Principal Component Analysis: is this an effective way to deal with over-identification in GMM panel data estimation?*
[Preliminary draft]

Maria Elena Bontempi†
Irene Mammi‡

September 30, 2013

Abstract

The problem of instrument proliferation and its consequences (overfitting of endogenous variables, bias of estimates, weakening of Sargan/Hansen test) are well known. The literature provides little guidance on how many instruments is too many. It is common practice to report the instrument count and to test the sensitivity of results to the use of more or fewer instruments. Strategies to alleviate the instrument proliferation problem are the lag-depth truncation and/or the collapse of the instrument set (the latter being an horizontal squeezing of the instrument matrix). However, such strategies involve either a certain degree of arbitrariness (based on the ability and the experience of the researcher) or of trust in the restrictions implicitly imposed (and hence untestable) on the instrument matrix. The aim of the paper is to introduce a new strategy to reduce the instrument count. The technique we propose is statistically founded and purely data-driven and, as such, it can be considered a sort of benchmark solution to the problem of instrument proliferation. We apply the principal component analysis (PCA) on the instrument matrix and exploit the PCA scores as the instrument set

*The authors would like to thank Jerry Hausman, Jacques Mairesse, Whitney Newey, Franco Peracchi and participants in the 5th Italian Congress of Econometrics and Empirical Economics (Genova, 2013) for comments and encouragement. The authors also thank Roberto Golinelli and Carla Rampichini for suggestions. The usual disclaimer applies.

†Dep. of Economics, University of Bologna. Email: mariaelena.bontempi@unibo.it
‡Dep. of Economics, University of Bologna. Email: irene.mammi@unibo.it

for the panel generalized method-of-moments (GMM) estimation. Through Monte Carlo simulations for static and dynamic model specifications and under alternative characteristics of persistence of the variables, we compare the performance of the Difference GMM, Level and System GMM estimators when lag truncation, collapsing and our principal component-based IV reduction (PCIVR henceforth) are applied to the instrument set. The same comparison has been carried out with two empirical applications on real data: the first replicates the estimates of Blundell and Bond [1998]; the second exploits a new and large panel data-set in order to assess the role of tangible and intangible capital on productivity. Results show that PCIVR is a promising strategy of instrument reduction.

JEL classification: C13, C15, C33, C36, C63.
Keywords: Panel data, generalized method of moments, proliferation of instruments, principal component analysis, persistence.

1 Introduction

Dynamic panel data (DPD) have become very popular in the last two decades, thanks in particular to the increasing availability of panel datasets 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: in fact, it is now a common practice to estimate dynamic models in empirical analysis in most microeconomic fields.

In particular, the generalized method-of-moments (GMM) estimator, in the Holtz-Eakin, Newey and Rosen [1988], Arellano and Bond [1991], Arellano and Bover [1995] and Blundell and Bond [1998] 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 well known DPD bias (see Nickell [1981]) and from 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. Another advantage is the availability of "internal" instruments (lags of the endogenous variables), a noticeable point when finding instruments is not an easy task. The implementation of ad hoc procedures in

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

many statistical softwares and the consequent 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 to DPD models and needs to receive more attention than what it has been done so far. The potential distortions in the estimates by instrumental variables (IV) and GMM estimators when the instrument count gets larger and larger have been treated extensively in the literature[2], but not enough attention has been paid to this issue in Difference, Level and System GMM estimation of DPD (DIF GMM, LEV GMM and SYS GMM henceforth).

Though these 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 with T and the dimension, m, of the vector of endogenous regressors other than the lagged dependent variable; this number can get rapidly large relative to the sample size. Consequently, the excessive number of instruments can create a trade-off between bias (overfitting of endogenous variables) and efficiency (additional moment conditions), give an imprecise estimate of the variance/covariance matrix of the moments, lower the power of specification tests (Sargan [1958] / Hansen [1982] test of over-identifying restrictions) and exacerbate the weak instruments problem.

Unfortunately, the problem of instrument proliferation is only rarely detected and addressed in empirical analyses with the consequent risk of drawing misleading conclusions about the coefficient estimates. In many empirical papers, GMM is often applied with unclear specification of the estimator concerning initial weighting matrix, onestep or twostep estimate and, in particular, the selection of instruments: different results emerge as a consequence of different choices of the instrument matrix (for example, how many lags are included) and it becomes difficult to interpret such results as robustness checks, as they are based on a certain degree of arbitrariness, ability or experience of the researcher.

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

The paper has two aims. The first one is to introduce a data-driven technique for the reduction of the instrument count in GMM estimation of DPD with explana-

[2]See, among the others, Ziliak [1997] and Bowsher [2002].

tory endogenous variables. We extract the principal components from the instrument matrix through the principal component analysis (PCA) and use the PCA scores as a new set of instruments (we call this procedure principal component-based IV reduction, PCIVR, henceforth). In doing so, we aspire to answer the question "How many moment conditions can be used and still expect to be able to obtain valid inference when estimating by GMM?". Since, in the words of Hall and Peixe [2003, p. 271], "It is impossible to verify a priori which elements of the candidate [instruments] set satisfy [the] conditions [orthogonality, identification, efficiency, and non-redundancy] for a given data set", we suggest a statistically founded rule for the selection of non redundant IVs, based on the characteristics of the empirical problem at hand. In doing so, we extend the analysis of Doran and Schmidt [2006] who simulate a simple autoregressive DPD and compare results for different autoregressive parameter values, different variance of individual effects, different sample sizes N and T.[3] We simulate both static and dynamic models with endogenous explanatory variables, in addition to the lagged dependent variable; we consider an eigenvalue-eigenvector decomposition of the instrument matrix and compare different selection criteria for the PCA scores.

The second aim of the paper is to fill the gap in the literature by comparing the performance of the Difference, Level and System GMM estimators when various instrument reduction techniques are adopted. In order to do so, we both run Monte Carlo experiments and we estimate economic models on real data, allowing for the presence of endogenous variables (together with the lagged dependent variable), and checking the effects of various persistence characteristics of the stochastic processes, of different sample sizes N and T, and of the use of Windmeijer [2005] finite sample correction.

Along with the PCIVR method, the other techniques to reduce the number of moment conditions we compare are the two usually employed in the empirical literature: the collapsing of the instrument matrix (Roodman [2009b]) and the reduction of the lag depth of the instruments. Both solutions make the instrument count linear in T: the former creates different instruments for each lag but not also for each time period; the latter consists of the inclusion as instruments of only few lags instead of all the available ones. Both techniques, separately or combined together, have gained popularity thanks to their direct implementability in the statistical softwares and are now commonly, and often blindly, used in empirical works.[4] However, collapsing and lag depth truncation involve a certain

[3] Mehrhoff [2009] sketches the idea of applying the PCA in the Difference GMM framework for a simple AR(1) process but he arbitrarily selects the number of components to be retained.

[4] Other suggestions by the literature had less following in the applied works: the projection-

degree of arbitrariness as they ask either to trust the restrictions that are imposed when the instrument matrix is collapsed or to choose how many lags to include among the instruments. 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 of a DPD with other endogenous variables besides the lagged dependent variable. Our paper aims to fill this gap.[5]

Our results confirm that PCIVR is a general, data-driven technique to reduce overidentification problems that can be fruitfully applied to any overidentified GMM problem. Having tried alternative criteria in order to select the number of retained components (keep only the components whose eigenvalues are larger than the average eigenvalue or retain only the components that explain a given predetermined portion of the original variance), we suggest, as a selection criterion, the explanation of 90% of the original variance.

In the remainder of the work we proceed as follows: in section 2, after reviewing the collapsing and limiting, we illustrate the extraction of principal components from the instrument matrix and discuss the rationale of applying the PCA on the instrument set; the comparison of a number of instrument reduction techniques is presented by replicating the Blundell and Bond [1998] estimates for the labour demand in the UK and by exploiting extensive Monte Carlo simulations (in section 3); in section 4 we present an empirical application that estimates a production function with three inputs - labour, tangible and intangible capital - for a large panel data-set; section 5 draws the conclusions and indicate practical hints for the empirical analysis; the Appendix runs through the technical details of the PCA.

2 Reducing the instrument count in GMM estimation

Consider the general one-way error component DPD model:

$$y_{it} = \alpha y_{it-1} + \boldsymbol{\beta}' \mathrm{x}_{i,t} + \phi_t + \upsilon_{it} \; , \; \upsilon_{it} = \eta_i + \varepsilon_{it}, \tag{1}$$

restricted IV estimation of Arellano [2003] and the canonical correlations and information criteria of Hall and Peixe [2003].

[5] Roodman [2009b] presents only a Monte Carlo experiment limited to an autoregressive model to compare the collapsing and lag-truncation 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 GMM estimator, that is less exposed to instrument proliferation dangers.

where $i = 1, .., N$, $t = 1, .., T$, $\mathbf{x}$ is a m-dimensional vector of potentially endogenous regressors, the ϕ_t are the time effects (usually considered deterministic), the η_i are the individual effects and ε_{it} is a zero-mean idiosyncratic error, allowed to be heteroskedastic but not serially correlated. The standard assumptions are: $E[\eta_i]=E[\varepsilon_{it}]=E[\eta_i\varepsilon_{it}]=0$ and predetermined initial conditions $E[y_{i1}\varepsilon_{it}]=0$.

The Arellano-Bond and Arellano-Bover / Blundell-Bond estimators are linear GMM estimators for the model in first differences (DIF GMM) or in levels (LEV GMM) or both (SYS GMM) where the instrument matrix $\mathbf{Z}$ includes the lagged values of the endogenous variables only or also the lagged first differences of the endogenous variables[6]. In the standard framework of DIF and SYS GMM, the columns of the instrument matrix $\mathbf{Z}$ correspond respectively to two different sets of meaningful moment conditions.

In particular, the Arellano-Bond DIF GMM estimator exploits, for each endogenous variable, the following $(T-2)(T-1)/2$ moment conditions for the equation (1) in first differences:[7]

$$E[(\mathbf{Z_i}^{\text{dif}})'\Delta\upsilon_i] = E[(\mathbf{Z_{it-l}}^{\text{dif}})'\Delta\upsilon_{it}] = 0 \text{ for } t \geq 3, l \geq 2 \tag{2}$$

For the sake of simplicity suppose m=1; the instrument matrix $\mathbf{Z}^{\text{dif}}$, that satisfies the moment restrictions in (2), contains an IV for each endogenous variable, time period and lag distance and it has the well known form:

$$\mathbf{Z}_i^{\text{dif}} = \begin{pmatrix} y_{i1} & x_{i1} & 0 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ldots & \ldots & 0 \\ 0 & 0 & \ldots & 0 & y_{i1} & \ldots & y_{iT-2} & x_{i1} & \ldots & x_{iT-2} \end{pmatrix} \tag{3}$$

The Blundell-Bond SYS GMM estimator also exploits, for each endogenous variable, the additional non-redundant $T-2$ orthogonality conditions for the equation (1) in levels:

$$E[(\mathbf{Z_i}^{\text{lev}})'\upsilon_i] = E[(\mathbf{Z_{is}}^{\text{lev}})'\upsilon_{iT}] = 0 \text{ for } s = 2, ..., T-1 \tag{4}$$

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

[7] Suitably lagged x-variables can also be used as IVs when the x-variables are predetermined or strictly exogenous: for predetermined x-variables we have $l \geq 1$ and $(T-2)(T+1)/2$ moment conditions; if they are instead strictly exogenous $l = 0$ and the moment conditions are $T(T-2)$.

where, again for m=1, the instrument matrix is:[8]

$$\mathrm{Z}_i^{\text{lev}} = \begin{pmatrix} \Delta y_{i2} & \Delta x_{i2} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \dots & 0 \\ 0 & 0 & \dots & \Delta y_{iT-1} & \Delta x_{iT-1} \end{pmatrix} \tag{5}$$

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

$$\mathrm{Z}_i^{\text{sys}} = \begin{pmatrix} \mathrm{Z}_i^{\text{dif}} & 0 \\ 0 & \mathrm{Z}_i^{\text{lev}} \end{pmatrix}. \tag{6}$$

Since usually lags of the explanatory variables are used as IVs, "the phenomenon of moment condition proliferation is far from being a theoretical construct and arises in a natural way in many empirical econometric settings" (Han and Phillips [2006, p. 149]). The dimension of the GMM-type instrument matrix grows as the number of time periods and regressors expands, even if the time span of the panel is of moderate size.

2.1 Collapsing and limiting the instrument set

As discussed in Roodman [2009], when we collapse the instrument set we impose the same condition for all t and we create an instrument for each endogenous variable and lag distance rather than for each endogenous variable, time period and lag distance. The collapsed instrument matrix for the equation in first differences has the form, for m=1:

$$\mathrm{Z}_i^{\text{dif, C}} = \begin{pmatrix} y_{i1} & 0 & x_{i1} & 0 & 0 & \dots \\ y_{i2} & y_{i1} & x_{i2} & x_{i1} & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{7}$$

with $(T-2)$ moment conditions for each endogenous.

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

$$\mathrm{Z}_i^{\text{lev, C}} = \begin{pmatrix} \Delta y_{i2} & \Delta x_{i2} \\ \Delta y_{i3} & \Delta x_{i3} \\ \vdots & \vdots \end{pmatrix} \tag{8}$$

[8]The LEV GMM estimation considers, for each endogenous variable, time period and lag distance, all the available lags of the first differences as instrument for the equation in levels because they are non redundant. See Bond [2002] and Bun and Windmeijer [2010] for further discussion on this issue.

The collapsed matrix for the system estimator will thus be:

$$\mathbf{Z}_i^{\text{sys, C}} = \begin{pmatrix} \mathbf{Z}_i^{\text{dif, C}} & 0 \\ 0 & \mathbf{Z}_i^{\text{lev, C}} \end{pmatrix}. \tag{9}$$

with $(T-2)+1$ moment conditions for each endogenous variable.

