha \Delta y_{i,t-1})\right] = 0 \quad \text{for} \quad t = 3, \ldots, T. \tag{1.4.16}$$

The 2SLS estimator is generally defined as:

$$\hat{\alpha}_{IV} = [X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}Z'y = (X'P_Z X)^{-1}X'P_Z y \tag{1.4.17}$$

where Z is the instrument matrix and $P_Z = Z(Z'Z)^{-1}Z'$ is the idempotent projection matrix of Z.

In this framework, for the i^{th} individual and with each row corresponding respectively to $t = 1, 2, \ldots, T$, we have:

$$y_i = \begin{pmatrix} . \\ y_{i2} - y_{i1} \\ y_{i3} - y_{i2} \\ y_{i4} - y_{i3} \\ y_{i5} - y_{i4} \\ \vdots \\ y_{iT} - y_{i,T-1} \end{pmatrix} = \begin{pmatrix} . \\ \Delta y_{i2} \\ \Delta y_{i3} \\ \Delta y_{i4} \\ \Delta y_{i5} \\ \vdots \\ \Delta y_{iT} \end{pmatrix}, X_i = \begin{pmatrix} . \\ . \\ y_{i2} - y_{i1} \\ y_{i3} - y_{i2} \\ y_{i4} - y_{i3} \\ \vdots \\ y_{i,T-1} - y_{i,T-2} \end{pmatrix} = \begin{pmatrix} . \\ . \\ \Delta y_{i2} \\ \Delta y_{i3} \\ \Delta y_{i4} \\ \vdots \\ \Delta y_{i,T-1} \end{pmatrix}. \tag{1.4.18}$$

Z_i is the instrument matrix in which we include $y_{i,t-2}$ as instrument for

[10]This is the so called Anderson-Hsiao level estimator, as we instrument the equation in first differences with lagged levels.

Δy_{it}, namely:

$$Z_i = \begin{pmatrix} \cdot \\ \cdot \\ y_{i1} \\ y_{i2} \\ \vdots \\ y_{i,T-2} \end{pmatrix}, \qquad (1.4.19)$$

where, again, the first row refers to $t = 1$, the last one to $t = T$. We need at least observations for three periods for the Anderson-Hsiao estimator, as $y_{i,t-2}$ is available only from $t = 3$ onwards.

This formulation of the Anderson-Hsiao estimator is consistent for large T.

An alternative strategy proposed by Anderson and Hsiao [1982] is to use $\Delta y_{i,t-2}$ as an instrument for $\Delta y_{i,t-1}$[11]: in fact, also $\Delta y_{i,t-2}$ is orthogonal to the error term and is therefore a valid instrument. This second Anderson-Hsiao 2SLS estimator exploits the single moment condition

$$E\left[\sum_{t=3}^{T} \Delta y_{i,t-2}(\Delta y_{i,t} - \alpha \Delta y_{i,t-1})\right] = 0. \qquad (1.4.20)$$

This estimator requires the availability of at least four periods, as $\Delta y_{i,t-2}$ is exploitable as instrument only from $t = 4$ onwards.

The Z_i matrix has thus the form

$$Z_i = \begin{pmatrix} \cdot \\ \cdot \\ \cdot \\ y_{i,2} - y_{i,1} \\ y_{i,3} - y_{i,2} \\ \vdots \\ y_{i,T-2} - y_{i,T-3} \end{pmatrix} = \begin{pmatrix} \cdot \\ \cdot \\ \cdot \\ \Delta y_{i,2} \\ \Delta y_{i,3} \\ \vdots \\ \Delta y_{i,T-2} \end{pmatrix}. \qquad (1.4.21)$$

One more observation for each individual is lost when $\Delta y_{i,t-2}$ is used instead of $y_{i,t-2}$ as instrument: this could result in an important loss of information, especially in short panels.

[11] This is the difference Anderson-Hsiao estimator, as it is instruments $\Delta y_{i,t-1}$ with a first-differenced instrument.

Anderson and Hsiao [1981] suggest that the 2SLS estimator is more efficient when the instrument is $\Delta y_{i,t-2}$ rather than $y_{i,t-2}$. Also the second specification of the 2SLS appears to be consistent, no matter whether T or N go to infinity.

When either Anderson-Hsiao estimators is used, the model is exactly identified, as one endogenous regressor is instrumented with one instrumental variable. However, these estimators are not actually efficient: they do not use all the available valid instruments.

In order to improve efficiency, we can use more lags of the dependent variable as instruments when observations for at least four periods are available. For example, we could use $y_{i,t-2}$, $y_{i,t-3}$, $y_{i,t-4}, \ldots$, $y_{i,T-2}$ together to instrument $\Delta y_{i,t-1}$ when the observations for $t \geq 5$ are available. More information can be exploited in this case, so that we expect a gain in efficiency: the model is now overidentified with L, the number of instruments, greater than the number of endogenous regressors K.

Unfortunately, as discussed in Roodman [2009a], we face a trade-off between the information introduced in the model and the number of observations we can use. In 2SLS estimation in fact the observations for which lagged values of the variables are missing are dropped from the sample. The larger the lag depth chosen to be used to instrument the dependent variable, the smaller the sample available.

If, for example, we use two lags, instead of only one, as instruments in 2SLS estimation, the Z_i matrix becomes

$$Z_i = \begin{pmatrix} \cdot & \cdot \\ \cdot & \cdot \\ y_{i,1} & \cdot \\ y_{i,2} & y_{i,1} \\ y_{i,3} & y_{i,2} \\ \vdots & \vdots \\ y_{i,T} & y_{i,T-1} \end{pmatrix}. \tag{1.4.22}$$

Compared to the case of a single instrument, here also the third row can not be exploited, as we have missing values. The sample size further decreases.

If we use all the available lags as instruments in 2SLS estimation, we have

$$Z_i = \begin{pmatrix} \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ y_{i,1} & \cdot & \cdot & \dots & \cdot \\ y_{i,2} & y_{i,1} & \cdot & \dots & \cdot \\ y_{i,3} & y_{i,2} & y_{i,1} & \dots & \cdot \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{i,T} & y_{i,T-1} & y_{i,T-2} & \dots & y_{i,1} \end{pmatrix}. \tag{1.4.23}$$

In this extreme case, only the last row is kept in 2SLS estimation.

Panel data GMM estimators allow to use these full sets of instruments exploiting, for each instrument, a set of meaningful moment conditions.

Holtz-Eakin, Newey and Rosen [HENR henceforth, 1988] and Arellano and Bond [1991] first suggest the use of all the available lags at each period in time as instruments for the endogenous first-differenced lagged dependent variable in the FD equation (1.4.7).

HENR first propose a *GMM-style* instrument matrix that, when only $y_{i,t-2}$ is used as instrument for $\Delta y_{i,t-1}$, has the form:

$$Z_i = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ y_{i,1} & 0 & \dots & 0 \\ 0 & y_{i,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & y_{i,t-2} \end{pmatrix}, \tag{1.4.24}$$

where the first row correspond to $t = 1$ and the last one to $t = T$. When all the available lags are used as instruments, Z_i has the form:

$$Z_i = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & \dots & 0 \\ y_{i,1} & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & \dots & 0 \\ 0 & y_{i,2} & y_{i,1} & 0 & 0 & 0 & 0 & \dots & 0 & \dots & 0 \\ 0 & 0 & 0 & y_{i,3} & y_{i,2} & y_{i,1} & 0 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & y_{i,T-2} & \dots & y_{i,1} \end{pmatrix}. \tag{1.4.25}$$

In both cases, each row gives the instruments used for the first-difference equations respectively for periods $t = 1, 2, 3, \dots, T$. We have here zeroes whenever the lag chosen as instrument is not available in the data.

Each column of Z_i is now interpreted in terms of a meaningful IV moment conditions of the form:

$$E\left[y_{i,t-2}(\Delta y_{i,t} - \alpha \Delta y_{i,t-1})\right] = 0 \text{ for } t \geq 3 \tag{1.4.26}$$

where $\Delta u_{it} = \Delta y_{i,t} - \alpha \Delta y_{i,t-1}$ is the transformed residual.

Each column of Z_i is therefore orthogonal to the transformed residual. Exploiting the orthogonality conditions in (1.4.26), we have at least as many instrumental variables as the endogenous regressors in the right-hand side of the model in first-differences, so that the model is identified: the parameters of interest can be therefore estimated.

The HENR formulation of the instrument matrix eliminates the trade-off between the number of lags of the dependent variables that can be used as instruments and the sample size: in fact no row of the matrix is deleted anymore, as the missing values are replaced by zeroes.

As no observation is left out of the estimates even when some of the lags are not available, it becomes meaningful to use all the available lags of y_{it} as instruments for $\Delta y_{i,t-1}$, starting from $y_{i,t-2}$ and going back to $y_{i,T-2}$.

This strategy will be the core of the difference-GMM estimation for DPD models first developed by Arellano and Bond [1991].

As the GMM estimation is the gist of our whole work, it is worth dedicating a separate section to the general linear GMM estimator before presenting how it is translated in DPD estimation.

1.5 The linear GMM estimator

The GMM estimator was originally introduced in the economic literature by Hansen [1982] and Hansen and Singleton [1982]. It has soon become one of the most widely used econometric techniques in both cross-section and panel estimation, as it is very flexible and requires only weak assumptions.

The GMM estimator is a general IV estimator that is particularly useful when the model is over-identified. The core of GMM estimation is the formulation of meaningful moment conditions that, when satisfied, allow the parameters of the model to be estimated consistently. By applying the analogy principle, we move from population moment conditions to their sample counterparts and we use the latter ones to estimate the parameters of the model.

We start with an intuitive presentation of the method of moments (MM)[12] and we then explore the general framework of GMM estimation.

We can consider the general linear regression model

$$y = x'\beta + \epsilon \tag{1.5.27}$$

where x and β are $K \times 1$ vectors and $\mathrm{E}(\epsilon\epsilon') = \Omega$. As usual the $\epsilon = y - x'\beta$ are the residuals: they are the hearth of the GMM estimation. We assume zero conditional mean for the error term ϵ, that is

$$\mathrm{E}[\epsilon|x] = 0. \tag{1.5.28}$$

By the law of iterated expectation we have K unconditional moment conditions

$$\mathrm{E}[x\epsilon] = 0. \tag{1.5.29}$$

The vector β thus satisfies

$$E[x(y - x'\beta)] = 0. \tag{1.5.30}$$

By the analogy principle, we want to solve the following sample moment condition:

$$\frac{1}{N}\sum_{i=1}^{N} x_i\epsilon_i = \frac{1}{N}\sum_{i=1}^{N} x_i(y_i - x_i'\beta) = 0. \tag{1.5.31}$$

The method of moment (MM) estimator is the solution we obtain by solving the previous moment condition. It thus is:

$$\hat{\beta}_{MM} = \left(\sum_{i=1}^{N} x_i x_i'\right)^{-1} \sum_{i=1}^{N} x_i y_i \tag{1.5.32}$$

[12] The following presentation follows Cameron and Trivedi [2005], Wooldridge [2010] and Baum et al. [2003].

or, in matrix notation,

$$\hat{\beta}_{MM} = \left(X'X\right)^{-1} X'y \tag{1.5.33}$$

where $X'X$ is required to have full rank.

It is straightforward to notice that equation (1.5.31) is the normal equation, i.e. the first order condition, for the Least Squares estimation and $\hat{\beta}_{MM}$ is the OLS estimator. The OLS estimator can therefore be seen as nothing but a special case of MM estimation.

When some variables in the vector x are endogenous we need to identify a vector of instruments z such that $E(\varepsilon|z) = 0$.

The vector β needs therefore to satisfy

$$E[z(y - x'\beta)] = 0. \tag{1.5.34}$$

The MM estimator is the one that solves the sample analog of the previous moment condition, namely

$$\frac{1}{N}\sum_{i=1}^{N} z_i(y_i - x_i'\beta) = 0. \tag{1.5.35}$$

When the model is exactly identified the MM estimator is the linear IV estimator

$$\hat{\beta}_{MM(IV)} = \left(\sum_{i=1}^{N} z_i x_i'\right)^{-1} \sum_{i=1}^{N} z_i y_i \tag{1.5.36}$$

or, in matrix notation,

$$\hat{\beta}_{MM(IV)} = (Z'X)^{-1}Z'y. \tag{1.5.37}$$

When the model is overidentified, there are more instruments available than endogenous regressors ($L > K$). In this case, we have L equations in K unknowns therefore we do not have a unique solution to the moment conditions in equation (1.5.35). For an overidentified model, in fact, not all the moment conditions in equation (1.5.35) can be satisfied together in a finite sample. In other terms, they will not be all set exactly to zero at the same time.

We could get consistent estimates of the parameters by using different subsets of the moment conditions corresponding each time to K instruments, but we would have a loss in efficiency, as we are not exploiting all the available information. Moreover, the estimates would differ according on which K instruments we are considering.

The correct strategy to adopt is to satisfy the moment conditions as best as we can: this is done by making the vector of the empirical moments $\frac{1}{N}\sum_{i=1}^{N} z_i(y_i - x_i\beta) = \frac{1}{N}Z'e$, where e is the vector of the empirical residuals, as small as possible.

We can take an $L \times L$ symmetric positive-definite weighting matrix W and use it to define a quadratic function of the moment conditions that has the form:

$$\begin{aligned} J(\hat{\beta}) &= \left[\frac{1}{N}\sum_{i=1}^{N} z_i(y_i - x_i'\beta)\right]' W \left[\frac{1}{N}\sum_{i=1}^{N} z_i(y_i - x_i'\beta)\right] \\ &= \left[\frac{1}{N}\sum_{i=1}^{N} z_i e_i\right]' W \left[\frac{1}{N}\sum_{i=1}^{N} z_i e_i\right] \\ &= \frac{1}{N^2} e'ZWZ'e \end{aligned} \tag{1.5.38}$$

where the e_i are the empirical residuals[13].

$J(\hat{\beta})$ is defined as the GMM criterion function.

The GMM estimator is the one that minimize $J(\hat{\beta})$: it is

$$\begin{aligned} \hat{\beta}_{GMM} &= \arg\min J(\hat{\beta}) \\ &= \left[\left(\frac{1}{N}\sum_{i=1}^{N} x_i z_i'\right) W \left(\frac{1}{N}\sum_{i=1}^{N} z_i x_i'\right)\right]^{-1} \times \\ &\quad \left[\left(\frac{1}{N}\sum_{i=1}^{N} x_i z_i'\right) W \left(\frac{1}{N}\sum_{i=1}^{N} z_i y_i\right)\right] \\ &= \left(\frac{1}{N^2} X'ZWZ'X\right)^{-1} \frac{1}{N^2} X'ZWZ'y \\ &= \left(X'ZWZ'X\right)^{-1} X'ZWZ'y. \end{aligned} \tag{1.5.39}$$

[13] Since we want to minimize $J(\hat{\beta})$, we can omit the term $\frac{1}{N^2}$, as it does not affect the value of β that minimizes $J(\hat{\beta})$. We can thus equivalently write $J(\hat{\beta}) = e'ZWZ'e$.

The GMM estimators are as many as the possible choices about the weighting matrix W. Also the asymptotic covariance matrix will differ according on W. However, the GMM estimator will be the same for matrices W that differ only by a constant of proportionality.

When the model is exactly identified, i.e. $L = K$, the $\hat{\beta}_{GMM}$ reduces to the simple IV estimator $(Z'X)^{-1}Z'y$, no matter what weighting matrix W is chosen.

Consistency is reached under mild assumptions and regardless the choice of W[14].

The main concern is about the efficiency of the GMM estimator.

When the model is overidentified, the choice of the weighting matrix is crucial and we need to identify the optimal one.

The general formula for the asymptotic variance[15] of the GMM estimator is

$$A\,\text{var}(\hat{\beta}_{GMM}) = N(X'ZWZ'X)^{-1}(X'ZWSWZ'X)(X'ZWZ'X)^{-1} \tag{1.5.40}$$

where S is the covariance matrix of the moment conditions:

$$S = \frac{1}{N}E(Z'\epsilon\epsilon'Z) = \frac{1}{N}E(Z'\Omega Z). \tag{1.5.41}$$

We want to find the optimal W such that the GMM estimator is efficient, i.e. has the smallest asymptotic variance.

Hansen [1982] finds that the optimal weighting matrix that makes the GMM estimator efficient is the one that uses weights for each moments that are the inverse of their variances/covariances. In other words, we weight the moments by the inverse of their variance matrix[16]. The optimal choice is therefore $W = S^{-1}$, so that

$$W_{EGMM} = S^{-1} = (Z'\Omega Z)^{-1}. \tag{1.5.42}$$

[14] Amemiya [1985] provides a formal proof of consistency of the GMM estimator under mild assumptions.

[15] See Cameron and Trivedi [2005] and Baum et al. [2003] for the derivation of the GMM asymptotic variance.

[16] $W = S^{-1}$ or $W = NS^{-1}$ or $W = N^2S^{-1}$ are all equivalent as they differ only by a constant of proportionality. The choice does not affect the GMM estimator, so, for the sake of simplification, we can get rid of the term N in the formula.

The resulting efficient GMM is thus:

$$\begin{aligned}\hat{\beta} &= (X'Z(Z'\Omega Z)^{-1}Z'X)^{-1}X'Z(Z'\Omega Z)^{-1}Z'y \\ &= (X'ZS^{-1}Z'X)^{-1}X'ZS^{-1}Z'y.\end{aligned} \tag{1.5.43}$$

The crucial problem at this stage is that Ω is unknown and therefore S is not known either. The efficient GMM (EGMM) is thus not feasible. We therefore need to make some assumptions on the covariance matrix of ϵ, Ω, in order to obtain a feasible estimator for $E(Z'\Omega Z)$.

If the assume that the errors are i.i.d., then $\Omega = \sigma^2 I$, from which

$$NS = E(Z'\Omega Z) = \sigma^2 E(Z'Z). \tag{1.5.44}$$

As the term σ^2 is a scalar that does not affect the GMM estimator we can ignore it and we can estimate $E(Z'Z)$ by $Z'Z$ so that

$$\hat{W} = \left(\frac{1}{N}Z'Z\right)^{-1}. \tag{1.5.45}$$

The EGMM estimator thus simplifies to

$$\hat{\beta}_{EGMM} = [X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}Z'y, \tag{1.5.46}$$

which is the well known 2SLS.

This estimator is also called the *one-step* GMM estimator as it can be calculated in one step.

When the errors are not necessarily assumed to be i.i.d, the general formulation of the feasible efficient GMM estimator is instead

$$\hat{\beta}_{FEGMM} = (X'Z(Z'\hat{\Omega}Z)^{-1}Z'X)^{-1}X'Z(Z'\hat{\Omega}Z)^{-1}Z'y \tag{1.5.47}$$

where $\hat{\Omega}$ is a consistent estimate of Ω.

In general, we can not assume homoskedasticity and we need to replace the optimal weighting matrix $W = S^{-1}$ by a consistent estimate $\hat{S}^{-1}$.

A consistent estimator for S will be

$$\hat{S} = \frac{1}{N}\left(Z'\hat{\Omega}Z\right) = \frac{1}{N}\left(\sum_{i=1}^{N}\hat{\varepsilon}_i^2 z_i z_i'\right). \tag{1.5.48}$$

The $\hat{\varepsilon}_i$ are the residuals obtained from an initial estimate of the model by any consistent estimator, no matter whether it is efficient or not. The $\hat{\varepsilon}_i$ incorporate the properties we assume for the errors.

We can choose any full rank matrix W for the initial estimate. However, the most common choice is $W = (Z'HZ)^{-1}$, where H is an estimate of Ω that incorporates the assumptions we make on the error terms, such as the form of the heteroskedasticity.

In practice, we adopt a *two-step* GMM procedure in which we estimate the following GMM regression at the first stage

$$\hat{\beta}_{GMM\ \text{first step}} = (X'Z(Z'HZ)^{-1}Z'X)^{-1}X'Z(Z'HZ)^{-1}Z'y \qquad (1.5.49)$$

from which we obtain the residuals $\hat{\varepsilon}_i$. We then use the residuals to form the optimal weighting matrix $\hat{W} = \hat{S}^{-1} = \frac{1}{N}\left(\sum_{i=1}^{N} \hat{\varepsilon}_i^2 z_i z_i'\right)^{-1}$ which will give the efficient GMM estimator, when used in equation (1.5.47). The asymptotic variance of the estimator is

$$A\,\text{var}(\hat{\beta}) = NX'Z(Z'\hat{\Omega}_\beta Z)^{-1}(Z'X)^{-1}. \qquad (1.5.50)$$

The optimal (two-step) GMM estimator has been shown to suffer small sample bias [17]. It does not necessarily perform better that the one-step GMM estimator, though it is more efficient. The intuition behind this is that, if the weighting matrix used in the one-step estimator is independent of the estimates of the parameters, the optimal W matrix in the two-step estimator requires a consistent estimate of the covariance matrix of the moment conditions and implies the replacement of S^{-1} with a consistent estimate $\hat{S}^{-1}$. This replacement is safe asymptotically, while this approximation could lead to inconsistent standard errors in finite samples.

As deeply discussed in Roodman [2009b], the variances of the empirical moments could be poorly estimated in finite samples, with the consequence that there could be an incorrect weighting of the observations. The estimates of the coefficients remain unbiased but the standard errors could appear erroneously good.

[17] See, among others, Altonji and Segal [1996] and Arellano and Bond [1991].

From the point of view of the empiricist, huge differences between the one-step and the two-step estimates may be a signal of a finite sample bias and the question should be addressed further.

Windmeijer [2005] proposes a feasible finite-sample correction for the standard errors in the two-step GMM estimation[18] that now makes it safe do to inference on the results of the two-step estimation. Through Monte Carlo experiments, Windmeijer [2005] shows that the two-step estimator gives less biased estimates and lower standard errors than the one-step estimator: he also finds that his correction improves the two-step standard errors, making them more reliable for inference purposes.

1.6 Testing overidentifying restrictions

As said above, when the equation is exactly identified, all the moment conditions are set exactly to zero by construction, no matter whether the restrictions are truly valid for the population. It is therefore not possible to verify the restrictions we are imposing.

On the contrary, when the model is overidentified we have more moment restrictions L than parameters K, and only K conditions are set equal to zero. We expect to minimize the distance of the remaining $L-K$ restrictions from zero.

If the moment conditions are sufficiently close to zero we have an evidence in support of the exogeneity of the instruments, while if the joint validity of the identifying restrictions is in doubt, we could have a signal of invalid instruments.

We can exploit a specification test proposed by Hansen [1982][19], that tests whether all the restrictions imposed by the model are jointly satisfied. Under the null hypothesis, we assume the joint validity of all the identifying restrictions. In other words, we assume that the empirical moments $\frac{1}{N}\sum_{i=1}^{N} z_i e_i$ are randomly distributed around 0.

[18]See Windmeijer [2005] for technical details on the correction and Roodman [2009a, 2009b] for further discussion.

[19]The IV counterpart of the Hansen test is the well known Sargan [1958] test for the validity of the instruments.

The Hansen J statistic is the value of the GMM objective function presented above

$$J(\hat{\beta}) = \left(\frac{1}{N} \sum_{i=1}^{N} z_i e_i \right)' W \left(\frac{1}{N} \sum_{i=1}^{N} z_i e_i \right) \tag{1.6.51}$$

evaluated at the efficient GMM estimator $\hat{\beta}_{EGMM}$, that is

$$J(\hat{\beta}_{EGMM}) = \left(\frac{1}{N} \sum_{i=1}^{N} z_i e_i \right)' \hat{S}^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} z_i e_i \right) \sim \chi^2_{L-K} \tag{1.6.52}$$

If the null hypothesis is rejected, the specification of the model is not valid, as the observations in the sample do not suit to all the moment restrictions jointly. In other words, the instruments do not satisfy the orthogonality conditions required for them to be valid.

As discussed in Baum et al. [2003], the instruments in this case may either be not truly exogenous or mistakenly excluded from the regression.

Unfortunately, we are not able to determine, on the basis of the Hansen test, which instruments are valid and which ones are not. If the empirical moment conditions are rejected, the GMM is an inconsistent estimator of the parameters of the model.

The Hansen test will be an useful tool to identify a serious problem of instrument proliferation in the GMM estimation of DPD models.

1.7 GMM estimation of DPD models

After having given a general overview of the GMM estimator, we can go back to the AR(1) panel model in equation (1.4.4) and present how the GMM estimator has been applied in DPD estimation.

The GMM estimator is a very useful tool in the context of IV estimation of DPD, as it requires very few general assumptions on the DGP and it is suitable for dynamic models that include fixed effects, potentially endogenous regressors and idiosyncratic errors that are uncorrelated across individuals[20], though it is computationally expensive. The main peculiarity of the GMM estimator for DPD is that it exploits internal instruments, namely the lags of the endogenous variables, still also allowing the use of classical instrumental variables.

1.7.1 The Difference GMM estimator

Moving from the GMM-style instrument matrix proposed by HENR [1988], illustrated in (1.4.24) and (1.4.25), Arellano and Bond [1991] propose a GMM estimator which uses, for the estimation of the model in first-differences

$$\Delta y_{it} = \alpha \Delta y_{i,t-1} + \Delta u_{it}, \tag{1.7.53}$$

all the lags of the dependent variable available at each period as instruments for $\Delta y_{i,t-1}$. The estimator is the Arellano-Bond *Difference GMM* (DIFF) estimator. The DIFF estimator exploits the $(T-1)(T-2)/2$ IV moment conditions[21]:

$$\mathrm{E}[y_{i,t-2}(\Delta y_{it} - \alpha \Delta y_{i,t-1})] = \mathrm{E}(y_{i,t-2}\Delta u_{it}) = 0 \quad \text{for} \quad t = 3, \ldots, T \tag{1.7.54}$$

to construct a sparse instrument matrix for the i^{th} individual of the form

$$Z_{\mathrm{D},\,i} = \begin{pmatrix} y_{i,1} & 0 & 0 & 0 & 0 & 0 & 0 & \ldots & 0 & \ldots & 0 \\ 0 & y_{i,1} & y_{i,2} & 0 & 0 & 0 & 0 & \ldots & 0 & \ldots & 0 \\ 0 & 0 & 0 & y_{i,1} & y_{i,2} & y_{i,3} & 0 & \ldots & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & 0 & \ldots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots & y_{i,1} & \ldots & y_{i,T-2} \end{pmatrix} \tag{1.7.55}$$

where we have one row for each period for which some instrument is available. The first row corresponds to the instruments that are available for the first-differenced equation for period $t = 3$, the last row is for $t = T$. The $Z_{\mathrm{D},\,i}$ matrix has $T-2$ rows with non-zero elements and $(T-2)(T-1)/2$ columns.

[20]The GMM estimator does not require homoskedasticity neither absence of autocorrelation in the residuals.

[21]It is obvious that $\Delta\varepsilon_{it}$ and Δu_{it} are equivalent. In fact, ε_{it} and u_{it} only differ by the individual effect μ_i that is canceled out by the FD transformation.

We have here separated instruments for each lag and for each time period: each column represents a different instrument. The model is obviously overidentified.

The difference moment condition can be written compactly as

$$\mathrm{E}[Z'_{\mathrm{D},i}(\Delta y_{it} - \alpha \Delta y_{i,t-1})] = \mathrm{E}[Z'_{\mathrm{D},i}(\Delta u_i)] = 0 \text{ for } i = 1, \ldots, N \qquad (1.7.56)$$

where $\Delta u_i = (\Delta u_{i3}, \Delta u_{i4}, \ldots, \Delta u_{iT})'$.

The GMM criterion function we need to minimize to find an asymptotically efficient GMM (AEGMM) estimator based on the previous moment conditions is

$$J(\hat{\alpha}) = \left(\frac{1}{N} \sum_{i=1}^{N} \Delta u_i' Z_{\mathrm{D},i} \right) W_N \left(\frac{1}{N} \sum_{i=1}^{N} Z'_{\mathrm{D},i} \Delta u_i \right) \qquad (1.7.57)$$

where W_N is the optimal weighting matrix.

As discussed above, in a two-step procedure, the optimal weighting matrix is

$$W_N = \left(\frac{1}{N} \sum_{i=1}^{N} (Z'_{\mathrm{D},i} \Delta \hat{u}_i \Delta \hat{u}_i' Z_{\mathrm{D},i}) \right)^{-1} \qquad (1.7.58)$$

where the estimates of the first-difference residuals are consistent and come from the initial estimation of the model by a consistent GMM estimator.

As the optimal weighting matrix does not include unknown parameters, when we assume homoskedasticity and absence of serial correlation, we can obtain an asymptotically equivalent GMM estimator in one step by using the weighting matrix

$$W_{\text{one step}} = \left(\frac{1}{N} \sum_{i=1}^{N} (Z'_{\mathrm{D},i} H Z_{\mathrm{D},i}) \right)^{-1}. \qquad (1.7.59)$$

The matrix $W_{\text{one step}}$ does not depend on any previous initial estimate. The H matrix for the Arellano-Bond estimator is the square $(T-2) \times$

$(T-2)$ matrix

$$H = \begin{pmatrix} 2 & -1 & 0 & \dots & 0 & 0 \\ -1 & 2 & -1 & \dots & 0 & 0 \\ 0 & -1 & 2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 2 & -1 \\ 0 & 0 & 0 & \dots & -1 & 2 \end{pmatrix}. \qquad (1.7.60)$$

In fact, if we assume that the residuals are homoskedastic and not serially correlated, we have that $\mathrm{E}(\Delta u_i \Delta u_i') = \sigma_u^2 H$. H is the estimated variance matrix that makes W_N to be the optimal weighting matrix and the estimator efficient. In any case, the DIFF estimator is consistent provided that we set W_N to be a positive definite matrix[22].

The resulting Arellano-Bond GMM estimator estimator is

$$\hat{\alpha}_{GMM\ \mathrm{diff}} = \left((\Delta y_{-1}' Z_D) W_N (Z_D' \Delta y_{-1})\right)^{-1} (\Delta y_{-1} Z_D) W_N (Z_D' \Delta y) \qquad (1.7.61)$$

where $\Delta y_i = (\Delta y_{i,3}, \Delta y_{i,4}, \dots, \Delta y_{i,T})'$, $\Delta y_{i,-1}$ is the vector that includes the first lag of Δy_i and $Z'\Delta y = \sum_{i=1}^{N} Z_{D,i} y_i$[23].

The DIFF GMM estimator is consistent for $N \to \infty$ and for fixed T and, in general[24], it is consistent also for $T \to \infty$.

Moreover, the DIFF GMM is easy to compute as the orthogonality conditions in (1.7.54) are all linear in α[25].

However, the number of orthogonality conditions in (1.7.54) rapidly gets large as T increases: the growth in the number of the moment conditions is quadratic in T, so that the instrument count in the DIFF GMM becomes quickly very high.

Arellano and Bond [1991] show that the DIFF estimator has only a very limited finite sample bias and has smaller variance than the IV Anderson and Hsiao [1981, 1982] estimators. A drawback they stress, how-

---

[22]For example, by choosing $W_N = I_N$, that is the identity matrix, we preserve consistency. Blundell and Bond [1998a] set H=I.

[23]See Blundell, Bond and Windmeijer [2000] for the full derivation of equation (1.7.61).

[24]See Alvarez and Arellano [2003].

[25]See Bond [2002] for a thorough analysis of the advantages of having linear moment conditions.

ever, is that the two-step DIFF estimator gives downward biased estimated standard errors, especially in finite samples. The two-step estimator takes advantage of the Windmeijer [2005] correction for the two-step estimated asymptotic standard errors.

Anyway, the DIFF GMM estimator has been found to suffer a severe finite sample bias, in particular when the lags of the dependent variable are only weakly correlated with the endogenous first differences at a subsequent period[26].

We have in this case an identification problem, as the instruments for the first-differenced equation are weak. The problem shows up in an AR(1) panel model when the series is very persistent, i.e. when the autoregressive parameter gets close to one, or when the variance of the individual effects μ_i is much larger than the variance of the idiosyncratic error term u_{it}[27].

Moreover, Blundell and Bond [1998a] show on simulated and on real data that the weak instruments problem is particularly accentuated when the temporal dimension T is small. Bond, Hoeffler and Temple [2001] analyze empirically this problem in the context of growth regressions and compare the performance of the DIFF GMM with that of standard estimators for the AR(1) panel model, such as OLS and WG, and confirm the findings of Blundell and Bond [1998a].

1.7.2 The System GMM estimator

It is possible to add more restrictive assumptions on the initial conditions in order to formulate additional orthogonality conditions that make more valid instruments available and efficiency gains possible.

Blundell and Bond [1998a] show that, under a mild stationarity assumption for the initial conditions y_{i1}, we can exploit a system GMM (SYS GMM) estimator that uses the lagged first-difference of the y_{it} series, Δy_{it}, as instruments for the equation (1.4.4) in levels, in addition to the use of the lagged levels of y_{it} as instrument for the first-differences equations.

[26]See Blundell and Bond [1998a].

[27]See also Bond [2002] for further discussion.

More specifically, under the assumption that the deviations of the initial conditions y_{i1} from their long-run mean $\mu_i/(1-\alpha)$ are uncorrelated with the long-run mean $\mu_i/(1-\alpha)$ itself[28], we have that $E(\Delta y_{i,2}\eta_i) = 0$ for $i = 1, 2, ..., N$[29].