When instead we limit the lag depth, we truncate the moment restrictions and exploit the conditions in equation (2) 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:

$$\mathbf{Z}_i^{\text{dif, L}} = \begin{pmatrix} y_{i1} & x_{i1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & y_{i2} & y_{i1} & x_{i2} & x_{i1} & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & y_{i3} & y_{i2} & x_{i3} & x_{i2} & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{10}$$

The number of instruments is $\frac{(T-2)(T-1)}{2} - \frac{(T-2-M)(T-1-M)}{2}$ and the instrument count depends on the number of endogenous variables, on T and on M. 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 The static model

When we assume that the coefficient α in equation 1 is equal to zero, the general dynamic panel data model reduces to the following static panel model

$$y_{it} = \boldsymbol{\beta}' \mathbf{x}_{i,t} + \phi_t + \upsilon_{it} \ , \ \upsilon_{it} = \eta_i + \varepsilon_{it}, \tag{11}$$

in which the vector $\mathbf{x}$ is a m-dimensional vector of potentially endogenous regressors. We obviously get rid of the endogeneity problems due to the presence of the lagged dependent variable, but we could still face the need to instrument endogenous variables in $\mathbf{x}$ with their lags. The moment conditions presented above are still valid for the variables in $\mathbf{x}$ and the insrument matrices are the same as above except for the fact that the columns relative to the y variable disappear.

In our Monte Carlo experiments, we will also simulate a static model whose results will be presented in section 3.2.

2.3 Extracting principal components from the matrix of instruments

In order to face the problem of instrument proliferation, we propose a strategy that involves a orthogonal transformation of the instrument set: we extract the principal components from the instrument matrix **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 macroeconomic fields of analysis. The main use of factors 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[9]. 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 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[10].

In the stream that uses factors as instruments, the main novelty of what we do here is that we consider a DPD model with endogenous explanatory variables and extract principal components allowing for two strategies: (1) we apply PCA to a large set of lags of each instrument considered separately (what we call PCIV); (2) we apply PCA to a large set of lags of all the different instruments taken together (what we call PCIVT). The idea is that of identifying the most meaningful basis to re-express the information conveyed by the **Z**, avoiding multicollinearities in the instrument set. This new basis should filter out the noise component of the moment conditions[11] and reveal the signal delivered by the instrument set (coming from the mean of the sample moment conditions); most important, the

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

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

[11] The degree of variation over the sample moment conditions increases as the number of moment conditions raises

noise reduction is the result of a data-driven procedure.

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 endogenous variables in GMM estimates (PCIVR).

In practice, defined $\mathbf{Z}$ as the general p-columns GMM-style instrument matrix[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) $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_p$. Our new instruments will be the scores from PCA that are defined as:

$$\mathbf{s}_k = \mathbf{Z}\mathbf{u}_k \text{ for } k = 1, 2, ..., p. \tag{12}$$

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:

$$\mathbf{s}_k = u_{k1}\mathbf{z}_1 + u_{k2}\mathbf{z}_2 + ... + u_{kp}\mathbf{z}_p \tag{13}$$

where u_{kj} is the j^{th} element of the principal component $\mathbf{u}_k$.

Since the aim of the PCA is dimension reduction, it would not help to keep all the p scores in the analysis as this would imply no decrease in the number of instruments; only in the first application of section 3 we will check the impact of PCIVR on estimation results when all the p components are retained. In general we suggest to retain only $(m+1) \leq q < p$ principal components; 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

[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. The PCA on the covariance matrix can be used when the variables are in commensurable units and have similar variances, as it is generally the case in Monte Carlo experiments. In estimating economic models the PCA on the correlation matrix is instead preferable. We always use PCA on the correlation matrix.

[14] $\mathbf{Z}$ can be $\mathbf{Z}^{\text{dif}}$, $\mathbf{Z}^{\text{sys}}$, $\mathbf{Z}^{\text{dif,C}}$, $\mathbf{Z}^{\text{sys,C}}$, $\mathbf{Z}^{\text{dif,L}}$, $\mathbf{Z}^{\text{sys,L}}$, according to the notation adopted in the previous sections. Remember that, in the simplified case of a balanced panel with $T_i = T \; \forall i$, and m endogenous variables plus the lagged dependent variable, we have: $\mathbf{Z}^{\text{dif}}$ has $p = ((T-2)(T-1)/2)(m+1)$ columns, $\mathbf{Z}^{\text{dif,C}}$ has $p = (T-2)(m+1)$ columns, $\mathbf{Z}^{\text{dif,L}}$ a number of columns depending also on the lag truncation. In system GMM estimation, further $(T-2)(m+1)$ columns are added in $\mathbf{Z}^{\text{sys}}$ and in $\mathbf{Z}^{\text{sys,L}}$, while only $m+1$ are added to $\mathbf{Z}^{\text{sys,C}}$.

the components to be retained.[15] In line with Doran and Schmidt [2006, p. 406], we propose the variability criterion; in particular, we retain the components that explain 90% of the original variance. With this criterion, the leading eigenvectors from the eigen decomposition of the correlation matrix of the instruments describe a series of uncorrelated linear combinations of the instruments that contain most of the variance. Compared to alternative criteria to select the eigenvalues of interest, we think that retaining principal components that explain a given predetermined portion of the original variance better avoids the magnification of sampling errors in the process of inversion of the variance matrix of the moment conditions. This should decrease the variance of the estimated weighting matrix and improve finite sample performance of the GMM estimator.[16]

Defined the matrix of PCA loadings as $\mathbf{V} = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_p]$ and the matrix of PCA scores as $\mathbf{S}$, we have that $\mathbf{S} = \mathbf{ZV}$. Instead of the moment conditions in (2), we will therefore exploit the following restrictions:

$$E[(\mathbf{S}^{\text{dif}})'\Delta\boldsymbol{\upsilon}] = E[(\mathbf{Z}^{\text{dif}}\mathbf{V})'\Delta\boldsymbol{\upsilon}] = 0. \tag{14}$$

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

$$E[(\mathbf{S}^{\text{lev}})'\boldsymbol{\upsilon}] = E[(\mathbf{Z}^{\text{lev}}\mathbf{V})'\boldsymbol{\upsilon}] = 0. \tag{15}$$

In both cases, the number of moment restrictions depends on the number of components we retain in the analysis that, in turn, depends on the nature of the data at hand. 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 PCIVR is to use, instead of the untransformed instruments, linear combinations of the original instruments that are properly weighed according to the PCA loadings: no available instrument is actually dropped, but its influence might be rescaled after the PCA. It is also worth noticing that none of the instruments 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.

[15] The criteria are discussed in the Appendix.

[16] According to alternative selection criteria, the smallest eigenvalue or the two or three smallest ones can be arbitrarily dropped; alternatively, one could retain the eigenvalues higher than the average eigenvalue or a fixed number of the highest ones. Results that compare the performance of PCIVR when several of such criteria are applied, as well as under various alternative settings, are available in Mammi [2011].

A further advantage of PCA is that we can extract principal components not only from the untransformed instrument matrix but also from any transformation we think could be useful; for example, applying PCA to the limited or collapsed instrument matrix would retain all the information each matrix conveys and thus further reduce the number of instruments. As another example, we could apply multistep PCA (see e.g. D'Alessio [1989]) to highlight structural aspects of the data at hand, like persistence or heterogeneity among clusters of individuals.

3 Comparing the instrument reduction techniques

3.1 Monte Carlo experiments: a multivariate dynamic panel data model

In our set of Monte Carlo simulations we estimate a multivariate DPD whose settings are in the spirit of Blundell et al. [2000] and Hayakawa and Nagata [2012]. The model of interest is:

$$y_{it} = \alpha y_{it-1} + \beta x_{it} + \eta_i + v_{it} \tag{16}$$

$$x_{it} = \rho x_{it-1} + \tau \eta_i + \theta v_{it} + e_{it}$$

where $\eta_i \sim \mathcal{N}(0, \sigma^2_\eta)$ are the fixed effects; $v_{it} \sim \mathcal{N}(0, \sigma^2_v)$ and $e_{it} \sim \mathcal{N}(0, \sigma^2_e)$ are the idiosyncratic shocks. Initial observations are drawn from a covariance stationary distribution such that

$$E\left[\left(x_{i1} - \frac{\tau\eta_i}{1-\rho}\right)\tau\eta_i\right] = 0 \tag{17}$$

$$E\left[\left(y_{i1} - \frac{\beta\left(\frac{\tau\eta_i}{1-\rho}\right) + \eta_i}{1-\alpha}\right)\eta_i\right] = 0. \tag{18}$$

The x_{it} process is positively correlated with η_i and the value of θ is negative to mimic the effects of measurement error. The setting of the parameters in the simulation model is as follows:

Table 1: Setting of the parameters in the main simulation model

α	**0.5, 0.95**
ρ	**0.5, 0.95**
Iterations	100, 1000
N	500
T	5, 8, 20
β	1
τ	0.25
θ	-0.1
σ_η^2	1
σ_v^2	1
σ_e^2	0.16

In Tables 3 to 11 we consider 500 individuals and T=8[17]; each experiment consists of 1000 iterations; reported estimates are the two-step DIF, LEV and SYS GMM estimators, with standard errors robust to heteroskedasticity and with the Windmeijer [2005] finite sample correction. We consider different degrees of persistence for y_{it} and x_{it}, as captured by the autoregressive coefficients α and ρ. The displayed results are as follows: *mean* is the mean of estimates; *sd* is the standard deviation of estimates, *p5* and *p95* are the 5^{th} and the 95^{th} percentiles of estimates; when the Hansen test is considered, we report the p-values corresponding to the 5% and 10% of the nominal size. The main aim of these simulations is to show that the PCIVR statistical approach gives results in line with the most appropriate estimation method, that depends on the parameters' setting and on the temporal length T. Compared to collapsing and limiting instruments reduction techniques, PCIVR magnifies the good or bad performance of an estimation method, without altering the core of the results. In the case of stationarity of both variables DIF and SYS GMM provide close results, confirmed by the PCIVR. With respect to the instrument count, the count is 42 for GMM DIF, 12 for GMM collapse, 22 for GMM limited, from 10 to 19 for PCIVRV90, from 5 to 17 for PCIVRV90T, from 10 to 12 for PCIVRA and 6 for PCIVRAT. It is evident the effectiveness of PCIVR in reducing the number of overidentifying

[17] Results for T=5 and T=20 are not reported but are available upon request.

restrictions: this is particularly true when all the instruments are considered together, as in the case of PCIVT, where the reduction process driven by the characteristics of the simulated data. While collapsing and limiting a priori fix the number of moment conditions, the PCIVR presents a range of overidentifying restrictions which is the wider the larger is T.

As we move towards the near unit root case of one or of both variables, SYS GMM provides less biased and more precise estimates. It is particularly remarkable that the collapsing gives the highest standard errors in the case of persistence: this loss in the precision of the estimates is due to non-acceptable constraints on the dynamic structure of the instrument set. PCIVR is generally safer than collapsing and limiting as it provides estimates closer to the true parameters. The only not convincing performance is that of PCIVT in the case of DIF GMM under persistent stochastic processes: in addition to the problems of near unit root in the variables, we have here also an artificial and not economically-grounded correlation structure among the variables that further negatively affects the procedure of principal component extraction. In section 4, we will see that, on the contrary, PCIVT on a set of variables that have an economically-founded relationship has a better and more convincing performance.

3.1.1 Monte Carlo experiments: additional results

In Tables 12 to 20 we change the hypotheses about σ_e^2, the variance of the idiosyncratic shock, and we set it to 1. The setting of the simulation is the same as in the main simulation model except for σ_e^2. The increase of the variance of the idiosyncratic shock e leads the downward bias to disappear for the case α=0.95 and ρ=0.5.

Hence, in the Figures, we focus on the case σ_e^2=1. The first two figures summarize GMM DIF and SYS results of the Tables; Figures 3 and 4 consider the case in which we set the ratio $\frac{\sigma_\eta^2}{\sigma_v^2}$ to 4 while Figures 5 and 6 set the ratio to 0.25. We confirm the bad performance of the estimates when the ratio $\frac{\sigma_\eta^2}{\sigma_v^2}$ is high.

In Figures 7 and 8 present the coefficient estimates for the case T=13 and N=50; Figures 11 and 12 report the Hansen p-value for such case. When there are more moment conditions than observations (132 instruments compared to 50 cross-sectional individuals in the DIF GMM case), the usual estimate of the variance matrix of the moment conditions is singular, and so the usual optimal weighting matrix cannot be calculated. The problem cannot be solved by using the generalized inverse of the estimated variance matrix (Satchachai and Schmidt

[2008]). In such a case, one can drop moment conditions, but the question is which ones to drop. We think that the principal component analysis is an useful strategy in cases of near-singularity.

Considerations on the Sargan/Hansen test to be added.

3.2 The static model

We consider again the model

$$y_{it} = \alpha y_{it-1} + \beta x_{it} + \eta_i + v_{it} \tag{19}$$

in which we now assume that $\alpha = 0$ and $\beta = 1$ so that it reduces to the following static model

$$y_{it} = x_{it} + \eta_i + v_{it} \tag{20}$$

where $\eta_i \sim \mathcal{N}(0, \sigma_\eta^2)$ are the fixed effects; $v_{it} \sim \mathcal{N}(0, \sigma_v^2)$ are the idiosyncratic shocks.

We now assume that $E(x_{it}v_{it}) \neq 0$ so that x_{it} is endogenous and therefore needs to be instrumented.

The DGP for the x_{it} is the following

$$x_{it} = \rho x_{i,t-1} + \pi_1 z_{1it} + \pi_2 z_{2it} + \pi_3 z_{3it} + \tau\eta_i + \theta v_{it} + e_{it} \tag{21}$$

where z_{1it}, z_{2it} and z_{3it} are three potential instrumental variables for x_{it} and $\theta = 0.5$ ensures that x_{it} is endogenous to y_{it}.

The three instruments have the following characteristics:

- $z_{1it} \sim \mathcal{N}(0, \sigma^2_{z_{1it}})$ is a valid instrument, such that $E(z_{1it}v_{it}) = 0$, and also relevant with $E(z_{1it}x_{it})$ with $\pi_1 = 0.8$

- $z_{2it} \sim \mathcal{N}(0, \sigma^2_{z_{2it}})$ is a valid instrument, such that $E(z_{2it}v_{it}) = 0$, but weak with $E(z_{2it}x_{it})$ with $\pi_1 = 0.02$

- $z_{3it} = \gamma v_{it} + e_{z_{3it}} + \delta z_{3it-1}$ with $e_{z_{3it}} \sim (0, \sigma_{ez3})$ is an invalid instrument such that $E(z_{3it}v_{it}) \neq 0$: $\gamma = 0.5$ ensures that z_{3it} is endogenous and δ=0.75 ensures that the endogeneity problem is not solved even when lags far away from t are used. The coefficient π_3 is set to 0.8 when we want z_{3it} to be an invalid but relevant instrument and to 0.02 when we want it weak.

Table 2: Instruments for the static model

π_1	π_2	π_3	included IVs
0	0	0	0 (x_{it} treated as exogenous)
0	0	0.8	1 invalid
0	0.02	0.8	1 weak, 1 invalid
0.8	0.02	0.8	1 relevant, 1 weak, 1 invalid
0.8	0	0.8	1 relevant, 1 invalid
0.8	0	0	1 relevant
0.8	0.02	0	1 relevant, 1 weak
0	0.02	0	1 weak

We simulate 100 panel dataset and estimate the model in equation 20 by setting σ_η^2, $\sigma_{z_{1it}}^2$, $\sigma_{z_{2it}}^2$, $\sigma_v^2$and σ_{ez3} to 1. We include different sets of instruments according to the values of the coefficients π_1, π_2 and π_3. Our choice about the instruments is reported in table 2.

RESULTS AND COMMENTS TO BE ADDED

4 Empirical examples

4.1 Old and new panel data methods applied to the controversial issue of production function estimates

In order to compare the performance of alternative instrument reduction techniques in the estimation of an economic model on real data, we use a production function specification with three inputs - labour, tangible and intangible capital stocks - on a large and unbalanced panel of Italian manufacturing companies over the period 1982-2010. Two main reasons drive our choice. As first motivation, the estimation of production functions from company panel data has become puzzling for panel data estimation methods (e.g. Mairesse and Sassenou [1991], Griliches [1998]). Pooled OLS regressions yield plausible parameter estimates, in line with factor shares and generally consistent with constant return to scale. However these estimates should be biased by omitted heterogeneity and endogeneity issues. Attempts to control for unobserved heterogeneity with within or first-difference transformations tend to yield less satisfactory parameter estimates: "In empirical practice, the application of panel data methods to micro-data produced rather unsatisfactory results: low and often insignificant capital coefficients and unreasonably low estimates of returns to scale" (Griliches and Mairesse [1998] p. 177; see also the discussion in Mairesse and Hall [1995]). The endogeneity issue arises from the simultaneous choice of output and inputs by the decision maker and from the correlations between firm-effects (efficiency levels of the companies, unknown to the econometrician) and the explanatory variables. It also arises from possible measurement errors in variables: omission of labour and capital intensity-of-utilisation variables - such as hours of work per employees and hours of operation per machine; problems in capital stocks construction (changes in the accounting normative, choice of depreciation rates); lack of distinction between blue and white collars in the labour input; lack of firm-specific prices. Noticeable is the fact that GMM methods are usually applied on first differenced equations using appropriately lagged levels of explanatory variables as instruments, with lag-depth truncation at t-3 (Mairesse and Hall [1996] for France and US; Mairesse and Jaumandreu [2005] for France and Spain; Bontempi and Mairesse [2008] for Italy). The second motivation is that our data-set is a large unbalanced panel with a considerable temporal span and our specification model includes three endogenous explanatory variables. Since the number of available instruments depends on the length of the panel and on the number of endogenous explanatory variables, and it changes from cross-section to cross-section, the GMM estimation procedures become very complex, calling for a fruitful use of PCIVR techniques in reducing overfitting problems. Table 21 shows the by-year and by-industry sample composition. Data are drawn from the CADS (Company Accounts Data Service of Centrale dei Bilanci), which is highly representative of the population of Italian companies, covering over 50% of the value-added produced by those companies included in the Italian Central Statistical Office's Census (further details, cleaning rules and definitions of variables are in Bontempi and Mairesse [2008]). The total number of observations, more than 717,000, is roughly equally splitted between services and manufacturing companies; the total number of individuals is 73,072, with the availability of minimum 4 years and of maximum 29 years. In order to produce estimation results in line with those of the literature on production function estimates and to preserve the handiness of the empirical framework, we proceed with only the manufacturing companies. We also split the temporal span in two periods, 1982–1993 and 1995-2010, so that we can check the robustness of our findings to changes in the macroeconomic context.[18]

The standard model proposed by the literature is the Cobb-Douglas production function with multiplicative specification of the total capital and constant (but non-unity) elasticity of substitution:

$$Q_{it} = A_i B_t L_{it}^{\beta} C_{it}^{\alpha} K_{it}^{\gamma} e_{it}^{\varepsilon} \tag{22}$$

where Q indicates the value added; the terms A_i and B_t respectively capture efficiency (unmeasurable firm-specific characteristics, like management ability) and the state of technology (the macroeconomic events that affect all companies, like business cycle and "disembodied technical changes" i.e. changes over time in the rates of productivity growth); labels C, K and L are tangible and intangible capital stocks and labour, respectively, with the associated parameters measuring the output elasticity to each input; ε_{it} is the usual idiosyncratic shocks, allowed to be heteroskedastic and within-firm autocorrelated.[19]

By taking the logarithms of equation 22, and defining all the variables per employee, the multiplicative production function specification becomes:

$$(q_{it} - l_{it}) = a_i + b_t + (\mu - 1) l_{it} + \alpha (c_{it} - l_{it}) + \gamma (k_{it} - l_{it}) + \varepsilon_{it} \tag{23}$$

[18]It is worthy to be noted the change of the accounting standards - particularly for the capital stock - following the implementation of the Fourth European Commission Directive since 1993.

[19]Note that we assume a one-period gestation lag before intangible and tangible stocks become fully productive; beginning-of-period capital measures avoid the simultaneous correlation between capital inputs and the disturbance term.

where lower-case letters denote logarithms; a_i and b_t are the usual individual and time effects. Table 22 reports, over the columns, the main statistics of the variables in model 23. In line with the Italian manufacturing division, the data-set is mainly characterized by small and medium-sized firms (with a median number of employees equal to 46 units; about 113 units on average).[20] Input variables are characterized by outliers causing departures of non-parametric measures of spread (inter-quartile range, iqr) from parametric ones (standard deviation, sd). This is particularly evident in intangible capital stock, suggesting that large intangible stocks are concentrated in relatively few companies, and that zeros more prevail here than in the other two inputs. The decomposition of standard deviation in its between, within and residual components shows that the across companies variability prevails, with shares higher than 60% (in line with the findings in Griliches [1988]). Table 23 presents correlations among the variables of equation 23 and tangible and intangible gross investments (inv and $iinv$, respectively); we shall return to this point below, in discussing the role of "internal" (lags of endogenous explanatory variables) and "external" (variables not included in the equation of interest but suggested by the economic structure of the problem at hand) instruments in GMM applications. For now, we note that investments are highly correlated with the endogenous variables of equation 23.

Table 24 presents estimation results for the sub-period 1982-1993. The first three columns report, as benchmarks, pooled OLS estimates (biased by the omission of firm-specific effects, correlated with explanatory variables), and within and first-differences estimates, both accounting for cross-sectional heterogeneity. The first-differences estimates are affected by random year-by-year noise that hides the signal of data (Griliches and Hausman [1986]); its effect is particularly evident in the elasticity of labour, and produces disappointing decreasing returns to scale. The following five columns of Table 24 compare DIF GMM estimates with usual "internal" instruments: it is noticeable the lack of robustness in estimation results accordingly to the different technique used to reduce the number of moment conditions and the rejection of overidentifying restrictions by the Hansen test; PCIVR and PCIVRT produce the best results. Estimates further improve as we move towards the last five columns of the Table, in which "external" instruments are used: particularly in the case of PCIVRT, overidentifying restrictions are not rejected and, at least, elasticities of the output to the capital stocks go in the direction of more sensible results. We prefer the "external" instrument to the "internal" ones, for at least one reason: the lags of the explanatory variables

[20] The average Italian limited liability company employs 44 workers.

may be affected by the same measurement error (possibly correlated over time) that we are trying to tackle. In general, however, the difficulty with DIF GMM estimates is that the past levels of variables are poor instruments for the current differences of the explanatory variables; this even in a large cross-sectional dimension, as in our case, see Bound et al. [1995]. Under covariance stationarity assumptions of the variables in equation 23 we use past differences of investment as ("external") instruments for the levels of productive inputs; accordingly to the above cited literature, LEV GMM - more than DIF GMM - keeps the relevant information in the variables of interest. Results are presented in Table 25 for the two 1982-1993 and 1995-2010 sub-periods. The estimates are encouraging, because robust to changes in the sample periods and in the temporal span, with a non-rejection by the Hansen test that is more evident in the most recent period; moreover, previous disappointing decreasing returns to scale have vanished in favour of constant returns to scale (from an economic point of view, in the first period, or both in economic and statistical terms in the second period).[21]. It is also remarkable the good performance of PCIVRT in an economic context in which the reduced form behind the production function contemplates the possibility of complementarities among productive inputs (which are magnified by the principal components extraction when the instruments and their lags are putted together). Compared to PCIVR, collapsing and lag truncation present worse results: estimated elasticities for some inputs are less in line with not-reduced GMM and PCIVR, and present lower precision. The not-convincing result obtained with lag-depth truncation of the instrument set should be paid a particular attention, as this reduction strategy is commonly adopted in the literature on productivity.

4.2 The application of PCIVR technique to Blundell and Bond [1998] model

In this section we apply our PCIVR technique to the Blundell and Bond [1998] dynamic labour demand equation of p. 135:

$$n_{it} = \alpha n_{it-1} + \beta_0 w_{it} + \beta_1 w_{it-1} + \gamma_0 k_{it} + \gamma_1 k_{it-1} + \phi_t + \eta_i + v_{it} \tag{24}$$

where n_{it} is the log of employment in firm i in year t, w_{it} is the log of the real product wage and k_{it} is the log of the capital stock. The sample is an unbalanced panel of 140 UK listed manufacturing companies with between 7 and 9 annual

[21]These estimates of elasticities of output with respect to inputs are consistent with evidence for other countries obtained by using constrained models - like the total factor productivity approach - to avoid endogeneity and GMM estimating problems

observations over the period 1976-1984. Results are reported in Tables 26, 27 and 28 for DIF, SYS and LEV GMM, respectively; in particular, the first column of Tables 26 and 27 replicate DIF and SYS GMM estimates of the last two columns of Table 4 in Blundell and Bond [1998]. Table 28 adds also LEV GMM estimates. The other columns of Tables 26, 27 and 28 present collapsing (DIFc, SYSc and LEVc), limiting (DIFl, SYSl and LEVl) and PCIVR on each variable separately and on the variables together (DPCIV90, DPCIVT90; SPCIV90, SP-CIVT90; LPCIV90, LPCIVT90). Reported estimates are the one-step GMM ones with standard errors robust to heteroskedasticity. The retain of the scores that are able to explain 90% of the original variance (PCIV90) in DIF GMM makes evident the problem of near unit root characterizing the data at hand: lagged wage is no more significant, and Hansen and residuals second-order autocorrelation tests present lower p-values. This a signal of weak instruments due to persistence that specially affect DIF GMM. These problems are exacerbated by PCIVT: putting together all the instruments and their lags, the PCA operates a sort of reduced form between near unit root stochastic processes and therefore, compared to collapse and lag truncation, casts light on the inappropriateness of the instruments. The overfitting of the model with troublesome moment conditions produces a downwards bias of the estimates (in the direction of Within-Group estimates) and a general increase in the variance. Moving to SYS GMM we note that the weak instruments problem due to persistence is reduced, as suggested by Blundell and Bond. Now PCIV90 delivers estimation results that are in line with original SYS GMM more than the other instrument reduction techniques, like collapsing and lag truncation. Compared to original SYS GMM, however, the reduced number of moment conditions implied by PCIVR reveals the rejection of the orthogonality conditions through the Hansen test. This rejection can be explained by the use of moment conditions in levels for the equation in differences; when we look at the LEV GMM, in which moment conditions in first differences are used for equations in level, we note how estimation results are close each other and that the Hansen test does not reject the overidentifying restrictions (the persistence of instruments is solved by the first-difference transformation).

5 Conclusions

This paper 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 (PCIVR), and compares the alternative instrument reduction techniques through Monte Carlo simulations and empirical applications.

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 both use empirical applications and run Monte Carlo simulations of multivariate DPD model with endogenous variables additional to the lagged dependent one. We found that the extraction of principal components from the instrument matrix tends to improve GMM results when the assumptions under DIF or LEV/SYS GMM are valid.

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 extraction of principal components from 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. Our results confirm that PCIVR is a general, data-driven technique to reduce overidentification problems that can be fruitfully applied to any overidentified GMM problem. We suggest the researcher on always reporting the number of instruments and not to adopt an instrument reduction technique a priori, as every strategy could have serious drawbacks if some assumptions do not hold.

Proper procedures to extract principal components from the instrument matrix have been programmed by the authors in the software Stata: these are based on the preliminary construction of the instrument matrices. This availability could facilitate the researchers in presenting the estimates obtained with alternative GMM estimators with and without data-driven instrument reduction techniques.

Further developments can go in the direction of merging our PCIVR with statistical tests on the validity of the moment conditions. The reduction in the number of overidentifying restrictions should improve the reliability of tests on instruments' validity. In particular, we are going in the direction of multi-step principal components analysis, which involves the identification of "reference" matrices of instruments that enlighten aspects of the data at hand that are problematic for the validity of the instruments; among these, the characteristics of persistence of the instruments.