Exploiting the standard mild assumption that $E(\Delta u_{it}\mu_i) = 0$ for $i = 1, 2, \ldots, N$ and $t = 3, 4, \ldots, T$, which requires that any change in the idiosyncratic term is not correlated with the individual effect, we have $T-2$ additional linear orthogonality conditions:

$$E\left(\Delta y_{i,t-1}(\mu_i + u_{it})\right) = E\left(\Delta y_{i,t-1}\varepsilon_{it})\right) = 0 \text{ for } t = 3, 4, \ldots, T. \quad (1.7.62)$$

These new moment conditions allow to use the lagged first difference of $y_{i,t}$, i.e. Δy_{t-1} for $t = 3, \ldots, T$, as instrument for $y_{i,t-1}$ in the equation (1.4.4) in levels.

The instrument matrix for the equations in levels can be defined as

$$Z_{\text{L},\,i} = \begin{pmatrix} \Delta y_{i2} & 0 & 0 & \ldots & 0 & \ldots & \vdots \\ 0 & \Delta y_{i3} & \Delta y_{i2} & \ldots & 0 & \ldots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \Delta y_{i,T-1} & \ldots & \Delta y_{i2} \end{pmatrix}, \quad (1.7.63)$$

where the first row corresponds to $t = 3$ and the last one to $t = T$.

The additional moment restrictions can be written compactly as

$$E(Z'_{\text{L},\,i}\varepsilon_i) = 0 \text{ with } \varepsilon_i = \begin{pmatrix} \varepsilon_{i3} \\ \varepsilon_{i4} \\ \vdots \\ \varepsilon_{iT} \end{pmatrix}. \quad (1.7.64)$$

The GMM estimator for the level equation is

$$\hat{\alpha}_{\text{level}} = \left(y'_{-1}Z_L W_N Z'_L y_{-1}\right)^{-1}(y'_{-1}Z_L)W_N Z'_L y \quad (1.7.65)$$

[28] In other words we want $E\left[\left(y_{i1} - \frac{\mu_i}{1-\alpha}\right)\mu_i\right] = 0$. The deviations of the initial conditions from the long-run means are required not to be correlated with the fixed effects. See Blundell and Bond [1998a] for the original formulation of the assumption, Baltagi [2008], Bond et al. [2001] and Roodman [2009b] for further discussion.

[29] The stationarity assumption above implies that y_{it} converges to its long-run mean from period $t = 2$ onward. Blundell and Bond [1998a] and Roodman [2009b] provide a proof of this relation. As a consequence, we have that $E(\Delta y_{i,2}\eta_i) = 0$.

where y_{-1} is the vector which include the first lag of y_i for each period.

Blundell and Bond [1998a] suggest a system GMM estimator which exploits both sets of moment conditions in equations (1.7.54) and (1.7.62).

The full set of conditions is therefore

$$\begin{aligned} E\left(y_{i,t-2}\Delta\varepsilon_{it}\right) &= 0 \text{ for } t = 3, \ldots, T \\ E\left(\varepsilon_{it}\Delta y_{i,t-1}\right) &= 0 \text{ for } t = 3, \ldots, T. \end{aligned} \tag{1.7.66}$$

The system consists of $(T-2)$ stacked equations in first-differences and $(T-2)$ stacked equations in levels for $t = 3, \ldots, T$.

The full instrument matrix is therefore

$$Z_{\mathrm{SYS},\,i} = \begin{pmatrix} Z_{\mathrm{D},\,i} & 0 & 0 & \ldots & 0 \\ 0 & \Delta y_{i2} & 0 & \ldots & 0 \\ 0 & 0 & \Delta y_{i3} & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \ldots & \Delta y_{i,T-2} \end{pmatrix} = \begin{pmatrix} Z_{\mathrm{D},\,i} & 0 \\ 0 & Z_{\mathrm{L},\,i} \end{pmatrix}. \tag{1.7.67}$$

The system GMM estimator is therefore

$$\hat{\alpha}_{SYS} = \left(q'_{-1}Z_{SYS}W_N Z'_{SYS}q_{-1}\right)^{-1}\left(q'_{-1}Z_{SYS}WZ'_{SYS}q_i\right) \tag{1.7.68}$$

where $q_i = (\Delta y_i', y_i')$[30].

The additional moment conditions for the equations in levels and hence the validity of the lagged differences as instruments were originally proposed by Arellano and Bover [1995]. Blundell and Bond [1998a] develop the Arellano and Bover [1995] approach further, show that the additional moment restrictions remain valid also in case of weak instruments and explicitly define the assumptions on the initial conditions that need to be satisfied for the SYS estimator to be valid.

In the SYS GMM estimator, we use only the most recent lag of the first differences as instruments for the equations in levels, as we are already using the lagged levels as instruments for the equations in first

[30]See Blundell, Bond and Windmeijer [2000] for the full derivation of equations (1.7.65) and (1.7.68).

differences. The use of additional lagged first-differences would result in redundant moment condition[31].

Blundell and Bond [1998a] show that the SYS GMM estimator is largely more efficient than the DIFF GMM, especially as $\alpha \to 1$ and as the ratio $\sigma^2_\mu / \sigma^2_\varepsilon$ increases.

When the series are persistent, the lagged levels of y_{it} are only weak instruments for the equations in first-differences, while the lagged first-differences remain informative, and hence valid instruments, for the equations in levels.

Blundell, Bond and Windmeijer [2000] show that, in DPD models that includes also weakly exogenous regressors, the SYS GMM performs much better than the DIFF GMM, both in terms of precision and finite sample bias. Recently, however, Bun and Windmeijer [2010] find that a weak instruments problem could also arise in the level GMM estimator, and hence in the SYS estimator.

A great attention has been paid in the literature to the choice of the weighting matrix for the one-step GMM procedure and for the first-step in a two-step procedure [32].

Blundell, Bond and Windmeijer [2000] e.g. use

$$W = \left(\frac{1}{N} \sum_{i=1}^{N} (Z'_{\text{sys}} M Z_{\text{sys}}) \right)^{-1} \tag{1.7.69}$$

where $M = \begin{pmatrix} H & 0 \\ 0 & I_{T-1} \end{pmatrix}$.

Blundell and Bond [1998a] instead use in first-step estimation

$$M = \begin{pmatrix} I_{T-1} & 0 \\ 0 & I_{T-1} \end{pmatrix} = I_{2T-2}, \tag{1.7.70}$$

which gives the standard IV estimator.

Both the weighting matrix are easily implementable in Stata.

Blundell and Bond [1998a] show in Monte-Carlo simulations that the two-step and the one-step SYS GMM estimators give relatively similar

[31] See Arellano and Bover [1995], Blundell and Bond [1998a], Bond [2002] and Roodman [2009a] for a deep discussion on the issue.

[32] See, among the others, Windmeijer [2005] and Kiviet [2007a].

results and suggest that, when homoskedasticity is assumed, the choice of the M matrix in the weighting matrix W is not crucially relevant.

On the other hand, as a warning to the empiricist, it is worth noticing that the choice of the weighting matrix is not irrelevant in empirical analysis on real data: different matrices can lead to very different results[33].

1.8 Instrument proliferation

As mentioned above, the GMM approach is particularly appealing in the estimation of DPD models, as it involves very few and general assumptions about the DGP, accounts for unobserved heterogeneity, makes "internal" instruments available and performs better than the standard estimators for DPD as FE, OLS and IV.

However, the number of moment conditions generated under both the DIFF GMM and the SYS GMM rapidly increases and the instrument count hence quickly becomes large relative to the sample size.

In particular, when the panel AR(1) in (1.4.4) is considered and no other regressor is included in the model, we have $(T-2)(T-1)/2$ instruments in levels in the DIFF GMM matrix $Z_{D,\,i}$ and $(T-2)$ additional instruments in first-difference in the SYS GMM instrument matrix.

Even when only the DIFF GMM instruments are used, their number grows quadratically as T increases.

The issue of instrument proliferation in GMM estimation is not new and has been paid a particular attention by Ziliak [1997], Altonji and Segal [1996] and Bowsher [2002].

Ziliak [1997] finds that, as the number of moment conditions expands, the GMM estimator for panel data is downward biased; the gain in efficiency, made possible by the adoption of a GMM approach, is negatively balanced by the induced bias.

Bowsher [2002] analyzes the effects of the use of too many moment restrictions in DPD: it is shown that the Sargan test of overidentifying restrictions based on the full DIFF GMM instrument matrix under-rejects

[33] As a practical strategy, it is enough to change the option `h()` in the `xtabond2` command in Stata and to check how the results are sensitive to the weighting matrix chosen.

the null hypothesis of joint validity of the instruments when an excessive number of moment restrictions is being tested. Bowsher also gives a practical indication: reducing the instrument count for the differenced equations is a simple and effective strategy to improve the performance of the Sargan test, especially in small samples.

Roodman [2009b] returns to the problem of instrument proliferation and suggests alternative solutions to face it, especially in applied works. He presents two kinds of practical problems that arise from the exploitation of too many moment conditions in small samples and from the use of a very sparse instrument matrix.

On one hand, there is the severe and well-known risk of an overfitting of the endogenous variables when the instruments are too many. Roodman [2009b] provides an intuitive insight on this issue. Ziliak [1997] presents the problem referring to the GMM estimation of panel data. He focuses on static models and finds that, when the number of instruments gets large, the GMM estimator is biased toward the Pooled OLS estimator[34]. Arellano [2003] treats analytically the overfitting bias in dynamic panel models when T grows. When the instrument matrix is so large and sparse, the instrument may be very weak, thus implying that some relevant effects may not be identified.

Unfortunately, it is not possible to test for the presence of an overfitting bias: it therefore becomes crucial to check whether the results of the GMM estimation are robust to the reduction in the number of the moment conditions exploited.

The second sort of problems is specific to the two-step FEGMM. The optimal weighting matrix is estimated using sample moments: when the number of moment restrictions imposed is large, it is hard to estimate the elements of the optimal weighting matrix, so that the estimates are often imprecise. The estimates remain however consistent, even though efficiency is often affected. When the optimal weighting matrix is imprecisely estimated, the standard errors in the two-step GMM are seriously

[34]See also Tauchen [1986] and Windmeijer [2005] for a simulation-based discussion on the problem.

biased downward when the instrument count is high[35]. The Windmeijer [2005] correction is necessary in order to make reliable inference on the two-step results.

The most evident symptom of instrument proliferation is a not reliable p-value of the Hansen test for overidentifying restrictions. As stressed in Bowsher [2002] and Andersen and Sorensen [1996], an increase in the number of moment restrictions often invalidates the test. Roodman [2009b] provides an extensive discussion on the effects of instrument proliferation on the Hansen test. It is not uncommon to find an implausible p-value of 1 when T is moderately large. The risk intrinsic in this implausible value is that of a false positive: in fact, we would be inclined to consider the moment restrictions as jointly satisfied when, in reality, the results are vitiated by biased estimates of the two-step standard errors that enter the Hansen test formula. In fact, also the Hansen test is based on the covariance matrix of the moment conditions: when this is estimated imprecisely, the test is strongly affected.

1.9 Solutions to the problem of instrument proliferation

As just said, instrument proliferation is a common risk in GMM estimation of DPD models, and it is particularly serious in small samples.

The use of the whole set of instruments in GMM estimation gives significant efficiency gains, as we are exploiting all the information we have, but at the huge costs we have just discussed.

However, no formal test is available to detect an instrument proliferation problem. It is therefore difficult to draw practical indications on which is a safe number of instruments. A rule of thumb is to keep the instrument count lower than the number of individuals in the sample.

Roodman [2009a, 2009b] suggests a second rule of thumb: we should worry whenever we get a p-value for the Hansen test greater than 0.25.

[35] See Roodman [2009b] and Windmeijer [2005]: the former contribution focuses on the risk of severe bias in the estimate of the two-step standard errors, especially in case of small samples.

Two main techniques have been proposed in the literature to stem the risks implied by instrument proliferation. Both of them can be seen in the light of interpretable meaningful moment conditions or in terms of deterministic transformations of the instrument matrices[36].

The first strategy is the use of a limited instrument set. It provides that only certain lags, say l lags, of the dependent variable are used as instruments rather than all the available lags.

As Roodman [2009a, 2009b] clarifies, we still have separated instruments for each period and for each lag from 2 to l, but we limit the number of lags so that the instrument count is linear in T. In other words, if we want to limit the lag depth of the instruments to be used for $\Delta y_{i,t-1}$ to the first and the second lag available, say $l = 2$, we have an instrument matrix of the form

$$Z^{\mathrm{L}}_{\mathrm{D},\, i} = \begin{pmatrix} y_{i1} & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & y_{i2} & y_{i1} & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & y_{i3} & y_{i2} & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{1.9.71}$$

where the first row refers to $t = 3$.

We can think to the limited instrument matrix as the result of the multiplication of the matrix $Z^{\mathrm{D}}_{\mathrm{D},\, i}$ by a transformation matrix T^{L}, which is a block matrix of identity matrices of dimension l (maximum lag depth) that are separated by rows of zeroes corresponding to the excluded lags:

$$Z^{\mathrm{L}}_{\mathrm{D},\, i} = Z_{\mathrm{D},\, i}T^{\mathrm{L}} = Z_{\mathrm{D},\, i} \left(\begin{array}{c|cc|cc|cc|c} 1 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ \hline 0 & 1 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & \dots \\ \hline 0 & 0 & 0 & 1 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & \dots \\ \hline \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \end{array}\right). \tag{1.9.72}$$

By limiting the lag depth, the number of instruments is reduced from $C = (T-2)(T-1)/2$ in the untransformed matrix to $C_{\mathrm{L}} = C - (T-2-L)(T-1-l)/2$.

[36]The transformation matrix consists in both cases of zeroes and ones. See below and Mehrhoff [2009] for the exact form of the transformation matrices.

This strategy obviously does not affect the number of instruments in first-differences for the equation in levels: in fact, in order to avoid redundancy in the instruments, we keep only one available first-difference.

The truncation of the instrument set in the lag depth has become very common in empirical works and it has been the strategy for checking the robustness of the estimates to the instrument reduction [37]. Alvarez and Arellano [2003] show that the DIFF GMM estimator is still consistent when the instrument set is truncated. Alfaro [2008] finds that this strategy affects the efficiency of the estimates and increases the impact of small samples biases.

We should be aware that, though very simple to be implemented, the truncation of the instrument set is generally arbitrary. The lag depth is commonly limited to one or two first available lags. This choice in general has not a meaningful explanation: it is simply an arbitrary strategy to reduce the instrument set. It should be a common practice in empirical works to always discuss the choice of the maximum lag depth and give an economic explanation to it.

The second technique provides the use of a collapsed instrument matrix. The instruments are aggregated in smaller sets by squeezing horizontally the instrument matrix, through addition over the columns.

As explained in Roodman [2009b], we formulate different orthogonality conditions instead of those in (1.7.54) and (1.7.62):

$$E(y_{i,t-l}\Delta\varepsilon_{it}) = 0 \text{ for each } l \geq 2. \tag{1.9.73}$$

We impose the same condition for each $t = 3, ..., T$, so that the moment restrictions do not need to hold for any period anymore, but only for each lag. We restrict some subsets of the instrument matrix, corresponding to the same lag, to have the same coefficient. Intuitively, we impose that the same lag, say e.g. $y_{i,t-2}$, has always the same impact on y_{it}, no matter the exact point in time t we are considering. We make one instrument for each lag distance and we substitute any missing value with zero.

The idea of applying each moment condition to all the available periods instead of using each moment condition only for a given period

[37] See Windmeijer [2005] and Bowsher [2002], among the others.

is first suggested in Calderon, Chong and Loayza [2002]. Their strategy provides that the number of orthogonality conditions is independent of T, while in the standard approach, with a separate moment condition, and hence a separate instrument, for each time period, the number of restriction is more than proportional with respect to T.

This approach is then used in Beck and Levine [2004] in order to reduce the overfitting bias due to instrument proliferation.

Roodman [2009b] independently suggests the same strategy and devises a procedure to implement collapsed GMM in Stata, which is now a common practice in empirical works[38].

When this strategy is adopted, the collapsed matrix for the equation in first-differences is the following:

$$Z^{C}_{D,\,i} = \begin{pmatrix} 0 & 0 & 0 & 0 & \dots \\ y_{i1} & 0 & 0 & 0 & \dots \\ y_{i2} & y_{i1} & 0 & 0 & \dots \\ y_{i3} & y_{i2} & y_{i1} & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{1.9.74}$$

with the first row referring to $t = 2$.

We can think to $Z^{C}_{D,\,i}$ as the result of the multiplication of $Z_{D,\,i}$ by a transformation matrix that consists of identity matrices of increasing dimension stacked one over the other with blocks of zero matrices to the right:

$$Z^{C}_{D,\,i} = Z_{D,\,i}T^{C} = Z_{D,\,i} \begin{pmatrix} 1 & 0 & 0 & 0 & \dots \\ 1 & 0 & 0 & 0 & \dots \\ 0 & 1 & 0 & 0 & \dots \\ 1 & 0 & 0 & 0 & \dots \\ 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{1.9.75}$$

The instrument count reduces from $C = (T-2)(T-1)/2$ to $C^{C} = T - 2$.

[38]See Roodman [2009a, 2009b] for a discussion about the rationale of the collapsing and for instructions about the implementability of the procedure in Stata.

Similarly, with respect to the instruments in first-differences for the equation in levels, we have a collapsed instrument matrix of the kind

$$Z^{C}_{L,\,i} \begin{pmatrix} 0 \\ \Delta y_{i2} \\ \Delta y_{i3} \\ \Delta y_{i4} \\ \vdots \end{pmatrix} . \tag{1.9.76}$$

The collapsing makes the instrument count linear in T: it has the advantage, over the truncation of the lag depth, that it retains much more information as it does not involve the drop of any lag. Provided that we consider the belief imposed by the new moment conditions as realistic, we do not do any arbitrary choice about the instruments to be kept in the estimation.

It is also possible to combine the two strategies by collapsing the instrument matrix where the lag depth has already been limited.

If the maximum lag depth is $l = 2$ we obtain, e.g.

$$Z^{LC}_{D,\,i} = \begin{pmatrix} 0 & 0 \\ y_{i1} & 0 \\ y_{i2} & y_{i1} \\ y_{i3} & y_{i2} \\ y_{i4} & y_{i3} \\ \vdots & \vdots \end{pmatrix} . \tag{1.9.77}$$

The $Z^{LC}_{D,\,i}$ can be thought as the result of the multiplication of $Z_{D,\,i}$ by a transformation matrix of the form

$$T^{LC} = \left(\begin{array}{cc} 1 & 0 \\ \hline 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ \hline 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ \hline \vdots & \end{array}\right) . \tag{1.9.78}$$

The number of instruments is now equal to the maximum lag depth we set: it is therefore linear in the lag depth but invariant to T.

In the absence of formal tests to detect and address the problem of instrument proliferation, we believe that robustness checks based on the response of the estimators to the transformation of the instrument matrix could be able to give some "rule of thumb" indication.

It should be stressed once again that both these strategies involve a certain degree of arbitrariness and the convincement that the beliefs about the new orthogonality condition are reasonable.

1.10 Conclusions

This chapter set out to analyze the state of the art in estimation of DPD models. We reviewed the standard estimators for DPD models and we illustrated in detail the GMM approach to DPD estimation. We tried to present the evolution of the methodologies for DPD in terms of a continuous improvement in the ability to identify potentially risky econometric issues and to develop strategies to tackle them.

Nowadays, most of the empirical literature estimates DPD models by GMM: as discussed in our presentation, this success is mainly due to the flexibility of the estimator, to the availability of internal instruments and to the easy implementation of GMM estimation in the most popular econometric softwares.

We aimed at showing that GMM estimation, though reasonably very appealing, is not free of faults and it is not the panacea to all the issues that arise in this context. The facets of this estimators are many and can lead to very different estimates in empirical applications.

In particular, the GMM estimator for DPD is not safe from severe problems due to the instrument proliferation as the number of periods gets moderately large. We discussed extensively the problems we are forced to face when the instrument count gets large and the strategies presented in the literature to address this issue.

We hope that we were able to cast some light on the complexity of the GMM approach and on the risks that may arise in this context if we take these issues too lightly.

Aware of all the potential drawbacks we could face in GMM estimation of DPD, we suggest on doing as many robustness checks as we can about the sensitivity of the estimates to the alternative specifications of the estimator in empirical analysis.

What is easy implementable in a software is not always safe and devoid of consequences.

Chapter 2

Instrument Proliferation in GMM Estimation of Dynamic Panel Data: new strategies for applied research

2.1 Introduction

Dynamic panel data (DPD) have become very popular in the last two decades, thanks in particular to the increasing availability of panel dataset both at a micro level (e.g. data for individuals, households or firms) and at a macro level (e.g. data for Regions or Countries). The use of dynamic models in macroeconomics dates back to many decades ago, while it is relatively recent in microeconomics. The possibility of including some kind of dynamics also in a microeconomic framework has become very appealing and it is now a common practice to estimate dynamic models in empirical analysis in most microeconomic fields.

In particular, the GMM estimator, in the Arellano-Bondell [1991] and Arellano-Bover [1995] / Blundell-Bond [1998a] formulations, has gained a leading role among the DPD estimators, mainly due to its flexibility and to the very few assumptions about the data generating process it requires. Most of all, while preventing from the rising of the well known DPD bias (see Nickell [1981]) and of the trade off between lag depth and sample size[1], the GMM estimator also gives the opportunity to account for individual time-invariant effects and for potential endogeneity of regressors. The implementations of ad hoc procedures in many statistical softwares and consequently the availability of "buttons to push" have done the rest of the job.

The GMM estimator however is not the panacea for all the drawbacks of the previously proposed DPD estimators: it is in fact not free of faults. Instrument proliferation, among the others, is a severe issue in the application of the GMM estimator for DPD models and needs to receive more attention than what it has been done so far. The potential distortions in the estimates by OLS, IV estimators and GMM when the instrument count gets larger and larger has been treated extensively in the literature[2], but not particular attention has been paid to this issue in Difference and System GMM estimation of DPD.

Though both versions of the GMM estimator are designed for a *large N-small T* framework, and though the time dimension in panel datasets remains well below that of a typical time series, it is well-known that the number of moment conditions increases exponentially with T and gets rapidly large relative to the sample size. Consequently, as extensively discussed in Roodman [2009b], the excessive number of instruments can overfit endogenous variables, gives an imprecise estimated variance matrix of the moments and lowers the power of useful specification tests such as the Sargan [1958] / Hansen [1982] test of overidentifying restrictions.

Unfortunately, the problem of instrument proliferation is only rarely detected and addressed in empirical analysis with the consequent risk of drawing misleading conclusions about the coefficient estimates.

[1] This former problem is instead an intrinsic and unavoidable characteristic of the Anderson-Hsiao [1981, 1982] 2SLS estimator for DPD.

[2] See, among the others, Ziliak [1997] and Bowsher [2002]. See also Chapter 1 for an extensive discussion on the issue.

Moreover, there is not a clear indication on how many instruments are too many and on which is a reasonable number of instruments to be used in empirical works.

Roodman [2009b] presents two alternative and complementary techniques to reduce the number of instruments in GMM applications: the collapsing of the instrument matrix and the reduction of the lag depth of the instruments included. Both solutions make the instrument count linear in T: through the former, we include different instruments for each lag but for all the available periods together; the latter consists of the inclusion as instruments of only few lags among all the available ones. Both techniques, separately or combined together, have gained popularity thanks to their direct implementability in the statistical software STATA through the `xtabond2` command[3] and are now commonly used in empirical works.

Both these techniques involve a certain degree of arbitrariness from the researcher which is asked either to choose how many lags to include among the instruments or to trust the restrictions that are imposed when the instrument matrix is collapsed[4].

Despite some attempts to investigate the performance of the GMM estimators when instrument reduction techniques are employed, the literature in this fields lacks of exhaustive experiments that compare extensively these strategies and their robustness to different settings of the parameters in the simulation model. Roodman [2009b] presents only a limited Monte Carlo experiment to compare the two techniques but restricts the analysis to the System GMM estimator and to a specific parameter setting. Mehrhoff [2009] instead bounds his experiment to the Difference estimator, that is less exposed to instrument proliferation dangers.

The first aim of the chapter is to fill the gap in the literature by running extensive Monte Carlo experiments in order to compare the performance of the difference and system GMM estimators when instrument reduction techniques are adopted.

[3]This is the user-written command for dynamic panel data estimation developed by D. Roodman. See Roodman [2009a] for details on the use and the syntax of the command.

[4]The two instrument reduction techniques are presented in details in Chapter 1.

In addition to these consolidated techniques, drawing on Mehrhoff [2009], we introduce a different data-driven technique for the reduction of the instrument count in DPD GMM estimation: we factorize the instrument matrix through the principal component analysis (PCA) and use the PCA scores as a new set of instruments[5]. We adopt alternative criteria in order to select the number of retained components and test the performance of this reduction technique relative to the alternative solutions proposed by Roodman [2009b].

A further contribution of the chapter is therefore an extensive investigation, through wide Monte Carlo simulations, of the properties of the GMM estimators when this new strategy is adopted.

In the remainder of the chapter we proceed as follows: in section 2 we explain the meaning of the factorization of the instrument matrix and discuss the rationale of applying the PCA on the instrument set; section 3 reports the results of an extensive Monte Carlo simulations through which we compare a number of instrument reduction techniques; in section 4 we present an empirical application that closely follows Roodman [2009b]; section 5 draws the conclusions and indicate practical hints for the empirical analysis; the Appendix runs through the technical details of the PCA.

2.2 Reducing the instrument count in GMM estimation

We consider here the standard general DPD model:

$$y_{it} = \alpha y_{i,t-1} + \mathbf{x}'\boldsymbol{\beta} + \varepsilon_{it} \tag{2.2.1}$$

$$\varepsilon_{it} = \eta_i + v_{it}. \tag{2.2.2}$$

This is a one-way error component model where $\mathbf{x}$ is a vector of potentially endogenous regressors, the η_i are the fixed effects (time-invariant

[5]Mehrhoff [2009] sketches the idea of applying the PCA on the GMM-style instrument matrix but he only limits its analysis to the difference GMM framework with no additional endogenous regressors and arbitrarily choses the number of components to be kept in the analysis.

individual effects) and v_{it} is a zero-mean idiosyncratic error.

For the sake of simplicity, we assume here that the only endogenous regressor is the lagged dependent variable so that the variables in **x**, if any, are strictly exogenous[6]. What follows can be easily generalized to the case of additional endogenous regressors in **x**.

The Arellano-Bond and Arellano-Bover / Blundell-Bond estimators[7] are linear GMM estimators for the model in first differences or in levels where the instrument matrix **Z** includes the lagged values of the endogenous variables only or also the lagged first differences of the endogenous variables[8]. In the standard framework, the columns of the instrument matrix **Z** correspond respectively to two different sets of meaningful moment conditions.

In particular, the Arellano-Bond DIFF GMM estimator exploits the following $(T-2)(T-1)/2$ moment conditions for the equation (2.2.2) in first differences:

$$E[(\mathbf{Z_i}^{\text{diff}})'\Delta v_i] = E[y_{i,t-l}(\Delta y_{it} - \alpha\Delta y_{i,t-1} - \Delta\mathbf{x}'\beta)] = 0 \text{ for } t \geq 3, l \geq 2 \tag{2.2.3}$$

where the instrument matrix $\mathbf{Z}^{\text{diff}}$, that satisfies the moment restrictions in (2.2.3), has the well known form:

$$\mathbf{Z}_i^{\text{diff}} = \begin{pmatrix} y_{i,1} & 0 & 0 & 0 & \dots & 0 & \dots & 0 \\ 0 & y_{i,1} & y_{i,2} & 0 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & y_{i,1} & \dots & y_{i,T-2} \end{pmatrix} \begin{matrix} (t=3) \\ (t=4) \\ \vdots \\ (t=T-2). \end{matrix} \tag{2.2.4}$$

[6]It is common in DPD analysis to consider only the lagged dependent variable as truly endogenous and therefore to instrument it with its lags. When some additional regressors are predetermined but not strictly endogenous, a large practice is to instrument them with themselves, so that they are treated as exogenous variables. Arellano and Bond [1991] and Blundell and Bond [1998a] exploit lags as instruments only for the lagged dependent variable, while they instrument the other regressors by themselves so that they treat them as exogenous variables, though they acknowledge they are likely not to be strictly exogenous.

[7]These estimators are presented in details in Chapter 1.

[8]We use **Z** to define a general instrument matrix for DPD GMM estimation. **Z** can stand for the untransformed matrix, the collapsed matrix or the limited matrix of instrument. When we need to indicate more precisely the matrix we are considering, we use specific superscripts to denote it.

The Blundell-Bond SYS GMM estimator also exploits the additional $T-2$ orthogonality conditions for the equation (2.2.2) in levels:

$$E[\Delta y_{i,t-1}v_{it}] = E[\Delta y_{i,t-1}(y_{it} - \alpha y_{i,t-1} - \mathbf{x}'\beta)] = 0 \text{ for } t \geq 3 \qquad (2.2.5)$$

that give origin to the instrument matrix for the equation in levels[9]:

$$\mathbf{Z}_i^{\text{lev}} = \begin{pmatrix} \Delta y_{i2} & 0 & \dots & 0 \\ 0 & \Delta y_{i3} & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & \Delta y_{i,T-2} \end{pmatrix} \begin{matrix} (t=3) \\ (t=4) \\ \vdots \\ (t=T-1). \end{matrix} \qquad (2.2.6)$$

The full instrument matrix for the system GMM estimator will thus be:

$$\mathbf{Z}_i^{\text{sys}} = \begin{pmatrix} \mathbf{Z}_i^{\text{diff}} & 0 & 0 & \dots & 0 \\ 0 & \Delta y_{i2} & 0 & \dots & 0 \\ 0 & 0 & \Delta y_{i3} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \dots & \Delta y_{i,T-2} \end{pmatrix} = \begin{pmatrix} \mathbf{Z}_i^{\text{diff}} & 0 \\ 0 & \mathbf{Z}_i^{\text{lev}} \end{pmatrix}. \qquad (2.2.7)$$

2.2.1 Limiting and collapsing the instrument set

As discussed in Roodman [2009b], when we collapse the instrument set we impose the following moment restrictions for the model in first differences, instead of those in equation (2.2.3):

$$E[y_{i,t-l}\Delta v_{it}] = 0 \text{ for } l \geq 2 \qquad (2.2.8)$$

so that we impose the same condition for all t and we create an instrument for each lag distance rather than for each time period and each lag.

[9]It is standard in SYS GMM estimation to consider only the first lagged difference for each time period as instrument for the equation in levels rather than all the available lags of the first differences. This is to avoid redundancy, as all the lags already enter the estimation process through the instrument set for the equations in first differences. See Roodman [2009b] and Bond [2002] for further discussion on this issue.

The collapsed instrument matrix for the equation in first differences has the form:

$$\mathbf{Z}_i^{\text{diff, C}} = \begin{pmatrix} 0 & 0 & 0 & 0 & \dots \\ y_{i1} & 0 & 0 & 0 & \dots \\ y_{i2} & y_{i1} & 0 & 0 & \dots \\ y_{i3} & y_{i2} & y_{i1} & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{matrix} (t=2) \\ (t=3) \\ (t=4) \\ (t=5) \\ \vdots \end{matrix} \tag{2.2.9}$$

Similarly, the collapsed matrix for the equation in levels is:

$$\mathbf{Z}_i^{\text{lev, C}} = \begin{pmatrix} 0 \\ \Delta y_{i2} \\ \Delta y_{i3} \\ \Delta y_{i4} \\ \vdots \end{pmatrix} \begin{matrix} (t=2) \\ (t=3) \\ (t=4) \\ (t=5) \\ \vdots \end{matrix} \tag{2.2.10}$$

The collapsed matrix for the system estimator will thus be:

$$\mathbf{Z}_i^{\text{sys, C}} = \begin{pmatrix} \mathbf{Z}_i^{\text{diff, C}} & 0 \\ 0 & \mathbf{Z}_i^{\text{lev, C}} \end{pmatrix}. \tag{2.2.11}$$

When instead we limit the lag depth, we truncate the moment restrictions and exploit the conditions in equation (2.2.3) only for $2 \leq l \leq M$ where M is the maximum lag depth we consider. The limited instrument matrix for the equation in first differences will be:

$$\mathbf{Z}_i^{\text{diff, L}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ y_{i1} & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & y_{i2} & y_{i1} & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & y_{i3} & y_{i2} & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{matrix} (t=2) \\ (t=3) \\ (t=4) \\ (t=5) \\ \vdots \end{matrix} \tag{2.2.12}$$

The truncation in the lag depth has no impact on $\mathbf{Z}_i^{\text{lev}}$, as it already includes only the first lag available. By limiting arbitrarily the lag depth, we drop from the instrument set $\mathbf{Z}$ all the information about the lags greater than M; by collapsing the instrument matrix, we retain a lot more information as none of the lags is actually dropped, though restrictions are imposed on the coefficients of subsets of instruments so that we only generate a single instrument for each lag.

2.2.2 Extracting principal components from the matrix of instruments

In order to face the problem of instrument proliferation, we can also use a strategy that involves a stochastic transformation of the instrument set: we can extract the principal components from the instrument matrix $\mathbf{Z}$.