Table 3: DIFF GMM estimates for α in the main simulation model

		Estimates for α						
α, ρ	Statistics	DIF	DIFc	DIFl	DPCIV90	DPCIVT90	DPCIVA	DPCIVAT
α=0.5, ρ=0.5	mean	0.4801	0.4903	0.4783	0.4684	0.4599	0.4633	0.3695
	sd	0.0374	0.0463	0.0477	0.0627	0.0887	0.0897	0.3822
	p5	0.4212	0.4112	0.3978	0.3607	0.3103	0.3159	-0.3276
	p95	0.5416	0.5682	0.5580	0.5715	0.6033	0.6071	0.9081
	N	1000.0000	1000.0000	1000.0000	1000.0000	1000.0000	1000.0000	1000.0000
α=0.5, ρ=0.95	mean	0.4877	0.4944	0.4758	0.4593	0.1454	0.4605	0.2879
	sd	0.0239	0.0292	0.0392	0.0909	0.5246	0.0892	0.3826
	p5	0.4486	0.4462	0.4115	0.3016	-0.7767	0.3043	-0.4610
	p95	0.5284	0.5420	0.5398	0.5991	0.9603	0.5938	0.7747
	N	1000.0000	1000.0000	1000.0000	1000.0000	1000.0000	1000.0000	1000.0000
α=0.95, ρ=0.5	mean	0.5665	0.4904	0.4236	0.4418	0.3983	0.3958	0.1898
	sd	0.1664	0.3282	0.2381	0.2884	0.3466	0.3652	0.5249
	p5	0.2841	-0.0671	0.0343	-0.0560	-0.1788	-0.2368	-0.6836
	p95	0.8247	0.9737	0.8170	0.8858	0.9526	0.9548	1.0554
	N	1000.0000	1000.0000	1000.0000	1000.0000	1000.0000	1000.0000	1000.0000
α=0.95, ρ=0.95	mean	0.9199	0.9169	0.9108	0.9098	0.8064	0.9103	0.8322
	sd	0.0277	0.0537	0.0434	0.0666	0.2977	0.0661	0.2399
	p5	0.8732	0.8289	0.8366	0.7942	0.2419	0.7968	0.3749
	p95	0.9648	0.9983	0.9759	1.0077	1.1300	1.0074	1.0776
	N	1000.0000	1000.0000	1000.0000	1000.0000	1000.0000	1000.0000	1000.0000

Table 4: DIFF GMM estimates for β in the main simulation model

		Estimates for β						
α, ρ	Statistics	DIF	DIFc	DIFl	DPCIV90	DPCIVT90	DPCIVA	DPCIVAT
α=0.5, ρ=0.5	mean	0.9220	0.9673	0.9164	0.9291	0.9508	0.9382	1.2447
	sd	0.1377	0.1607	0.1670	0.1746	0.1712	0.2006	0.8660
	p5	0.6903	0.7070	0.6455	0.6376	0.6695	0.6043	-0.0483
	p95	1.1455	1.2475	1.1998	1.2312	1.2354	1.2611	2.6352
	N	1000	1000	1000	1000	1000	1000	1000
α=0.5, ρ=0.95	mean	0.8752	0.9302	0.6577	0.7396	0.5578	0.7456	0.5073
	sd	0.2237	0.3181	0.5121	0.6053	1.7291	0.5986	1.4544
	p5	0.5085	0.4232	-0.2143	-0.2876	-2.0675	-0.2607	-1.7762
	p95	1.2291	1.4428	1.4888	1.6574	2.8481	1.6300	2.3972
	N	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.5	mean	0.2579	0.1041	-0.0237	0.0471	0.0450	-0.0104	0.0410
	sd	0.3292	0.6625	0.4700	0.5662	0.6433	0.6871	0.8842
	p5	-0.2882	-0.9890	-0.7960	-0.9375	-0.9975	-1.1856	-1.4220
	p95	0.7719	1.0941	0.7481	0.9307	1.0641	1.0492	1.3833
	N	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.95	mean	0.4807	0.4103	0.3134	0.3171	-0.0065	0.3240	0.0866
	sd	0.4681	0.9409	0.7426	1.1154	2.3012	1.1074	1.8096
	p5	-0.3016	-1.2200	-0.9215	-1.6136	-3.7010	-1.6136	-2.6870
	p95	1.2123	1.8259	1.4483	1.9524	3.5865	1.9521	2.9837
	N	1000	1000	1000	1000	1000	1000	1000

Table 5: Hansen p-value for DIFF estimates in the main simulation model

		Hansen p-value						
α, ρ	**Statistics**	**DIF**	**DIFc**	**DIFl**	**DPCIV90**	**DPCIVT90**	**DPCIVA**	**DPCIVAT**
α=0.5, ρ=0.5	p5	0.0554	0.0562	0.0436	0.0560	0.0589	0.0498	0.0624
	p10	0.1039	0.1008	0.0940	0.1064	0.1030	0.0986	0.1289
	p95	0.9440	0.9568	0.9468	0.9450	0.9389	0.9435	0.9526
	N	1000	1000	1000	1000	1000	1000	1000
α=0.5, ρ=0.95	p5	0.0567	0.0569	0.0485	0.0608	0.0817	0.0618	0.0596
	p10	0.1023	0.1106	0.1074	0.1097	0.1436	0.1095	0.1295
	p95	0.9324	0.9451	0.9467	0.9470	0.9595	0.9505	0.9534
	N	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.5	p5	0.0535	0.0602	0.0519	0.0554	0.0653	0.0581	0.0668
	p10	0.0888	0.1122	0.1035	0.1170	0.1418	0.1146	0.1302
	p95	0.9364	0.9624	0.9448	0.9555	0.9561	0.9642	0.9606
	N	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.95	p5	0.0776	0.0691	0.0622	0.0665	0.1032	0.0663	0.0838
	p10	0.1232	0.1251	0.1351	0.1371	0.1719	0.1389	0.1600
	p95	0.9527	0.9626	0.9646	0.9643	0.9680	0.9633	0.9662
	N	1000	1000	1000	1000	1000	1000	1000

Table 6: LEV GMM estimates for α in the main simulation model

		Estimates for α						
α, ρ	Statistics	LEV	LEVc	LEVl	LPCIV90	LPCIVT90	LPCIVA	LPCIVAT
α=0.5, ρ=0.5	mean	0.5207	0.5066	0.5174	0.5394	0.5321	0.5962	0.5690
	sd	0.0323	0.0315	0.0328	0.0452	0.0436	0.0849	0.0781
	p5	0.4677	0.4527	0.4651	0.4657	0.4602	0.4524	0.4409
	p95	0.5744	0.5577	0.5704	0.6152	0.6046	0.7296	0.6925
	N	1000	1000	1000	1000	1000	1000	1000
α=0.5, ρ=0.95	mean	0.5197	0.5090	0.5199	0.5249	0.5293	0.5296	0.5254
	sd	0.0286	0.0376	0.0285	0.0317	0.0379	0.0532	0.0611
	p5	0.4747	0.4436	0.4703	0.4737	0.4651	0.4425	0.4252
	p95	0.5663	0.5682	0.5662	0.5763	0.5887	0.6154	0.6275
	N	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.5	mean	0.9812	0.9746	0.9806	0.9815	0.9816	0.9824	0.9804
	sd	0.0068	0.0144	0.0085	0.0076	0.0079	0.0101	0.0108
	p5	0.9698	0.9500	0.9660	0.9678	0.9681	0.9654	0.9621
	p95	0.9925	0.9968	0.9941	0.9938	0.9938	0.9985	0.9961
	N	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.95	mean	0.9593	0.9579	0.9590	0.9593	0.9592	0.9585	0.9588
	sd	0.0021	0.0052	0.0028	0.0023	0.0026	0.0047	0.0041
	p5	0.9556	0.9485	0.9545	0.9553	0.9546	0.9505	0.9519
	p95	0.9624	0.9659	0.9633	0.9628	0.9629	0.9643	0.9639
	N	1000	1000	1000	1000	1000	1000	1000

Table 7: LEV GMM estimates for β in the main simulation model

		Estimates for β						
α, ρ	Statistics	LEV	LEVc	LEVl	LPCIV90	LPCIVT90	LPCIVA	LPCIVAT
α=0.5, ρ=0.5	mean	1.0209	1.0074	1.0278	1.0474	1.0278	1.1591	1.0025
	sd	0.1517	0.1527	0.1521	0.1785	0.1844	0.2960	0.2870
	p5	0.7672	0.7596	0.7687	0.7583	0.7253	0.6652	0.5465
	p95	1.2821	1.2634	1.2728	1.3466	1.3360	1.6367	1.4893
	N	1000	1000	1000	1000	1000	1000	1000
α=0.5, ρ=0.95	mean	1.0061	0.9956	1.0074	0.9971	0.9805	0.9792	0.9824
	sd	0.0602	0.0803	0.0597	0.0637	0.0757	0.1086	0.1336
	p5	0.9071	0.8633	0.9086	0.8947	0.8611	0.8006	0.7552
	p95	1.1010	1.1358	1.1051	1.1037	1.1099	1.1605	1.1904
	N	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.5	mean	0.9293	0.9706	0.9461	0.8869	0.8691	0.8081	0.8957
	sd	0.1052	0.1237	0.1138	0.1407	0.1560	0.2280	0.1966
	p5	0.7647	0.7725	0.7641	0.6658	0.6195	0.4292	0.5701
	p95	1.1015	1.1906	1.1307	1.1278	1.1183	1.1799	1.2127
	N	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.95	mean	0.9697	0.9751	0.9719	0.9683	0.9692	0.9696	0.9690
	sd	0.0224	0.0331	0.0245	0.0235	0.0243	0.0365	0.0323
	p5	0.9347	0.9258	0.9339	0.9302	0.9325	0.9233	0.9264
	p95	1.0069	1.0333	1.0132	1.0081	1.0101	1.0282	1.0213
	N	1000	1000	1000	1000	1000	1000	1000

Table 8: Hansen p-value for LEV estimates in the main simulation model

		Hansen p-value						
α, ρ	**Statistics**	**LEV**	**LEVc**	**LEVl**	**LPCIV90**	**LPCIVT90**	**LPCIVA**	**LPCIVAT**
α=0.5, ρ=0.5	p5	0.0335	0.0558	0.0272	0.0185	0.0245	0.0228	0.0173
	p10	0.0732	0.0971	0.0744	0.0427	0.0548	0.0543	0.0453
	p95	0.9189	0.9529	0.9379	0.9184	0.9163	0.9155	0.9244
	N	1000	1000	1000	1000	1000	1000	1000
α=0.5, ρ=0.95	p5	0.0180	0.0590	0.0121	0.0135	0.0185	0.0199	0.0248
	p10	0.0541	0.1157	0.0339	0.0376	0.0401	0.0516	0.0574
	p95	0.9226	0.9617	0.9273	0.9221	0.9187	0.9332	0.9410
	N	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.5	p5	0.0283	0.0359	0.0328	0.0356	0.0299	0.0454	0.0425
	p10	0.0532	0.0739	0.0708	0.0678	0.0616	0.0807	0.0790
	p95	0.9116	0.9616	0.9321	0.9273	0.9177	0.9371	0.9393
	N	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.95	p5	0.0517	0.0513	0.0534	0.0494	0.0450	0.0563	0.0536
	p10	0.0833	0.1166	0.0936	0.0869	0.0919	0.1071	0.1072
	p95	0.9203	0.9616	0.9388	0.9469	0.9411	0.9484	0.9500
	N	1000	1000	1000	1000	1000	1000	1000

Table 9: SYS GMM estimates for α in the main simulation model

		Estimates for α										
α, ρ	Stats	SYS	SYSc	SYSl	SPC90	SPCT90	SPCA	SPCAT	SPCLD90	SPCTLD90	SPCLDA	SPCLDAT
α=0.5, ρ=0.5	mean	0.5098	0.5003	0.5108	0.5089	0.5108	0.5086	0.6165	0.1629	0.1392	0.1412	0.1218
	sd	0.0260	0.0293	0.0279	0.0294	0.0324	0.0322	0.1542	0.0576	0.0578	0.0719	0.0682
	p5	0.4677	0.4528	0.4636	0.4609	0.4568	0.4553	0.3802	0.0665	0.0414	0.0217	0.0023
	p95	0.5542	0.5496	0.5578	0.5557	0.5642	0.5599	0.8790	0.2534	0.2325	0.2516	0.2295
	N	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
α=0.5, ρ=0.95	mean	0.5086	0.4999	0.5094	0.5069	0.5187	0.5059	0.5915	0.1851	–0.1006	0.3307	–0.1565
	sd	0.0159	0.0163	0.0176	0.0215	0.0344	0.0222	0.1134	0.0539	0.1033	0.2727	0.1189
	p5	0.4837	0.4738	0.4816	0.4718	0.4590	0.4704	0.4170	0.0976	–0.2852	0.0463	–0.3660
	p95	0.5352	0.5269	0.5386	0.5426	0.5738	0.5426	0.7751	0.2733	0.0554	0.9444	0.0297
	N	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.5	mean	0.9777	0.9407	0.9785	0.9778	0.9784	0.9775	0.9803	0.9721	0.9817	0.9862	0.9792
	sd	0.0105	0.0824	0.0118	0.0136	0.0153	0.0159	0.0240	0.0198	0.0378	0.0235	0.0475
	p5	0.9599	0.7977	0.9574	0.9545	0.9534	0.9513	0.9425	0.9390	0.9202	0.9477	0.9052
	p95	0.9931	1.0005	0.9965	0.9989	1.0021	1.0023	1.0153	1.0000	1.0466	1.0211	1.0562
	N	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.95	mean	0.9573	0.9519	0.9573	0.9577	0.9586	0.9573	0.9583	0.9609	0.9921	0.9614	0.9931
	sd	0.0036	0.0086	0.0039	0.0043	0.0051	0.0047	0.0080	0.0016	0.0059	0.0022	0.0065
	p5	0.9504	0.9369	0.9500	0.9498	0.9499	0.9489	0.9454	0.9582	0.9834	0.9575	0.9837
	p95	0.9622	0.9634	0.9630	0.9643	0.9664	0.9646	0.9697	0.9636	1.0027	0.9649	1.0041
	N	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table 10: SYS GMM estimates for β in the main simulation model

		Estimates for β										
α, ρ	Stats	SYS	SYSc	SYSl	SPC90	SPCT90	SPCA	SPCAT	SPCLD90	SPCTLD90	SPCLDA	SPCLDAT
α=0.5, ρ=0.5	mean	0.9943	0.9930	1.0063	1.0118	1.0025	1.0091	0.6542	4.9381	5.0962	5.1316	5.2334
	sd	0.1193	0.1267	0.1280	0.1226	0.1307	0.1266	0.5066	0.3769	0.3838	0.4555	0.4490
	p5	0.7944	0.7841	0.7881	0.8154	0.7908	0.8071	-0.2119	4.3434	4.4867	4.4187	4.5424
	p95	1.1895	1.2112	1.2090	1.2108	1.2141	1.2124	1.3921	5.5918	5.7356	5.8731	5.9621
	N	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
α=0.5, ρ=0.95	mean	1.0174	0.9984	1.0173	1.0210	0.9964	1.0232	0.8790	1.9426	2.6256	1.5962	2.7600
	sd	0.0381	0.0526	0.0399	0.0557	0.0686	0.0592	0.1876	0.1302	0.2481	0.6525	0.2858
	p5	0.9545	0.9095	0.9513	0.9307	0.8846	0.9272	0.5834	1.7354	2.2515	0.1320	2.3091
	p95	1.0773	1.0781	1.0800	1.1133	1.1091	1.1240	1.1888	2.1521	3.0648	2.2836	3.2615
	N	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.5	mean	0.9543	0.9627	0.9617	0.9908	0.9750	0.9969	0.9452	1.6570	1.0857	0.8203	1.2108
	sd	0.1010	0.1903	0.1136	0.1136	0.1243	0.1168	0.1673	1.1682	2.2386	1.4194	2.8249
	p5	0.7930	0.6599	0.7741	0.8048	0.7676	0.8058	0.6758	-0.0190	-2.7460	-1.2942	-3.3355
	p95	1.1192	1.2029	1.1490	1.1760	1.1747	1.1890	1.2273	3.5657	4.7187	3.1486	5.5460
	N	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.95	mean	0.9843	0.9891	0.9851	0.9803	0.9752	0.9861	0.9761	0.9372	0.2073	0.9257	0.1843
	sd	0.0263	0.0657	0.0276	0.0313	0.0339	0.0388	0.0525	0.0375	0.1399	0.0517	0.1521
	p5	0.9457	0.9168	0.9460	0.9354	0.9261	0.9267	0.8950	0.8744	-0.0354	0.8456	-0.0789
	p95	1.0311	1.0695	1.0360	1.0336	1.0285	1.0541	1.0586	0.9977	0.4154	1.0135	0.4040
	N	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table 11: Hansen p-value for SYS estimates in the main simulation model

		Hansen p-value										
α, ρ	Stats	SYS	SYSc	SYSl	SPC90	SPCT90	SPCA	SPCAT	SPCLD90	SPCTLD90	SPCLDA	SPCLDAT
α=0.5, ρ=0.5	p5	0.0551	0.0575	0.0497	0.0478	0.0519	0.0470	0.0479	0.0014	0.0073	0.0122	0.0178
	p10	0.0958	0.1209	0.0943	0.1021	0.1097	0.0951	0.0942	0.0067	0.0196	0.0316	0.0390
	p95	0.9173	0.9541	0.9428	0.9502	0.9397	0.9482	0.9553	0.7834	0.8940	0.8982	0.9185
	N	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
α=0.5, ρ=0.95	p5	0.0250	0.0605	0.0248	0.0158	0.0190	0.0225	0.0360	0.0000	0.0003	0.0000	0.0012
	p10	0.0548	0.1151	0.0608	0.0495	0.0435	0.0483	0.0708	0.0000	0.0014	0.0000	0.0040
	p95	0.9154	0.9480	0.9310	0.9354	0.9396	0.9293	0.9412	0.0000	0.6642	0.0000	0.8645
	N	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.5	p5	0.0370	0.0612	0.0398	0.0443	0.0426	0.0435	0.0571	0.0000	0.0000	0.0000	0.0000
	p10	0.0763	0.1204	0.0781	0.0828	0.0894	0.0852	0.1098	0.0000	0.0000	0.0000	0.0000
	p95	0.9204	0.9559	0.9387	0.9379	0.9391	0.9409	0.9589	0.0000	0.0000	0.0643	0.0010
	N	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.95	p5	0.0448	0.0507	0.0455	0.0383	0.0562	0.0418	0.0554	0.0000	0.0000	0.0000	0.0000
	p10	0.0948	0.1088	0.0859	0.0897	0.1197	0.1005	0.1176	0.0000	0.0000	0.0000	0.0000
	p95	0.9365	0.9489	0.9422	0.9419	0.9526	0.9433	0.9542	0.0000	0.2001	0.2189	0.3966
	N	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table 12: DIFF GMM estimates for α with $\sigma_e^2 = 1$

		Estimates for α						
α, ρ	Statistics	DIF	DIFc	DIFl	DPCIV90	DPCIVT90	DPCIVA	DPCIVAT
α=0.5, ρ=0.5	mean	0.4922	0.4967	0.4917	0.4906	0.4921	0.4906	0.4878
	sd	0.0289	0.0345	0.0345	0.0386	0.0401	0.0607	0.0898
	p5	0.4440	0.4388	0.4356	0.4284	0.4266	0.3893	0.3364
	p95	0.5400	0.5525	0.5478	0.5532	0.5563	0.5869	0.6285
	N	1000	1000	1000	1000	1000	1000	1000
α=0.5, ρ=0.95	mean	0.4962	0.4987	0.4957	0.4964	0.4775	0.4968	0.4955
	sd	0.0140	0.0161	0.0180	0.0412	0.1429	0.0380	0.0585
	p5	0.4734	0.4723	0.4676	0.4272	0.2716	0.4325	0.4046
	p95	0.5179	0.5229	0.5253	0.5644	0.6650	0.5595	0.5884
	N	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.5	mean	0.8248	0.8355	0.7640	0.8206	0.8202	0.8022	0.7946
	sd	0.0886	0.1632	0.1463	0.1299	0.1422	0.1841	0.1930
	p5	0.6685	0.5149	0.5134	0.5855	0.5688	0.4842	0.4459
	p95	0.9631	1.0742	0.9891	1.0252	1.0297	1.0814	1.0769
	N	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.95	mean	0.9450	0.9444	0.9429	0.9432	0.9411	0.9434	0.9411
	sd	0.0109	0.0229	0.0172	0.0290	0.0564	0.0281	0.0564
	p5	0.9266	0.9054	0.9150	0.8972	0.8552	0.8981	0.8552
	p95	0.9630	0.9794	0.9701	0.9889	1.0136	0.9888	1.0136
	N	1000	1000	1000	1000	1000	1000	1000

Table 13: DIFF GMM estimates for β with $\sigma_e^2 = 1$

		Estimates for β						
α, ρ	Statistics	DIF	DIFc	DIFl	DPCIV90	DPCIVT90	DPCIVA	DPCIVAT
α=0.5, ρ=0.5	mean	0.9962	0.9999	0.9917	1.0012	1.0015	1.0060	1.0154
	sd	0.0518	0.0560	0.0561	0.0523	0.0537	0.0755	0.1298
	p5	0.9119	0.9094	0.8994	0.9183	0.9161	0.8792	0.8040
	p95	1.0816	1.0941	1.0842	1.0860	1.0880	1.1245	1.2347
	N	1000	1000	1000	1000	1000	1000	1000
α=0.5, ρ=0.95	mean	0.9747	0.9907	0.9608	0.9879	0.9676	0.9889	0.9861
	sd	0.0907	0.1132	0.1449	0.1578	0.4666	0.1527	0.3511
	p5	0.8216	0.7982	0.7304	0.7361	0.3650	0.7484	0.4834
	p95	1.1123	1.1706	1.1884	1.2279	1.6022	1.2219	1.4874
	N	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.5	mean	0.8003	0.8093	0.6853	0.7914	0.7918	0.7718	0.7720
	sd	0.1485	0.2818	0.2512	0.2214	0.2404	0.2973	0.3017
	p5	0.5440	0.2645	0.2669	0.3966	0.3556	0.2462	0.2247
	p95	1.0375	1.2007	1.0709	1.1252	1.1290	1.1951	1.1988
	N	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.95	mean	0.9150	0.9017	0.8796	0.8829	0.8890	0.8858	0.8890
	sd	0.1813	0.3938	0.2929	0.4859	0.8423	0.4740	0.8423
	p5	0.6086	0.2532	0.4001	0.1128	-0.2909	0.1315	-0.2909
	p95	1.2022	1.4916	1.3489	1.6357	2.1791	1.6296	2.1791
	N	1000	1000	1000	1000	1000	1000	1000

Table 14: Hansen p-value for DIFF estimates with $\sigma_e^2 = 1$

		Hansen p-value						
α, ρ	Statistics	DIF	DIFc	DIFl	DPCIV90	DPCIVT90	DPCIVA	DPCIVAT
α=0.5, ρ=0.5	p5	0.0515	0.0617	0.0411	0.0544	0.0594	0.0530	0.0536
	p10	0.1050	0.1033	0.0906	0.0997	0.0985	0.0911	0.1003
	p95	0.9398	0.9378	0.9460	0.9426	0.9479	0.9310	0.9470
	N	1000	1000	1000	1000	1000	1000	1000
α=0.5, ρ=0.95	p5	0.0446	0.0521	0.0433	0.0510	0.0581	0.0502	0.0556
	p10	0.1034	0.1016	0.1063	0.0945	0.1193	0.0990	0.1041
	p95	0.9391	0.9367	0.9463	0.9375	0.9450	0.9340	0.9385
	N	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.5	p5	0.0536	0.0525	0.0635	0.0528	0.0492	0.0472	0.0464
	p10	0.0927	0.1077	0.1221	0.0928	0.1011	0.0964	0.1031
	p95	0.9373	0.9537	0.9595	0.9557	0.9524	0.9486	0.9601
	N	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.95	p5	0.0579	0.0702	0.0655	0.0709	0.0853	0.0683	0.0853
	p10	0.1124	0.1288	0.1331	0.1262	0.1498	0.1311	0.1498
	p95	0.9523	0.9613	0.9605	0.9536	0.9639	0.9536	0.9639
	N	1000	1000	1000	1000	1000	1000	1000

Table 15: LEV GMM estimates for α with $\sigma_e^2 = 1$

		Estimates for α						
α, ρ	Statistics	LEV	LEVc	LEVl	LPCIV90	LPCIVT90	LPCIVA	LPCIVAT
α=0.5, ρ=0.5	mean	0.5108	0.5040	0.5101	0.5202	0.5160	0.5566	0.5220
	sd	0.0275	0.0288	0.0276	0.0339	0.0379	0.0716	0.0548
	p5	0.4668	0.4547	0.4621	0.4633	0.4535	0.4373	0.4341
	p95	0.5564	0.5505	0.5550	0.5750	0.5754	0.6752	0.6184
	N	1000	1000	1000	1000	1000	1000	1000
α=0.5, ρ=0.95	mean	0.5084	0.5037	0.5089	0.5085	0.5078	0.5056	0.5057
	sd	0.0164	0.0216	0.0180	0.0166	0.0178	0.0232	0.0191
	p5	0.4821	0.4682	0.4786	0.4811	0.4789	0.4680	0.4747
	p95	0.5348	0.5395	0.5385	0.5355	0.5377	0.5441	0.5363
	N	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.5	mean	0.9729	0.9623	0.9712	0.9717	0.9693	0.9672	0.9657
	sd	0.0075	0.0113	0.0087	0.0082	0.0095	0.0106	0.0109
	p5	0.9603	0.9429	0.9560	0.9577	0.9534	0.9492	0.9475
	p95	0.9842	0.9771	0.9852	0.9846	0.9841	0.9832	0.9822
	N	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.95	mean	0.9575	0.9535	0.9562	0.9563	0.9560	0.9550	0.9553
	sd	0.0021	0.0040	0.0029	0.0029	0.0031	0.0042	0.0035
	p5	0.9537	0.9463	0.9510	0.9508	0.9504	0.9470	0.9489
	p95	0.9607	0.9599	0.9603	0.9605	0.9606	0.9609	0.9605
	N	1000	1000	1000	1000	1000	1000	1000

Table 16: LEV GMM estimates for β with $\sigma_e^2 = 1$

		Estimates for β						
α, ρ	Statistics	LEV	LEVc	LEVl	LPCIV90	LPCIVT90	LPCIVA	LPCIVAT
α=0.5, ρ=0.5	mean	0.9846	0.9919	0.9835	0.9720	0.9650	0.9688	0.9579
	sd	0.0732	0.0804	0.0737	0.0843	0.1001	0.1194	0.1219
	p5	0.8589	0.8625	0.8610	0.8337	0.7997	0.7702	0.7467
	p95	1.1031	1.1273	1.1051	1.1113	1.1189	1.1680	1.1493
	N	1000	1000	1000	1000	1000	1000	1000
α=0.5, ρ=0.95	mean	0.9931	0.9950	0.9920	0.9917	0.9910	0.9928	0.9931
	sd	0.0306	0.0386	0.0307	0.0309	0.0327	0.0444	0.0364
	p5	0.9428	0.9296	0.9419	0.9396	0.9380	0.9198	0.9339
	p95	1.0433	1.0575	1.0409	1.0424	1.0436	1.0652	1.0564
	N	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.5	mean	0.9419	0.9751	0.9467	0.9367	0.9441	0.9396	0.9563
	sd	0.0509	0.0555	0.0497	0.0579	0.0580	0.0731	0.0595
	p5	0.8581	0.8886	0.8676	0.8402	0.8491	0.8216	0.8585
	p95	1.0283	1.0656	1.0312	1.0292	1.0378	1.0584	1.0511
	N	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.95	mean	0.9778	0.9899	0.9820	0.9813	0.9817	0.9835	0.9838
	sd	0.0118	0.0181	0.0138	0.0144	0.0153	0.0206	0.0169
	p5	0.9607	0.9629	0.9636	0.9613	0.9596	0.9555	0.9599
	p95	0.9994	1.0236	1.0085	1.0066	1.0101	1.0226	1.0153
	N	1000	1000	1000	1000	1000	1000	1000