The adoption of principal components analysis (PCA) or factor analysis to extract a small number of factors from a large set of variables has become popular in other fields of economic analysis. The seminal works by Stock and Watson [1998, 2002a, 2002b] develop the use of static principal components to identify common factors when the number of variables in the dataset gets very large, while Forni et al.[2000, 2004, 2005] propose the use of dynamic principal components. Stock and Watson [2002a] prove consistency of the factors as the number of original variables gets sufficiently large, so that the principal components are estimated precisely enough to be used as data instead of the original variables in subsequent regressions. The factors are mainly applied in macroeconomic contexts: their main use is in forecasting in second stage regressions, but they are also employed as instrumental variables in IV estimation, in augmented VAR models and in DSGE models[10].

The idea of using principal components or factors as instrumental variables is not so new in the literature. Kloek and Mennes [1960] and Amemiya [1966] first proposed the use of principal components in instrumental variable (IV) estimation. In this stream of literature, we find, among the others, important contributions by Kapetanios and Marcellino [2010], Groen and Kapetanios [2009] and by Bai and Ng [2010] that rely on factor-IV or factor-GMM estimation[11].

In the stream that uses factor as instruments, the main novelty of what we do here is that we consider a dynamic panel data model and extract principal components not from a large set of different economic

[10]Stock and Watson [2010] provide an extensive survey on the use of estimated factors in economic analysis.

[11]A review of the literature on Factor-IV and Factor-GMM estimations is in the introduction of Kapetanios and Marcellino [2010].

variables (in order to identify a common set of underlying factors) but instead we factorize a large set of lags of the same variable (the dependent variable in this context) in order to summarize the information conveyed by the y_{it} and avoid multicollinearities in the instrument set.

Through the PCA we extract the largest eigenvalues from the estimated covariance[12] or correlation matrix[13] of $\mathbf{Z}$ and, by combining the relative eigenvectors, we obtain the loading matrix and the score matrix. We then use the PCA scores as new instrumental variables for the lagged dependent variable.

In practice, defined $\mathbf{Z}$ as the general p-columns GMM-style instrument matrix, that can be untransformed, limited or collapsed[14], we extract p eigenvalues $\lambda_1, \lambda_2, ..., \lambda_p \geq 0$ from the covariance matrix of $\mathbf{Z}$, ordered from the largest to the smallest, and derive the corresponding eigenvectors (principal components) $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, ..., \boldsymbol{\alpha}_p$. Our new instruments will be the scores from PCA that are defined as:

$$\mathbf{s}_k = \mathbf{Z}\boldsymbol{\alpha}_k \text{ for } k = 1, 2, ..., p. \tag{2.2.13}$$

If we write $\mathbf{Z} = [\mathbf{z}_1 \quad \mathbf{z}_2 \quad ... \quad \mathbf{z}_p]$ with $\mathbf{z}_j$ being the j^{th} column of the instrument matrix, the score $\mathbf{s}_k$ corresponding to the k^{th} component can therefore be rewritten as:

$$\mathbf{s}_k = \alpha_{k1}\mathbf{z}_1 + \alpha_{k2}\mathbf{z}_2 + ... + \alpha_{kp}\mathbf{z}_p \tag{2.2.14}$$

where α_{kj} is the j^{th} element of the principal component $\boldsymbol{\alpha}_k$.

Since the aim of the PCA is data reduction, it would not help to keep all the p scores in the analysis as this would imply no decrease in the

[12] An unbiased estimator of the covariance matrix of a p-dimensional vector $\mathbf{x}$ of random variables is given by the sample covariance matrix $\mathbf{C} = \frac{1}{N-1}\mathbf{X}'\mathbf{X}$ where $\mathbf{X}$ is a $N \times p$ zero mean design matrix.

[13] There is not a clear indication in the theoretical literature on which is the preferable matrix among the two. The PCA is scale dependent and the components that are extracted from either matrices are different. In empirical analysis, the PCA on the covariance matrix is better when the variables are in commensurable units and have similar variances, as it is the case in our Monte Carlo experiments.

[14] $\mathbf{Z}$ can be $\mathbf{Z}^{\text{diff}}$, $\mathbf{Z}^{\text{sys}}$, $\mathbf{Z}^{\text{diff,C}}$, $\mathbf{Z}^{\text{sys,C}}$, $\mathbf{Z}^{\text{diff,L}}$, $\mathbf{Z}^{\text{sys,L}}$. We consider here balanced panels with $T_i = T \; \forall i$. $\mathbf{Z}^{\text{diff}}$ has $p = (T-2)(T-1)/2$ columns, $\mathbf{Z}^{\text{diff,C}}$ has $p = T-2$ columns, $\mathbf{Z}^{\text{diff,L}}$ a variable number of columns. In system GMM estimation, further $T-2$ columns are added in $\mathbf{Z}^{\text{sys}}$ and in $\mathbf{Z}^{\text{sys,L}}$, while only one is added to $\mathbf{Z}^{\text{sys,C}}$.

number of instruments. Only q principal components will therefore be retained and, as a consequence, only the q corresponding score vectors will form the new transformed instrument matrix. Alternative criteria can be applied in order to select the component to be retained: we will adopt here the two most commonly used strategies, the average criterion and the variability criterion[15].

Defined the matrix of PCA loadings as $\mathbf{V} = [\boldsymbol{\alpha}_1 \quad \boldsymbol{\alpha}_2 \quad ... \quad \boldsymbol{\alpha}_p]$ and the matrix of PCA scores as $\mathbf{S}$, we have that $\mathbf{S} = \mathbf{ZV}$. Instead of the moment conditions in (2.2.3), we will therefore exploit the following restrictions:

$$E[(\mathbf{S}^{\text{diff}})'\Delta\boldsymbol{v}] = E[(\mathbf{Z}^{\text{diff}}\mathbf{V})'\Delta\boldsymbol{v}] = 0 \text{ for } t \geq 3. \tag{2.2.15}$$

Similarly, in the system GMM we will also exploit the additional orthogonality conditions

$$E[(\mathbf{S}^{\text{lev}})'\boldsymbol{v}] = E[(\mathbf{Z}^{\text{lev}}\mathbf{V})'\boldsymbol{v}] = 0 \text{ for } t \geq 3. \tag{2.2.16}$$

In both cases, the number of moment restrictions depends on the number of components we retain in the analysis. As our starting point is that instruments are orthogonal to the error term, a linear combination of the original instruments will also obviously be orthogonal to the error term.

The rationale of this procedure is to use, instead of the untransformed instruments, linear combinations of the original instruments properly weighted according to the PCA loadings: no available lag is actually dropped, but its influence might be rescaled after the PCA. It is also worth noticing that none of the lags that are not in the original matrix $\mathbf{Z}$ will enter the linear combinations which forms the columns of the new instrument matrix. PCA thus preserves all the information in the original instrument set.

A further advantage of PCA is that we can factorize not only the untransformed instrument matrix but also the limited and collapsed instrument matrix, retaining all the information each matrix conveys and thus further reducing the number of instruments. It has to be recognized that,

[15] The criteria are discussed in the Appendix.

in order to apply PCA on an already transformed matrix, we need to trust the restrictions imposed on the coefficients of the original instrument matrix, even when we believe them too arbitrary: however, the instrument proliferation problem could persist after a first instrument reduction and a further decrease in the number of instruments, obtained through factorization, could be helpful.

In the following section we compare all these alternative or complementary strategies through Monte Carlo simulations and we check how far it is safe to go in instrument reduction.

2.3 Comparing the instrument reduction techniques

2.3.1 The simple panel AR(1)

In the first set of Monte Carlo simulations we estimate the standard panel AR(1) process with fixed effects and without additional explanatory variables, namely:

$$y_{it} = \alpha y_{i,t-1} + \varepsilon_{it} \tag{2.3.17}$$

$$\varepsilon_{it} = \mu_i + v_{it}$$

$$E[\mu_i] = E[v_{it}] = E[\mu_i v_{it}] = 0.$$

In the case of the AR(1) model, as we do not add any other regressor in the model, instruments proliferate purely because T gets large and not because more variables are to be instrumented.

We generate the fixed effects μ_i and the idiosyncratic terms v_{it} as following a $\mathcal{N}(0,1)$ distribution.

In order for the moment restrictions in equation (2.2.5) to hold, we need to impose a mild stationarity restriction on the process which generates the initial conditions for y_{it}, i.e. y_{i1}, as showed in Blundell and Bond [1998a].

The Blundell-Bond assumption on initial conditions imposes that the time-invariant individual effects in the model, μ_i, are uncorrelated with the deviations of the initial conditions from the long-run mean of y_{it}. The steady state level for y_{it} is defined by $\frac{\mu_i}{1-\alpha}$, from which we have that the deviation from the long-run mean can be defined as $u_i = y_{i1} - \frac{\mu_i}{1-\alpha}$. The Blundell-Bond stationarity conditions on y_{i1} requires therefore that:

$$E\left[\left(y_{i1} - \frac{\mu_i}{1-\alpha}\right)\mu_i\right] = E[u_i\mu_i] = 0 \tag{2.3.18}$$

so that y_{it} converges toward its long-run mean $\frac{\mu_i}{1-\alpha}$ for each individual for every period $t \geq 2$. [16]

Following Roodman [2009b], we simulate the initial value of y_{it}, namely y_{i1}, according to the following process:

$$\begin{aligned} &\mu_i, w_i \sim \mathcal{N}(0,1) \\ &u_i = \sigma_u(\sqrt{1-\rho^2}w_i + \rho\mu_i) \\ &y_{i1} = \frac{\mu_i}{1-\alpha} + u_i. \end{aligned} \tag{2.3.19}$$

The form we give to u_i allows us to take into account different levels of violation of the Blundell-Bond condition defined in equation (2.3.18): when $\rho = 0$ we have that the deviation from the long-run mean is totally uncorrelated with the individual effects, so that the Blundell-Bond restriction on initial conditions holds; when instead $\rho = 1$, the u_i is perfectly correlated with the individual effects thus yielding a severe violation of the condition in (2.3.18). The functional form chosen for u_i is particularly convenient because it implies a direct interpretation of the

[16] If the stationarity condition is satisfied in some given period, the first one in this framework, it will hold also for all the subsequent periods. See Blundell and Bond [1998a] and Roodman [2009b] for a proof of this condition.
The condition in equation (2.3.18) yields the condition

$$E[\Delta y_{i,t-1}\mu_i] = 0$$

that, together with the following usual assumption

$$E[\Delta v_{it}\mu_i] = 0 \text{ for } t \geq 3,$$

yields the additional $T-2$ linear moment restrictions:

$$E[\Delta y_{i,t-1}(\mu_i + v_{it})] = E[\Delta y_{i,t-1}\varepsilon_{it}] = 0$$

that allow to use lagged first difference of y_{it} as instrument for the level equation in the system GMM estimator.

parameters ρ and σ_u: as the two Normal variables w_i and μ_i are independent one of the other, it can be shown that the correlation coefficient between u_i and μ_i is exactly ρ and that the variance of u_i is constant and equal to σ_u^2.

In Table (2.1) we summarize the settings of the parameters in our Monte Carlo experiment for the panel AR(1) model.

Table 2.1: Setting of the parameters in the simulation model

Iterations	1000
N	100
T	5, 10, 15, 20
σ_u	2
α	0.2, 0.5, 0.9
ρ	0, 0.9

In the baseline simulations, we consider 100 individuals and different time lengths in the sample, i.e. $T = 5, 10, 15, 20$[17]. We consider different degrees of persistence for y_{it}, as captured by the autoregressive coefficient α. We set $\rho = 0$ when we want that the assumption for the SYS GMM to be valid holds; we set instead $\rho = 0.9$ when we want the Blundell-Bond assumption to be violated and therefore the SYS GMM to be invalid. In line with Roodman [2009b], we set $\sigma_u = 2$ so that the violation of the Blundell-Bond conditions is particularly problematic also with a limited number of time periods[18].

Each experiments consists of 1000 iterations[19].

For each combination of α, ρ and T, we estimate the coefficient α in equation (2.3.17) on 1000 simulated panels with several variants of the DIFF GMM estimator and of the SYS GMM estimator.

[17]This is exactly the sample size of the Monte Carlo experiments in Roodman [2009b]. We will discuss below some results for larger samples.

[18]As discussed in Roodman [2009b], the violation of the stationarity assumption on the initial conditions becomes less problematic over time. $\sigma_u = 2$ is a value large enough to make the violation of the assumption dangerous even as T approaches 20.

[19]The choice of running only 1000 iteration is unfortunately due to the fact that each experiment is very time consuming. Anyway, Roodman [2009b] runs only 500 iterations.

We apply the two GMM estimators on the untransformed instrument matrix (*UNTR* GMM henceforth), on the collapsed (*COL* GMM), on the limited (*LIM* GMM, where we sometimes specify 1 or 2 to indicate a lag depth of 1 or 2) and on the factorized set of instruments and on sets where the instrument reduction techniques are combined (two abbreviations are also combined, e.g. LIMCOL). With respect to the factorized set of instruments, on one hand we keep the principal components whose eigenvalues are above the average of the eigenvalues (average criterion, *FACTA* henceforth); on the other hand, we keep the components that explain the 90% of the variability in the instrument matrix (variance criterion, *FACTV* hencefort). In what follows, we apply the PCA analysis on the covariance matrix of $\mathbf{Z}$. In line with Roodman [2009b], all the estimates are two-steps and are made robust to heteroskedasticity in the residuals; we apply the Windmeijer [2005] correction for two-step estimation in small samples.

Estimation results for $\alpha = 0.2$

In Table (2.2) we report the estimation results for the AR(1) model in (2.3.17) when the autoregressive process for y_{it} is not persistent, i.e. $\alpha = 0.2$, a scenario that is generally safe from weak instrument problems in DPD models. From this extensive set of results, some interesting tendencies may be underlined: the SYS UNTR estimator has overall a better performance than the DIFF UNTR when $\rho = 0$: in fact, it is less biased and has a slightly lower variance, in particular as T gets larger. In this context, the Hansen test is not safe, according to Roodman [2009b]'s suggestions[20], even when $T = 5$ as it is systematically above 0.4. When $\rho = 0.9$ the performance of the UNTR estimator is reversed, the best one being the DIFF UNTR: this means that the violation of the Blundell-Bond condition remains problematic also for large T.

No matter which instrument reduction technique is adopted, the variance of the estimators increases. The COL estimates are less biased than

[20]As a rule of thumb, we should be suspicious whenever we get a Hansen p-value as high as 0.25.

the UNTR ones; when the SYS GMM is invalid, not surprisingly the performance of the DIFF COL is better than that of the SYS COL[21].

Interestingly, the DIFF LIM, when the lag depth is bounded to lag 1 only or to lags 1 and 2[22], performs better than the DIFF UNTR while the SYS LIM is worse overall, with larger bias and variance, even when $\rho = 0$ and though the Hansen p-value is significantly lower when $T = 20$. The LIMCOL estimator gives the least biased estimates overall, though the its variance is slightly larger than that of the UNTR GMM. The violation of the Blundell-Bond condition is not very problematic for the LIM2COL estimator, while it remains very troublesome for the LIM1COL GMM even when T is large.

When FACT GMM is used, the results are sensitive to the criterion we apply to choose the number of components to be retained. The FACTA GMM is more parsimonious than the FACTV estimator in terms of number of instruments. Such parsimony is beneficial for the DIFF FACT estimator while it is not always the case for the SYS FACT. Overall, the DIFF FACT estimator performs better than the DIFF UNTR, though it is more biased than the DIFF COL; notwithstanding this, the SYS FACT has performance similar to the SYS UNTR. The DIFF FACT COLL estimator has a very good performance while the SYS FACT COLL estimates are worse than the SYS COLL and the SYS FACT ones.

Overall, when $\alpha = 0.2$, the estimates are generally close each other and none of the estimators gives misleading results when the Blundell-Bond condition holds. The Hansen p-values always give rise to doubts about an instrument proliferation problem. However, only when $T = 20$ the reduction techniques significantly lower the p-value, though to a value that is still above 0.25. When the SYS GMM is not valid, the SYS estimates are generally more biased; however, the Hansen p-value is extraordinary low, so that we are better at detecting the violation of the assumption.

---

[21] The violation of the Blundell-Bond assumption does not affect the moment restrictions for the DIFF GMM estimator, while it is certainly problematic for the SYS GMM estimator.

[22] The truncation of the lag depth to one or two lags is the most common strategy to reduce the instrument count in the empirical literature.

As a general conclusion, especially in the case of the SYS GMM, the instrument proliferation does not seem such a big issue as the untransformed estimates are pretty good and instrument reduction techniques do not yield significant improvements.

Estimation results for $\alpha = 0.9$

In Table (2.3) we report the results for the case in which y_{it} is highly persistent, i.e. $\alpha = 0.9$. At a first sight, we see that the estimates are much more heterogeneous than in the previous scenario and that are overall more biased.

The DIFF GMM estimates when $\rho = 0$ are systematically very biased and present a very high variance, no matter which is the specification of the estimator we adopt. The results are particularly misleading when T is short, as the estimates are very far from the true value of the coefficient. We interpret these results as deriving from a weak instrument problem in DIFF GMM estimation when the series are highly persistent, and thus as due to a weak identification issue, as showed in Blundell and Bond [1998a]. With $\alpha = 0.9$ in fact, we expect the lags of y_{it} in levels to be only weakly correlated with the lagged first-differenced dependent variable: in this case, they are very weak instruments for the endogenous regressors. DIFF estimates become less biased as T increases: DIFF LIM COL estimates are very good. When instruments are weak, a drastic limitation in their number is beneficial: by limiting the lag depth we only keep the least weak ones. The DIFF FACT estimator performs slightly worse than the DIFF COL, while the DIFF FACT COL gives the best estimates among the DIFF ones. The SYS estimates are systematically pretty close to the true value of α, though differences among the estimators still persist. The SYS COL and the SYS LIM COL are the least biased SYS estimates: however the SYS UNTR estimator, which exploits the highest number of instruments, is not far from the SYS COL and presents a smaller variance. The SYS FACT estimator performs better than the SYS LIM but not as well as the SYS COL: differently from the DIFF case, the combination of collapsing and PCA does not imply an improvement in the estimates. When the instruments are weak, as in the DIFF case, the stronger the reduction in the instrument count, the better the estimates. When the instruments are valid, a further reduction, beyond the one operated by the collapsing, can be harmful. Note that the Hansen p-values tend to be similar among the estimators that reduce someway the instruments: this similarity has not a direct counterpart in the performances of the estimator. We therefore need to be cautious in the interpretation of the Hansen p-value in terms of a symptom of instrument proliferation. Moreover, when the Blundell-Bond condition is violated, the DIFF GMM estimator systematically performs better than the SYS estimator: this is in line with the findings of Hayakawa [2009] in his analysis on the effect of non-stationary initial conditions in DPD models. Furthermore, when $\rho = 0.9$ the SYS COL and the SYS LIM estimators tend to perform worse than the SYS UNTR; the SYS FACT is the best, though its variance is larger than the one of the SYS UNTR.

Some robustness checks on the AR(1) model

We present now the results of some robustness checks with respect to the choice between the covariance and the correlation matrix for the PCA or between the one-step and two-step estimator, with respect to the sample size, the threshold of explained variance and non-stationary initial conditions[23].

We performed the same estimates as above applying the PCA on the correlation matrix instead of the covariance matrix. We found outcomes that are specular to those discussed above: as the instruments in a pure AR(1) model are the lags of the dependent variables, we obviously do not have any discrepancy in the measurement units of the instruments, so that the choice of the matrix is not relevant.

We also run the estimates with the one-step GMM and we found no significant differences in the results. In the literature, the two-step estimator is generally found to be more efficient than the one-step but also to

[23]We keep the presentation of additional results as parsimonious as possible. All the results discussed in this section are available upon request.

suffer small sample biases: we see that the Windmeijer [2005] correction works very well in this experimental context and we obtain mirror-like estimates in our approach.

With respect to the sample size, we found that by doubling the number of individuals, i.e. by taking $N = 200$, we do not have relevant differences in the estimates with respect to the baseline experiment. However, the Hansen p-value remains well below 0.5 even when we estimate the model on the full set of instruments. By increasing the number of individuals, we have that N is systematically above the number of instruments[24], that depends instead on T, so that the risk of instrument proliferation is alleviated and it is in general less important to adopt instrument reduction techniques.

We also changed the threshold explained variance in the variability criterion. If we lower the portion of explained variance to the 70%, we keep less principal components than those kept using the average criterion. However, this choice overall worsens the estimates. The parsimony deriving from the factorization of the instruments is generally beneficial provided that the the number of retained components is not reduced too far ahead. In general, we prefer the average criterion to select the components to be retained as it generally reduces the instruments to a pretty "safe" number.

Finally, we keep the same setting of the baseline experiment except for the fact the we generate non-stationary initial conditions for y_{it}. As argued in Blundell and Bond [1998a], the mild stationarity assumption for the SYS GMM estimator to be valid is not without content: it is violated, for example, if $y_{i1} = k \ \forall i$ or $y_{i1} \sim \text{i.i.d.}(0, \sigma^2_{y_1})$. The violation of the Blundell-Bond assumption is here induced by random initial conditions rather than by the imposed correlation between the individual effects and the deviations of the y_{i1} from their long-run mean. This threat to the validity of the SYS GMM could be real, e.g., in empirical analysis that consider datasets beginning after a war or a severe structural

[24] N is 200, while the maximum instrument count is 190 for the untransformed SYS GMM when $T = 20$.

break or young workers or recently-born firms[25]. Following Blundell and Bond [1998a], we generate here non-stationary initial conditions as $y_{i1} \sim$ i.i.d$\mathcal{N}$ $(0, 16/3)$. For the sake of brevity, we only present the estimates for $\alpha = 0.9$. In the light of the previous discussion, we select here the principal components according only to the average criterion. The results are reported in Table (2.4). The non-stationarity of the initial conditions seriously affects the SYS GMM estimates of α: the coefficient is in fact estimated above 1 in most cases by the SYS estimator. The DIFF GMM performs significantly better in this context, even though the series is highly persistent, as it does not require any stationarity assumption. In this, we confirm the finding of Hayakawa [2009] that, when the initial conditions are not stationary, the DIFF estimator may be less biased than the SYS estimator. COL and FACT estimates, either DIFF or SYS, are more biased than UNTR ones and have a larger variance. However, the FACT estimates are generally less biased than the COL ones and have a smaller variance. It is impressive to notice how low is the Hansen p-value in FACT or COL estimates, even below 0.1 in the FACT estimates: this however comes along with worse estimates than the UNTR one, which instead reaches an instrument count of 190 and a p-value of 1. The implication we draw is that we need to be careful in addressing the issue of instrument proliferation and in taking the Hansen p-value too seriously. In empirical analyses, we should investigate whether the Blundell-Bond assumption may not hold, as a violation could come along with misleading estimates.

Final remarks on the AR(1) simulations

The results from the estimation of the simple AR(1) are not particularly revealing. The AR(1) is the most exploited model in Monte Carlo simulations of DPD but it comes out not to be very informative, as the process it models is likely to be too simplified and therefore not very useful. It is anyway worth paying attention to this model as it is the priv-

[25]See Hayakawa [2009] and Arellano [2003] for a discussion about the scenarios in which the stationarity assumption could be at risk.

ileged tool in the literature to study the dynamics and the persistence of relevant economic variables. What emerges from alternative scenarios, only differing in the degree of persistence[26], is that the results seem to be mostly driven by the value of the autoregressive coefficient α, and by whether the Blundell-Bond condition hold or not, rather than by distortions caused by instrument proliferation. It seems really difficult to detect symptoms of instrument proliferation in this framework.

If the only source of endogeneity is the inclusion of the lagged dependent variable, it is not really meaningful to use a simple AR(1) model to try to detect problems such as the overfitting of endogenous variables. What we can say, however, is that the strategy of collapsing the instrument set is not harmful in this context, while the choice of keeping only few lags in the analysis can be controversial. The performance of the factorized estimator is similar to that of the collapsed one, but the lack of exploitable information in this context, together with the fact that the untransformed instrument matrix is very sparse, implies a weak covariance structure among the variables and affects the potentialities of a purely statistical technique such as the PCA. We see that, when we deal with potentially non-stationary initial conditions, the safest strategy is to estimate the model by GMM UNTR because the COL, LIM and FACT estimators can give misleading results.

2.3.2 A multivariate dynamic panel data model

Here we extend the pure AR(1) by adding an endogenous explanatory variable x_{it}. The model of interest becomes:

$$y_{it} = \alpha y_{i,t-1} + \beta x_{it} + \mu_i + v_{it} \tag{2.3.20}$$

where $\mu_i \sim \mathcal{N}(0,1)$ are the fixed effects and the $v_{it} \sim \mathcal{N}(0,1)$ are the idiosyncratic shocks.

[26]We have also estimated the simple AR(1) for $\alpha = 0.5$: as the results do not show significant difference with those above, we prefer to report the estimates for the two more interesting cases. All the results are however available upon request.

We generate the variable x_{it} according to the following process:

$$x_{it} = d_t + \gamma x_{i,t-1} + \xi \mu_i + \phi v_{i,t-1} + u_{it} \qquad (2.3.21)$$

where the $d_t \sim \mathcal{N}(0,1)$ are time dummies, the μ_i are the same fixed effects that appear in equation (2.3.20) and so also the idiosyncratic shocks v_{it}, that are here lagged one period. The additional regressor x_{it} is in this case predetermined but not strictly exogenous with respect to v_{it}.

We generate x_{it} so that the process is covariance stationary and the following initial conditions hold:

$$E\left[\left(x_{i1} - \frac{\xi\mu_i}{1-\gamma}\right)\xi\mu_i\right] = 0. \qquad (2.3.22)$$

The initial conditions for y_{it} in order for the Blundell-Bond assumption to hold are:

$$y_{i1} = \frac{\beta\left(\frac{\xi\mu_i}{1-\gamma}\right) + \mu_i}{1-\alpha}. \qquad (2.3.23)$$

As discussed in Blundell and Bond [1998b], the joint stationarity of the processes y_{it} and x_{it} is sufficient for the use of the lagged first differences as instruments in the equation in levels.

We still keep the values of α as in Table (2.1) and we set $\beta = 1$. Furthermore, we set $\xi = 0.5$ and $\phi = 0.5$[27]. In Table (2.5) we report the results for the model in (2.3.20) from a Monte Carlo experiment on 1000 simulated samples when $\alpha = 0.9$ and $\beta = 1$. We report the estimates obtained by the DIFF and SYS UNTR, COL, LIM and FACT estimation[28].

The DIFF GMM estimates for both α and β are very similar across the alternative specifications; the reduction of the instrument count does not give a significant improvement in the estimates in terms of bias and variance reduction, though it drastically lowers the Hansen p-values for

[27] The model in equations (2.3.20) and (2.3.21) is similar, though a bit more complex, to the model simulated in Blundell and Bond [1998b] and Bond and Windmeijer [2002].

[28] In the light of the simulation results of the previous section, and to simplify the presentation, we omit here the estimates on the limited instrument matrix as they have proved not to be very reliable especially for $\alpha = 0.9$. Our main interest here is the comparison between the COL and FACT estimates. We also omit the estimates for the cases in which the instrument reduction techniques are combined. All the results are available upon request.

$T = 15$ and $T = 20$. The DIFF COL exploits 37 instruments when $T = 20$ instead of 360 used by the DIFF UNTR: however, this reduction does not come along with a sensible improvement in the estimates. The DIFF FACT estimator performs pretty well overall and in general has a lower variance than the DIFF COL, though the smallest variance is that of the DIFF UNTR. The DIFF GMM estimator, no matter which is the specification we choose, tends to underestimate the coefficient of interest: the reason, as stressed above, is the weak instrument problem when the series are persistent.

With respect to the SYS estimates, we do not highlight important differences in the estimates of α and β among the different estimators. We have a significant decrease in the Hansen p-value when the instruments are reduced through collapsing or factorization but this does not come along with significantly different estimation results. The SYS UNTR estimator is again the one with the smallest variance.

What emerges here is that, despite the fact that the number of instruments gets incredibly large also in this context, it is very hard to detect an instrument proliferation problem when the models are so simplified. It is clear however that, in the light of the two sets of simulations, the collapsing and the factorization of the instrument matrix are both safe instrument reduction techniques: they generally slightly improve the estimates of the coefficients though they tend to be characterized by a larger variance than the untransformed estimator.

The application of the PCA on the instrument matrix seems a valid and effective strategy to reduce the instrument count because it performs well despite the fact that the instrument matrix is very sparse and the correlation among the lags could be weak. We prefer this last technique rather than the collapsing as, other conditions being equal, the PCA is a purely data-driven procedure and does not impose any restriction on the moment conditions.

2.4 An empirical example

So far we have presented results obtained from the simulation of very simple models. We now exploit the same empirical example used in Roodman [2009b] in order to compare the estimates obtained by alternative specifications of the GMM estimator on a real dataset. We replicate the analysis of Forbes [2000] on the effect of inequalities in income, as captured by the Gini coefficient, on economic growth. We use the reproduction dataset built by Roodman [2009b][29]. We deal with a panel dataset with 5-year averages for 45 Countries in the period 1975-95. The dependent variable in the model of interest is the growth rate of income per capita ($Growth_{it}$) while the regressors include the Gini coefficient lagged one period ($Gini$) as a measure of inequality, the income per capita lagged one period ($Income$), the years of secondary schooling for men lagged one period ($MaleEduc$), the years of secondary schooling for women lagged one period ($FemaleEduc$), the price level of investments lagged on period ($InvPrice$); the model allows for individual and time fixed effects. We can summarize the model as follows:

$$\begin{aligned} Growth_{it} = {} & \alpha Gini_{i,t-1} + \beta Income_{i,t-1} \\ & + \gamma MaleEduc_{i,t-1} + \delta FemaleEduc_{i,t-1} \\ & + \xi InvPrice_{i,t-1} + \mu_i + \tau_t + \varepsilon_{it}. \end{aligned} \quad (2.4.24)$$

We treat all the regressors as endogenous and we instrument them all with GMM-style instruments. As we deal with a very low number of periods for each Country, here the instrument proliferation depends on a large number of endogenous regressors rather than on a large T. In Table (2.6) we report the estimates of the model obtained by DIFF and SYS GMM estimation on the full set of instruments, on the limited and collapsed sets and on factorized instruments. In column 1 we report the original results presented in Forbes [2000]; the DIFF estimates on the un-

[29] Forbes [2000] provides extensive details on the sources of the data and summary statistics for the variables of interest. Roodman [2009b] replicates the original dataset by drawing from the cited sources the data that were available at the time the original paper was published. As usual, some differences between the two datasets occur.

transformed, limited and collapsed instruments replicate the results presented in Roodman [2009b]; we add here the SYS estimates on the untransformed, limited and collapsed set of instruments and both the DIFF and SYS estimates on the factorized instruments.

A priori we expect that the GDP per capita is a highly persistent series, so that the lags of the GDP per capita could be weak instruments and, consequently, the DIFF GMM estimator may be biased. We believe therefore that it is safer to estimate the model and investigating the potential instrument proliferation problem by adopting a SYS GMM approach.

The most natural choice would be the SYS estimator but, when we look at the table, we see that the twice-lagged residuals come out to be autocorrelated whatever specification of the SYS estimator we adopt, except for the SYS COL estimates. We have therefore doubts about the validity of the lags as instruments in this context and about the choice of this application as a didactic example.

What emerges from the results is that there is a huge heterogeneity in the estimates, according on whether we use DIFF or SYS GMM and on whether we apply any instrument reduction strategy: this is a typical outcome when the model is misspecified.

The lagged income per capita, found very significant in the original paper, is significant at least at a 10% level only in the DIFF UNTR and the SYS LIM COL estimates[30]. Worryingly the SYS LIM COL estimator gives a positive coefficient for the lagged income, despite the fact that it is very parsimonious in the number of instruments (only 16 against the 80 in the former case)[31]. The inequality variable is significant in the DIFF UNTR estimates but, surprisingly, also in two additional cases: when the

[30] We recall here that, in a two-sided t-test, the critical values are 1.96 and 1.65 respectively for a significance level of 5% and 10%.

[31] In growth regressions, with the growth rate of income per capita as dependent variable, we expect a negative coefficient for the initial income and we interpret it as a sign of convergence. A negative significant coefficient in this scenario, where additional regressors are included in the model, would imply a conditional convergence tendency in the sample. In the case of an autoregressive coefficient for the income very close to 1, β would be close to 0, rather than negative. The reader interested in growth regressions can refer, among others, to Barro and Sala-i-Martin [2003] and Durlauf et al. [2005].

estimator is the least parsimonious in terms of number of instruments, the SYS UNTR, and when we are in the most parsimonious case, the SYS LIM COL estimates. The estimates of ξ remain significant across all the specifications of the GMM estimator. When we adopt a SYS GMM approach, we contradict the tendencies underlined by Roodman [2009b]: the SYS estimates tend to become more significant as the number of instrument decreases. This could be due to the fact that the instruments are not valid and a reduction could alleviate the problem. In this context, it is not safe to trust the Hansen p-value too much, though it systematically decreases as the number of instruments gets smaller. In the light of the results in the previous section, we would be inclined to choose the SYS COL estimator as the safest in this context: in this case, none of the main variable comes out to be significant. As already found in our Monte Carlo experiment, the FACT estimates are not really far from the COL ones, neither in terms of biasedness nor in terms of significance of the coefficients and Hansen p-values. The FACT estimator could suffer the fact that the instruments are likely to be invalid in this context and to have weak correlations with the endogenous variables.