Table 17: Hansen p-value for LEV estimates with $\sigma_e^2 = 1$

		Hansen p-value						
α, ρ	**Statistics**	**LEV**	**LEVc**	**LEVl**	**LPCIV90**	**LPCIVT90**	**LPCIVA**	**LPCIVAT**
α=0.5, ρ=0.5	p5	0.0409	0.0534	0.0455	0.0289	0.0406	0.0257	0.0401
	p10	0.0839	0.1030	0.0896	0.0602	0.1001	0.0574	0.0751
	p95	0.9453	0.9519	0.9433	0.9369	0.9480	0.9201	0.9483
	N	1000	1000	1000	1000	1000	1000	1000
α=0.5, ρ=0.95	p5	0.0367	0.0449	0.0358	0.0409	0.0434	0.0487	0.0513
	p10	0.0694	0.1038	0.0720	0.0755	0.0829	0.0904	0.1004
	p95	0.9372	0.9443	0.9324	0.9329	0.9445	0.9502	0.9450
	N	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.5	p5	0.0063	0.0149	0.0084	0.0084	0.0101	0.0162	0.0132
	p10	0.0176	0.0537	0.0262	0.0288	0.0274	0.0342	0.0334
	p95	0.8840	0.9452	0.8899	0.8958	0.9249	0.9097	0.9350
	N	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.95	p5	0.0171	0.0236	0.0160	0.0184	0.0207	0.0222	0.0211
	p10	0.0353	0.0647	0.0410	0.0487	0.0481	0.0540	0.0518
	p95	0.8974	0.9379	0.9079	0.9177	0.9141	0.9365	0.9320
	N	1000	1000	1000	1000	1000	1000	1000

Table 18: SYS GMM estimates for α with $\sigma_e^2 = 1$

		Estimates for α										
α, ρ	Stats	SYS	SYSc	SYSl	SPC90	SPCT90	SPCA	SPCAT	SPCLD90	SPCTLD90	SPCLDA	SPCLDAT
α=0.5, ρ=0.5	mean	0.5049	0.5000	0.5049	0.5043	0.5053	0.5053	0.5181	0.3758	0.3519	0.3528	0.4515
	sd	0.0197	0.0224	0.0210	0.0219	0.0231	0.0253	0.0574	0.0539	0.0555	0.0725	0.0836
	p5	0.4732	0.4644	0.4692	0.4686	0.4680	0.4648	0.4189	0.2836	0.2529	0.2239	0.3156
	p95	0.5373	0.5362	0.5386	0.5399	0.5439	0.5458	0.6059	0.4569	0.4342	0.4624	0.5821
	N	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
α=0.5, ρ=0.95	mean	0.5017	0.4999	0.5017	0.5034	0.5076	0.5028	0.5079	0.6882	0.7134	0.9269	0.7090
	sd	0.0077	0.0077	0.0082	0.0140	0.0223	0.0133	0.0265	0.1033	0.1261	0.1149	0.1242
	p5	0.4889	0.4872	0.4883	0.4807	0.4695	0.4821	0.4644	0.5083	0.4993	0.7169	0.4977
	p95	0.5143	0.5124	0.5149	0.5262	0.5416	0.5244	0.5515	0.8411	0.9134	1.0802	0.9027
	N	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.5	mean	0.9655	0.9492	0.9658	0.9654	0.9638	0.9648	0.9671	0.9893	1.0034	0.9916	1.0039
	sd	0.0100	0.0207	0.0108	0.0107	0.0119	0.0126	0.0161	0.0069	0.0069	0.0107	0.0073
	p5	0.9483	0.9167	0.9467	0.9463	0.9433	0.9442	0.9397	0.9775	0.9919	0.9755	0.9915
	p95	0.9794	0.9727	0.9809	0.9806	0.9804	0.9824	0.9908	1.0003	1.0139	1.0088	1.0155
	N	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.95	mean	0.9532	0.9505	0.9532	0.9538	0.9545	0.9536	0.9552	0.9607	0.9639	0.9608	0.9640
	sd	0.0030	0.0032	0.0030	0.0034	0.0039	0.0035	0.0054	0.0006	0.0007	0.0007	0.0007
	p5	0.9480	0.9452	0.9478	0.9480	0.9476	0.9475	0.9450	0.9597	0.9627	0.9596	0.9629
	p95	0.9579	0.9558	0.9577	0.9592	0.9601	0.9592	0.9630	0.9616	0.9651	0.9620	0.9652
	N	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table 19: SYS GMM estimates for β with $\sigma_e^2 = 1$

		Estimates for β										
α, ρ	Stats	SYS	SYSc	SYSl	SPC90	SPCT90	SPCA	SPCAT	SPCLD90	SPCTLD90	SPCLDA	SPCLDAT
α=0.5, ρ=0.5	mean	0.9942	1.0003	0.9942	0.9973	0.9951	0.9935	0.9707	3.2693	3.4444	3.6717	2.8138
	sd	0.0496	0.0550	0.0549	0.0501	0.0512	0.0591	0.1096	0.3687	0.3852	0.4493	0.5777
	p5	0.9118	0.9125	0.9037	0.9194	0.9126	0.8963	0.7977	2.6760	2.8469	3.0050	1.8756
	p95	1.0752	1.0899	1.0830	1.0790	1.0772	1.0938	1.1571	3.8819	4.1166	4.4749	3.7535
	N	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
α=0.5, ρ=0.95	mean	1.0011	0.9998	1.0011	0.9976	0.9902	0.9990	0.9914	0.7119	0.6559	0.1536	0.6661
	sd	0.0180	0.0185	0.0183	0.0275	0.0360	0.0267	0.0374	0.2393	0.2940	0.2695	0.2896
	p5	0.9682	0.9669	0.9683	0.9524	0.9333	0.9552	0.9279	0.3578	0.1928	–0.2077	0.2109
	p95	1.0299	1.0289	1.0303	1.0415	1.0503	1.0414	1.0531	1.1231	1.1560	0.6469	1.1585
	N	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.5	mean	0.9761	0.9929	0.9744	0.9841	0.9848	0.9852	0.9795	0.5984	–0.2181	0.4710	–0.2853
	sd	0.0458	0.0502	0.0485	0.0446	0.0444	0.0489	0.0493	0.3850	0.3929	0.6424	0.4126
	p5	0.8979	0.9141	0.8934	0.9116	0.9120	0.9033	0.9021	–0.0239	–0.8103	–0.5893	–0.9119
	p95	1.0486	1.0707	1.0524	1.0557	1.0589	1.0656	1.0635	1.2667	0.3883	1.4465	0.3977
	N	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.95	mean	0.9939	0.9979	0.9940	0.9900	0.9875	0.9933	0.9875	0.9424	0.8763	0.9396	0.8741
	sd	0.0148	0.0171	0.0147	0.0163	0.0164	0.0187	0.0189	0.0127	0.0154	0.0162	0.0154
	p5	0.9716	0.9701	0.9715	0.9656	0.9643	0.9658	0.9623	0.9215	0.8501	0.9121	0.8479
	p95	1.0198	1.0278	1.0197	1.0186	1.0187	1.0262	1.0236	0.9633	0.9004	0.9661	0.8989
	N	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Table 20: Hansen p-value for SYS estimates with $\sigma_e^2 = 1$

		Hansen p-value										
α, ρ	Stats	SYS	SYSc	SYSl	SPC90	SPCT90	SPCA	SPCAT	SPCLD90	SPCTLD90	SPCLDA	SPCLDAT
α=0.5, ρ=0.5	p5	0.0458	0.0564	0.0502	0.0523	0.0507	0.0582	0.0403	0.0000	0.0000	0.0001	0.0000
	p10	0.0879	0.1041	0.0908	0.0897	0.1098	0.1104	0.0803	0.0000	0.0000	0.0005	0.0000
	p95	0.9376	0.9467	0.9502	0.9311	0.9402	0.9372	0.9524	0.3787	0.6250	0.8357	0.6963
	N	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
α=0.5, ρ=0.95	p5	0.0476	0.0488	0.0502	0.0398	0.0311	0.0428	0.0349	0.0000	0.0000	0.0000	0.0000
	p10	0.0889	0.1053	0.0931	0.0860	0.0636	0.0874	0.0738	0.0000	0.0000	0.0000	0.0000
	p95	0.9313	0.9425	0.9343	0.9335	0.9388	0.9396	0.9329	0.0000	0.0011	0.0002	0.0010
	N	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.5	p5	0.0206	0.0569	0.0215	0.0190	0.0275	0.0275	0.0327	0.0000	0.0000	0.0000	0.0000
	p10	0.0543	0.0957	0.0482	0.0554	0.0600	0.0577	0.0722	0.0000	0.0000	0.0000	0.0000
	p95	0.9160	0.9514	0.9240	0.9231	0.9130	0.9265	0.9436	0.0000	0.1167	0.2976	0.2326
	N	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
α=0.95, ρ=0.95	p5	0.0455	0.0468	0.0457	0.0378	0.0274	0.0364	0.0391	0.0000	0.0014	0.0000	0.0013
	p10	0.0940	0.1045	0.0856	0.0786	0.0552	0.0842	0.0922	0.0000	0.0052	0.0000	0.0046
	p95	0.9325	0.9410	0.9426	0.9399	0.9353	0.9329	0.9404	0.0000	0.7436	0.5728	0.7274
	N	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