In conclusion, the choice of this empirical application by Roodman [2009b] is not particularly lucky: on one hand, the replication of the original dataset is not satisfactory, as the estimates are very different; the number of observations is really small, data are averaged over 5-years periods, so that we loose a lot of information, and the periods are dangerously few; on the other hand, the series are persistent and call for the use of a SYS approach, but the instruments appear to be invalid. When the lack of information is so serious, it is very difficult to compare the performances of the estimators and to detect instrument proliferation problems; in particular, it is difficult to say what is more problematic between the proliferation of instruments and the lack of information that complicates the dangers intrinsic in short panels estimations. Moreover, we are not in a proper context to assess the potentialities of the factorization of the instrument set as this technique requires more information and more solid covariances among the instruments.

2.5 Conclusions

This chapter introduces a new strategy to reduce the number of instruments in the GMM estimation of dynamic panel data, namely the extraction of principal components from the instrument matrix, and compares the alternative instrument reduction techniques through Monte Carlo simulations.

First, we discussed the rationale of applying the PCA on the instrument matrix stressing that it involves a purely data-driven procedure which does not require particular assumptions on the coefficient of the matrix: it is instead the most information-preserving technique among those we discuss here.

Secondly, we run extensive Monte Carlo simulations of a pure AR(1) panel model and of a multivariate dynamic panel model. We argued that the models we considered in the analysis are extremely simplified and have relevant drawbacks in this context: anyway it is reasonable to start from the simplest models to have a proper control on what we are doing. We found that the collapsing and the factorization of the instrument matrix give similar results and that the GMM estimates on the untransformed set are still generally safe. We also warned about the fact that the Blundell-Bond assumption on initial conditions could be much more involved than what it seems: not taking into account potential threats to the validity of the SYS GMM estimator can lead to misleading results and conclusions, especially when the instrument set is transformed someway. Overall, the Hansen p-value by itself does not seem a safe criterion to assess a risk of instrument proliferation, thought it is generally indicative.

We extended the empirical application chosen by Roodman [2009b], and we were inclined to warn about the drawbacks of this example: in a context were the number of periods is incredibly short and there is a severe lack of information, it can be misleading to pay an excessive attention to the proliferation of the instruments rather than to look for additional information.

In the light of the previous findings, we are able to suggest some indications for applied research and to sketch some potential extensions of this work.

Overall, the factorization of the instrument set seems to be a promising approach to the issue of instrument proliferation: in fact it appears reasonable to exploit the correlations between the instruments to summarize the original information. However, additional tests on its potentialities and performances would be useful. In particular, it would be interesting to simulate more complex dynamic models that include both fixed effects and several endogenous regressors to test the response of the estimates to the collapsing or the factorization of the instrument matrix[32].

At this stage, a relevant cost to be paid in order to factorize the instrument matrix is the need of good programming skills, as the instrument sets are to be constructed by hand. It would be useful to develop automatic procedures to directly extract the principal components from the matrix of lags[33]. On the other hand, the `xtabond2` command already includes a `collapse` option that automatically transforms the instrument set: most of the popularity of the collapsing is due to the fact that is easily implementable. For the two approaches to be really comparable, they need to be equally easy to implement.

In the light of our findings, we suggest the researcher on always reporting the number of instruments and to worry when it exceeds the number of observations in the sample. We should not adopt an instrument reduction technique a priori, as every strategy could have serious drawbacks if some assumptions do not hold. The best strategy is the presentation of the estimates obtained with alternative GMM estimators with and without instrument reduction techniques.

[32]Chapter 3 will present a first attempt to simulate a more complex model where there is more than one source of endogeneity and where the number of instruments gets rapidly very large.

[33]Stata provide the `pca` command that we exploit here but we need to build a priori the matrices of lags presented above.

Table 2.2: Estimates for the AR(1) with $\alpha = 0.2$

		$\rho = 0.0$ (SYS GMM valid)				$\rho = 0.9$ (SYS GMM invalid)			
Time dimension		5	10	15	20	5	10	15	20
Untransformed instr.									
DIFF GMM	Estimated α mean	0.192	0.189	0.187	0.183	0.194	0.192	0.190	0.185
	Estimated α st. dev.	0.070	0.047	0.032	0.027	0.057	0.041	0.029	0.025
	Hansen p-value avg.	0.483	0.414	0.445	1.000	0.487	0.415	0.443	1.000
	Instrument count	6	36	91	171	6	36	91	171
SYS GMM	Estimated α mean	0.206	0.204	0.203	0.198	0.347	0.250	0.228	0.222
	Estimated α st. dev.	0.058	0.041	0.032	0.029	0.067	0.040	0.031	0.028
	Hansen p-value avg.	0.472	0.397	0.610	1.000	0.000	0.050	0.601	1.000
	Instrument count	10	45	105	190	10	45	105	190
Collapsed instr.									
DIFF GMM	Estimated α mean	0.198	0.198	0.197	0.198	0.198	0.199	0.198	0.198
	Estimated α st. dev.	0.072	0.046	0.035	0.029	0.056	0.039	0.032	0.027
	Hansen p-value avg.	0.495	0.479	0.484	0.458	0.501	0.472	0.480	0.458
	Instrument count	3	8	13	18	3	8	13	18
SYS GMM	Estimated α mean	0.200	0.200	0.198	0.199	0.272	0.186	0.185	0.187
	Estimated α st. dev.	0.073	0.046	0.035	0.030	0.085	0.044	0.034	0.029
	Hansen p-value avg.	0.492	0.477	0.480	0.454	0.000	0.015	0.048	0.087
	Instrument count	5	10	15	20	5	10	15	20
Limited instr. (Lag 1 only)									
DIFF GMM	Estimated α mean	0.195	0.197	0.199	0.198	0.195	0.198	0.199	0.199
	Estimated α st. dev.	0.073	0.052	0.043	0.038	0.058	0.044	0.038	0.035
	Hansen p-value avg.	0.494	0.478	0.480	0.480	0.485	0.476	0.479	0.482
	Instrument count	3	8	13	18	3	8	13	18
SYS GMM	Estimated α mean	0.213	0.214	0.213	0.212	0.314	0.261	0.245	0.237
	Estimated α st. dev.	0.073	0.051	0.041	0.035	0.060	0.047	0.039	0.035
	Hansen p-value avg.	0.481	0.472	0.470	0.474	0.001	0.010	0.037	0.084
	Instrument count	7	17	27	37	7	17	27	37
Limited instr. (Lags 1 and 2)									

Continued on next page

Table 2.2: Estimates for the AR(1) with $\alpha = 0.2$ - continued

		$\rho = 0.0$ (SYS GMM valid)				$\rho = 0.9$ (SYS GMM invalid)			
Time dimension		5	10	15	20	5	10	15	20
DIFF GMM	Estimated α mean	0.193	0.196	0.199	0.198	0.193	0.198	0.200	0.199
	Estimated α st. dev.	0.069	0.047	0.039	0.034	0.057	0.041	0.035	0.032
	Hansen p-value avg.	0.477	0.453	0.458	0.454	0.476	0.454	0.461	0.458
	Instrument count	5	15	25	35	5	15	25	35
SYS GMM	Estimated α mean	0.205	0.207	0.207	0.207	0.343	0.274	0.251	0.238
	Estimated α st. dev.	0.060	0.045	0.038	0.033	0.066	0.046	0.038	0.033
	Hansen p-value avg.	0.465	0.444	0.449	0.449	0.000	0.010	0.052	0.134
	Instrument count	9	24	39	54	9	24	39	54
Limited and coll. instr. (Lag 1 only)									
DIFF GMM	Estimated α mean	0.202	0.202	0.201	0.201	0.200	0.201	0.201	0.200
	Estimated α st. dev.	0.074	0.051	0.041	0.035	0.059	0.045	0.038	0.033
	Hansen p-value avg.								
	Instrument count	1	1	1	1	1	1	1	1
SYS GMM	Estimated α mean	0.206	0.203	0.201	0.201	0.217	0.126	0.139	0.151
	Estimated α st. dev.	0.083	0.055	0.043	0.036	0.081	0.050	0.041	0.036
	Hansen p-value avg.	0.497	0.511	0.498	0.495	0.000	0.001	0.002	0.001
	Instrument count	3	3	3	3	3	3	3	3
Limited and coll. instr. (Lags 1 and 2)									
DIFF GMM	Estimated α mean	0.200	0.202	0.201	0.201	0.198	0.201	0.201	0.200
	Estimated α st. dev.	0.073	0.048	0.038	0.032	0.056	0.041	0.034	0.030
	Hansen p-value avg.	0.474	0.487	0.504	0.499	0.505	0.500	0.516	0.503
	Instrument count	2	2	2	2	2	2	2	2
SYS GMM	Estimated α mean	0.202	0.203	0.202	0.201	0.273	0.194	0.193	0.193
	Estimated α st. dev.	0.074	0.048	0.038	0.032	0.087	0.050	0.040	0.034
	Hansen p-value avg.	0.479	0.499	0.496	0.497	0.000	0.002	0.003	0.003
	Instrument count	4	4	4	4	4	4	4	4
PCA untransformed avg									
DIFF GMM	Estimated α mean	0.194	0.193	0.193	0.193	0.195	0.196	0.196	0.196
	Estimated α st. dev.	0.069	0.049	0.040	0.035	0.056	0.042	0.036	0.033

Continued on next page

Table 2.2: Estimates for the AR(1) with $\alpha = 0.2$ - continued

		$\rho = 0.0$ (SYS GMM valid)				$\rho = 0.9$ (SYS GMM invalid)			
Time dimension		5	10	15	20	5	10	15	20
	Hansen p-value avg.	0.501	0.485	0.481	0.481	0.500	0.479	0.473	0.475
	Instrument count	3	9.93	20.36	30.80	3	8	13	18
SYS GMM	Estimated α mean	0.196	0.195	0.197	0.198	0.419	0.315	0.281	0.263
	Estimated α st. dev.	0.068	0.048	0.040	0.035	0.062	0.048	0.042	0.038
	Hansen p-value avg.	0.502	0.490	0.480	0.474	0.000	0.002	0.012	0.035
	Instrument count	5.01	12.93	25.37	38.12	5.01	10.97	18.03	25.26
PCA untransformed var									
DIFF GMM	Estimated α mean	0.193	0.187	0.185	0.189	0.195	0.189	0.186	0.189
	Estimated α st. dev.	0.069	0.049	0.037	0.028	0.056	0.041	0.034	0.027
	Hansen p-value avg.	0.500	0.479	0.467	0.494	0.500	0.489	0.477	0.475
	Instrument count	4.17	21.06	51.74	95.07	3.00	16.04	43.02	82.46
SYS GMM	Estimated α mean	0.206	0.202	0.201	0.204	0.369	0.281	0.244	0.228
	Estimated α st. dev.	0.061	0.046	0.036	0.029	0.065	0.047	0.037	0.030
	Hansen p-value avg.	0.486	0.464	0.449	0.771	0.000	0.021	0.156	0.508
	Instrument count	8.17	29.15	64.72	112.07	3.00	24.13	56.00	99.46
PCA collapsed avg									
DIFF GMM	Estimated α mean	0.200	0.202	0.201	0.201	0.198	0.201	0.200	0.201
	Estimated α st. dev.	0.074	0.049	0.038	0.032	0.056	0.042	0.034	0.030
	Hansen p-value avg.		0.506	0.507	0.483		0.456	0.507	0.481
	Instrument count	1.00	1.43	2.02	2.79	1.00	1.14	2.00	2.62
SYS GMM	Estimated α mean	0.205	0.207	0.207	0.208	0.423	0.329	0.290	0.267
	Estimated α st. dev.	0.070	0.048	0.038	0.033	0.054	0.047	0.041	0.036
	Hansen p-value avg.	0.509	0.515	0.513	0.503	0.000	0.001	0.003	0.006
	Instrument count	3.01	4.43	7.03	10.12	3.01	4.11	7.03	9.88
PCA collapsed var									
DIFF GMM	Estimated α mean	0.198	0.199	0.197	0.197	0.198	0.199	0.197	0.197
	Estimated α st. dev.	0.072	0.045	0.035	0.029	0.056	0.038	0.031	0.026
	Hansen p-value avg.	0.495	0.499	0.497	0.476	0.501	0.488	0.498	0.478
	Instrument count	3.00	6.01	9.08	11.74	3.00	6.00	9.17	12.09

Continued on next page

Table 2.2: Estimates for the AR(1) with $\alpha = 0.2$ - continued

		$\rho = 0.0$ (SYS GMM valid)				$\rho = 0.9$ (SYS GMM invalid)			
Time dimension		5	10	15	20	5	10	15	20
SYS GMM	Estimated α mean	0.210	0.211	0.210	0.209	0.376	0.279	0.251	0.237
	Estimated α st. dev.	0.068	0.046	0.036	0.030	0.068	0.045	0.036	0.030
	Hansen p-value avg.	0.485	0.496	0.478	0.469	0.000	0.003	0.015	0.036
	Instrument count	7.00	14.10	22.06	28.74	7.00	14.09	22.15	29.09

Table 2.3: Estimates for the AR(1) with $\alpha = 0.9$

		$\rho = 0.0$ (SYS GMM valid)				$\rho = 0.9$ (SYS GMM invalid)			
Time dimension		5	10	15	20	5	10	15	20
Untransformed instr.									
DIFF GMM	Estimated α mean	0.266	0.638	0.759	0.776	0.853	0.816	0.841	0.843
	Estimated α st. dev.	0.514	0.169	0.076	0.053	0.354	0.090	0.045	0.034
	Hansen p-value avg.	0.477	0.399	0.442	1.000	0.507	0.398	0.441	1.000
	Instrument count	6	36	91	171	6	36	91	171
SYS GMM	Estimated α mean	0.940	0.946	0.942	0.909	1.025	0.997	0.966	0.936
	Estimated α st. dev.	0.099	0.040	0.030	0.039	0.071	0.026	0.022	0.027
	Hansen p-value avg.	0.451	0.363	0.611	1.000	0.422	0.266	0.606	1.000
	Instrument count	10	45	105	190	10	45	105	190
Collapsed instr.									
DIFF GMM	Estimated α mean	0.248	0.587	0.747	0.818	0.856	0.873	0.885	0.888
	Estimated α st. dev.	0.846	0.398	0.232	0.137	0.349	0.104	0.059	0.042
	Hansen p-value avg.	0.502	0.471	0.463	0.455	0.505	0.473	0.470	0.464
	Instrument count	3	8	13	18	3	8	13	18
SYS GMM	Estimated α mean	0.847	0.884	0.892	0.896	1.032	1.011	0.971	0.932
	Estimated α st. dev.	0.233	0.096	0.059	0.045	0.176	0.075	0.061	0.051
	Hansen p-value avg.	0.505	0.479	0.472	0.462	0.464	0.310	0.240	0.259
	Instrument count	5	10	15	20	5	10	15	20
Limited instr. (Lag 1 only)									
DIFF GMM	Estimated α mean	0.219	0.558	0.739	0.807	0.821	0.868	0.882	0.885
	Estimated α st. dev.	0.818	0.404	0.237	0.151	0.309	0.100	0.059	0.045
	Hansen p-value avg.	0.496	0.457	0.467	0.477	0.473	0.463	0.476	0.485
	Instrument count	3	8	13	18	3	8	13	18
SYS GMM	Estimated α mean	0.960	0.965	0.966	0.966	1.020	1.002	0.989	0.979
	Estimated α st. dev.	0.077	0.036	0.028	0.023	0.057	0.028	0.021	0.018
	Hansen p-value avg.	0.464	0.444	0.436	0.443	0.482	0.333	0.310	0.323
	Instrument count	7	17	27	37	7	17	27	37
Limited instr. (Lags 1 and 2)									

Continued on next page

Table 2.3: Estimates for the AR(1) with $\alpha = 0.9$ - continued

		$\rho = 0.0$ (SYS GMM valid)				$\rho = 0.9$ (SYS GMM invalid)			
Time dimension		5	10	15	20	5	10	15	20
DIFF GMM	Estimated α mean	0.242	0.459	0.640	0.736	0.764	0.842	0.869	0.876
	Estimated α st. dev.	0.569	0.340	0.247	0.175	0.292	0.106	0.062	0.045
	Hansen p-value avg.	0.489	0.444	0.451	0.452	0.469	0.445	0.460	0.463
	Instrument count	5	15	25	35	5	15	25	35
SYS GMM	Estimated α mean	0.945	0.950	0.954	0.956	1.026	1.007	0.991	0.977
	Estimated α st. dev.	0.102	0.045	0.033	0.026	0.073	0.033	0.025	0.020
	Hansen p-value avg.	0.455	0.414	0.405	0.414	0.416	0.300	0.271	0.290
	Instrument count	9	24	39	54	9	24	39	54
Limited and coll. instr. (Lag 1 only)									
DIFF GMM	Estimated α mean	0.522	-4.565	0.930	0.950	0.972	0.906	0.902	0.899
	Estimated α st. dev.	10.519	175.563	1.096	1.950	1.121	0.123	0.089	0.071
	Hansen p-value avg.								
	Instrument count	1	1	1	1	1	1	1	1
SYS GMM	Estimated α mean	0.892	0.886	0.890	0.894	1.065	1.051	1.029	0.993
	Estimated α st. dev.	0.205	0.101	0.072	0.058	0.173	0.073	0.062	0.074
	Hansen p-value avg.	0.517	0.516	0.492	0.497	0.443	0.279	0.155	0.092
	Instrument count	3	3	3	3	3	3	3	3
Limited and coll. instr. (Lags 1 and 2)									
DIFF GMM	Estimated α mean	0.397	0.765	0.867	0.875	0.870	0.899	0.898	0.897
	Estimated α st. dev.	1.414	0.915	0.388	0.239	0.331	0.116	0.081	0.063
	Hansen p-value avg.	0.539	0.522	0.505	0.503	0.478	0.490	0.492	0.494
	Instrument count	2	2	2	2	2	2	2	2
SYS GMM	Estimated α mean	0.865	0.885	0.890	0.894	1.043	1.035	1.016	0.987
	Estimated α st.dev.	0.220	0.103	0.071	0.053	0.170	0.064	0.053	0.065
	Hansen p-value avg.	0.495	0.505	0.493	0.502	0.437	0.281	0.157	0.099
	Instrument count	4	4	4	4	4	4	4	4
PCA untransformed avg									
DIFF GMM	Estimated α mean	0.266	0.561	0.724	0.800	0.849	0.869	0.880	0.883
	Estimated α st. dev.	1.282	0.416	0.254	0.156	0.349	0.104	0.060	0.044

Continued on next page

Table 2.3: Estimates for the AR(1) with $\alpha = 0.9$ - continued

		$\rho = 0.0$ (SYS GMM valid)				$\rho = 0.9$ (SYS GMM invalid)			
Time dimension		5	10	15	20	5	10	15	20
	Hansen p-value avg.	0.507	0.470	0.464	0.467	0.505	0.477	0.474	0.475
	Instrument count	2.43	7.42	12.43	17.42	2.97	8.00	13.00	18.00
SYS GMM	Estimated α mean	0.887	0.938	0.942	0.942	1.005	0.987	0.967	0.953
	Estimated α st. dev.	0.306	0.085	0.061	0.049	0.165	0.063	0.043	0.034
	Hansen p-value avg.	0.491	0.455	0.449	0.452	0.461	0.356	0.338	0.351
	Instrument count	4.91	12.33	19.77	27.15	5.44	12.88	20.28	27.70
PCA untransformed var									
DIFF GMM	Estimated α mean	0.248	0.464	0.623	0.712	0.853	0.838	0.858	0.865
	Estimated α st. dev.	0.826	0.486	0.340	0.252	0.354	0.173	0.105	0.075
	Hansen p-value avg.	0.502	0.472	0.464	0.471	0.507	0.469	0.463	0.471
	Instrument count	3.00	6.00	10.00	13.76	3.00	6.00	10.00	14.00
SYS GMM	Estimated α mean	0.941	0.952	0.956	0.956	1.028	1.014	1.001	0.989
	Estimated α st. dev.	0.108	0.054	0.042	0.035	0.084	0.045	0.037	0.033
	Hansen p-value avg.	0.475	0.449	0.436	0.432	0.430	0.369	0.330	0.323
	Instrument count	7.00	14.58	23.00	30.76	7.00	14.58	23.00	31.00
PCA collapsed avg									
DIFF GMM	Estimated α mean	-1.322	0.897	0.857	0.894	0.954	0.903	0.896	0.896
	Estimated α st. dev.	81.804	3.287	0.945	0.128	0.429	0.118	0.064	0.046
	Hansen p-value avg.			0.529	0.513			0.512	0.506
	Instrument count	1.00	1.00	2.00	2.00	1.00	1.00	2.00	2.00
SYS GMM	Estimated α mean	0.929	0.960	0.962	0.962	1.012	0.999	0.984	0.969
	Estimated α st. dev.	0.318	0.069	0.048	0.040	0.136	0.051	0.038	0.032
	Hansen p-value avg.	0.530	0.502	0.481	0.448	0.483	0.396	0.317	0.269
	Instrument count	3.47	5.91	9.34	11.73	3.47	5.88	9.28	11.70
PCA collapsed var									
DIFF GMM	Estimated α mean	-1.467	0.851	0.857	0.893	0.905	0.898	0.893	0.896
	Estimated α st. dev.	81.588	0.894	0.945	0.113	0.391	0.109	0.064	0.044
	Hansen p-value avg.	0.539	0.541	0.528	0.503	0.522	0.511	0.510	0.501
	Instrument count	1.78	2.00	2.00	2.97	2.00	2.00	2.24	3.00

Continued on next page

Table 2.3: Estimates for the AR(1) with $\alpha = 0.9$ - continued

		$\rho = 0.0$ (SYS GMM valid)				$\rho = 0.9$ (SYS GMM invalid)			
Time dimension		5	10	15	20	5	10	15	20
SYS GMM	Estimated α mean	0.948	0.962	0.966	0.965	1.026	1.009	0.996	0.982
	Estimated α st. dev.	0.096	0.046	0.034	0.029	0.073	0.033	0.027	0.024
	Hansen p-value avg.	0.486	0.469	0.448	0.412	0.444	0.362	0.281	0.224
	Instrument count	5.78	10.58	15.00	19.97	6.00	10.58	15.24	20.00

Table 2.4: Estimates for the AR(1) model with $\alpha = 0.9$ and non stationary initial conditions

Time dimension		5	10	15	20
Untransformed instr.					
DIFF GMM	Estimated α mean	0.883	0.889	0.892	0.879
	Estimated α st. dev.	0.086	0.028	0.016	0.021
	Hansen p-value avg.	0.477	0.410	0.442	1.000
	Instrument count	6	36	91	171
SYS GMM	Estimated α mean	1.139	1.042	0.992	0.969
	Estimated α st. dev.	0.152	0.028	0.017	0.017
	Hansen p-value avg.	0.055	0.111	0.599	1.000
	Instrument count	10	45	105	190
Collapsed instr.					
DIFF GMM	Estimated α mean	0.865	0.816	0.857	0.871
	Estimated α st. dev.	0.214	0.165	0.103	0.075
	Hansen p-value avg.	0.490	0.473	0.474	0.484
	Instrument count	3	8	13	18
SYS GMM	Estimated α mean	1.306	1.230	1.137	1.094
	Estimated α st. dev.	0.219	0.042	0.028	0.020
	Hansen p-value avg.	0.294	0.352	0.281	0.274
	Instrument count	5	10	15	20
PCA untransformed avg					
DIFF GMM	Estimated α mean	0.854	0.790	0.879	0.884
	Estimated α st. dev.	0.583	0.210	0.068	0.053
	Hansen p-value avg.	0.483	0.470	0.469	0.473
	Instrument count	3	9	18	27
SYS GMM	Estimated α mean	1.073	1.127	1.084	1.063
	Estimated α st. dev.	0.376	0.141	0.050	0.033
	Hansen p-value avg.	0.136	0.064	0.078	0.134
	Instrument count	5	13	24	35

Table 2.5: Estimates for the multivariate dynamic model with $\alpha = 0.9$ and $\beta = 1$

Time dimension		5	10	15	20
Untransformed instr.					
DIFF GMM	Estimated α mean	0.829	0.865	0.877	0.883
	Estimated α st. dev.	0.094	0.031	0.017	0.012
	Estimated β mean	0.884	0.960	0.977	0.985
	Estimated β st.dev	0.159	0.045	0.025	0.019
	Hansen p-value avg.	0.465	0.429	1.000	1.000
	Instrument count	15	80	195	360
SYS GMM	Estimated α mean	0.919	0.915	0.911	0.905
	Estimated α st. dev.	0.029	0.013	0.011	0.011
	Estimated β mean	0.994	1.002	1.001	0.996
	Estimated β st.dev	0.053	0.028	0.021	0.020
	Hansen p-value avg.	0.463	0.461	1.000	1.000
	Instrument count	22	97	222	397
Collapsed instr.					
DIFF GMM	Estimated α mean	0.804	0.870	0.885	0.892
	Estimated α st. dev.	0.160	0.058	0.028	0.020
	Estimated β mean	0.831	0.955	0.981	0.990
	Estimated β st.dev	0.277	0.094	0.044	0.032
	Hansen p-value avg.	0.506	0.478	0.476	0.481
	Instrument count	7	17	27	37
SYS GMM	Estimated α mean	0.901	0.898	0.898	0.899
	Estimated α st. dev.	0.046	0.021	0.015	0.012
	Estimated β mean	0.990	0.997	0.999	1.000
	Estimated β st.dev	0.071	0.034	0.024	0.019
	Hansen p-value avg.	0.491	0.481	0.480	0.477
	Instrument count	10	20	30	40
PCA untransformed avg					
DIFF GMM	Estimated α mean	0.798	0.871	0.885	0.889
	Estimated α st. dev.	0.165	0.055	0.026	0.019
	Estimated β mean	0.824	0.958	0.980	0.986
	Estimated β st.dev	0.285	0.084	0.040	0.029
	Hansen p-value avg.	0.506	0.479	0.473	0.481
	Instrument count	6.30	21.08	39.84	61.16
SYS GMM	Estimated α mean	0.905	0.908	0.907	0.906
	Estimated α st. dev.	0.048	0.021	0.016	0.012
	Estimated β mean	0.991	1.002	1.000	1.001
	Estimated β st.dev	0.084	0.036	0.027	0.021
	Hansen p-value avg.	0.485	0.488	0.476	0.473
	Instrument count	10.02	28.16	50.19	74.76
PCA untransformed var					
DIFF GMM	Estimated α mean	0.830	0.876	0.886	0.888
	Estimated α st. dev.	0.118	0.041	0.021	0.014
	Estimated β mean	0.881	0.968	0.983	0.987
	Estimated β st.dev	0.203	0.060	0.033	0.023
	Hansen p-value avg.	0.490	0.479	0.480	0.596

Continued on next page

Table 2.5: Estimates for the multivariate dynamic model with $\alpha = 0.9$ and $\beta = 1$ - continued

Time dimension		5	10	15	20
	Instrument count	9.54	29.99	61.38	102.28
SYS GMM	Estimated α mean	0.919	0.917	0.916	0.914
	Estimated α st. dev.	0.028	0.015	0.012	0.009
	Estimated β mean	0.995	1.002	1.003	1.005
	Estimated β st.dev	0.052	0.031	0.022	0.017
	Hansen p-value avg.	0.483	0.453	0.467	0.974
	Instrument count	16.54	45.02	84.87	134.05

Table 2.6: Estimation of Forbes [2000]'s model on the effect of inequality on growth

Dependent variable: $Growth_{it}$ (GDP per capita growth)														
Variable		Original	ABall	BBall	ABlim	BBlim	ABcol	BBcol	ABlimcol	BBlimcol	ABpcaa	ABpcav	BBpcaa	BBpcav
L.Gini	*coeff*	0.0013	0.0032	0.0015	0.0026	0.0012	0.0032	0.0023	0.0026	0.0033	0.0037	0.0022	0.0015	0.0013
	sd	0.001	0.001	0.001	0.002	0.001	0.003	0.002	0.005	0.002	0.003	0.002	0.001	0.001
	t	2.17	2.12	1.88	1.25	1.26	1.09	1.5	0.57	2.15	1.37	0.96	1.2	1.62
L.Income	*coeff*	-0.0470	-0.0538	-0.0042	-0.0533	0.001	-0.0188	0.0037	0.0574	0.054	-0.0451	-0.0298	0.007	-0.0024
	sd	0.008	0.029	0.007	0.036	0.008	0.041	0.026	0.053	0.029	0.059	0.069	0.009	0.007
	t	5.88	-1.89	-0.6	-1.49	0.12	-0.46	0.14	1.08	1.84	-0.76	-0.43	0.78	-0.34
L.MaleEduc	*coeff*	-0.0080	0.0049	0.0229	-0.0016	0.0322	-0.0162	0.0365	0.0512	0.0157	0.0053	0.0068	0.022	0.0271
	sd	0.022	0.023	0.016	0.036	0.016	0.035	0.027	0.112	0.024	0.059	0.056	0.015	0.016
	t	0.36	0.21	1.42	-0.05	1.97	-0.47	1.37	0.46	0.66	0.09	0.12	1.45	1.65
L. FemaleEduc	*coeff*	0.0740	0.0183	-0.0062	0.0271	-0.0192	0.0472	-0.0123	-0.0269	-0.022	0.05	0.0212	-0.0115	-0.0144
	sd	0.018	0.021	0.014	0.035	0.012	0.031	0.021	0.095	0.021	0.053	0.054	0.015	0.016
	t	4.11	0.86	-0.44	0.78	-1.57	1.5	-0.58	-0.28	-1.03	0.94	0.4	-0.74	-0.93
L. InvPrice	*coeff*	-0.0013	-0.0007	-0.0005	-0.0008	-0.0005	-0.0008	-0.001	-0.0011	-0.001	-0.0003	-0.0005	-0.0004	-0.0004
	sd	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	t	13.00	-3.72	-3.45	-5.26	-2.67	-2.13	-3.32	-2.87	-4.18	-1.28	-2.35	-1.45	-2.06
Obs. $(N \times T)$		135	138	228	138	228	138	228	138	228	138	138	228	228
$\overline{T}$			3.0667	3.3529	3.0667	3.3529	3.0667	3.3529	3.0667	3.3529	3.0667	3.0667	3.3529	3.3529
Instruments			80	106	30	56	30	36	10	16	25	27	38	46
hansenp			0.9999	0.9997	0.5339	0.3786	0.1188	0.3294		0.6497	0.2842	0.1754	0.3749	0.3242
Autocorrelation			0.2708	0.0365	0.2504	0.0410	0.1306	0.1976	0.2123	0.0522	0.2119	0.0626	0.0230	0.0289

Notes:
a. In the first column we report the estimates of Forbes [2000]. The other columns show the estimates on the reproduction dataset by Roodman [2009b].
b. AB and BB stand, respectively, for the Arellano-Bond DIFF and the Blundell-Bond SYS two-step estimators. all indicates we are exploiting the full instrument set; lim stands for the limited set of instruments where only the second-lag is kept; col is the collapsed instrument matrix; limcol means that instruments are both limited and collapsed; pcaa and pcav stand, respectively, for the factorized instrument set where the principal components are retained according to the average criterion and the variability criterion.
c. We report the estimates for the coefficients (coeff), the standard deviation (sd) and the t-statistic (sig) for the significance of the coefficients.
d. hansenp is the p-value for the Hansen overidentifying restriction test (robust but weakened by many instruments; see Roodman [2009b]).
e. Autocorrelation reports the p-value of the residuals' second-order autocorrelation test.

Appendix

The principal component analysis (PCA)

The PCA is a statistical tool which is used for data reduction according to a data-driven procedure. Intuitively, what PCA does is to find several orthogonal linear combinations of the original variables ordering them on the basis of the portion of the variance in the original data they account for. A principal component is therefore a linear combination of observed variables that is obtained by exploiting a set of optimal weights for each original variable. The first principal component (PC) will be the linear combination of the original variables that has the largest variance among all the possible linear combinations of the original variables. The second PC will be the linear combination, orthogonal to the first PC, that accounts for the largest portion of the residual variance once the first PC has been extracted, and so on. All the principal components taken together contain all the information conveyed by the original data.

In other words, through PCA we aim at reducing the dimension of the data while retaining, at the same time, as much of the original variability in the data as possible.

More formally, if we define $\mathbf{C}$ as the p x p covariance or correlation matrix of the p original variables in the data, the k^{th} principal component $\mathbf{pc}_k$ for $k = 1, 2, ..., p$ is obtained as

$$\mathbf{pc}_k = \mathbf{u}_k'\mathbf{x} \tag{2.6.25}$$

where $\mathbf{x}$ is the vector of the p variables in the sample, $\mathbf{u}_k$ is the k^{th} eigenvector of $\mathbf{C}$ corresponding to the k^{th} largest eigenvalue λ_k subject to the normalization constraints:

$$\mathbf{u_k}'\mathbf{u_k} = 1 \tag{2.6.26}$$

$$\mathbf{u_k}'\mathbf{u_j} = 0 \text{ for } i \neq j. \tag{2.6.27}$$

$\mathbf{pc}_1 = \mathbf{u}_1'\mathbf{x}$ is therefore the linear combination of the p variables orthogonal to all other combinations that, subject to the above constraints, has the maximum variance. Similarly $\mathbf{pc}_2$ is the linear combination, orthogonal to pc_1, that maximizes the residual variance.

In matrix notation, we can interpret the principal components in the light of the eigenvalue-eigenvector decomposition of the correlation or the covariance matrix $\mathbf{C}$:

$$\mathbf{C} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}' = \sum_{i=1}^{p} \lambda_i \mathbf{v}_i \mathbf{v}_i' \tag{2.6.28}$$

where $\mathbf{V}$ is the matrix consisting of the eigenvectors (principal components) of $\mathbf{C}$, $\mathbf{\Lambda}$ is the diagonal matrix that has as element kk the eigenvalue λ_k corresponding to the eigenvector $\mathbf{v}_k$. The elements v_{kj} of the eigenvector $\mathbf{v}_k$, namely the coefficients of each linear combination, are the *loadings*, that represent the contribution of each original value to the PC: in other words, they can be interpreted as the weights of the j^{th} variable in $\mathbf{pc}_k$.

Subject to the conditions in equations (2.6.27) and (13), that is if $\mathbf{u}_k$ is such to have unit length, the variance of the k^{th} principal component, $\mathrm{var}(\mathbf{pc}_k)$, is given by λ_k. The total variance of all the principal components will be equal to the variance of the original variables so that:

$$\sum_{k=1}^{p} \lambda_k = \mathrm{tr}(\mathbf{C}). \tag{2.6.29}$$

As a consequence, each principal component will account for a portion of the variance of the original data equal to:

$$P_k = \frac{\lambda_k}{\mathrm{tr}(\mathbf{C})}. \tag{2.6.30}$$

By multiplying each original variable by its loading in each PC, we obtain the matrix of the principal component *scores* defined as follows:

$$\mathbf{S} = \mathbf{X}\mathbf{V} \tag{2.6.31}$$

where $\mathbf{X}$ is the original data matrix and $\mathbf{V}$ is the same as above. In other terms, the scores s_j indicate the influence of a PC on a specific sample. The matrix $\mathbf{S}$ can be used in the analysis in the place of $\mathbf{X}$: in fact, the matrix $\mathbf{S}$ contains the original data matrix in a rotated coordinate system. Clearly the original matrix of data can be written as:

$$\mathbf{X} = \mathbf{V}'\mathbf{S} \tag{2.6.32}$$

where **V** and **S** are orthogonal.

The number of eigenvalues and eigenvectors, and thus of the principal components, obviously equals the number of variables in the original data.

As the aim of PCA is a reduction of the data dimension through a maximization of the variance explained by the first components and the elimination of multicollinearities in the data, that imply potential problems in inverting the original matrix, we will want to select and keep a number of components q which is smaller than p: we will therefore select the q eigenvectors corresponding to the q largest eigenvalues of **C** such that they explain most of the variability in the data. The q largest principal components will account for the following portion of the original variance:

$$\frac{\sum_{k=1}^{q} \lambda_k}{\mathrm{tr}(\mathbf{C})}. \tag{2.6.33}$$

Accordingly, in the matrix **V** only q eigenvectors will be retained and the scores will be computed form the reduced **V** matrix.

It is then possible to exploit directly the scores from the PCA by using them instead of the original variables.

A relevant issue is how to choose the the q principal components to be retained in the analysis. Two criteria are generally adopted in the literature: the first implies that only the components that explain a given predetermined portion, usually between 70% and 90%, of the original variance are to be retained; the second one keeps only the components whose eigenvalues are larger than the average eigenvalue which obviously is the average variance in the original data.

Chapter 3

GMM estimation of fiscal reaction functions: Monte Carlo experiments and empirical tests

3.1 Introduction

The empirical estimation of fiscal reaction functions has gained in popularity in the last few decades and it is now particularly diffused in the fiscal policy literature. The estimation of fiscal rules, enhanced by the seminal works of Bohn [1998] and Taylor [2000], has been mainly motivated by the question whether discretionary fiscal actions by the policy makers act pro-cyclically or counter-cyclically.

Finding an answer about the reaction of budgetary policies to the economic cycle has become particularly important within the European Monetary Union where the constraints imposed by the Maastricht Treaty and the Stability Growth Pact potentially affect the response of fiscal variables to the economic cycle and could weaken the autonomy of the national policy maker in determining discretionary fiscal actions.

In a very influential paper, Galì and Perotti [2003] find that discretionary policy in the EMU Countries has become more countercyclical over time: their work has given a further boost to a lively debate on the response of fiscal policies to the cycle that has led to many recent contributions which estimate fiscal reaction function in the EU or the OECD area. Among the most relevant works it is worth numbering in particular the contributions of Ballabriga and Martinez Mongay [2002], Balassone and Francese [2004], Forni and Momigliano [2004], Wyplosz [2006], Debrun and Kumar [2007], Golinelli and Momigliano [2006] and Beetsma and Giuliodori [2008].

What emerges from this stream of literature, however, is a strong lack of consensus on whether discretionary fiscal policy behaves procyclically or counter-cyclically: the results are often conflicting, even in the case they focus on similar, or even same, samples of Countries and on a comparable time span. Golinelli and Momigliano [2009] provide an extensive survey of the empirical works on fiscal response functions and explain the enormous heterogeneity in empirical findings as due to differences in the specifications of the fiscal rule, in the estimation methodologies and in the samples covered.

In a context where the empirical findings have proved not to be robust to different estimation methodologies, we believe it can be very useful to investigate the sensitivity of the estimates of the fiscal variables to alternative choices about the estimation techniques.

Our interest is therefore mainly focused on the comparison of alternative estimators that can be adopted in order to estimate fiscal response functions.

The dimension of the datasets in fiscal policy literature is generally limited both in terms of the number of years available and in terms of Countries considered: it is therefore common habit to consider the sample as a panel and to adopt panel data techniques in order to estimate fiscal rules. The fiscal reaction function is by its nature a dynamic model which also include unobservable fixed effects, potentially endogenous additional regressors and in which the series of interest are likely to be highly persistent. The choice of the methodology should therefore account for these relevant econometric issues and try to fix them properly. Unfortunately, in spite of these problems, it is still common to estimate the fiscal reaction function with OLS or fixed-effect estimators[1] that may not be appropriate in this context.

A common approach in the literature to deal with the potentially dangerous econometric issues in this context is to use a linear GMM estimator: in particular, the most popular choice is the Arellano-Bond Difference GMM estimator [Arellano and Bond, 1991] for dynamic panel data models[2]. In the context of fiscal policy, however, where the series are often very persistent, we argue that the Blundell-Bond System GMM estimator [Blundell and Bond, 1998] could be a more appropriate choice, as it allows to overcome a potential weak instrument problem in the estimation procedure. Only few empirical works have estimated fiscal reaction functions by System GMM[3].

Overall there is still very little guidance on which is the safest strategy to adopt in order to estimate fiscal response function.

Our main purpose is therefore to compare the performances of different dynamic panel data estimators in the estimation of fiscal rules and, in particular, to investigate the sensitiveness of the estimates to alternative settings of the GMM estimators. We will pay a particular attention to the issues of endogeneity and instrument proliferation and we will adopt specific techniques in order to reduce the instrument count: we will thus check the robustness of the estimates to these alternative specifications of the estimators.

Monte Carlo simulations are the privileged tool to study the statistical properties of alternative estimators for the coefficients in a fiscal rule: in fact they allow us to simulate a model for which we have control over all the parameters.

[1]Taylor [2000], Galì and Perotti [2003], Forni and Momigliano [2004], Wyplosz [2006], Debrun et al. [2008], among others, all adopted these strategies.

[2]This approach has been followed, among the others, by Balassone and Francese [2004], Forni and Momigliano [2004], Debrum and Kumar [2007].

[3]Golinelli and Momigliano [2006, 2009] and Bernouth et al.[2008] are among the few ones.

The present work contributes to the literature in fiscal policies in several ways.

First, we simulate the most popular fiscal rule used in the literature, the so called cyclically-adjusted primary balance ($CAPB$) model[4], and we use Monte Carlo experiments in order to assess the performance of alternative estimators. To our knowledge, this is the first Monte Carlo experiment on fiscal rule models that aims at comparing various estimators and at giving practical indications on the safest methodologies to use in this framework. As far as we know, only Celasun and Kang [2006] have used Monte Carlo simulations in the context of the estimation of fiscal reaction functions but their experiment design is very different from ours and has different purposes[5].

Secondly, we estimate on simulated data also a different fiscal rule, the primary balance (PB) model[6], where the dependent variable is the unadjusted primary balance, in order to check whether the widespread habit of assessing the discretionary fiscal response to the cycle directly from the estimates of the PB model is a safe strategy.

Third, we estimate the $CAPB$ model on real data for the EMU and we present the estimates of the model with all the alternative estimators: we do not choose a methodology a priori, as commonly and mistakenly done in the literature, but we interpret all the results in the light of the findings in Monte Carlo simulations.

The remainder of the paper proceeds as follows. In section 2 we

[4]In a nutshell, the $CAPB$ model aims at explaining the discretionary fiscal policy (i.e. not due to automatic stabilizers) on the basis of an economic cycle indicator and of the initial conditions of both deficit and debt.

[5]Celasun and Kang [2006] in fact adopt a fiscal rule where the lagged dependent is the primary balance, and not the cyclically-adjusted primary balance, and where the lagged dependent variable is not included among the regressors. On one hand, they aim at assessing the bias of the OLS and of the Least Squares Dummy Variable (LSDV) estimators when the are used to estimate the fiscal rule with respect to the bias in a standard AR(1) model for debt; on the other hand they want to check the robustness of the estimates obtained with alternative estimators to different parameter settings in the model when the dynamics is left out of the model. We strongly believe, and most of the authors in the literature with us, that the dynamics is an intrinsic feature of fiscal rules and that it can not be disregarded.

[6]Differently from the $CAPB$, the PB model aims at explaining the overall fiscal policy, instead of only the discretionary policy.

present in details the simulation model and we discuss the setting of the parameters in the fiscal rule. Section 3 introduces relevant econometric issues and reports the estimates of the simulated $CAPB$ model. In section 4 we illustrate the PB model and its links with the $CAPB$ model, we estimate it on simulated data and we present a relatively safe strategy to derive the discretionary adjustments to the cycle from the estimates of the PB model parameters. Section 5 is devoted to the estimation of the $CAPB$ model on real data. Section 6 draws conclusive indications and sketches potential addresses for future research.

3.2 Simulation model

We adopt here the fiscal reaction function used in most empirical works in fiscal policy[7]: the **CAPB model**, where the dependent variable, namely the change in the cyclically-adjusted primary balance ($\Delta CAPB$), measures the discretionary fiscal actions that can be taken by policy makers. $\Delta CAPB$ is explained by the lagged values of the cyclically-adjusted primary balance ($CAPB$) and the stock of public debt ($DEBT$), that represent the state of public finances, and by the economic cycle as captured by the lagged level of the output gap (GAP)[8].

The fiscal rule we want to estimate is therefore the following:

$$\Delta CAPB_{it} = \phi_c CAPB_{i,t-1} + \phi_d DEBT_{i,t-1} + \phi_g GAP_{i,t-1} + \mu_i + \varepsilon_{it} \quad (3.2.1)$$

where the $\mu_i \sim \mathcal{N}(0, \sigma^2_\mu)$ are the fixed effects for the $CAPB$ and the $\varepsilon_{it} \sim \mathcal{N}(0, \sigma^2_\varepsilon)$ are fiscal policy shocks.

[7]The $CAPB$ model is chosen, among the others, by Galì and Perotti [2003], Forni and Momigliano [2004], Wyplosz [2006] and Golinelli and Momigliano [2009].

[8]We take here the output gap lagged one period as common in the literature. It is worth noticing however that some authors include the simultaneous output gap instead of the lagged one. At the same way, some authors prefer the actual level of the $CAPB$ as dependent variable instead of its change. Golinelli and Momigliano [2009], in their extensive review of the literature on the empirical estimation of fiscal reaction functions, discuss the different specifications of the fiscal models, the links between them and how the interpretation of the estimation results changes according to the model chosen in the analyses.

A positive coefficient ϕ_g implies that discretionary budgetary actions are counter-cyclical, while a negative coefficient indicates pro-cyclicality.

It follows from equation (3.2.1) that the AR process for the $CAPB$ is:

$$CAPB_{it} = (\phi_c + 1)CAPB_{i,t-1} + \phi_d DEBT_{i,t-1} + \phi_g GAP_{i,t-1} + \mu_i + \varepsilon_{it}. \tag{3.2.2}$$

We need to simulate all the variables that appear in equations (3.2.2).

It will be clear soon that, in order to dispose of the series for the gap, the debt and the cyclically-adjusted primary balance, it is not enough to simulate only these series of interest: we need to actually simulate many more variables from which the variables of interest are derived, according to accountancy rules.

We generate the output gap as an autoregressive process that is independent of the other fiscal variables, except for the presence of fiscal policy shocks that also affect other variables; the cyclically-adjusted primary balance is generated according to the process in equation (3.2.2) and the remaining variables follow well-known public accountancy rules.

Our simulation model is thus as follows.

Cyclically-adjusted primary balance

We initialize the $CAPB$ by generating its initial conditions so that they are stationary and satisfy the Blundell-Bond conditions for the system-GMM estimation[9]. Therefore we have:

$$CAPB_{i0} = \frac{\mu_i}{1 - \phi_g} + u_{i0} \tag{3.2.3}$$

where the $u_{i0} \sim \mathcal{N}(0, 1)$ are the random deviations from the long-run mean of the cyclically-adjusted primary balance $\mu_i / (1 - \phi_g)$.

Once we have generated the initial value for $CAPB$ and the initial values of the variables that appear in equation (3.2.2), we will be able to generate $CAPB_{i1}$ and so on, iteratively, period by period.

[9]In Chapter 1, we extensively discuss the content of these assumptions and their implications in GMM estimation.

Output gap

We generate stationary initial conditions also for the gap as follows:

$$GAP_{i0} = \frac{\eta_i}{1-\alpha} + z_{i0} \tag{3.2.4}$$

where the $\eta_i \sim \mathcal{N}(0, \sigma_\eta^2)$ are fixed effects for the output gap and $z_{i0} \sim \mathcal{N}(0,1)$ are the deviations from the long-run mean of the output gap.

We then generate the rest of the series as follows:

$$GAP_{it} = \alpha GAP_{i,t-1} + \eta_i + \xi \varepsilon_{i,t-1} + v_{it} \tag{3.2.5}$$

where the ε_{it} are again the fiscal policy shocks and the $v_{it} \sim \mathcal{N}(0, \sigma_v^2)$ are idiosyncratic shocks.

Nominal growth rate

$$n_t \equiv \left[e^{\ln(1+\dot{p}_{it})+\ln(1+pg_{it})+\Delta\ln(1+GAP_{it})}\right] - 1 \tag{3.2.6}$$

where $\dot{p}_{it}$ is the inflation rate and pg_{it} is the potential growth. The inflation rate follows the process $\dot{p}_{it} = \dot{p}_{i,t-1} + \zeta GAP_{i,t-1}$, with $\zeta = 0.05$, while the potential growth is assumed to be centered around the 2% according to the process $pg = 2 + \gamma_{it}$ with $\gamma_{it} \sim \mathcal{N}(0,1)$ being a random shock.

Primary balance

$$PB_{it} \equiv \frac{CAPB_{it}}{1+GAP_{it}} + \omega \frac{GAP_{it}}{1+GAP_{it}} \tag{3.2.7}$$

with ω being the elasticity of the overall budget that represents the effect of the automatic stabilizers.

Interest payments

$$INT_{it} \equiv R_{it} \frac{DEBT_{i,t-1}}{1+n_{it}} \tag{3.2.8}$$

where R_{it} is the average cost of debt and is assumed to move together with the nominal growth according to the process $R_{it} = n_{it} + \iota_{it}$ with $\iota_{it} \sim \mathcal{N}(0, 1)$ being a random shock.

Overall fiscal balance

$$B_{it} \equiv PB_{it} - INT_{it}. \tag{3.2.9}$$

Public debt

$$DEBT_{it} = \frac{DEBT_{i,t-1}}{1 + n_{it}} - B_{it}. \tag{3.2.10}$$

In this simulation process, the variables listed above can be generated sequentially period by period for each individual unit. We start setting the initial conditions for the output gap and for the cyclically-adjusted primary balance, we then initialize the other variables for the first period and we simulate them all as a cascade process following the relationships presented above.

3.2.1 Baseline setting for the parameters

We simulate a panel dataset for a fixed number of periods T and for a fixed number of economies N. We want to replicate realistic profiles for the series and, as a reference framework, we have in mind the advanced economies of the European Union or of the most developed OECD Countries[10]. A a consequence, we fix the number of Countries to $N = 15$ as this is a very reasonable number of units in the empirical estimation of fiscal rules for the European Union or the OECD: in fact in most of the empirical literature on fiscal reaction functions the sample size is generally not far from 15.

[10]Our choice is motivated by the fact that the European Union, and in particular the Euro area, is a privileged context for the estimation of fiscal reaction functions in the fiscal policy literature and it will also be our sample in the empirical analysis in the prosecution of the chapter.

We also fix the number of time periods to $T = 15$, as this a reasonable time span for a dynamic panel analysis and it is about the number of years for which data are available, in the EU context, after the Maastricht agreement. Moreover, if we extended further T we would move toward a context where it would be more reasonable to consider individual time series rather than a panel, as doubts could rise about parameters' poolability. In order to have 15 periods available we generate data for a time span of 50 periods and we allow for different pre-samples, i.e. 1, 15 and 35 years. Independently of the length of the pre-sample, we always generate the same 50-years samples and we then keep time by time only the years we are interested in: what changes is simply the point in history we capture in the analysis.

Through calibration, we set the variances and the parameters in the model to values that make the simulated scenario as more realistic as possible so that the profiles of the simulated series are similar to the observed ones, or are plausible realizations of actual data for advanced economies. In order to succeed in this task, we match as close as possible these parameter values with those observed in our real data in order to replicate the empirical moments of the dataset we will use later.

In Table (3.1) we summarize the baseline setting of the parameters that we will use in the core of our simulation exercise. Deviations of several parameters from the baseline values have also been considered to check for the robustness of the results[11].

We generate the values of the variables as percentages of the GDP so that, e.g., a value for $DEBT_{it}$ equal to 50 indicates that the stock of public debt of Country i in period t is the 50% of the GDP; therefore, according to the same scaling, a standard deviation equal to 1 has to be seen as a standard deviation of 1%.

With regards to our baseline setting, we can notice from Table (3.1) that the standard deviations for the fixed effects of the output gap and

[11] The results are not presented here but are available upon request. In particular, we have run the experiments with larger variances for the fixed effects of $CAPB$ and GAP and the fiscal policy shocks, that is $\sigma_\mu = \sigma_\eta = \sigma_\varepsilon = 1$. The profile of the series, with this modified setting, become explosive and the degree of heterogeneity introduced in the data is completely unrealistic. We see this trial as a robustness check for our preferred setting.

Table 3.1: Baseline setting of the parameters in the simulation model

N	15
T	15
σ_μ	0.15
σ_η	0.30
σ_ε	0.30
σ_v	1
$\phi_c + 1$	0.8 or 0.1
ϕ_g	0.10, -0.10 or 0
ϕ_d	0.15
α	0.8 or 0.1
ξ	0.2
ω	0.5

the cyclically-adjusted primary balance are relatively low. This is justified by the fact that, in our period by period simulation process, the fixed effects are cumulated year after year through the autoregressive processes of the two variables. Standard deviations between 0.15 and 0.30 are high enough to introduce in the model a realistic degree of heterogeneity that matches closely the standard deviation (around 0.40) of the fixed effects estimated on actual data. Higher variances for the $CAPB$ and GAP effects would imply an induced non stationarity in the series and an explosive profile for the debt[12].

We keep also the variance of the fiscal policy shocks relatively low, as we generate the shocks separately and independently for each period and each Country. In order not to complicate the model further, we do not simulate here common shocks for the different Countries and we do not introduce any transmission mechanism of the shocks[13]. The simultaneous fiscal policy shock appears in equation (3.2.2) for $CAPB$, while the lagged fiscal shock shows up in the equation for GAP, scaled by a coefficient $\xi = 0.2$, thus making the output gap predetermined but not

[12]In this latter case, unrealistic values for the fixed effects variances generate a trend in the series.

[13]In fact, it would then be unrealistic to assume that very severe fiscal shocks could affect only one Country without affecting the other economies

strictly exogenous.

In our baseline simulation, we set the autoregressive coefficients for GAP and $CAPB$ both to 0.8 as the actual series are very persistent[14]. As a pure robustness check for our results, we will also consider an opposite extreme scenario in which the dynamics of the gap is very fast ($\alpha = 0.1$) and the inertia of fiscal policies is lower, i.e. $\phi_c + 1 = 0.1$. We acknowledge that such an alternative setting is certainly far from being realistic but it can be very useful for robustness checks.

We fix ω, the effect of the automatic stabilizers, to 0.5 as constant over time and across Countries: the actual value of this elasticity is around 0.5 on average in the Euro area and in the advanced OECD economies[15] so this is the most natural choice in our simulation, as commonly used in the literature.

A crucial issue is the setting of the parameters of the variables in the fiscal rules, as they represent the reaction of discretionary fiscal actions, as measured by $\Delta CAPB$, to the economic cycle and to the initial fiscal conditions. When the policies are sticky, i.e. $\phi_c + 1 = 0.8$, the implied ϕ_c in equation (3.2.1) is -0.2; when policies are instead completely flexible, i.e. $\phi_c + 1 = 0.1$, we have $\phi_c = -0.9$.

We expect ϕ_g and ϕ_d not to be very high in absolute values, as the effect of discretionary actions is by its nature limited when compared to the effect of automatic stabilizers.

The chosen values for ϕ_g and ϕ_d are respectively 0.1 and 0.005[16].

The value of ϕ_g is also a relevant issue. A value of $\phi_g = 0.10$ obvi-

[14] The dynamics of the economic cycle, measured by α, is generally very slow; the inertia of the fiscal policies, measured by the autoregressive coefficient of the $CAPB$, i.e. $\phi_c + 1$, is commonly very strong

[15] The average ω for the Old Europe Members is generally between 0.45 and 0.50 and it is on average around 0.45 for the New Members of the EU. See IMF Fiscal Monitor [2011] and OECD Economic Outlook [2011] for further details.

[16] It is also worth noticing that, despite the fact that the adjustment to the output gap and to initial fiscal conditions seems small at a first sight, these parameters imply a long run adjustment relationship as follows:

$$CAPB^* = \frac{0.1}{0.2} GAP^* + \frac{0.005}{0.2} DEBT^* = 0.5 GAP^* + 0.025 DEBT^* \tag{3.2.11}$$

that is far from being economically irrelevant.

ously implies that we are arbitrarily assuming discretionary fiscal decisions to be strongly counter-cyclical, when, in reality, they could also be pro-cyclical or even a-cyclical. In our simulation exercise we aim at comparing the performances of alternative estimators rather than at checking whether the fiscal policies are cyclically symmetric or not. What matters in this context is therefore the absolute value of the output gap coefficient, rather than its sign: however, as a robustness check, we will also run few simulations with $\phi_g = -0.10$. A different strategy is to set $\phi_g = 0$ to check whether the alternative estimators are able to detect the a-cyclicality of the policies when it really takes place.

3.3 Monte Carlo simulations of the CAPB model

3.3.1 Relevant econometric issues

The model we want to estimate, as it is specified in equation (3.2.1), is a dynamic panel data model, as it presents the lagged dependent variable among the regressors. It is characterized by the presence of fixed effects that can be problematic when not taken properly into account.

The output gap, though generated independently of the other variables, comes out to be predetermined but not strictly exogenous, due to the presence in its generating process of the lagged fiscal shocks that also appears in the fiscal rule but this time in their realization contemporaneous to $\Delta CAPB_{it}$. Specific remedies are therefore needed to account for the endogeneity of the output gap. Also from a purely economic point of view, this endogeneity issue is likely to be even more serious in the real world: it is in fact reasonable to think that the output gap is potentially endogenous to the shocks that affect the primary balance.

In a real economy, also the public debt can hardly be exogenous to shocks in the $CAPB$: in our simulated model, the debt is by construction endogenous and needs therefore to be instrumented.

A different kind of endogeneity is due to the dynamic specification of the fiscal rule: the lagged dependent variable is correlated with the fixed effects in the model, so that it is not strictly exogenous, giving rise to the well-known dynamic panel bias. Both the within transformation and the first difference transformation, that are strategies commonly applied in this framework, - the latter being also a pillar of GMM estimation - induce a correlation between the transformed lagged dependent variable and the transformed error terms and thus imply endogeneity problems in the model. An instrumental variable approach is therefore needed in this framework.

In the GMM-estimation of the model, we will therefore instrument all the regressors. We need to be aware that the recourse to instruments for all the regressors can easily lead to a problem of instrument proliferation that could bias the GMM estimates[17] and weaken the tests for overidentifying restrictions. It is therefore opportune to try to detect this problem and adopt proper solutions to face it.

Another relevant issue is that the more realistic simulated scenario, with $\phi_g = 0.8$ and $\phi_c + 1 = 0.8$ implies a high degree of persistence of the series of interest. In this context, the weak instrument issue could easily arise in the difference-GMM estimation of the model as the instruments in levels are only weakly correlated to the lagged first differenced endogenous variables[18].

3.3.2 Estimation results with simulated data

We now present the outcomes of the Monte Carlo exercise we have performed in order to analyze the performance of alternative estimators for the coefficients of the fiscal rule in equation (3.2.1).

Econometric specifications for the estimates

For each setting of the parameters in the simulation model, we run 1000 iterations on samples with 225 observations and report the mean

[17]See Chapter 1 for more details and Ziliak [1997], Roodman [2009b] and Bowsher [2002] for further references.

[18]See Blundell and Bond [1998] for the discussion of this well-known issue.

of each estimated coefficient over all the iterations (*mean*), the standard deviation of the estimated coefficient (*sd*), the power of the t-test for each coefficient (*sig*)[19] and the average Hansen-test p-value (*hansenp*).

Equation (3.2.1) is estimated with the following panel data estimators (the abbreviations used in the tables are within brackets)[20]:

- Ordinary least squares (*OLS*);
- Fixed effects estimator (*FE*);
- Anderson-Hsiao Instrumental variable estimator (*AH*): we consider here only the specification with instruments in levels;
- Arellano-Bond difference-GMM estimator with a full set of instruments (*ABa*): all the potentially available lags from $t-2$ backward are used as instruments in levels for the endogenous regressors in period t;
- Blundell-Bond system-GMM estimator with a full instrument set (*BBa*); all the potentially available lags from $t-2$ backward are used as instruments in levels for the model in first differences and the lagged first difference of the endogenous regressors is used as instrument for the equation in levels;
- Arellano-Bond difference-GMM estimator with a limited set of instruments (*ABl*): with respect to *ABa*, we use here only the $t-2$ and $t-3$ lags of the endogenous regressors as instruments;
- Blundell-Bond system-GMM estimator with a limited set of instruments (*BBl*): with respect to *BBa*, we use here only the $t-2$ and $t-3$ lags of the endogenous regressors as instruments for the equation in first differences;

[19]For each coefficient, we test the null hypotesis $H_0 : \beta = 0$ against the alternative hypothesis $H_a : \beta \neq 0$ at a 5% significance level. The power of the test gives us the probability of rejecting H_0 when it is false. Only in the case of simulated data for which $\phi_g = 0$, *sig* should be interpreted as the size of the test, as it gives the probability of rejecting H_0 when it is true.

[20]Chapter 1 provides an extensive overview of these estimators.

- Arellano-Bond difference-GMM estimator with a collapsed set of instuments (ABc): the full instrument set is collapsed following Roodman [2009b];

- Blundell-Bond system-GMM estimator with a collapsed set of instruments (BBc): the full set of instruments is collapsed following Roodman [2009b];

- Arellano-Bond difference-GMM estimator with a limited collapsed set of instruments ($ABlc$): the limited set of instruments is also collapsed;

- Blundell-Bond system-GMM estimator with a limited collapsed set of instruments ($BBlc$): the limited set of instruments is also collapsed;

- Arellano-Bond difference-GMM estimator with a factorized instrument set[21] according to the average criterion ($ABpcaa$)[22];

- Arellano-Bond difference-GMM estimator with a factorized instrument set according to the variability criterion ($ABpcav$)[23];

- Blundell-Bond system-GMM estimator with a factorized set of instruments according to the average criterion ($BBpcaa$);

- Blundell-Bond system-GMM estimator with a factorized set of instruments according to the variability criterion ($BBpcav$).

The dependent variable is always the change in the cyclically-adjusted primary balance, namely $\Delta CAPB_{it}$. Since all the regressors are considered as endogenous, they are all instrumented.

[21] The application of the principal component analysis (PCA) on the GMM-style instrument matrix is the core of Chapter 2. Please refer to Chapter 2 for a detailed discussion and for further references.

[22] We keep in the analysis only the principal components whose eigenvalues are above the average of the eigenvalues. See Chapter 2 for further details.

[23] We keep a number of principal components such that the explained variance is the 70% of the total variance in the original data. The variability criterion is also discussed in details in Chapter 2.

All the estimates are made robust to the potential heteroskedasticity of the residuals. With respect to GMM estimators, only one-step estimates of the parameters are reported[24].

The true parameters for the lagged $DEBT$ and GAP are respectively ϕ_d and ϕ_g whose values are set as in Table (3.1); with respect to the lagged value of the dependent variable instead, its true parameter will be $\phi_c = -0.2$ when we set $\phi_c + 1 = 0.8$ and $\phi_c = -0.9$ when $\phi_c + 1 = 0.1$.

The core of our simulations is run on a 15-year pre-sample so that the length of the pre-sample equals the number of periods considered in the analysis.

The estimation results for the $CAPB$ model are presented through Table (3.2) to Table (3.7).

For the sake of brevity, in the comments to the estimates we will refer to DIFF and SYS GMM to indicate, respectively, the system and the difference GMM. UNTR, COL and LIM will indicate respectively that the estimators exploit the untransformed, collapsed or limited[25] set of instruments. FACT stands for the factorized set of instruments[26]. When two reduction techniques are combined, two abbreviations are also combined (e.g. LIMCOL).

A preliminary robustness check: alternative pre-sample lengths

We adopt a 15-year pre-sample in our experiments. As a preliminary robustness check, however, Table (3.2) and Table (3.3) present the estimation results when we set either a 1-year pre-sample or a 35-year pre-sample. The setting of the parameter is the baseline one, with $\phi_c = 0.8$ and $\phi_g = 0.8$.

[24]Though more efficient in large samples and well performing in simulated samples, the two-step estimator suffers a poor finite sample behaviour on actual data and it is rarely used in the empirical analyses. We also estimated all the coefficient by two-step GMM and found larger biases and variances due to a pretty small sample. For the sake of consistency with the empirical estimations that will follow, we prefer reporting only one-step estimates already by now. Two-steps results are available upon request.

[25]We sometimes specify 1 or 2 to indicate a lag depth of 1 or 2.

[26]The additional letters A and V will indicate, respectively, that the principal components are retained according to the average and the variability criteria.

In general, what emerges in this first scenario is that the behavior of the alternative estimators is sufficiently in line with what we could expect in a dynamic panel data framework, both with a minimum and a very long pre-sample: the GMM estimators perform better than the OLS estimator, as the latter does not take into account the presence of fixed effects, while the FE estimator, that accounts for this heterogeneity, is close to the GMM estimator, to the Arellano-Bond in particular.

With respect to the estimates of the parameter of interest ϕ_g, we find that different specifications of the DIFF estimator tend to perform generally slightly better than the alternative specifications of the SYS estimator when the pre-sample is shorter, though the variance is smaller for the SYS GMM estimators: the behaviour is reversed with a longer pre-sample, with the SYS GMM estimator still having the lowest variance. The SYS estimator, however, seems to have the best performance overall: this could be due to the fact that, once the initial conditions satisfy the Blundell-Bond assumptions, it is less sensitive to the accumulation of shocks and individual heterogeneity over time that is intrinsic in our simulation process. An evident feature is the under-rejection of H_0 for ϕ_g by the DIFF estimator: this problem becomes more serious as we decrease the number of instruments and we thus drop useful information from our data. In DIFF GMM frameworks, the instruments are weak if the series are persistent and a reduction in their count comes out to be problematic.

With respect to ϕ_c, we have here the classical textbook behaviour of the alternative estimators: the OLS estimator overestimates the coefficient, the DIFF GMM is systematically biased downwards, while the SYS estimator is the best performing, being less biased, in particular when the pre-sample is short.

It is worth stressing the good performance of alternative SYS FACT estimators, especially for the estimate of ϕ_g, and, on the contrary, the very poor behaviour of the DIFF FACT estimators. In a dynamic model with very persistent series, the factorization of the instrument matrix of DIFF GMM further weakens the already weak instruments. When the instruments are valid, as for the SYS GMM estimator, their factorization does not worsen the estimates.

As far as ϕ_d is concerned, in both the scenarios the most evident feature is the very misleading under-rejection of H_0: only in the minority of cases the coefficient is detected as significant, as it actually is. Overall, the adjustment of fiscal policy to the initial public debt is either overestimated or underestimated and it is in general estimated far from the true value of 0.005.