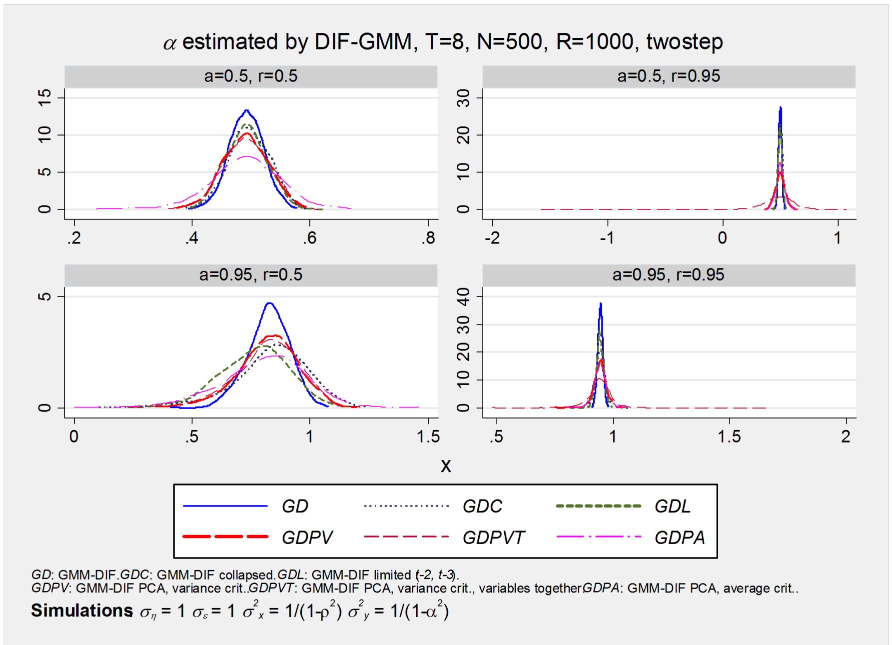

Figure 1: Figure 1

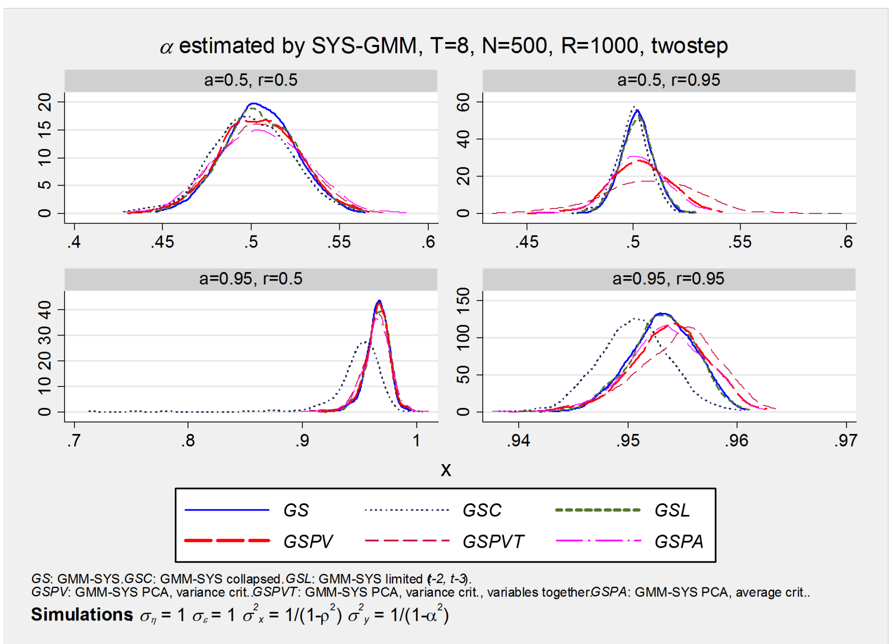

Figure 2: Figure 2

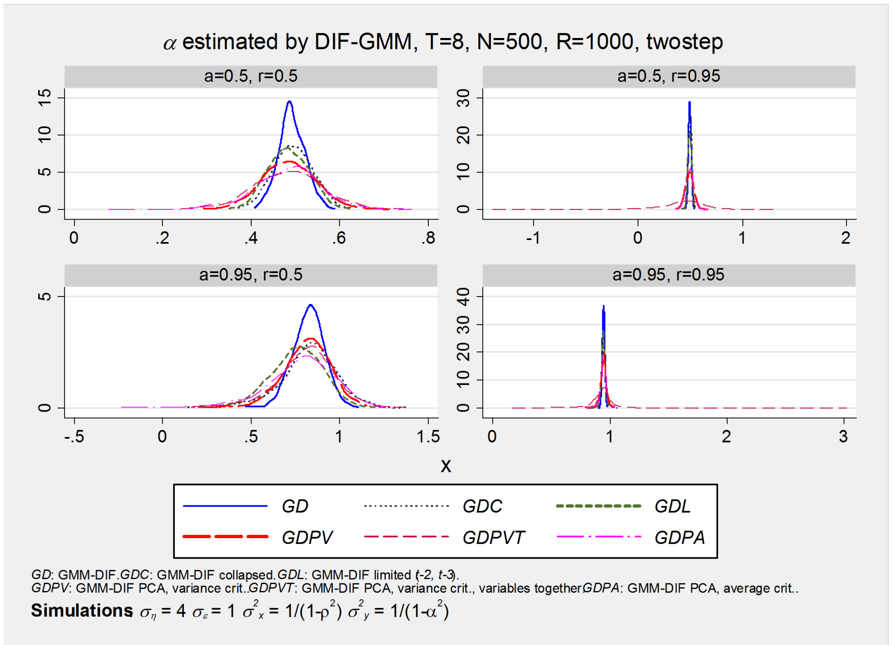

Figure 3: Figure 3

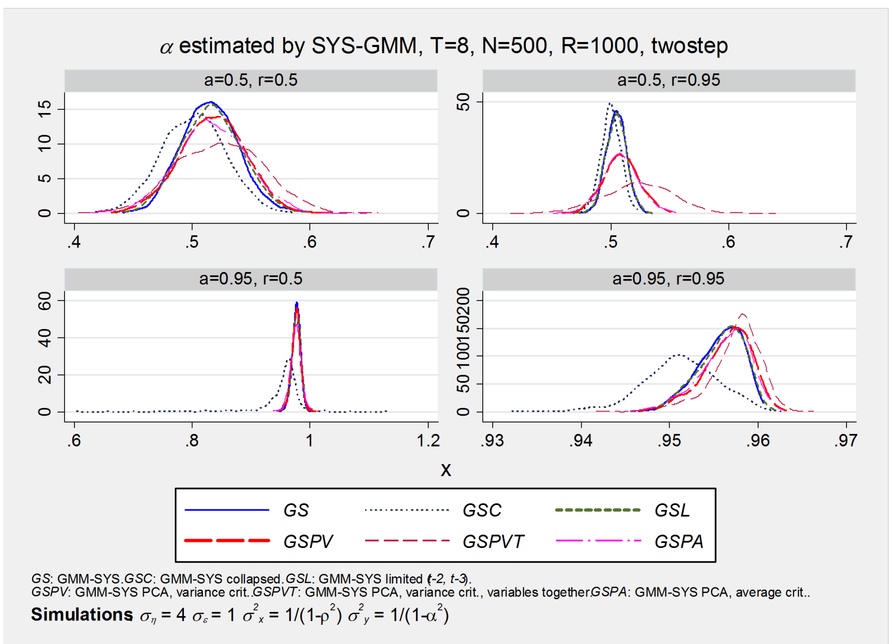

Figure 4: Figure 4

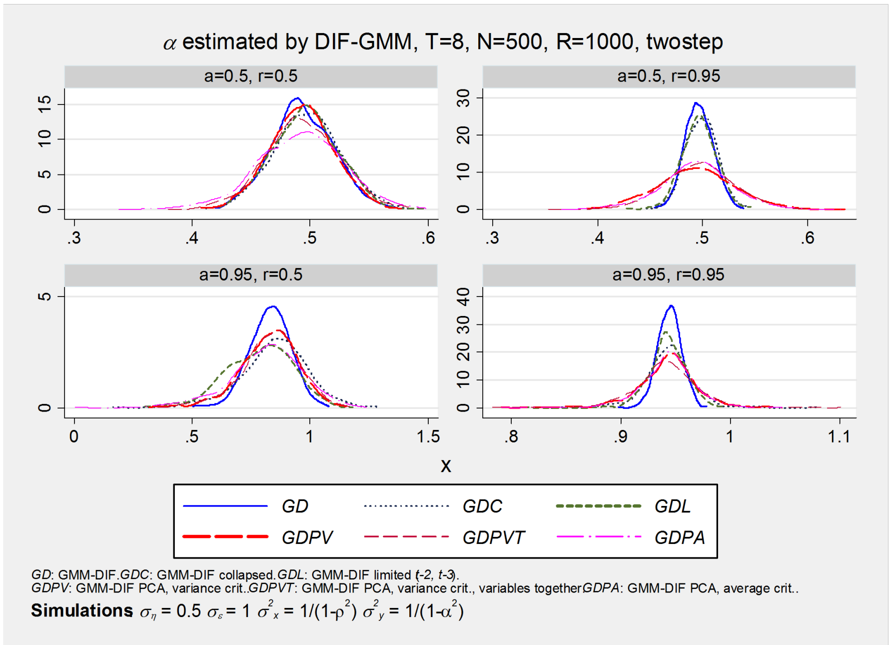

Figure 5: Figure 5

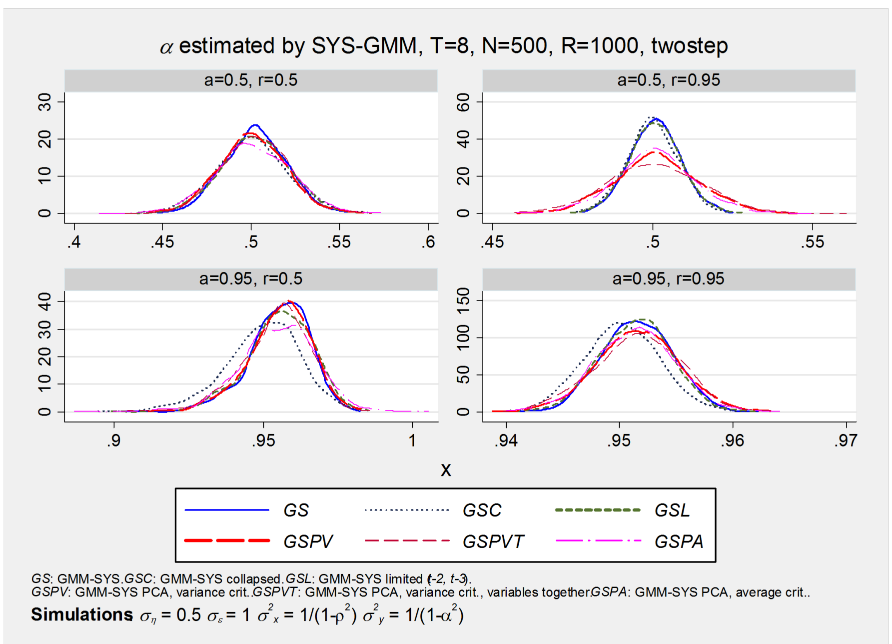

Figure 6: Figure 6

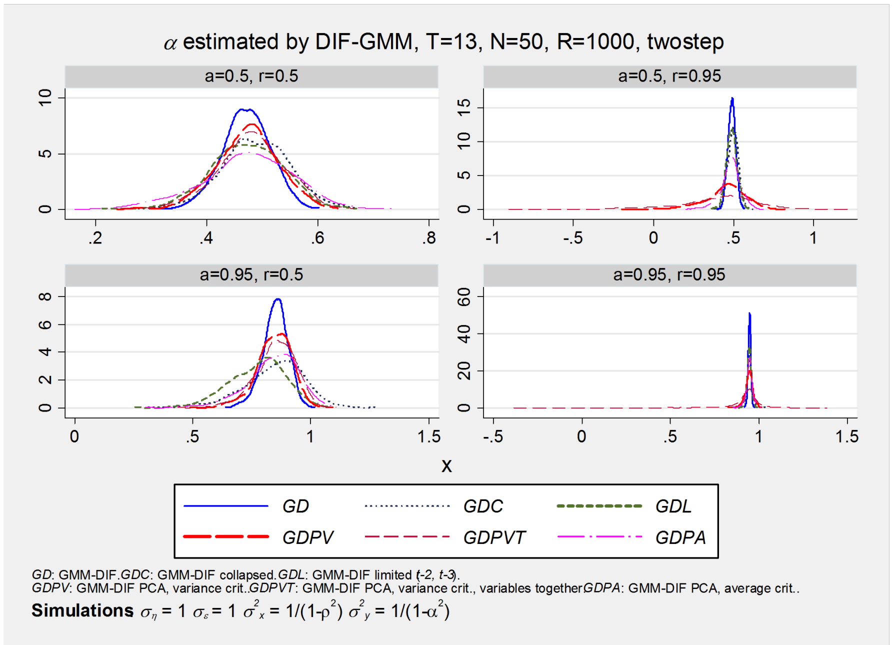

Figure 7: Figure 7

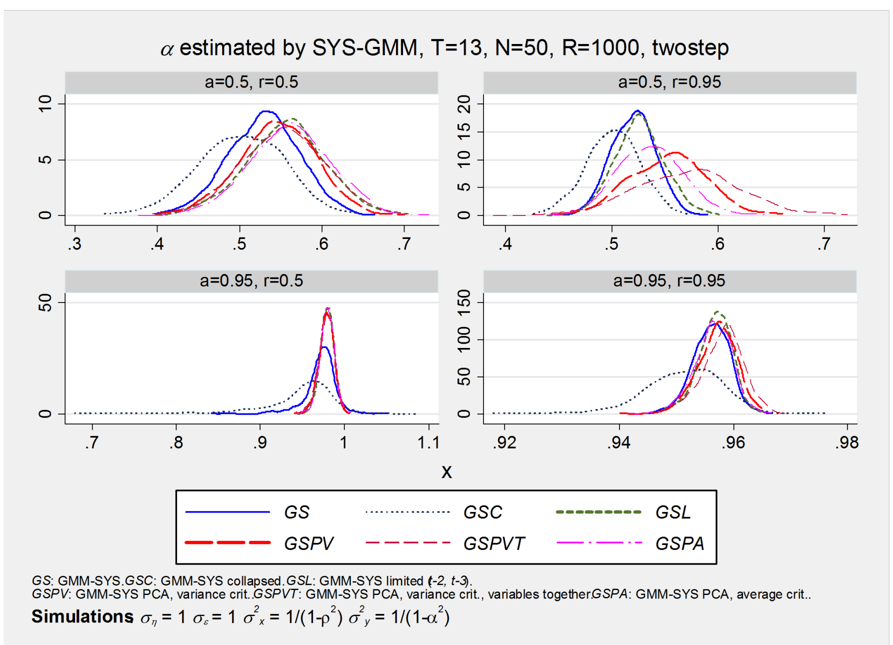

Figure 8: Figure 8

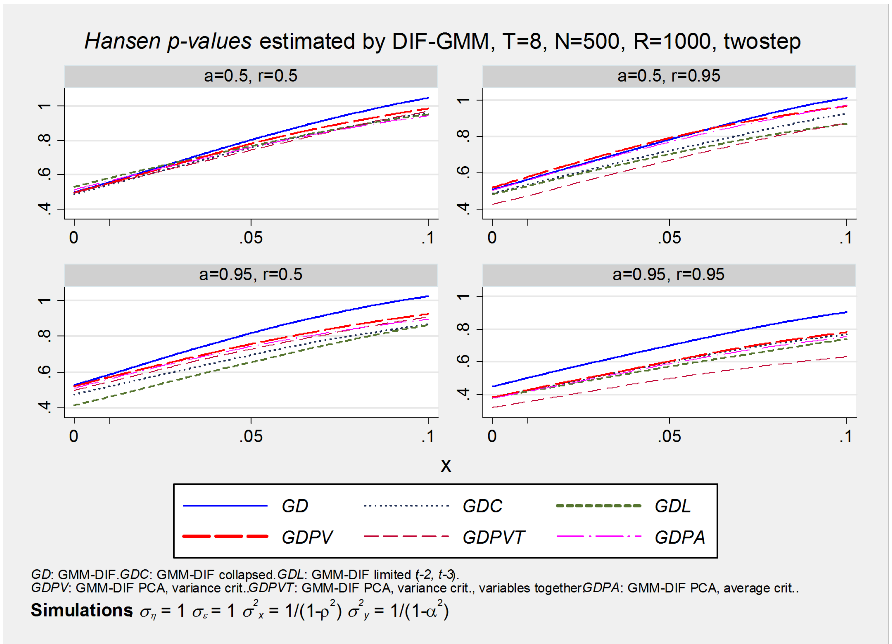

Figure 9: Figure 9

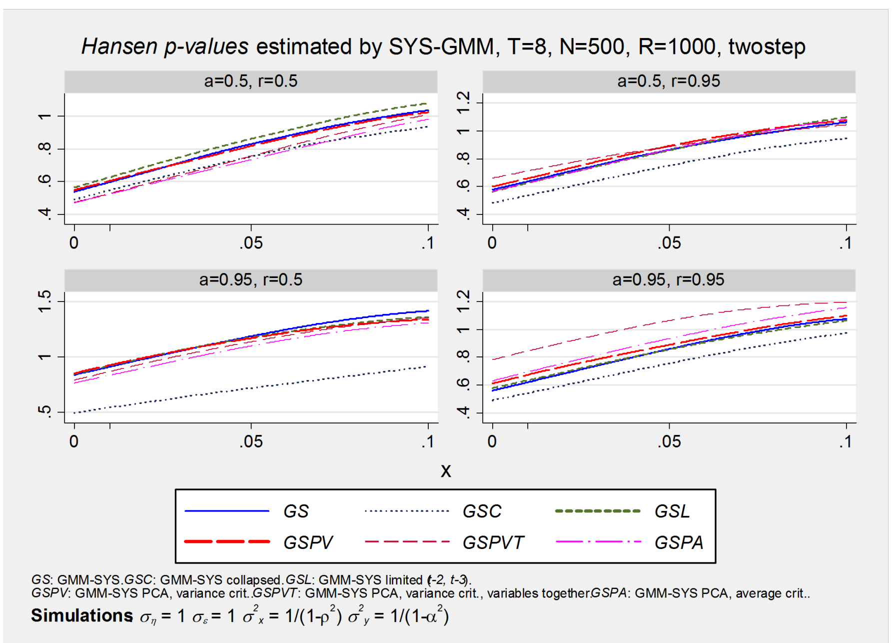

Figure 10: Figure 10

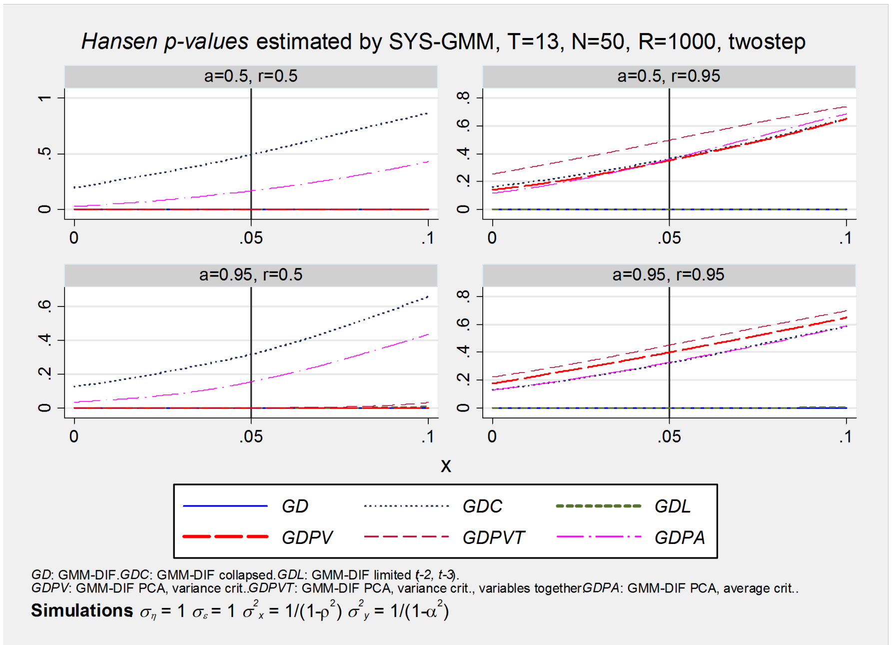

Figure 11: Figure 11

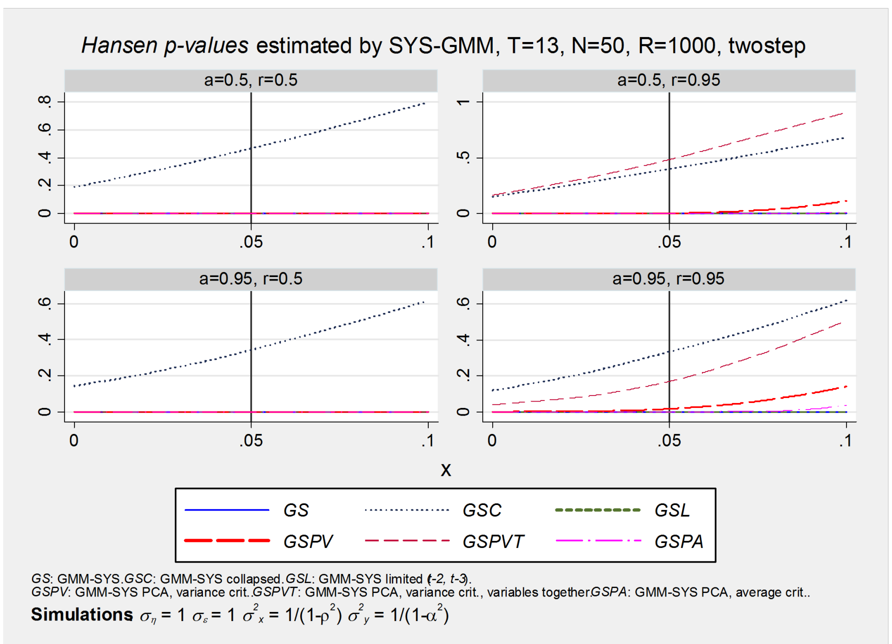

Figure 12: Grafico12

Table 21: Production function: sample size

Year	*Serv.*	*Manuf.*	Total	**Year**	*Serv.*	*Manuf.*	Total
1982	5,146	10,122	15,268	1997	14,075	15,749	29,824
1983	5,101	9,553	14,654	1998	13,786	15,398	29,184
1984	6,371	11,421	17,792	1999	14,251	15,532	29,783
1985	7,286	12,288	19,574	2000	14,394	15,331	29,725
1986	8,084	12,999	21,083	2001	14,138	14,456	28,594
1987	8,490	13,225	21,715	2002	13,276	13,716	26,992
1988	9,044	13,420	22,464	2003	16,469	16,173	32,642
1989	9,922	14,053	23,975	2004	16,875	16,365	33,240
1990	10,563	14,546	25,109	2005	15,929	14,824	30,753
1991	10,421	14,389	24,810	2006	15,088	13,676	28,764
1992	10,328	14,268	24,596	2007	14,115	12,709	26,824
1993	9,275	12,155	21,430	2008	13,226	12,136	25,362
1994	13,216	14,259	27,475	2009	11,958	11,179	23,137
1995	11,198	12,864	24,062	2010	10,529	10,081	20,610
1996	8,111	9,966	18,077	**Total**	**330,665**	**386,853**	**717,518**

Table 22: Production function: statistics

	mean	*p50*	*sd*	*iqr*	*between*	*within*	*residual*	*N*	$\overline{T}$
q_l	3.797	3.791	0.53	0.593	60.34	2.78	36.88	386853	10.13
c_l	3.458	3.488	1.032	1.294	79.54	3.66	16.8	284433	7.54
k_l	0.215	0.246	1.537	1.931	67.17	0.34	32.49	284433	7.54
l	3.908	3.829	1.06	1.242	91.59	0.62	7.8	386853	10.13

Table 23: Production function: pairwise correlations

	q_l	c_l	k_l	l	*inv*	*iinv*
q_l	1					
c_l	0.3612*	1				
k_l	0.1622*	0.0941*	1			
l	-0.0978*	-0.0687*	-0.0626*	1		
inv	0.1428*	0.3281*	0.0479*	-0.0760*	1	
iinv	0.1114*	0.0311*	0.3425*	-0.0316*	0.1117*	1

Table 24: Production function: benchmark and DIF GMM estimates 1982–1993

				Internal IVs					External IVs				
Var.	OLS	WI	FD	DIF	DIFc	DIFl	DPCIV90	DPCIVT90	DIF	DIFc	DIFl	DPCIV90	DPCIVT90
c_l	0.153	0.104	0.076	0.053	0.103	0.007	0.138	0.146	0.060	0.067	0.028	0.070	0.092
se	0.003	0.004	0.004	0.031	0.039	0.037	0.037	0.059	0.052	0.049	0.052	0.050	0.052
t	51.0	26.7	18.1	1.7	2.6	0.2	3.8	2.5	1.2	1.4	0.5	1.4	1.8
k_l	0.032	0.004	0.006	0.047	–0.132	0.008	–0.002	–0.009	0.014	0.011	0.010	0.013	0.012
se	0.002	0.002	0.002	0.041	0.098	0.056	0.050	0.065	0.006	0.006	0.006	0.006	0.006
t	18.8	2.4	2.7	1.2	–1.3	0.1	0.0	–0.1	2.2	1.8	1.6	2.1	1.9
l	–0.027	–0.212	–0.548	–0.435	–0.696	–0.640	–0.399	–0.378	–0.682	–0.603	–0.754	–0.667	–0.631
se	0.003	0.008	0.009	0.079	0.178	0.106	0.093	0.115	0.144	0.176	0.168	0.146	0.149
t	–10.8	–26.5	–60.2	–5.5	–3.9	–6.0	–4.3	–3.3	–4.7	–3.4	–4.5	–4.6	–4.2
H	–	–	–	211.7	63.9	113.6	108.0	92.9	115.5	24.3	55.7	70.2	82.6
Hp	–	–	–	0.000	0.000	0.000	0.001	0.001	0.031	0.110	0.008	0.537	0.126
Hdf	–	–	–	142	25	50	66	53	89	17	33	72	69
N	109738	109738	79519	79519	79519	79519	79519	79519	79519	79519	79519	79519	79519
$\overline{T}$	5.07	5.07	4.16	4.16	4.16	4.16	4.16	4.16	4.16	4.16	4.16	4.16	4.16

Table 25: Production function: LEV GMM estimates with external IVs

	1982–1993					1995–2010				
Var.	LEV	LEVc	LEVl	LPCIV90	LPCIVT90	LEV	LEVc	LEVl	LPCIV90	LPCIVT90
c_l	0.207	0.208	0.254	0.210	0.214	0.226	0.244	0.231	0.227	0.214
se	0.016	0.017	0.019	0.016	0.016	0.037	0.064	0.062	0.045	0.041
t	13.2	12.2	13.2	13.2	13.5	6.1	3.8	3.8	5.1	5.2
k_l	0.041	0.040	0.041	0.040	0.038	0.031	0.030	0.032	0.032	0.028
se	0.005	0.005	0.005	0.005	0.005	0.011	0.019	0.018	0.013	0.012
t	8.7	8.9	8.2	8.5	8.3	2.8	1.6	1.8	2.4	2.3
l	0.034	0.036	0.041	0.034	0.039	0.025	0.026	0.019	0.022	0.033