Interestingly, the Hansen test p-values are far from 1[27] only for LIM COL estimates, suggesting a severe and generalized problem of instrument proliferation and invalid overidentifying restrictions. However, this decrease in the Hansen test p-value does not come together with a significant improvement of estimators performance with respect to the UNTR estimates; rather, it comes along with an under-rejection of the null hypotheses for the coefficients.

We can safely use a pre-sample of 15 years, since the estimators performed similarly in the two scenarios.

Sticky policies and very persistent gap: $\phi_c + 1 = 0.8$ and $\alpha = 0.8$

In Table (3.4) we report the results of our baseline scenario with a pre-sample of 15 years. We have very persistent economic cycle and very persistent fiscal policies.

With respect to ϕ_g, the OLS estimator gives not surprisingly the most biased estimates because it neglects the fixed effects, that are by construction very relevant in this framework. The DIFF and SYS GMM UNTR estimators give similar estimates, though the SYS estimator has systematically a lower variance. When the instrument count is reduced, the performance of the DIFF estimators suffers a severe worsening, while the SYS estimators do not change significantly their behaviour. The bad performance of the DIFF estimators with transformed instrument sets is

[27] The implausible p-value of 1 is interpreted as a symptom of a potentially dangerous instrument proliferation problems, as argued in Roodman [2009b]. In the last sections of Chapter 1 we discuss the issue of instrument proliferation in GMM estimation; in Chapter 2 we compare alternative strategies to address it. Please refer to the Chapters 1 and 2 for an extensive discussion of the issue.

generally accompanied by a severe under-rejection of H_0 of $\phi_g = 0$. In parallel, the SYS FACT estimator gives the least biased estimates overall, while the DIFF FACT is the estimator that suffers the most. Applying the PCA on the DIFF GMM set of weak instruments further worsens the weak instrument problem due to very feeble covariances between the instruments. By adding the instruments in first differences, we overcome this weakness issue and we have valid instruments that can be safely reduced in their count: in fact, SYS FACT estimates are better than SYS UNTR estimates.

It is remarkable the fact that the Hansen test systematically gives an implausible p-value of 1 or very close to 1, except for LIM COL estimates: the more reliable p-value, though not really "safe" yet[28], does not come along with a sensible improvement in the estimates, that are worse than the LIM estimates. The collapsing of the instrument matrix does not seem very effective in this context and gives, instead, less robust estimates. In other terms, the number of instruments unavoidably becomes very large in this context, but this does not seem to affect significantly the estimates: the UNTR GMM or, at most, the LIM estimator[29] perform well and a further reduction in the instrument count by collapsing does not seem required or can be misleading at worst.

Considering the estimates for ϕ_c, we find a high degree of heterogeneity in the estimates and features that are typical in GMM estimation of dynamic panel data models: the OLS estimator, that ignores the individual heterogeneity, seriously underestimates the dynamics of the $CAPB$, by severely overestimating the coefficient; the DIFF estimators tend to systematically estimate a too negative coefficient and have very high standard deviations. Things improve a lot when we estimate the

[28]Roodman [2009b] suggests that we should worry every time we get a p-value above 0.25, as higher values are symptoms of a weakening of the tests of overidentifying restrictions due to instrument proliferation.

[29]The lag depth truncation is very common in the empirical literature whenever we have a large number of endogenous regressors. As a curiosity, it is nice to report that the statistical softwares sometime have not enough capacity to perform the system GMM-estimation on the full set of instruments: in such a circumstance, the limitation of the instruments to the first lags is an inescapable choice.

coefficients by SYS GMM, as this estimator is both less biased and more efficient. The SYS LIM and the SYS FACT estimates are the closest to the true value of ϕ_c. The collapsing of the instrument matrix comes along with more biased estimates and, when associated also to the limitation of the lags, with a marked under-rejection of the null hypothesis for ϕ_c. The factorization of the instrument set in DIFF GMM appears to be very dangerous because of the worsening of weak instrument problems.

It is evident and not particularly surprising that the tendencies already underlined for the output gap coefficient are exacerbate when we consider the coefficient of the lagged dependent variable, that is certainly the most problematic in this framework.

In a context characterized by stickiness of the fiscal policies and a high persistence of the output gap, it is relatively safe to estimate the adjustment effects by SYS UNTR or, at most, by SYS LIM GMM estimators: we should not worry too much about the instrument proliferation problem in fiscal rule estimation as the simulations do not show a true risk of overfitting of the endogenous variables, despite the controversial results of the Hansen test.

With respect to ϕ_d, it is confirmed that there is a general tendency to under-reject $H_0 : \phi_d = 0$ and that the estimates are often not sufficiently close to the true value, except for the SYS estimates, which are also those with the lowest variance.

Less persistent policies and gap: $\phi_c + 1 = 0.1$ and $\alpha = 0.1$

In Table (3.5) we consider a less realistic scenario were $\alpha = 0.1$, so that the economic cycle is very fast, and $\phi_c + 1 = 0.1$ such that the fiscal policies present almost no inertia at all. The aim of the simulation with this particular setting of the parameters is to check whether the previous findings are robust to different degrees of persistence of the series. The true values of ϕ_g and ϕ_c are respectively 0.10 and -0.90.

As far as ϕ_g is concerned, the behaviour of the alternative estimators is in line with that found in the baseline scenario, with SYS UNTR and the SYS FACT being the best performing estimators and the ones with the lowest variance. Overall, a slight difference is that the estimates for the output gap generally tend to be closer to the true value of ϕ_g: since here the output gap is set as not persistent at all, the weak instrument problem is mitigated a lot and the DIFF GMM estimates are less affected than in the scenario previously considered. Also the tendency to under-reject the null hypothesis on ϕ_g is alleviated and the COLL estimates, even in the case of the DIFF estimators, are closer to the true value.

In a less persistent scenario, the estimates are more robust to the specifications of the GMM estimators and they are not worsened by the adoption of instrument reduction techniques. It is remarkable however the bad performance of the DIFF FACT estimator in this case: if the series are close to being white noise, the lagged levels of the endogenous regressors have very weak covariances between them and are not suitable to be factorized.

In general, if the series were not persistent, every choice about the specification of the GMM estimators would be relatively safe and would give reliable estimates. When the series are very persistent, the issue of instrument weakness seem to overwhelm and dominate that of instrument proliferation.

If the instruments are weak, a reduction in their number through the collapsing of the instrument matrix can bring the estimates far from the true values of the coefficients. On the other hand, when the instruments are stronger, the reduction of the instrument count has a very little impact on the estimates.

The estimates of ϕ_c follows the same regularities as above: the OLS estimator overestimates ϕ_c and has a very high variance in the experiment; the DIFF estimators overestimate the effect of the lagged $CAPB$ by giving a too large coefficient. When the $CAPB$ is less persistent, ϕ_c is always detected as statistically significant.

In this last scenario, the estimation of ϕ_d comes out to be the most problematic, as the parameter of debt is very often stated not significant, especially when the SYS GMM estimator is used; the DIFF GMM tends to find the debt statistically significant many more times, but this is due to the fact that it generally overestimates the adjustment effects of the fiscal policies to the initial stock of debt.

What it is reassuring from the analysis of the results through Table (3.2) to Table (3.5) is that the sign of ϕ_g is always estimated correctly in GMM estimation; on the other hand, what can be worrying is the fact that, even in an almost perfectly controlled experiment, the estimates of ϕ_g too often point to an a-cyclicality of discretionary fiscal policies.

Alternative cyclicality of the policy stance: $\phi_g = -0.10$ **and** $\phi_g = 0$

As a further robustness check, in Tables (3.6) and (3.7) we consider again the scenario characterized by the high persistence of both the gap and the $CAPB$, but we set respectively $\phi_g = 0$ and $\phi_g = -0.10$, so that discretionary fiscal policies are assumed to be a-cyclical in the first case and pro-cyclical in the second one.

With respect to ϕ_c and ϕ_d we find that in both cases there are not relevant differences with respect to the results presented above. The same is also true for the Hansen test p-value that is still in line with the values found before.

What we are most interested in here is whether and how there are changes in the estimates of ϕ_g.

In the scenario in which $\phi_g = 0$, we can check for the size of alternative estimators through the frequency of rejection of the null hypothesis. We find here an over-rejection of H_0, as the size of the test is above the 5%. In particular, the estimates tend too often to a pro-cyclicality or to a counter-cyclicality of the fiscal policies when the truth is the a-cyclicality. The SYS COL estimator tends to exacerbate this problem. However, this tendency is not as marked as the one, detected above, to find a-cyclicality when the policies are actually not neutral to the economic cycle.

When instead $\phi_g = -0.10$, the estimates are generally in line with what happens if $\phi_g = 0.10$, though they are slightly less biased and the tendency to under-reject H_0 is less severe. In particular, the OLS estimator is much closer to the true parameter and the DIFF estimates are more similar to the SYS ones. It is confirmed that the SYS FACT estimator performs impressively well in a context where fiscal policies are very sticky.

The adoption of instrument reduction techniques does not necessarily improve the estimate, though it lowers the Hansen-test p-value.

Overall, the main lesson we draw from these Monte Carlo experiments is that even when the best performing estimator, the SYS GMM in all its variants, is used, we are not totally safe from detecting a misleading reaction of discretionary fiscal policies to the economic cycle. In fact, the perfect control on both the true parameters of the model and the generating processes of the variables of interest is not sufficient to always estimate correctly the adjustment effects of the fiscal policies in a context where strong persistence of the series and a severe endogeneity of the economic variables are serious issues. We should be aware that the risk of drawing incorrect conclusions on the effects of interest can be even more serious in empirical analysis on real data as we do not know the real data generating processes. Cautiousness remains the best strategy to adopt.

Table 3.2: Estimates of the CAPB model on simulated data (1 year pre-sample): $\alpha = 0.8$, $\phi_c + 1 = 0.8$, $\phi_g = 0.10$.

Dependent variable: $\Delta CAPB$

Variable		OLS	FE	AH	ABa	BBa	ABl	BBl	ABc	BBc	ABlc	BBlc	ABpcaa	ABpcav	BBpcaa	BBpcav
L.GAP	*mean*	0.0811	0.1023	0.1057	0.1011	0.1084	0.0960	0.1080	0.0971	0.1144	0.0977	0.1108	0.0794	0.0739	0.0996	0.0982
	sd	0.0141	0.0193	0.3337	0.0209	0.0165	0.0370	0.0188	0.0394	0.0209	0.0666	0.0232	0.0464	0.0791	0.0229	0.0271
	sig	1	0.997	0.378	0.995	1	0.797	1	0.745	1	0.461	0.997	0.477	0.242	0.993	0.952
L.CAPB	*mean*	-0.1165	-0.2554	-0.2092	-0.2585	-0.2041	-0.2479	-0.1893	-0.2300	-0.2088	-0.2089	-0.1956	-0.2872	-0.3887	-0.1917	-0.1898
	sd	0.0343	0.0404	0.2010	0.0426	0.0392	0.0530	0.0434	0.0513	0.0472	0.0602	0.0551	0.0859	0.1670	0.0484	0.0632
	sig	0.975	1	0.818	1	1	1	0.997	0.997	0.996	0.907	0.95	0.975	0.801	0.984	0.88
L.DEBT	*mean*	0.0054	0.0047	0.0045	0.0045	0.0039	0.0051	0.0045	0.0046	0.0032	0.0050	0.0034	0.0078	0.0102	0.0046	0.0045
	sd	0.0023	0.0028	0.0283	0.0029	0.0026	0.0034	0.0028	0.0035	0.0031	0.0053	0.0038	0.0059	0.0112	0.0030	0.0037
	sig	0.633	0.492	0.183	0.476	0.454	0.426	0.497	0.352	0.279	0.229	0.235	0.268	0.131	0.44	0.332
hansenp					1.0000	1.0000	1.0000	1.0000	0.9998	1.0000	0.4698	0.4425	1.0000	0.9228	1.0000	1.0000

Notes:

a. We run 1000 iterations in the Monte Carlo experiment.

b. We report the average point estimates for the coefficients (mean), the average estimated standard deviation (sd) and the power of the t-test (sig) for the significance of the coefficients.

c. Details on the abbreviations for the estimators are in section 3.2.1.

c. hansenp is the average p-value for the Hansen overidentifying restriction test (robust but weakened by many instruments; see Roodman [2009b].)

Table 3.3: Estimates of the CAPB model on simulated data (35 years pre-sample): $\alpha = 0.8$, $\phi_c + 1 = 0.8$, $\phi_g = 0.10$

Dependent variable: $\Delta CAPB$

Variable		OLS	FE	AH	ABa	BBa	ABl	BBl	ABc	BBc	ABlc	BBlc	ABpcaa	ABpcav	BBpcaa	BBpcav
L.GAP	*mean*	0.0679	0.1069	0.1400	0.1043	0.1029	0.0803	0.1006	0.0818	0.1063	0.0731	0.1029	0.0826	0.0918	0.0962	0.0940
	sd	0.0162	0.0182	1.4127	0.0198	0.0172	0.0400	0.0201	0.0431	0.0225	0.1496	0.0250	0.0478	0.0679	0.0213	0.0252
	sig	0.999	0.999	0.112	0.996	1	0.636	0.995	0.56	0.992	0.14	0.975	0.551	0.34	0.994	0.944
L.CAPB	*mean*	-0.1340	-0.2928	-0.1459	-0.3028	-0.2327	-0.3397	-0.2140	-0.2879	-0.2268	-0.2500	-0.2023	-0.3296	-0.4920	-0.2101	-0.2068
	sd	0.0369	0.0476	1.9612	0.0511	0.0440	0.0892	0.0510	0.0853	0.0590	0.2156	0.0697	0.1043	0.1846	0.0570	0.0668
	sig	0.997	1	0.215	1	1	0.999	0.997	0.962	0.975	0.368	0.848	0.941	0.816	0.973	0.895
L.DEBT	*mean*	0.0029	0.0084	-0.0014	0.0088	0.0041	0.0124	0.0041	0.0096	0.0041	0.0091	0.0043	0.0114	0.0158	0.0039	0.0038
	sd	0.0008	0.0041	0.2194	0.0044	0.0014	0.0073	0.0015	0.0075	0.0023	0.0234	0.0026	0.0104	0.0186	0.0015	0.0017
	sig	0.974	0.685	0.036	0.62	0.871	0.539	0.849	0.291	0.584	0.051	0.527	0.216	0.108	0.794	0.684
hansenp					1.0000	1.0000	1.0000	1.0000	0.9999	1.0000	0.5125	0.4721	1.0000	0.9450	1.0000	1.0000

Notes:

a. We run 1000 iterations in the Monte Carlo experiment.

b. We report the average point estimates for the coefficients (mean), the average estimated standard deviation (sd) and the power of the t-test (sig) for the significance of the coefficients.

c. Details on the abbreviations for the estimators are in section 3.2.1.

c. hansenp is the average p-value for the Hansen overidentifying restriction test (robust but weakened by many instruments; see Roodman [2009b].)

Table 3.4: Estimates of the CAPB model on simulated data (15 years pre-sample): $\alpha = 0.8$, $\phi_c + 1 = 0.8$, $\phi_g = 0.10$

Dependent variable: $\Delta CAPB$

Variable		OLS	FE	AH	ABa	BBa	ABl	BBl	ABc	BBc	ABlc	BBlc	ABpcaa	ABpcav	BBpcaa	BBpcav
L.GAP	*mean*	0.0686	0.1062	0.0699	0.1033	0.1052	0.0811	0.1026	0.0845	0.1085	0.0843	0.1043	0.0798	0.0882	0.0980	0.0962
	sd	0.0168	0.0189	1.1978	0.0206	0.0180	0.0383	0.0208	0.0405	0.0230	0.1216	0.0259	0.0496	0.0736	0.0224	0.0262
	sig	0.998	0.999	0.23	0.995	0.999	0.649	0.996	0.632	0.991	0.288	0.978	0.469	0.289	0.985	0.941
L.CAPB	*mean*	-0.1266	-0.2960	-0.2540	-0.3073	-0.2246	-0.3459	-0.2041	-0.2908	-0.2150	-0.2378	-0.1967	-0.3572	-0.5331	-0.2051	-0.1995
	sd	0.0372	0.0468	1.9647	0.0506	0.0425	0.0879	0.0495	0.0834	0.0553	0.2057	0.0658	0.1152	0.2058	0.0572	0.0690
	sig	0.98	1	0.288	1	1	0.998	0.994	0.968	0.982	0.456	0.867	0.953	0.82	0.958	0.844
L.DEBT	*mean*	0.0030	0.0079	0.0073	0.0083	0.0039	0.0104	0.0040	0.0082	0.0041	0.0066	0.0041	0.0111	0.0168	0.0039	0.0039
	sd	0.0011	0.0030	0.1325	0.0032	0.0017	0.0048	0.0017	0.0047	0.0019	0.0130	0.0022	0.0085	0.0160	0.0020	0.0022
	sig	0.85	0.852	0.07	0.805	0.715	0.704	0.72	0.463	0.65	0.11	0.556	0.276	0.128	0.596	0.496
hansenp					1.0000	1.0000	1.0000	1.0000	0.9998	1.0000	0.5045	0.4635	0.9999	0.9178	1.0000	1.0000

Notes:

a. We run 1000 iterations in the Monte Carlo experiment.

b. We report the average point estimates for the coefficients (mean), the average estimated standard deviation (sd) and the power of the t-test (sig) for the significance of the coefficients.

c. Details on the abbreviations for the estimators are in section 3.2.1.

c. hansenp is the average p-value for the Hansen overidentifying restriction test (robust but weakened by many instruments; see Roodman [2009b].)

Table 3.5: Estimates of the CAPB model on simulated data (15 years pre-sample): $\alpha = 0.1$, $\phi_c + 1 = 0.1$, $\phi_g = 0.10$

Dependent variable: $\Delta CAPB$

Variable		OLS	FE	AH	ABa	BBa	ABl	BBl	ABc	BBc	ABlc	BBlc	ABpcaa	ABpcav	BBpcaa	BBpcav
L.GAP	*mean*	0.0911	0.0941	0.1003	0.0931	0.1023	0.0888	0.1036	0.0943	0.1049	0.0980	0.1059	0.0699	0.0685	0.1032	0.1026
	sd	0.0221	0.0227	0.0407	0.0234	0.0225	0.0298	0.0243	0.0287	0.0245	0.0355	0.0260	0.0410	0.0464	0.0259	0.0277
	sig	0.984	0.983	0.716	0.976	0.992	0.867	0.976	0.898	0.982	0.8	0.968	0.486	0.407	0.974	0.945
L.CAPB	*mean*	-0.7818	-0.9698	-0.8953	-0.9750	-0.9187	-1.0000	-0.9182	-0.9472	-0.9110	-0.9069	-0.8920	-1.0309	-1.0462	-0.9195	-0.9269
	sd	0.0815	0.0649	0.1052	0.0664	0.0703	0.0777	0.0777	0.0777	0.0787	0.0905	0.0861	0.1061	0.1180	0.0839	0.0904
	sig	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
L.DEBT	*mean*	-0.0023	0.0126	0.0038	0.0128	0.0017	0.0136	0.0002	0.0087	0.0017	0.0056	-0.0001	0.0336	0.0347	-0.0004	-0.0002
	sd	0.0049	0.0091	0.0223	0.0098	0.0067	0.0136	0.0070	0.0128	0.0088	0.0190	0.0104	0.0299	0.0352	0.0087	0.0096
	sig	0.237	0.341	0.059	0.314	0.094	0.211	0.074	0.134	0.083	0.076	0.076	0.208	0.179	0.085	0.07
hansenp					1.0000	1.0000	1.0000	1.0000	0.9998	1.0000	0.4756	0.4614	1.0000	1.0000	1.0000	1.0000

Notes:

a. We run 1000 iterations in the Monte Carlo experiment.

b. We report the average point estimates for the coefficients (mean), the average estimated standard deviation (sd) and the power of the t-test (sig) for the significance of the coefficients.

c. Details on the abbreviations for the estimators are in section 3.2.1.

c. hansenp is the average p-value for the Hansen overidentifying restriction test (robust but weakened by many instruments; see Roodman [2009b].)

Table 3.6: Estimates of the CAPB model on simulated data (15 years pre-sample): $\alpha = 0.8$, $\phi_c + 1 = 0.8$, $\phi_g = 0$

Dependent variable: $\Delta CAPB$

Variable		OLS	FE	AH	ABa	BBa	ABl	BBl	ABc	BBc	ABlc	BBlc	ABpcaa	ABpcav	BBpcaa	BBpcav
L.GAP	*mean*	-0.0134	0.0005	-0.0043	-0.0005	0.0049	-0.0079	0.0034	-0.0047	0.0161	-0.0062	0.0084	-0.0180	-0.0147	-0.0026	-0.0037
	sd	0.0120	0.0197	0.0960	0.0216	0.0172	0.0387	0.0194	0.0381	0.0244	0.0609	0.0286	0.0495	0.0749	0.0215	0.0247
	sig	0.2	0.092	0.067	0.101	0.088	0.109	0.08	0.08	0.117	0.043	0.078	0.068	0.043	0.076	0.059
L.CAPB	*mean*	-0.1122	-0.3645	-0.2109	-0.3867	-0.2268	-0.5051	-0.1943	-0.3926	-0.2292	-0.2734	-0.1851	-0.5117	-0.6690	-0.1919	-0.1831
	sd	0.0337	0.0595	0.3071	0.0637	0.0537	0.1357	0.0633	0.1273	0.0852	0.2227	0.0984	0.1467	0.2208	0.0689	0.0809
	sig	0.977	1	0.214	1	0.999	0.999	0.95	0.945	0.875	0.343	0.543	0.955	0.843	0.838	0.652
L.DEBT	*mean*	0.0008	0.0107	0.0053	0.0113	0.0033	0.0155	0.0030	0.0112	0.0038	0.0077	0.0030	0.0220	0.0275	0.0027	0.0025
	sd	0.0012	0.0039	0.0142	0.0042	0.0025	0.0068	0.0025	0.0062	0.0031	0.0105	0.0035	0.0158	0.0304	0.0028	0.0032
	sig	0.0860	0.8950	0.1000	0.8150	0.3650	0.7430	0.2870	0.5060	0.2770	0.1380	0.1460	0.2740	0.0920	0.1790	0.1320
hansenp					1.0000	1.0000	1.0000	1.0000	0.9998	1.0000	0.4899	0.4612	0.9999	0.9096	1.0000	1.0000

Notes:

a. We run 1000 iterations in the Monte Carlo experiment.

b. We report the average point estimates for the coefficients (mean), the average estimated standard deviation (sd) and the size of the t-test (sig) for the significance of the coefficients.

c. Details on the abbreviations for the estimators are in section 3.2.1.

c. hansenp is the average p-value for the Hansen overidentifying restriction test (robust but weakened by many instruments; see Roodman [2009b].)

Table 3.7: Estimates of the CAPB model on simulated data (15 years pre-sample): $\alpha = 0.8$, $\phi_c + 1 = 0.8$, $\phi_g = -0.10$

Dependent variable: $\Delta CAPB$

Variable		OLS	FE	AH	ABa	BBa	ABl	BBl	ABc	BBc	ABlc	BBlc	ABpcaa	ABpcav	BBpcaa	BBpcav
L.GAP	*mean*	-0.0932	-0.1081	-0.1051	-0.1065	-0.0985	-0.0908	-0.0971	-0.0891	-0.0931	-0.0921	-0.0970	-0.0922	-0.0940	-0.1002	-0.1003
	sd	0.0144	0.0191	0.1967	0.0207	0.0185	0.0369	0.0209	0.0372	0.0274	0.0736	0.0312	0.0447	0.0677	0.0212	0.0247
	sig	1	0.999	0.415	0.997	0.999	0.776	0.998	0.73	0.93	0.474	0.874	0.427	0.183	0.998	0.978
L.CAPB	*mean*	-0.1620	-0.3082	-0.1920	-0.3238	-0.1983	-0.3854	-0.1820	-0.3146	-0.1778	-0.2402	-0.1613	-0.4781	-0.5937	-0.1942	-0.1931
	sd	0.0270	0.0490	0.5191	0.0534	0.0384	0.1062	0.0420	0.0951	0.0540	0.1954	0.0581	0.1374	0.2120	0.0482	0.0594
	sig	1	1	0.328	1	1	0.998	1	0.955	0.953	0.439	0.861	0.943	0.683	0.989	0.932
L.DEBT	*mean*	-0.0013	0.0106	0.0043	0.0109	0.0022	0.0129	0.0019	0.0095	0.0022	0.0066	0.0012	0.0403	0.0732	0.0016	0.0015
	sd	0.0016	0.0041	0.0271	0.0044	0.0029	0.0067	0.0028	0.0059	0.0032	0.0101	0.0039	0.0228	0.0603	0.0034	0.0038
	sig	0.186	0.85	0.105	0.753	0.179	0.621	0.138	0.416	0.13	0.125	0.074	0.215	0.097	0.105	0.086
hansenp					1.0000	1.0000	1.0000	1.0000	0.9998	1.0000	0.4941	0.4455	0.9999	0.9110	1.0000	1.0000

Notes:

a. We run 1000 iterations in the Monte Carlo experiment.

b. We report the average point estimates for the coefficients (mean), the average estimated standard deviation (sd) and the power of the t-test (sig) for the significance of the coefficients.

c. Details on the abbreviations for the estimators are in section 3.2.1.

c. hansenp is the average p-value for the Hansen overidentifying restriction test (robust but weakened by many instruments; see Roodman [2009b].)

3.4 Estimation of the PB model

So far, we have focused on the CAPB model that aims at estimating the cyclical reaction of discretionary fiscal policies.

Another approach commonly adopted in the literature is the estimation of an alternative fiscal rule that focuses on the overall budgetary reactions. The **PB model** aims at estimating directly the overall reaction of fiscal policies to the economic cycle, and not only the discretionary responses as the $CAPB$ model does. In the PB model specification, the dependent variable, that represents the decision of the fiscal policy maker, is the change in the unadjusted overall primary balance, while the explanatory variables are given by the lagged primary balance, the lagged debt and the lagged output gap. The PB model is therefore the following:

$$\Delta PB_{it} = \phi_p^{PB} PB_{i,t-1} + \phi_d^{PB} DEBT_{i,t-1} + \phi_g^{PB} GAP_{i,t-1} + u_{it} \qquad (3.4.12)$$

where u_{it} is an error terms that collects fixed effects, fiscal policy shocks and idiosyncratic errors.

The primary balance is, by definition:

$$PB_{it} \equiv CAPB_{it} + \omega_{it} GAP_{it}, \qquad (3.4.13)$$

that is the combination of the cyclically-adjusted primary balance, that captures the structural or discretionary budget, and a cyclical component given by the product of the output gap and the effect of automatic stabilizers. It is therefore straightforward that the change in the primary balance measures the full fiscal action.

Even in the case of the PB model, the main parameter of interest is ϕ_g^{PB}, as it gives indications on whether fiscal policies are pro-cyclical, if the coefficient has a negative sign, or counter-cyclical, in case of positive coefficient.

As the PB model only allows to estimates gross effects, it quite common in the fiscal policy literature to estimate the reaction of discretionary policies to the cycle in the following way: we subtract the coefficient ω[30],

[30] ω has generally a value around 0.5.

that is the average of ω_{it}, from the estimated coefficient of the output gap in the PB model (ϕ_g^{PB})[31]. The cyclicality of discretionary policies is thus estimated as:

$$\phi_g^{PB(discr)} \approx \phi_g^{PB} - \omega. \quad (3.4.14)$$

We estimate the PB model as in equation (3.4.12) on our simulated data generated according to the processes presented above. We again run 1000 iterations, we use the same estimators used for the $CAPB$ and consider all the regressors as endogenous so that we instrument them all. We keep the baseline settings used for the $CAPB$ model. In Tables (3.8) and (3.9) we report the estimates for the scenario in which both the gap and the cyclically-adjusted primary balance are very persistent and in which the two processes are almost not persistent at all. We know that the true coefficient for the output gap in the $CAPB$ is 0.10.

In the tables we also report the estimates of $\phi_g^{PB(discr)}$ that are implied by equation (3.4.14); we use $\omega = 0.5$ in all our experiments.

For the convenience of the reader, in the last rows of Tables (3.8) and (3.9) we report again from Tables (3.4) and (3.5) the estimates of ϕ_g^{CAPB} obtained in exactly the same experiment setting.

What we can immediately notice is that both in the more persistent and in the less persistent scenarios, the estimates of ϕ_g^{PB} are not very different from those obtained for ϕ_g^{CAPB}: when we use the alternative specifications of the estimator that has been shown above to be the safest in this context, the SYS GMM, we get estimates for ϕ_g^{PB} that are in the range $0.06 - 0.09$, not very far from those in the range $0.095 - 0.11$ obtained by estimating the CAPB model. In other words, the estimated ϕ_g^{PB} is far from being higher of around 0.5 than the estimated ϕ_g^{CAPB}. This is evident when we consider the estimates of $\phi_g^{PB(discr)}$: if the subtraction of ω from the estimates of ϕ_g^{PB} were a correct strategy in order to estimate the discretionary reaction in the PB model, we would expect the estimated $\phi_g^{PB(discr)}$ to have an average close to $\phi_g^{CAPB} = 0.10$. This is obviously not the case as we systematically get a very large negative es-

[31] See, among the others, Bouthevillain et al. [2001], Balassone and Franzese [2004] and Socol and Socol [2009].

timate of $\phi_g^{PB(discr)}$: this would point to a very strong pro-cyclicality of discretionary fiscal policies in a context in which we have set the policies as counter-cyclical. What is interesting is that these evidences are generally confirmed also in the more unrealistic less persistent framework.

We have argued that the most realistic profile is the one where the gap and the deficit are strongly persistent: we find here that, in this context, it can be very misleading to estimate the discretionary effects according to equation (3.4.14) as we could draw wrong conclusions about the cyclicality of the policies, no really matters which estimator we are using.

In order to estimate the reaction of discretionary policies to the economic cycle, the best strategy remain the estimation of ϕ_g^{CAPB}.

At this stage, we are not able to say anything very precise on the estimates of the PB model as we do not know the values of the coefficients in the model[32]. What we can only argue is that the estimates of the PB model come out to be very close to the those of the $CAPB$ model, but we do not know more on the relationship between the parameters in the two models, and that the estimate of the discretionary response in the PB model does not come as straightforwardly as it seemed from the estimated ϕ_g^{PB}. To say something more on the estimates of the PB model in comparison to those of the $CAPB$ model, it is worth investigating further the algebra of the relationships between the two models.

3.4.1 Algebraic links between the parameters of the PB model and the CAPB model

Aware of the evidences found in our preliminary analysis, we aim here at discovering the reasons why the estimates of ϕ_g^{CAPB} and of ϕ_g^{PB} do not differ as much as we would expect.

We recall here that the $CAPB$ model is:

$$\Delta CAPB_{it} = \phi_c^{CAPB} CAPB_{i,t-1} + \phi_d^{CAPB} DEBT_{i,t-1} + \phi_g^{CAPB} GAP_{i,t-1} + \mu_i + \varepsilon_{it}$$

[32]We recall here that we fix the setting for the $CAPB$ model and we then estimate both the $CAPB$ and the PB model on our simulated data.

and that the PB model is:

$$\Delta PB_{it} = \phi_p^{PB} PB_{i,t-1} + \phi_d^{PB} DEBT_{i,t-1} + \phi_g^{PB} GAP_{i,t-1} + u_{it}.$$

By exploiting the equivalence deriving from the consolidated accountancy rule $PB_{it} \equiv CAPB_{it} + \omega GAP_{it}$, we can derive the exact relationships between the coefficients in the two models.

By substituting the expression for PB_{it} into the PB model we have, through some algebraic manipulations:

$$\begin{aligned} \Delta CAPB_{it} = & \phi_{pb}^{PB} CAPB_{i,t-1} + \phi_d^{PB} DEBT_{i,t-1} + \\ & \phi_{pb}^{PB} \omega GAP_{i,t-1} + \phi_g^{PB} GAP_{i,t-1} - \\ & \omega GAP_{it} + \omega GAP_{i,t-1} + v_{it} \end{aligned} \tag{3.4.15}$$

It is already straightforward that the coefficient of $CAPB_{i,t-1}$ in the $CAPB$ model is the same as the coefficient of $PB_{i,t-1}$ in the PB model and that the coefficient of the lagged debt is the same in the two models. We thus have:

$$\phi_{pb}^{PB} = \phi_c^{CAPB} \tag{3.4.16}$$

$$\phi_d^{PB} = \phi_d^{CAPB}. \tag{3.4.17}$$

In order to derive an analogous link for the gap coefficient, we need to substitute $\alpha GAP_{i,t-1} + \eta_i + \xi \varepsilon_{i,t-1} + v_{it}$ as in equation (3.2.5) for GAP_{it}. We get:

$$\begin{aligned} \Delta CAPB_{it} = & \phi_{pb}^{PB} CAPB_{i,t-1} + \phi_d^{PB} DEBT_{i,t-1} + \\ & + (\omega \phi_{pb}^{PB} + \phi_g^{PB} + \omega - \alpha\omega) GAP_{i,t-1} && (3.4.18) \\ & - \omega(\eta_i + \xi \varepsilon_{i,t-1} + v_{it}) + u_{it}. && (3.4.19) \end{aligned}$$

The relationship between the gap parameters in the two models is therefore the following:

$$\phi_g^{CAPB} = \omega \phi_{pb}^{PB} + \phi_g^{PB} + \omega(1 - \alpha). \tag{3.4.20}$$

Note also that both ϕ_{pb}^{PB} and α play a role in this relationship and are crucial in order to determine how close the estimates in the two models are.

We have just seen that $\phi_{pb}^{PB} = \phi_{c}^{CAPB}$ so that, if we set for example $\omega = 0.5$ and $\phi_{c}^{CAPB} + 1 = 0.8$, we will have $\phi_{pb}^{PB} = \phi_{c}^{CAPB} = -0.2$ and thus $\omega\phi_{pb}^{PB} = -0.1$. When the autoregressive coefficient of the output gap is set to be $\alpha = 0.8$, we have that $\omega\phi_{pb}^{PB} = \omega(1-\alpha) = 0.1$: in this case, we have that the output gap parameters in the two models are the same, i.e. $\phi_{g}^{CAPB} = \phi_{g}^{PB}$ as the first and the last terms cancel out. Every time we have similar dynamics for the primary balance and for the output gap, the two terms will tend to annul each other and we expect to obtain estimates of the coefficients for the gap in the two models very close each other.

In order for the relationship $\phi_{g}^{PB(discr)} \approx \phi_{g}^{PB} - \omega$ to be reasonably applicable, the dynamics of the two processes should be approximately the same in magnitude but with opposite signs[33]. It is evident that such a scenario is not realistic. Once again, we have the confirmation that it is not safe at all to estimate the discretionary reaction of the policies by applying the proxy (3.4.14).

In the light of what we have said so far, we can now go back to the estimation results for the PB model in Tables (3.8) and (3.9). In the tables we also report the implied ϕ_{g}^{CAPB} as deriving from equation (3.4.20)[34].

We can confirm here that, imposing the same dynamics for the primary balance and for the output gap, we obtain estimates of the gap coefficient in the PB model close to those in the $CAPB$ model, though they are not as close as we would expect according to equation (3.4.20). The estimates of the gap are more heterogeneous than the respective ones in the CAPB model and present a much higher variance.

[33]The relationship would hold, for example, if the primary balance were a static process and the output gap were a unit root process or if $\phi_{g}^{CAPB} + 1 = x$ and $\alpha = -x$.

[34]It is worth saying that here we generate the PB series according to equation (3.2.7), that is we scale both $CAPB$ and GAP by the term $1 + GAP$, so that the relationship in equation (3.4.13) does not perfectly hold: the denominator, however, is very close to 1, as the gap moves within the range $-5/5\%$, so we should not expect very significant deviations in the estimates.

When compared to the estimates for ϕ_c^{CAPB} in Tables (3.3) and (3.4), the estimates for ϕ_{pb}^{PB} are much more volatile than those in the $CAPB$ and are generally pretty far from the value of -0.2 we would expect according to equation (3.4.17). In the most persistent scenario, the behaviour of the estimators reflects the one we have underlined in the $CAPB$ estimation: the OLS estimator largely underestimate the effect by overestimating the coefficient, the DIFF GMM estimators gives too large coefficients, while the SYS GMM estimators are the safest and the ones that give the estimates closest to the true value. The SYS UNTR and the SYS FACT give estimates close to -0.2 and with the smallest variance overall.

In the case in which $\alpha = 0.1$ and $\phi_c^{CAPB} + 1 = 0.1$ we have less heterogeneity in the estimates and an overall slightly better performance of the DIFF estimators over the SYS ones: in a less persistent framework, the weak instrument problem is mitigated, if not inexistent, and parsimony in the number of instruments is beneficial.

Overall we find that the estimates of the coefficients in the PB model are much more biased and have a much larger variance than those for the $CAPB$ model: this evidence has a bad impact on the estimated of the implied ϕ_g^{CAPB} coefficient derived from the PB estimates. As both the estimates of ϕ_{pb}^{PB} and ϕ_g^{PB} enter equation (3.4.17), if the are biased and imprecise, the estimate of ϕ_g^{CAPB} will also be likely to be biased. This is exactly what happens in our experiment. If the estimates of ϕ_{pb}^{PB} and ϕ_g^{PB} were better, we would expect that the relationship $\phi_g^{PB} = \phi_g^{CAPB}$ would approximately hold, as the two terms $\omega\phi_{pb}^{PB}$ and $\omega(1-\alpha)$ would tend to cancel out. Here, instead, the dynamics of the primary balance and the response to the gap are estimated most imprecisely and the implied ϕ_g^{CAPB} is therefore a very weak, and often misleading, estimate of the discretionary response to the gap.

This larger bias and imprecision of the estimates in the PB model could be explained by the fact that the PB model estimates gross unadjusted effects and introduces much more noise in the variables, with respect to the $CAPB$ model that filters out the cycle effects from the primary balance. We can have an intuition of that by looking at the term $-\omega(\eta_i + \xi\varepsilon_{i,t-1} + v_{it})$ in equation (3.4.19): this noise is part of the error term in the $CAPB$ model, while it is part of the PB variable in the PB model. It is intuitive that overall measures of the deficit are more likely to be endogenous to the shocks in the model than already adjusted measures.

For the sake of completeness, we report in Tables (3.10) and (3.11) the results for the alternative mixed scenarios $\alpha = 0.8; \phi_c + 1 = 0.$ and $\alpha = 0.1; \phi_c + 1 = 0.8$, though they are not very realistic. In this cases we give the $CAPB$ and the GAP different dynamics and we expect to find estimates of ϕ_g^{PB} far from 0.10 as the terms $\omega\phi_{pb}^{PB}$ and $\omega(1-\alpha)$ do not cancel out anymore. This is exactly what happens, as the estimates of the gap coefficient are centered respectively around 0.30 and -0.20: in both cases the estimate of $\phi_g^{(PB)discr}$ is very misleading as it implies a very strong pro-cyclicality of the discretionary policies, especially in the second scenario, when they are instead set as countercyclical. The estimate of the implied ϕ_c^{CAPB} is instead more correct and centered on the true value of ϕ_c^{CAPB}, but only when the SYS GMM estimator is used. Neither in this case, however, the strategy of using equation (3.4.17) in order to estimate the discretionary response to the gap is completely safe, due to the large heterogeneity in the estimates of the coefficients ϕ_g^{PB} and ϕ_{pb}^{PB}.

In conclusion, what emerges from our experiments is that, if we want to assess how discretionary policies react to the cycle, the safest strategy is the estimation of ϕ_g^{CAPB} directly in the $CAPB$ model by SYS GMM estimation. Moreover, it is unreasonable to estimates $\phi_g^{(PB)discr}$ by subtracting ω from the estimated ϕ_g^{PB}: if we can not avoid the estimation of the PB model, it can be advisable to estimate the discretionary response by applying the strategy suggested in this section.

Table 3.8: Estimates of the PB model on simulated data (15 years pre-sample): $\alpha = 0.8$, $\phi_c + 1 = 0.8$, $\phi_g = 0.10$

Dependent variable: ΔPB

Variable		OLS	FE	AH	ABa	BBa	ABl	BBl	ABc	BBc	ABlc	BBlc	ABpcaa	ABpcav	BBpcaa	BBpcav
L.GAP	*mean*	0.0538	0.0969	-0.3544	0.0938	0.0764	0.0465	0.0650	0.0642	0.0805	0.0590	0.0831	0.0432	0.0468	0.0759	0.0687
	sd	0.0433	0.0629	14.9057	0.0681	0.0636	0.0999	0.0746	0.0957	0.0831	0.1444	0.0967	0.1380	0.2293	0.0753	0.0896
	sig	0.285	0.408	0.074	0.356	0.32	0.135	0.217	0.157	0.224	0.105	0.199	0.095	0.068	0.235	0.175
$\hat{\phi}_g - \omega$	*mean*	-0.4462	-0.4031	-0.8544	-0.4062	-0.4236	-0.4535	-0.4350	-0.4358	-0.4195	-0.4410	-0.4169	-0.4568	-0.4532	-0.4241	-0.4313
Implied ϕ_g^{CAPB}	*mean*	0.0860	0.0235	-0.6535	0.0082	0.0803	-0.0992	0.0878	-0.0305	0.1018	0.0054	0.1133	-0.1052	-0.1722	0.0841	0.0827
L.PB	*mean*	-0.1356	-0.3470	-0.7982	-0.3712	-0.1921	-0.4913	-0.1543	-0.3894	-0.1574	-0.3071	-0.1394	-0.4969	-0.6379	-0.1836	-0.1720
	sd	0.0577	0.0879	17.4203	0.0978	0.0823	0.1844	0.0945	0.1701	0.1043	0.3862	0.1230	0.2157	0.3604	0.1084	0.1293
	sig	0.662	0.986	0.09	0.973	0.707	0.882	0.446	0.725	0.405	0.182	0.248	0.672	0.457	0.461	0.316
L.DEBT	*mean*	0.0004	0.0147	0.0403	0.0156	0.0012	0.0244	0.0019	0.0167	0.0025	0.0141	0.0034	0.0347	0.0500	0.0023	0.0020
	sd	0.0018	0.0062	1.0702	0.0068	0.0032	0.0120	0.0033	0.0112	0.0036	0.0261	0.0040	0.0184	0.0308	0.0038	0.0042
	sig	0.065	0.829	0.035	0.731	0.147	0.688	0.166	0.406	0.183	0.077	0.176	0.473	0.297	0.142	0.12
hansenp					1.0000	1.0000	1.0000	1.0000	0.9998	1.0000	0.4869	0.4553	0.9999	0.9314	1.0000	1.0000
L.GAP (CAPB)	*mean*	0.0686	0.1062	0.0699	0.1033	0.1052	0.0811	0.1026	0.0845	0.1085	0.0843	0.1043	0.0798	0.0882	0.0980	0.0962
	sd	0.0168	0.0189	1.1978	0.0206	0.0180	0.0383	0.0208	0.0405	0.0230	0.1216	0.0259	0.0496	0.0736	0.0224	0.0262
	sig	0.998	0.999	0.23	0.995	0.999	0.649	0.996	0.632	0.991	0.288	0.978	0.469	0.289	0.985	0.941

Notes:

a. We run 1000 iterations in the Monte Carlo experiment.

b. We report the average point estimates for the coefficients (mean), the average estimated standard deviation (sd) and the power of the t-test (sig) for the significance of the coefficients.

c. Details on the abbreviations for the estimators are in section 3.2.1.

c. $\hat{\phi}_g - \omega$ *is the measure of the cyclical response of discretionary policies according to equation 3.4.14*

d. Implied ϕ_g^{CAPB} *is the gap coefficient in the CAPB model computed using the estimates of the PB model according to equation 3.4.20*

e. hansenp is the average p-value for the Hansen overidentifying restriction test (robust but weakened by many instruments; see Roodman [2009b].)

f. The last rows of the table reports the estimated of the output gap in the CAPB model with the same setting of the parameters.

Table 3.9: Estimates of the PB model on simulated data (15 years pre-sample): $\alpha = 0.1$, $\phi_c + 1 = 0.1$, $\phi_g = 0.10$

Dependent variable: ΔPB

Variable		OLS	FE	AH	ABa	BBa	ABl	BBl	ABc	BBc	ABlc	BBlc	ABpcaa	ABpcav	BBpcaa	BBpcav
L.GAP	*mean*	0.0376	0.0430	0.0643	0.0457	0.0591	0.0543	0.0642	0.0572	0.0631	0.0594	0.0657	0.0124	0.0030	0.0736	0.0805
	sd	0.0746	0.0800	0.1137	0.0817	0.0838	0.0952	0.0924	0.0913	0.0941	0.1027	0.1025	0.1140	0.1292	0.0946	0.1028
	sig	0.078	0.091	0.108	0.136	0.154	0.155	0.16	0.149	0.142	0.124	0.139	0.077	0.073	0.166	0.1650
$\hat{\phi}_g - \omega$		-0.4624	-0.4570	-0.4357	-0.4543	-0.4409	-0.4457	-0.4358	-0.4428	-0.4369	-0.4406	-0.4343	-0.4876	-0.4970	-0.4264	-0.4195
Implied ϕ_g^{CAPB}		0.1188	0.0398	0.1128	0.0357	0.0881	0.0188	0.0921	0.0632	0.1025	0.0999	0.1174	-0.0352	-0.0476	0.0967	0.0956
L.PB	*mean*	-0.7376	-0.9064	-0.8030	-0.9201	-0.8421	-0.9710	-0.8442	-0.8881	-0.8211	-0.8190	-0.7965	-0.9951	-1.0012	-0.8538	-0.8697
	sd	0.1271	0.1323	0.2110	0.1342	0.1409	0.1600	0.1570	0.1556	0.1600	0.1832	0.1767	0.1797	0.2055	0.1599	0.1740
	sig	1	1	0.951	1	1	1	1	1	1	0.984	0.991	0.9990	0.995	1	0.998
L.DEBT	*mean*	-0.0115	0.0342	0.0011	0.0347	0.0015	0.0347	-0.0037	0.0176	-0.0001	0.0054	-0.0050	0.1105	0.1223	-0.0024	-0.0015
	sd	0.0070	0.0186	0.0431	0.0204	0.0130	0.0291	0.0129	0.0255	0.0172	0.0361	0.0191	0.0554	0.0628	0.0154	0.0176
	sig	0.488	0.58	0.044	0.489	0.108	0.276	0.121	0.125	0.095	0.051	0.097	0.541	0.4960	0.107	0.113
hansenp					1.0000	1.0000	1.0000	1.0000	0.9998	1.0000	0.4778	0.4690	1.0000	1.0000	1.0000	1.0000
L.GAP (CAPB)	*mean*	0.0911	0.0941	0.1003	0.0931	0.1023	0.0888	0.1036	0.0943	0.1049	0.0980	0.1059	0.0699	0.0685	0.1032	0.1026
	sd	0.0221	0.0227	0.0407	0.0234	0.0225	0.0298	0.0243	0.0287	0.0245	0.0355	0.0260	0.0410	0.0464	0.0259	0.0277
	sig	0.984	0.983	0.716	0.976	0.992	0.867	0.976	0.898	0.982	0.8	0.968	0.486	0.407	0.974	0.945

Notes:

a. We run 1000 iterations in the Monte Carlo experiment.

b. We report the average point estimates for the coefficients (mean), the average estimated standard deviation (sd) and the power of the t-test (sig) for the significance of the coefficients.

c. Details on the abbreviations for the estimators are in section 3.2.1.

c. $\hat{\phi}_g - \omega$ *is the measure of the cyclical response of discretionary policies according to equation 3.4.14*

d. Implied ϕ_g^{CAPB} *is the gap coefficient in the CAPB model computed using the estimates of the PB model according to equation 3.4.20*

e. hansenp is the average p-value for the Hansen overidentifying restriction test (robust but weakened by many instruments; see Roodman [2009b].)

f. The last rows of the table reports the estimated of the output gap in the CAPB model with the same setting of the parameters.

Table 3.10: Estimates of the PB model on simulated data (15 years pre-sample): $\alpha = 0.8$, $\phi_c + 1 = 0.1$, $\phi_g = 0.10$

Dependent variable: $\Delta CAPB$

Variable		OLS	FE	AH	ABa	BBa	ABl	BBl	ABc	BBc	ABlc	BBlc	ABpcaa	ABpcav	BBpcaa	BBpcav
L.GAP	*mean*	0.3332	0.3560	0.7235	0.3444	0.3931	0.2233	0.3991	0.2853	0.4091	0.2735	0.4122	0.2546	0.2626	0.3927	0.3924
	sd	0.0746	0.0813	9.6283	0.0843	0.0855	0.1195	0.0963	0.1212	0.0991	0.2984	0.1101	0.1405	0.2332	0.0982	0.1182
	sig	0.9960	0.9890	0.2700	0.9790	0.9950	0.5600	0.9870	0.6990	0.9830	0.3310	0.9630	0.4630	0.2880	0.9810	0.9270
$\hat{\phi}_g - \omega$		-0.1668	-0.1440	0.2235	-0.1556	-0.1069	-0.2767	-0.1009	-0.2147	-0.0909	-0.2265	-0.0878	-0.2454	-0.2374	-0.1073	-0.1076
Implied ϕ_g^{CAPB}		0.0821	0.0027	0.4817	-0.0149	0.0720	-0.1577	0.0840	-0.0667	0.0958	-0.0588	0.1106	-0.1256	-0.1671	0.0845	0.0841
L.PB	*mean*	-0.7024	-0.9066	-0.6836	-0.9187	-0.8421	-0.9622	-0.8302	-0.9039	-0.8268	-0.8645	-0.8032	-0.9603	-1.0593	-0.8165	-0.8167
	sd	0.1264	0.1307	3.5360	0.1338	0.1397	0.1544	0.1574	0.1533	0.1612	0.2106	0.1797	0.2098	0.3354	0.1608	0.1954
	sig	1.0000	1.0000	0.7150	1.0000	1.0000	1.0000	1.0000	1.0000	0.9990	0.9360	0.9910	0.9860	0.8670	0.9980	0.9860
L.DEBT	*mean*	-0.0016	0.0230	-0.0768	0.0246	-0.0006	0.0494	-0.0002	0.0305	0.0005	0.0318	0.0021	0.0414	0.0494	0.0002	-0.0002
	sd	0.0030	0.0105	2.4833	0.0118	0.0051	0.0241	0.0053	0.0226	0.0064	0.0601	0.0074	0.0240	0.0340	0.0054	0.0062
	sig	0.1060	0.7640	0.0060	0.6550	0.1160	0.7090	0.0950	0.3180	0.0870	0.0300	0.0900	0.4650	0.3060	0.1040	0.1040
hansenp					1.0000	1.0000	1.0000	1.0000	0.9998	1.0000	0.4869	0.4553	0.9999	0.9314	1.0000	1.0000

Notes:

a. We run 1000 iterations in the Monte Carlo experiment.

b. We report the average point estimates for the coefficients (mean), the average estimated standard deviation (sd) and the power of the t-test (sig) for the significance of the coefficients.

c. Details on the abbreviations for the estimators are in section 3.2.1.

c. $\hat{\phi}_g - \omega$ *is the measure of the cyclical response of discretionary policies according to equation 3.4.14*

d. Implied ϕ_g^{CAPB} *is the gap coefficient in the CAPB model computed using the estimates of the PB model according to equation 3.4.20*

e. hansenp is the average p-value for the Hansen overidentifying restriction test (robust but weakened by many instruments; see Roodman [2009b].)

f. The last rows of the table reports the estimated of the output gap in the CAPB model with the same setting of the parameters.

Table 3.11: Estimates of the PB model on simulated data (15 years pre-sample): $\alpha = 0.1$, $\phi_c + 1 = 0.8$, $\phi_g = 0.10$

Dependent variable: ΔPB

Variable		OLS	FE	AH	ABa	BBa	ABl	BBl	ABc	BBc	ABlc	BBlc	ABpcaa	ABpcav	BBpcaa	BBpcav
L.GAP	*mean*	-0.2522	-0.2011	-0.2951	-0.1901	-0.2608	-0.1362	-0.2787	-0.1820	-0.2774	-0.2466	-0.2827	-0.1491	-0.1840	-0.2390	-0.2320
	sd	0.0511	0.0645	0.2203	0.0669	0.0632	0.1031	0.0704	0.0965	0.0721	0.1655	0.0796	0.1057	0.1422	0.0760	0.0835
	sig	0.9990	0.8960	0.4830	0.8280	0.9860	0.3980	0.9780	0.5280	0.9700	0.5020	0.9340	0.2500	0.1510	0.8800	0.8010
$\hat{\phi}_g - \omega$		-0.7522	-0.7011	-0.7951	-0.6901	-0.7608	-0.6362	-0.7787	-0.6820	-0.7774	-0.7466	-0.7827	-0.6491	-0.6840	-0.7390	-0.7320
Implied ϕ_g^{CAPB}		0.1318	0.0600	0.1311	0.0562	0.0982	0.0247	0.1012	0.0673	0.1128	0.1068	0.1237	0.0121	-0.0233	0.1034	0.1028
L.PB	*mean*	-0.1320	-0.3778	-0.0475	-0.4074	-0.1820	-0.5780	-0.1401	-0.4014	-0.1197	-0.1931	-0.0872	-0.5776	-0.5786	-0.2151	-0.2303
	sd	0.0664	0.1016	0.5648	0.1100	0.0870	0.2321	0.0967	0.2122	0.0970	0.4093	0.1079	0.2114	0.2681	0.1225	0.1398
	sig	0.4870	0.9790	0.0700	0.9670	0.6200	0.8440	0.3620	0.5530	0.2740	0.1070	0.1530	0.7090	0.4530	0.4370	0.3890
L.DEBT	*mean*	-0.0021	0.0144	0.0008	0.0149	0.0015	0.0206	0.0013	0.0129	0.0016	0.0058	0.0018	0.0635	0.0985	0.0021	0.0019
	sd	0.0030	0.0073	0.0274	0.0078	0.0048	0.0122	0.0047	0.0107	0.0050	0.0196	0.0061	0.0386	0.0618	0.0062	0.0069
	sig	0.1310	0.6530	0.0590	0.5560	0.1110	0.5040	0.1000	0.2490	0.1100	0.0760	0.0820	0.2030	0.0510	0.0950	0.0850
hansenp					1.0000	1.0000	1.0000	1.0000	0.9998	1.0000	0.4778	0.4690	1.0000	1.0000	1.0000	1.0000

Notes:

a. We run 1000 iterations in the Monte Carlo experiment.

b. We report the average point estimates for the coefficients (mean), the average estimated standard deviation (sd) and the power of the t-test (sig) for the significance of the coefficients.

c. Details on the abbreviations for the estimators are in section 3.2.1.

c. $\hat{\phi}_g - \omega$ *is the measure of the cyclical response of discretionary policies according to equation 3.4.14*

d. $Implied\phi_g^{CAPB}$ *is the gap coefficient in the CAPB model computed using the estimates of the PB model according to equation 3.4.20*

e. hansenp is the average p-value for the Hansen overidentifying restriction test (robust but weakened by many instruments; see Roodman [2009b].)

f. The last rows of the table reports the estimated of the output gap in the CAPB model with the same setting of the parameters.

3.5 Empirical estimation of the CAPB model

After such an extensive Monte Carlo experiment, we have plenty of useful indications about the performance of alternative dynamic panel data estimators and, in the light of the previous findings, it is worth estimating the model of interest on real data in order to compare the performance of the estimators in a real context with their behaviour in a controlled experiment.

We estimate the $CAPB$ model on the dataset used in Golinelli and Momigliano [2009] and we adopt the alternative specifications of the GMM estimator already employed in our simulations.

The sample includes data for 11 Countries of the European Monetary Union over the period 1994-2008. Data for the cyclically adjusted primary balance and the output gap are taken from the OECD Economic Outlook [n. 83, June 2008]; data for the stock of public debt derive from the European Commission AMECO database [June 2008][35].

We choose the specification of the $CAPB$ model as in Table (4) of Golinelli and Momigliano [2009] where, in addition to the lagged dependent variable, debt and gap, we include two additional explanatory variables that are by now standard in empirical analysis of fiscal rules: an election variable ($Elect$)[36], that is a dummy variable equal to 1 in years in which regular elections take place (0 otherwise), and a Maastricht variable ($Maas$), capturing the role of EU rules, that is different from 0 when the deficit is above the 3 per cent target threshold[37].

The model we estimate on real data is therefore:

$$\begin{aligned}\Delta CAPB_{it} =& \phi_c CAPB_{i,t-1} + \phi_d DEBT_{i,t-1} + \phi_g GAP_{i,t-1} \\ &+ \beta Maas_{i,t-1} + \delta Elect_{it} + \mu_i + \tau_t + \varepsilon_{it} \end{aligned} \tag{3.5.21}$$

where μ_i are fixed effects, τ_t are time effects and ε_{it} is the idiosyncratic

[35] Details on the definitions of the variables and on the sources of data are in the Appendix of Golinelli and Momigliano [2009].

[36] Data are from the International Institute for Democracy and Electoral Assistance (IDEA).

[37] The Appendix of Golinelli and Momigliano [2009] provides technical details on the construction of the Maastricht variable.

term.

We estimate the model in equation (3.5.21) by GMM, according to the various specifications presented above, and by OLS, as they are still commonly considered when fiscal rules are to be estimated[38].

All the estimates include time dummies. In line with Golinelli and Momigliano [2009] and with the previous Monte Carlo experiments, we take the lagged dependent variable, the output gap and the Maastricht variable as endogenous: these three variables are therefore instrumented by their own lags in levels in the equation in first-differences and by their lagged first difference in the equation in levels.

With respect to the debt, we adopt two alternative strategies: first, following Golinelli and Momigliano [2006], we consider the debt as an exogenous regressor[39]; second, in line with what done in the simulations, we consider also the debt as endogenous and we instrument it with GMM-style instruments[40].

We keep the same abbreviations as in Section 3.2.1 for the alternative estimators in the tables: the only difference here is that, when we limit the instrument matrix, we consider both the case in which only one lag is kept ($l1$ and $lc1$ stand for the limited and collapsed matrix)[41] and the

[38]Our main interest is in GMM estimation of fiscal response functions with alternative specification of the GMM estimator. We therefore omit here the FE estimates, as they are supplanted by the Arellano-Bond estimator in empirical analysis, and Anderson-Hsiao estimates, as they have proved to be greatly biased also in a controlled experiment. These additional results are available upon request.

[39]In this perspective, the potential endogeneity of the debt in an economic sense does not translate into an endogeneity in a pure statistical sense, as it does instead in the case of the lagged dependent variable; therefore the debt does not necessarily need to be instrumented. This strategy is in line with the common habit of using GMM-style instruments only for the lagged dependent variable when other potential endogenous regressors are not endogenous in a statistical sense. See, e.g., Arellano and Bond [1991].

[40]In our simulations, the debt was endogenous by construction, as generated sequentially in such a way that it incorporated fiscal policy shocks, economic cycle shocks and individual effects, and it necessarily needed to be instrumented. In a real world, we are inclined to believe the debt to be truly correlated with fiscal policy shocks and to the economic cycle, and therefore necessarily correlated with the error term. We thus prefer to consider it endogenous.

[41]This choice depends on comparison reasons since Golinelli and Momigliano [2009] use only the first valid lag for endogenous regressors.

case, as above, in which we keep two lags of each endogenous variable ($l2$ and $lc2$ for the limited and collapsed matrix).

All the estimates are made robust to the potential heteroskedasticity of the residuals; GMM estimates are always one-step.

In each table, we report the estimate of the coefficient, the standard error and the t-statistic[42].

We want first to reproduce the results in column (4) of Table (4) in Golinelli and Momigliano [2009], then we estimate the model by all the alternative estimators. In the original paper, the estimates are not made robust to heteroskedasticity; they are obtained by one-step SYS GMM and use only the first available lag for the endogenous regressors[43]; as hinted above, the authors consider the debt exogenous.

Aiming at exactly reproducing the original results, we should introduce at this stage a pure technical but not irrelevant issue: the updating and the correction of bugs of the commands in the most widely-used econometric softwares. In our estimates, we use the command `xtabond2` for GMM estimation[44] in the latest available version for Stata10. The authors obtained their estimates in Stata9 with the version of the `xtabond2` command available at the time the paper was published.

In Table (3.12), in the first column we replicate the original estimate not robust to heteroskedasticity; in the second column we make it robust using the version of `xtabond2` exploited by the authors; in the third and the fourth columns we replicate the estimates of the first and the second ones by using the latest version of `xtabond2`.

In Table (3.12), obviously we do not find any difference in the point estimates of the coefficients, while we have discrepancies in the estimated standard errors only when the estimates are not made robust to heteroskedasticity of the errors. Luckily, the two versions of the command still estimate very similar standard errors so that the inference on the parameters does not change. It is instead worth noticing that the

[42]When the sample is large, in a two-sided test, the critical values are 1.96 and 1.65 respectively for a significance level of 5% and 10%.

[43]The original estimate is the $BBl1$ in our notation.

[44]This is the most used command for GMM estimation of dynamic panel data in Stata. Details about the syntax are in Roodman [2009a].

inference can be affected by whether or not we account for potential heteroskedaticity: in the case of the lagged dependent variable, e.g, we find a coefficient significant at a 10% significance level when estimates are not robust, not significant when they are robust. The main didactic aim of Table (3.12) is to warn that original results can sometimes be exactly reproduced only by going back to the versions of the commands available at the time the estimates were published.

We can now move to the whole set of estimation results on real data. In Table (3.13) we report the results for the model in equation (3.5.21) when the debt is assumed to be exogenous.

It is evident, when we analyze the estimates for ϕ_g, that there is a noticeable heterogeneity in the estimated coefficient of the output gap: there is not any agreement among the estimators on the sign of the coefficient and the response to the gap is never found to be statistically significant. The estimates are particularly imprecise and unreliable, as characterized by a very large standard error, mainly when some techniques are adopted in order to reduce the number of instruments. In line with the findings in our Monte Carlo simulations, the SYS GMM estimators tend to give more counter-cyclical coefficients than the DIFF GMM estimators and estimates with a smaller variance. This is particularly evident for the LIM COL estimates. When we look at the t-statistic, we have that all the estimators point to the a-cyclicality of the discretionary fiscal policies.

With respect to the coefficient of the lagged dependent variable, we find a general picture similar to that from the simulations: the DIFF estimators systematically tend to give larger coefficients than the SYS estimators and larger standard errors. What is particularly interesting here is that there is a strong tendency of the estimates to become less negative or even positive when we adopt LIM, COL or LIMCOL estimators. The smaller the transformed instrument set, the less negative the coefficient. This gets along with a propensity to go towards a non significance of the coefficient as more as we reduce the number of instruments: for the UNTR estimator we find an estimate of ϕ_c around -0.2 and the coefficient comes out to be strongly statistically significant; for the LIM2 GMM the estimates remain very close to the untransformed ones and strongly significant, both in the case of the SYS and DIFF GMM estimation; LIM1 GMM gives smaller estimated coefficients that also become not significant. The SYS COLL estimator returns not significant estimates. Things are worsened when the LIMCOL estimator is used: we have the extreme case for LIM1COL estimates, as in this case we get estimates of the coefficient with a positive sign[45], though not significant. Mindful of the results in the Monte Carlo experiment, we are inclined to believe that the "true" value of ϕ_g is in a neighborhood of $-0.15/-0.2$, close to the value estimated by the untransformed system-GMM estimator. As we know that the $CAPB$ is by its nature a dynamic and very persistent process, we expect a significant ϕ_c coefficient, and we are suspicious whenever the coefficient is detected as not significant.

The practical indication we draw from the analysis of the estimates is that a sensible reduction in the instrument count comes along with a too severe loss of information that results in an impossibility to obtain reliable estimates of the fiscal response function. The instrument reduction techniques proposed by Roodman [2009b] reveal themselves as potentially very dangerous in the current context where the lack of information in the data is a crucial issue: the assumptions we make on the coefficients of the instrument matrix in order to limit or collapse it seem to be too restrictive in this framework. On the other hand, it is worth noticing that, when we adopt a purely statistical technique (such as PCA) to reduce the number of instruments, we have a smaller reduction in the instrument count but we safeguard the significance of the coefficients and we get estimates in line with those obtained on the untransformed instrument set.

A significant feature is that the Hansen test gives systematically a p-value of 1 even when the instrument count is drastically reduced: in the light of the previous simulations we can argue that we should not worry too much about this value, as an excessive number of instruments

[45] It is worth reminding here that an estimate of ϕ_c with a positive sign would imply an autoregressive coefficient for the $CAPB$, given by $\phi_c + 1$, that is above 1: we would deal with the unrealistic scenario of an explosive primary balance series.

is probably less dangerous than the loss of information deriving from a reduction in the instrument count.

With respect to the debt coefficient, it is detected as significant only when estimated by DIFF GMM; this comes along with a larger absolute value of the coefficient when compared to the SYS GMM estimator.

The two additional regressors appear to have a relevant explanatory power as their coefficients are very large and are generally found to be strongly significant: in particular, the Maastricht variable is found to have a systematically significant effect that is estimated around -0.5/-0.7 by the safest GMM specifications; the regular elections appear to play a relevant role, suggesting a systematic fiscal loosening in the election years[46].

We have here a confirmation of the fact that we can not omit the elections and the EU rules in a fiscal response function and we have a clue on the fact that additional variables not included in the standard $CAPB$ model, such as institutional factors or budgetary rules, could play a relevant role in explaining the discretionary response of fiscal policies and that their omission could potentially affect the inference on the coefficients of the variables included.

We could argue that, in the EMU context, the discretionary action of the policy maker is driven more by the provisions of the Stability Growth Pact rather than by the output gap and the actual stock of debt.

In Table (3.14) we report the estimate for the $CAPB$ model in the case where also the debt is treated as an endogenous regressor. What immediately emerges is that things do not change much with respect to the previous case. Our main parameter of interest, ϕ_g, is again constantly not significant in all the cases except for $ABl1c$[47].

Also as far as ϕ_c is concerned, the estimates mirror those in Table (3.13), with a considerable tendency to be statistically significant and

[46] The coefficient looses in significance only when the instrument count is too drastically reduced.

[47] The statistical significance in this case comes along with an implausibly high estimated coefficient. The most likely explanation is again the bias due to a too drastic reduction in the number of instrument together with the well-known tendency of the DIFF estimator to overestimate the coefficients.

around -0.20 when the safest specifications of the SYS estimator are used, and to become implausibly positive and not significant when we reduce at the minimum level the columns of the instrument matrix. Again, we are inclined to believe the true parameter is not far from -0.20.

With respect to ϕ_d, we find a significant coefficient when we use a DIFF GMM approach and also when the instrument set is factorized or contemporaneously limited and collapsed. In contrast with what happens for the other endogenous regressors, whose significance is affected by a reduction in the instrument count, the coefficient of debt gets more significant as we reduce the number of instruments.

Overall, there is not a clear indication on the role of the initial state of finances, as captured by the debt, in determining discretionary fiscal actions: the estimates are not robust to the alternative choices about the instrument matrix. The interpretation of the results for ϕ_d becomes even more controversial if we recall the findings of the Monte Carlo simulations: also in a controlled experiment, in fact, we found that the alternative estimators failed to detect a significant effect of the debt in an enormously large number of iterations. We therefore argue that it is incredibly difficult to assess the role of the stock of public debt in a fiscal response function as the one specified in our model.

Even in this second case, the two explanatory variables are found to be very important in explaining the discretionary fiscal action and this evidence is robust to the estimator used.

The indication we draw from these results on real data is that, in this context, it is even possible that neither the gap nor the debt should be necessarily included among the explanatory variables while, on the other hand, there are potentially crucial additional factors that can not be excluded a priori. It is likely that we should worry more about the potential omission of relevant variables rather than on the risks deriving from instrument proliferation in GMM estimation of fiscal response functions.

Table 3.12: CAPB model on real data: reproduction of Golinelli and Momigliano [2009]

Dependent variable: $\Delta CAPB$					
Variable		BBl1 orig.	BBl1 orig. rob.	BBl1 not rob.	BB1l1 rob.
L.GAP	*coeff*	0.069	0.069	0.069	0.069
	sd	0.07	0.048	0.072	0.048
	t	1	1.44	0.96	1.44
L.CAPB	*coeff*	-0.123	-0.123	-0.123	-0.123
	sd	0.073	0.104	0.076	0.104
	t	-1.67	-1.18	-1.62	-1.18
L.DEBT	*coeff*	0.004	0.004	0.004	0.004
	sd	0.005	0.008	0.005	0.008
	t	0.76	0.46	0.74	0.46
L.Maas	*coeff*	-0.74	-0.74	-0.74	-0.74
	sd	0.176	0.212	0.182	0.212
	t	-4.21	-3.49	-4.07	-3.49
Elect	*coeff*	-0.466	-0.466	-0.466	-0.466
	sd	0.178	0.188	0.184	0.188
	t	-2.61	-2.47	-2.52	-2.47
Obs. $(N \times T)$		165	165	165	165
$\overline{T}$		15	15	15	15
Instruments		124	124	124	124
hansenp			1		1
Autocorrelation		0.044	0.038	0.051	0.038

Notes:
a. Data are for 11 EMU Countries over the period 1994-2008. The dataset is the same used by Golinelli and Momigliano [2009].
b. We report the point estimates for the coefficients (coeff), the estimated standard deviation (sd) and the t-statistic (sig) for the significance of the coefficients.
c. Details on the abbreviations for the estimators are in section 3.2.1.
d. In the first column we reproduce the original estimates of the authors; in column 2 we make standard errors robust; in column 3 and 4 we replicate the estimates in column 1 and 2 with the most recent version of the estimation commands in Stata10.
e. Instruments is the number of instruments used in the estimates.
f. hansenp is the p-value for the Hansen overidentifying restriction test (robust but weakened by many instruments; see Roodman [2009b].)
g. Autocorrelation reports the p-value of the residuals' second-order autocorrelation test.

Table 3.13: CAPB model on real data with exogenous debt

Dependent variable: $\Delta CAPB$

Variable		OLS	ABa	BBa	ABl1	BBl1	ABl2	BBl2	ABc	BBc	ABl1c	BBl1c	ABl2c	BBl2c	ABpcaa	ABpcav	BBpcaa	BBpcav
L.GAP	*coeff*	-0.008	-0.050	0.021	0.004	0.069	-0.055	0.016	-0.066	-0.008	0.025	0.258	-0.049	0.173	-0.049	-0.139	0.009	-0.021
	sd	0.062	0.065	0.037	0.135	0.048	0.078	0.047	0.073	0.072	0.27	0.229	0.173	0.191	0.069	0.13	0.045	0.047
	t	-0.13	-0.77	0.57	0.03	1.44	-0.7	0.34	-0.9	-0.11	0.09	1.13	-0.29	0.9	-0.7	-1.07	0.19	-0.45
L.CAPB	*coeff*	-0.156	-0.285	-0.198	-0.208	-0.123	-0.310	-0.211	-0.301	-0.126	0.177	0.325	-0.041	0.177	-0.325	-0.406	-0.239	-0.262
	sd	0.062	0.060	0.053	0.2	0.104	0.101	0.075	0.094	0.105	0.286	0.219	0.183	0.177	0.079	0.123	0.066	0.066
	t	-2.53	-4.74	-3.76	-1.04	-1.18	-3.07	-2.82	-3.21	-1.2	0.62	1.48	-0.23	1	-4.11	-3.3	-3.64	-3.96
L.DEBT	*coeff*	0.005	0.016	0.008	0.027	0.004	0.026	0.008	0.023	0.004	0.049	-0.016	0.038	-0.009	0.022	0.056	0.010	0.011
	sd	0.004	0.005	0.005	0.012	0.008	0.01	0.006	0.01	0.007	0.019	0.013	0.013	0.01	0.006	0.019	0.006	0.006
	t	1.1	3.10	1.41	2.25	0.46	2.65	1.33	2.36	0.59	2.53	-1.22	2.93	-0.92	3.54	3	1.72	1.84
L.Maas	*coeff*	-0.555	-0.450	-0.578	-0.782	-0.740	-0.547	-0.620	-0.471	-0.735	-1.261	-1.366	-0.927	-1.095	-0.463	-0.518	-0.537	-0.508
	sd	0.147	0.145	0.133	0.175	0.212	0.123	0.169	0.142	0.206	0.353	0.357	0.196	0.27	0.125	0.181	0.145	0.164
	t	-3.77	-4.09	-4.33	-4.47	-3.49	-4.45	-3.66	-3.32	-3.57	-3.57	-3.83	-4.73	-4.06	-3.7	-2.85	-3.71	-3.1
Elect	*coeff*	-0.426	-0.513	-0.468	-0.411	-0.466	-0.400	-0.460	-0.476	-0.398	-0.293	-0.390	-0.323	-0.379	-0.484	-0.430	-0.373	-0.433
	sd	0.177	0.168	0.156	0.213	0.188	0.184	0.166	0.181	0.182	0.225	0.225	0.217	0.213	0.166	0.2	0.178	0.198
	t	-2.4	-3.09	-3	-1.92	-2.47	-2.17	-2.77	-2.62	-2.19	-1.3	-1.73	-1.49	-1.78	-2.91	-2.15	-2.09	-2.19
Obs. $(N \times T)$		165	154	165	154	165	154	165	154	165	154	165	154	165	154	154	165	165
$\overline{T}$		15	14	15	14	15	14	15	14	15	14	15	14	15	14	14	15	15
Instruments			154	226	68	124	118	174	96	101	20	25	24	29	130	94	185	135
hansenp			1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Autocorrelation			0.037	0.039	0.030	0.038	0.034	0.037	0.037	0.042	0.065	0.085	0.045	0.069	0.035	0.031	0.037	0.030

Notes:

a. Data are for 11 EMU Countries over the period 1994-2008. The dataset is the same used by Golinelli and Momigliano [2009].

b. We report the point estimates for the coefficients (coeff), the estimated standard deviation (sd) and the t-statistic (sig) for the significance of the coefficients.

c. Details on the abbreviations for the estimators are in section 3.2.1.

d. Instruments is the number of instruments used in the estimates.

e. hansenp is the p-value for the Hansen overidentifying restriction test (robust but weakened by many instruments; see Roodman [2009b].)

f. Autocorrelation reports the p-value of the residuals' second-order autocorrelation test.

Table 3.14: CAPB model on real data with endogenous debt

Dependent variable: $\Delta CAPB$

Variable		OLS	ABa	BBa	ABl1	BBl1	ABl2	BBl2	ABc	BBc	ABl1c	BBl1c	ABl2c	BBl2c	ABpcaa	ABpcav	BBpcaa	BBpcav
L.GAP	*coeff*	-0.008	-0.034	-0.007	-0.001	0.036	-0.044	-0.001	-0.042	-0.004	0.493	0.063	0.187	0.063	-0.107	-0.151	0.010	-0.034
	sd	0.062	0.059	0.045	0.101	0.046	0.067	0.041	0.064	0.066	0.241	0.093	0.132	0.096	0.097	0.13	0.053	0.043
	t	-0.13	-0.57	-0.15	-0.01	0.78	-0.65	-0.03	-0.66	-0.06	2.05	0.68	1.42	0.65	-1.1	-1.16	0.19	-0.79
L.CAPB	*coeff*	-0.156	-0.267	-0.236	-0.287	-0.216	-0.311	-0.255	-0.276	-0.192	0.453	0.086	0.099	-0.012	-0.317	-0.417	-0.209	-0.275
	sd	0.062	0.058	0.062	0.152	0.099	0.083	0.075	0.082	0.101	0.224	0.237	0.162	0.217	0.103	0.138	0.077	0.069
	t	-2.53	-4.6	-3.79	-1.89	-2.19	-3.73	-3.4	-3.38	-1.9	2.02	0.36	0.61	-0.05	-3.08	-3.02	-2.7	-4.01
L.DEBT	*coeff*	0.005	0.014	0.006	0.007	0.002	0.016	0.006	0.019	-0.004	-0.070	-0.044	-0.033	-0.047	0.020	0.053	0.008	0.016
	sd	0.004	0.005	0.005	0.014	0.006	0.005	0.006	0.01	0.014	0.05	0.022	0.019	0.021	0.007	0.02	0.006	0.006
	t	1.1	2.68	1.11	0.46	0.37	3.01	1.14	1.9	-0.27	-1.39	-2.03	-1.72	-2.2	2.88	2.65	1.36	2.7
L.Maas	*coeff*	-0.555	-0.523	-0.524	-0.711	-0.612	-0.532	-0.539	-0.510	-0.607	-1.411	-1.119	-0.891	-0.841	-0.387	-0.497	-0.505	-0.526
	sd	0.147	0.117	0.137	0.166	0.164	0.117	0.149	0.131	0.168	0.357	0.393	0.221	0.293	0.125	0.143	0.134	0.145
	t	-3.77	-4.49	-3.82	-4.29	-3.73	-4.54	-3.62	-3.88	-3.62	-3.96	-2.84	-4.02	-2.87	-3.09	-3.49	-3.77	-3.62
Elect	*coeff*	-0.426	-0.496	-0.466	-0.497	-0.475	-0.466	-0.442	-0.413	-0.398	-0.451	-0.412	-0.425	-0.441	-0.455	-0.436	-0.422	-0.422
	sd	0.177	0.161	0.155	0.185	0.186	0.157	0.157	0.201	0.199	0.266	0.222	0.226	0.215	0.188	0.194	0.191	0.186
	t	-2.4	-3.08	-3.01	-2.7	-2.56	-2.98	-2.81	-2.05	-2	-1.7	-1.85	-1.88	-2.05	-2.42	-2.25	-2.21	-2.27
Obs. $(N \times T)$		165	154	165	154	165	154	165	154	165	154	165	154	165	154	154	165	165
$\overline{T}$		15	14	15	14	15	14	15	14	15	14	15	14	15	14	14	15	15
Instruments			154	239	81	151	144	214	109	115	20	26	25	31	105	100	158	151
hansenp			1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Autocorrelation			0.037	0.038	0.031	0.035	0.034	0.036	0.036	0.040	0.102	0.052	0.059	0.048	0.037	0.028	0.038	0.030

Notes:

a. Data are for 11 EMU Countries over the period 1994-2008. The dataset is the same used by Golinelli and Momigliano [2009].

b. We report the point estimates for the coefficients (coeff), the estimated standard deviation (sd) and the t-statistic (sig) for the significance of the coefficients.

c. Details on the abbreviations for the estimators are in section 3.2.1.

d. Instruments is the number of instruments used in the estimates.

e. hansenp is the p-value for the Hansen overidentifying restriction test (robust but weakened by many instruments; see Roodman [2009b].)

f. Autocorrelation reports the p-value of the residuals' second-order autocorrelation test.

3.6 Conclusions

This chapter compares alternative dynamic panel data estimators for the estimation of fiscal response functions. To this end, we run extensive Monte Carlo simulations where we allowed various parameters in the fiscal rule to change.

First, we estimated the $CAPB$ model by different estimators, always accounting for the dynamic structure of the model, for endogeneity of the regressors, and for the risk of instrument proliferation in GMM estimation. We found significant heterogeneity in the estimates in all the scenarios and we showed that the system GMM estimator is the safest in this context as it gives the least biased estimates and the lowest variance.

We came out with a strong propensity to the use of the system GMM estimator either on the untransformed instrument matrix or on the matrix where the lag depth of the instruments is reduced not too far ahead.

We also warned about the risk of getting very biased and misleading estimates if we worry too much about the instrument count and adopt strategies to drastically reduce the number of instruments. In case of multiple endogenous regressors, we could lower significantly the p-value of the Hansen test only by paying the risk of obtaining unreliable estimates. We found that the safest instrument reduction strategy is the purely data-driven one, namely the factorization of the instrument set.

A natural consequence of our Monte Carlo experiment is the recommendation to be cautious in the interpretation of the estimated coefficients in a fiscal rule: despite the fact that we were in a controlled experiment, too many times the coefficients were detected as not significant when they actually were; on the other hand, there was also a tendency to label a coefficient as significant when it actually was not.

Second, we estimated the PB model on simulated data and we cast new light on the risks of assessing the discretionary response of fiscal actions to the cycle by simply subtracting the effect of automatic stabilizers from the estimated coefficient of the output gap in the PB model. In fact, when the dynamics of the gap and of the $CAPB$ are close each other and the discretionary policies are set to be countercyclical, this strategy points systematically to a pro-cyclicality of the policies. We suggested an alternative way to estimate the discretionary response starting from the coefficients of the PB model and we found that it is generally safer than the commonly adopted one. However, we concluded that the best strategy in order to assess the response of discretionary policies to the cycle remains the direct estimation of the CAPB model.

Third, we estimated the $CAPB$ model on the dataset of Golinelli and Momigliano [2009] for the EMU Countries. We confirmed that the estimates obtained by SYS GMM estimation are the most reliable; a too drastic reduction in the instrument count through a huge limitation of the lag depth and the collapsing of the instrument set seemed harmful for the estimates as it led to a pretty implausible not significance of the coefficient of the lagged dependent variable. We found that the detected a-cyclicality of discretionary fiscal policies in this sample is robust to all the specifications of the estimators. Across all the estimators, we systematically found a relevant importance of the SGP rules and of the regular elections in explaining discretionary fiscal actions and we argued that we should care more about potentially relevant institutional factors that can not be omitted a priori.

3.6.1 A look ahead

There are many potential extensions of the present work both with respect to simulation experiments and to the estimation of fiscal rules on real data.

We should aim at adding new and useful information to the data as more information, and hence more variability, is always beneficial to the estimates.

We can add information in several ways.

On one hand, we can add more Countries or more time periods, thus enlarging the panel either in one or in both directions. By augmenting the number of Countries, we necessarily bring more heterogeneity in the sample and we need to account for it. So far, we have assumed the poolability of the parameters[48]. It is pretty reasonable to assume poolability in the EMU context where the Maastricht Treaty constraints aim at imposing the same restrictions to all Countries in terms of fiscal policies. However, the assumption could be problematic if we enlarged the sample and considered, e.g, the whole European Union or the OECD Countries.

In any case, it would be very worthy to spend some effort to study the poolability in the context of fiscal response functions.

By adding more time periods to the analysis, we certainly have more information about the long run dynamics of the fiscal variables but, at the same time, we would be in a scenario in which with longer time series it could be reasonable and safer to estimate fiscal reaction functions separately for each Country. In addition, with more time periods in the sample, we incur a severe risk of having structural breaks in the series[49]. In the context of the EMU, a structural break has been identified in the Maastricht agreement and the Stability Growth Pact and several empirical works have aimed at checking whether the Treaty has affected the cyclical response of fiscal policies by estimating the fiscal rule before and after the Maastricht Treaty[50]. If we were able to add information through an increase in the number of periods, on one hand we would have the opportunity to test for structural breaks and to assess their impact on fiscal responses, on the other hand we could move toward a time series context and estimate the reaction function for individual time series.

Another way to add information is to include more variables in the estimation. The estimates on the real data seem to give a very clear indication: 'Not only the gap!'. When fiscal response functions are estimated, most interest is devoted to the coefficient of the output gap, as it gives indications about the cyclical response of discretionary policy: however, as seen also in our empirical analysis, the gap coefficient is often identified as not significant, so that the estimates point to the a-cyclicality of the policies. We could argue, in the light of what seen above, that the

[48]We implicitly assume that all the Countries follow the same fiscal response functions or at least very similar ones.

[49]Structural breaks could alter the response of discretionary policies to the cycle.

[50]See, among others, Galì and Perotti [2003] and Wyplosz [2006].

standard $CAPB$ model, where no additional regressors are added to the the lagged dependent variable, the debt and the gap, is likely to be too simplified and to lack some explanatory power. The inclusion of additional variables, capturing institutional and legislative factors, has been shown to be very beneficial in this context. In continuity with the work of Debrun et al. [2008][51], it would be worthy to look for additional potentially relevant determinants of fiscal action and to try to identify a core of factors that can not be omitted a priori from the fiscal rule. It would be interesting then to check how the inclusion of additional regressors affects the estimates of the coefficients of the traditional fiscal variables.

It could also be attractive, though certainly not easy, to improve further the simulation model in order to make it even more realistic: for example, we could account for monetary variables that contribute to the determination of interest rates and inflation rates or model the fiscal policy shocks so that there are transmission mechanisms between Countries. We should be very cautious however as, by adding more random variables in the simulation model, we also introduce more noise that could partially hide the information conveyed by the data.

[51]The authors build a Fiscal Rule Index that aims at capturing institutional, political, geographical, fiscal factors that could play a role in determining fiscal rules.

Conclusions

After having run through the main literature on DPD model estimation, we illustrated in details the GMM approach to DPD estimation and warned that, though very appealing, it is not free of faults. Instrument proliferation, among other econometric issues, can be particularly dangerous in this context.

We introduced the extraction of principal components from the instrument matrix as an effective strategy to reduce the instrument count in DPD GMM. In Monte Carlo experiments on the panel AR(1) and on a multivariate panel models, we found that the collapsing and the factorization of the instrument matrix give similar estimates. We argued that it is difficult to detect instrument proliferation dangers in such a simplified framework. The Hansen test p-value came out not to be a safe criterion to assess a risk of instrument proliferation, though it was pretty indicative in most cases. Different violations of the Blundell-Bond assumption on the initial conditions were found to strongly affect the estimation results and to give very biased estimates when the instrument count is reduced someway.

Any instrument reduction strategy should not be adopted a priori, as every technique may have drawbacks if some assumptions do not hold.

We found that the system estimator is the best-performing in GMM estimation of fiscal response functions. The high instrument count did not come out to be particularly problematic in this context, so that the GMM estimator on the full instrument set is a safe choice. The estimates on the factorized instrument set were generally found less biased and less volatile than those obtained on the collapsed instruments. The reduction of the lag depth to only the first available lag gave very biased estimates. The collapsed estimator also proved to be particularly biased in this framework, despite it was generally the best in the AR(1) model estimation.

We estimated the $CAPB$ and the PB model on simulated panels and made the algebraic links between the parameters in the two models explicit. We also suggested an effective strategy to estimate the discretionary fiscal response from the coefficients of the PB model.

In the empirical test on a dataset for EMU Countries, we found that the detected a-cyclicality of discretionary policies is robust to all the specifications of the GMM estimator.

Overall, we suggested the researcher on always reporting the estimates obtained with alternative settings of the GMM estimators and on running as many robustness checks as possible. The estimates obtained on reduced set of instruments should always be compared with those obtained on the full set.

Which is the best estimator in DPD strongly depends on the model we are considering, on potential endogeneity of the regressors, on the collinearities among the explanatory variables and on whether crucial assumptions for the GMM estimator holds or not.

Cautiousness is always the best strategy.

Bibliography

[1] **Alfaro, R. (2008)** "Estimation of dynamic panel data: the case of corporate investments in Chile", *Central Bank of Chile*, Discussion Paper.

[2] **Altonji, J.G. and L.M. Segal (1996)** "Small-sample bias in GMM estimation of covariance structures", *Journal of Business and Economic Statistics*, vol.14, pp. 353-366.

[3] **Alvarez, J. and M. Arellano (2003)** "The time series and cross-section asymptotics of dynamic panel data estimators", *Econometrica*, vol. 71, pp. 1121-1159.

[4] **Amemiya, T. (1966)** "On the use of principal components of independent variables in two-stage least-squares estimation", *International Economic Review*, vol. 7, pp. 283-303.

[5] **Amemiya, T., (1985)** "Advanced econometrics", *Harvard University Press*, Cambridge.

[6] **Andersen, T.G. and B.E. Sorensen (1996)** "GMM estimation of a stochastic volatility model: a Monte Carlo study", *Journal of Business and Economic Statistics*, vol. 14, pp. 328-352.

[7] **Anderson, T.W., Hsiao, C. (1981)** "Estimation of dynamic models with error components", *Journal of the American Statistical Association*, vol. 76, pp. 598-606.

[8] **Anderson, T.W., Hsiao, C. (1982)** "Formulation and estimation of dynamic models using panel data", *Journal of Econometrics*, vol. 18, pp. 47-82.

[9] **Arellano, M. (2003)** "Panel data econometrics", *Oxford University Press*, Oxford.

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

[11] **Arellano, M., O. Bover (1995)** "Another look at the instrumental variables estimation of error-components models", *Journal of Econometrics*, vol. 68, pp. 29-51.

[12] **Bai, J. and S. Ng (2010)** "Instrumental variable estimation in a data rich environment", *Econometric Theory*, vol. 26, pp. 1577-1606.

[13] **Balassone, F. and F. Francese (2004)** "Cyclical asymmetries in fiscal policy, debt accumulation and the Treaty of Maastricht", *Bank of Italy*, Temi di Discussione n. 531.

[14] **Balassone, F., Francese, F. and S. Zotteri (2008)** "Cyclical asymmetries in fiscal variables", *Bank of Italy*, Temi di Discussione n.671.

[15] **Ballabriga, F. and C. Martinez-Mongay(2002)** "Has EMU shifted policy", *European Commission*,DG for Economic and Financial Affairs, Economic Paper n.166.

[16] **Baltagi, B.H. (2008)** "Econometric analysis of Panel Data. Fourth edition.", *John Wiley and Sons, Ltd*, Chichester.

[17] **Barro, R.J. and X. Sala-i-Martin (2003)** "Economic growth, Second edition", *MIT Press*, Cambridge.

[18] **Baum, C.F., Schaffer, M.E. and S. Stillman (2003)** "Instrumental variables and GMM: estimation and testing", *Boston College Working Papers in Economics*, n. 545.

[19] **Beck, T. and R. Levine (2004)** "Stock markets, banks and growth: panel evidence", *Journal of Banking and Finance*, vol. 28, pp. 423-442.

[20] **Beetsma, R. and M. Giuliodori (2008)** "Fiscal adjustment to cyclical developments in the OECD: an empirical analysis based on real-time data", *CEPR*, Discussion Paper n. 6692.

[21] **Bernoth, K., Hughes Hallett, A. and J. Lewis (2008)** "Did fiscal policy makers know what they were doing? Reassessing fiscal policy with real time data", *CEPR*, Discussion Paper n. 6758.

[22] **Blundell, R.W. and S.R. Bond (1998a)** "Initial conditions and moment restrictions in dynamic panel data models", *Journal of Econometrics*, vol. 87, pp. 115-143.

[23] **Blundell, R.W. and S.R. Bond (1998b)** "GMM estimation with persistent panel data: an application to production functions", *Institute for Fiscal Studies*, Working Paper 99/4.

[24] **Blundell, R.W., Bond, S.R. and F. Windmeijer (2000)** "Estimation in dynamic panel data models: improving on the performance of the standard GMM estimator", *Institute for Fiscal Studies*, Working Paper 00/12.

[25] **Bohn, H. [1998]** "The behavior of U.S. public debt and deficits", *Quarterly Journal of Economics*, vol. 113, pp. 949-963.

[26] **Bond, S. (2002)** "Dynamic panel data models: a guide to micro data methods and practice", *CEMMAP*, Working paper 09/02.

[27] **Bond, S.R., Hoeffler, A. and J. Temple (2001)** "GMM estimation of empirical growth models", *CEPR Discussion paper*, n. 3048.

[28] **Bond, S. and F. Windmeijer (2002)** "Finite sample inference for GMM estimators in linear panel data model", *CEMMAP*, Working paper 04/02.

[29] **Bouthevillain, C., Cour-Thimann, P., van den Dool, G., Hernandez de Cos, P., Langenus, G., Mohr, M., Momigliano, S., and M.**

Tujula (2001) “Cyclically adjusted budget balances: an alternative approach”, *European Central Bank*, Working Paper n. 77.

[30] **Bowsher, C. G. (2002)** “On testing overidentifying restrictions in dynamic panel data models”, *Economics Letters*, vol. 77, pp. 211-220.

[31] **Bun M.J.G. and F. Windmeijer (2010)** “The weak instrument problem of the system GMM estimator in dynamic panel data models”, *Econometrics Journal*, vol. 13, pp. 95-126.

[32] **Calderon, C.A., Chong, A. and N.V. Loyaza (2002)** “Determinants of current account deficits in developing countries”, *Contributions to Macroeconomics*, vol. 2.

[33] **Cameron, A.C. and P.K. Trivedi (2005)** “Microeconometrics: methods and applications”, *Cambridge University Press*, Cambridge.

[34] **Celasun, O., and J. S. Kang (2006)** “On the properties of various estimators for fiscal reaction functions”, *IMF*, Working Paper n. 182.

[35] **Debrun, X. and M. Kumar (2007)** “The discipline-enhancing role of fiscal institutions: theory and empirical evidence”, *IMF*, Working Paper n. 171.

[36] **Debrun, X., Moulin, L., Turrini, A., Ayuso-i-Casals, J., and M. Kumar (2008)** “National fiscal rules”, *Economic Policy*, April 2008, pp. 297-362.

[37] **Durlauf, S.N., Johnson, P.A. and J.R.W. Temple (2005)** “Growth econometrics”, in P. Aghion and S. N. Durlauf (eds.) *Handbook of Economic Growth*, Vol. 1A, North-Holland, Amsterdam, pp. 555-677.

[38] **Forbes, K. J. (2000)** “A reassessment of the relationship between inequality and growth”, *American Economic Review*, vol. 90, pp. 869-887.

[39] **Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000)** “The generalized factor model: identification and estimation”, *Review of Economics and Statistics*, vol. 82, pp. 540-554.

[40] **Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2004)** "The generalized factor model: consistency and rates", *Journal of Econometrics*, vol. 119, pp. 231-255.

[41] **Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2005)** "The generalized dynamic factor model: one-sided estimation and forecasting", *Journal of the American Statistical Association*, Vol. 100, pp. 830-839.

[42] **Forni, L. and S. Momigliano (2004)** "Cyclical sensivity of fiscal policies based on real-time data", *Applied Economics Qurterly*, vol. 50, pp. 299-326.

[43] **Galì, J. and R. Perotti (2003)** "Fiscal policy and monetary integration in Europe", *Economic Policy*, vol. 18, pp. 533-572.

[44] **Golinelli, R. and S. Momigliano (2006)** "Real-time determinants of the fiscal policies in the Euro area", *Journal of Policy Modeling*, vol. 28, pp. 943-964.

[45] **Golinelli, R. and S. Momigliano (2008)** "The cyclical response of fiscal policies in the Euro area: why do results of empirical research differ so strongly?", *Bank of Italy*, Temi di Discussione n. 654.

[46] **Golinelli, R. and S. Momigliano (2009)** "The cyclical reaction of fiscal policies in the Euro area: the role of modeling choices and data vintages", *Fiscal Studies*, vol. 30, n. 1.

[47] **Groen, J.J.J. and G. Kapetanios (2009)** "Parsimonious estimation with many instruments", *Federal Reserve Bank of New York*, Staff Report n. 386.

[48] **Hansen, L.P., (1982)** "Large sample properties of generalized methods of moments estimators", *Econometrica*, vol. 50, pp. 1029-1054.

[49] **Hansen, L. P., Singleton, K.J. (1982)** "Generalized instrumental variables estimation of nonlinear rational expectations models", *Econometrica*, vol. 50, pp. 1269-1286.

[50] **Hayakawa, K. (2009)** "On the effect of mean-nonstationarity in dynamic panel data model", *Journal of Econometrics*, vol. 153, pp. 133-135.

[51] **Holtz-Eakin, D., Newey, W., Rosen, H. S. (1988)** "Estimating vector autoregressions with panel data", *Econometrica*, vol. 56, pp. 1371-1395.

[52] **Hsiao, C. (2003)** "Analysis of panel data. Second edition", *Cambridge University Press*, Cambridge.

[53] **Hsiao, C. (2007)** "Panel data analysis - Advantages and challenges", *TEST: An Official Journal of the Spanish Society of Statistics and Operations Research*, vol. 16, pp. 1-22.

[54] **IMF (2011)** "Fiscal monitor", *International Monetary Fund*, April 2011.

[55] **Kapetanios, G. and M. Marcellino (2010)** "Factor-GMM estimation with large sets of possibly weak instruments manuscript", *Computational Statistics and Data Analysis*, vol. 54, pp. 2655-2675.

[56] **Kiviet, J.F. (2007a)** "Judging contending estimators by simulation: tournaments in dynamic panel data models", Ch. 11 in *The Refinement of Econometric Estimation and Test Procedures (eds.: G.D.A. Phillips and E. Tzavalis), Cambridge University Press*, Cambridge.

[57] **Kiviet, J.F. (2007b)** " On the optimal weighting matrix for the GMM system estimator in dynamic panel data models", *UvA Econometrics*, Discussion Paper 2007/2008.

[58] **Kloek, T., and L.B.M. Mennes (1960)** "Simultaneous equations estimation based on principal components of predetermined variables", *Econometrica*, vol. 28, pp. 45-61.

[59] **Mehrhoff, J., (2009)** "A solution to the problem of too many instruments in dynamic panel data GMM", *Discussion paper n. 1/2009*, Deutsche Bundesbank.

[60] **Nickell, S. (1981)** "Biases in dynamic models with fixed effects", *Econometrica*, vol. 49, pp. 1417-1426.

[61] **OECD (2011)** "Economic Outlook", *OECD*, n. 89, June 2011.

[62] **Roodman, D. (2009a)** "How to do xtabond2: an introduction to "Difference" and "System" GMM in Stata", *The Stata Journal*, vol. 9, pp. 86-136.

[63] **Roodman, D. (2009b)** "A Note on the theme of too many instruments", *Oxford Bulletin of Economics and Statistics*, vol. 71, pp. 135-158.

[64] **Sargan, J.D. (1958)** "The estimation of economic relationships using instrumental variables", *Econometrica*, vol. 26, pp. 393-415.

[65] **Socol, C. and A. G. Socol (2009)** "The analysis of fiscal policy management in Romania: lessons for emerging countries", *African Journal of Business Management*, vol. 3, pp. 240-247.

[66] **Stock, J.H. and M.W. Watson, (1998)** "Diffusion indexes", *NBER*, Working Paper 6702.

[67] **Stock, J.H. and M.W. Watson, (2002a)** "Forecasting using principal components from a large number of predictors", *Journal of the American Statistical Association*, vol. 97, pp. 1167-1179.

[68] **Stock, J.H. and M.W. Watson, (2002b)** "Macroeconomic forecasting using diffusion indexes", *Journal of Business and Economic Statistics*, vol. 20, pp. 147-162.

[69] **Stock, J.H. and M.W. Watson, (2010)** "Dynamic factor models", Chapter 2 in M. Clements and D. Hendry (eds), *Oxford Handbook of Economic Forecasting*.

[70] **Tauchen, G. (1986)** "Statistical properties of generalized method-of-moments estimators of structural parameters obtained from financial market data", *Journal of Business and Economic Statistics*, vol. 4, pp. 397-416.

[71] **Taylor, J. B. (2000)** "Reassessing discretionary fiscal policy", *Journal of Economic Perspectives*, vol. 14, n. 3, pp. 21-36.

[72] **Windmeijer, F. (2005)** "A finite sample correction for the variance of linear efficient two-step GMM estimators", *Journal of Econometrics*, vol. 126. pp. 25-51.

[73] **Wooldridge, J.M. (2010)** "Econometric analysis of cross section and panel data. Second edition.", *MIT Press*, Cambridge.

[74] **Wyplosz (2006)** "European Monetary Union: the dark sides of a major success", *Economic Policy*, vol. 21, pp. 208-261.

[75] **Ziliak (2006)** "Efficient estimation with panel data when instruments are predetermined: an empirical comparison of moment-condition estimators", *Journal of Business and Economic Statistics*, vol. 15, pp. 419-431.