se	0.016	0.014	0.016	0.016	0.016	0.028	0.049	0.046	0.033	0.031
t	2.2	2.5	2.6	2.2	2.5	0.9	0.5	0.4	0.7	1.1
H	118.7	38.4	42.5	106.9	97.8	198.3	32.9	67.2	157.0	155.6
Hp	0.019	0.002	0.124	0.014	0.034	0.180	0.132	0.043	0.154	0.145
Hdf	89	17	33	77	74	181	25	49	140	138
N	109738	109738	109738	109738	109738	156241	156241	156241	156241	156241
$\overline{T}$	5.07	5.07	5.07	5.07	5.07	6.00	6.00	6.00	6.00	6.00

Table 26: BB98 model: comparison between GMM DIF estimates

Variable		DIF	DIFc	DIFl	DPCIV90	DPCIVT90
n	*coeff*	0.707	0.840	0.787	0.802	0.508
	se	0.084	0.107	0.120	0.126	0.179
	p	0.000	0.000	0.000	0.000	0.005
w	*coeff*	-0.709	-0.971	-0.662	-0.862	-0.675
	se	0.117	0.290	0.193	0.210	0.269
	p	0.000	0.001	0.001	0.000	0.012
w_{t-1}	*coeff*	0.500	0.632	0.617	0.222	0.315
	se	0.111	0.163	0.130	0.294	0.235
	p	0.000	0.000	0.000	0.450	0.179
k	*coeff*	0.466	0.632	0.479	0.578	0.654
	se	0.101	0.215	0.139	0.225	0.209
	p	0.000	0.003	0.001	0.010	0.002
k_{t-1}	*coeff*	-0.215	-0.547	-0.438	-0.411	-0.200
	se	0.086	0.192	0.111	0.195	0.236
	p	0.012	0.004	0.000	0.035	0.397
Hansen		88.797	14.622	35.693	23.432	17.197
Hansenp		0.211	0.553	0.389	0.136	0.102
Hansen df		79	16	34	17	11
ar1p		0.000	0.000	0.000	0.001	0.055
ar2p		0.891	0.901	0.929	0.544	0.547
Obs.		751	751	751	751	751

Table 27: BB98 model: comparison between GMM SYS estimates

Variable		SYS	SYSc	SYSl	SPCIV90	SPCIVT90
n	*coeff*	0.811	0.777	0.841	0.902	0.857
	se	0.058	0.068	0.059	0.048	0.068
	p	0.000	0.000	0.000	0.000	0.000
w	*coeff*	-0.795	-0.875	-0.784	-0.742	-0.724
	se	0.097	0.260	0.148	0.154	0.150
	p	0.000	0.001	0.000	0.000	0.000
w_{t-1}	*coeff*	0.550	0.693	0.560	0.464	0.560
	se	0.152	0.255	0.179	0.195	0.180
	p	0.000	0.007	0.002	0.017	0.002
k	*coeff*	0.429	0.604	0.506	0.534	0.540
	se	0.076	0.210	0.078	0.096	0.098
	p	0.000	0.004	0.000	0.000	0.000
k_{t-1}	*coeff*	-0.280	-0.434	-0.380	-0.441	-0.414
	se	0.078	0.246	0.079	0.103	0.097
	p	0.000	0.078	0.000	0.000	0.000
Hansen		115.726	17.997	70.504	57.597	42.518
Hansenp		0.135	0.523	0.078	0.022	0.022
Hansen df		100	19	55	38	26
ar1p		0.000	0.000	0.000	0.000	0.000
ar2p		0.934	0.975	0.920	0.785	0.905
Obs.		891	891	891	891	891

Table 28: BB98 model: comparison between GMM LEV estimates

Variable		LEV	LEVc	LEVl	LPCIV90	LPCIVT90
n	*coeff*	0.944	0.893	0.934	0.944	0.927
	se	0.022	0.091	0.033	0.027	0.026
	p	0.000	0.000	0.000	0.000	0.000
w	*coeff*	-0.606	-0.730	-0.809	-0.776	-0.723
	se	0.167	0.239	0.166	0.162	0.158
	p	0.000	0.002	0.000	0.000	0.000
w_{t-1}	*coeff*	0.500	0.725	0.552	0.609	0.612
	se	0.177	0.221	0.175	0.159	0.164
	p	0.005	0.001	0.002	0.000	0.000
k	*coeff*	0.522	0.831	0.500	0.516	0.565
	se	0.062	0.126	0.068	0.065	0.060
	p	0.000	0.000	0.000	0.000	0.000
k_{t-1}	*coeff*	-0.477	-0.763	-0.444	-0.468	-0.510
	se	0.068	0.161	0.074	0.070	0.066
	p	0.000	0.000	0.000	0.000	0.000
Hansen		86.805	17.657	49.700	62.608	60.454
Hansenp		0.257	0.344	0.040	0.455	0.148
Hansen df		79	16	34	62	50
ar1p		0.135	0.380	0.037	0.018	0.092
ar2p		0.912	0.543	0.080	0.232	0.487
Obs.		891	891	891	891	891

References

[1] **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.

[2] **Arellano, M. (2003)** "Modelling optimal instrumental variables for dynamic panel data models", *CEMFI*, working paper 0310, Madrid.

[3] **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.

[4] **Arellano, M., O. Bover (1995)** "Another look at the instrumental variables estimation of error-components models", *Journal of Econometrics*, vol. 68, pp. 29-51.

[5] **Bai, J. and S. Ng (2010)** "Instrumental Variable Estimation In A Data Rich Environment", *Econometric Theory*, Vol. 26, pp. 1577-1606.

[6] **Barro, R.J. and X. Sala-i-Martin (2003)** "Economic Growth, Second edition", *MIT Press*, Cambridge.

[7] **Blundell, R.W. and S.R. Bond (1998)** "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models", *Journal of Econometrics*, vol. 87, pp. 115-143.

[8] **Blundell, R.W., S.R. Bond and F. Windmeijer (2000)** "Estimation in dynamic panel data models: improving on the performance of the standard GMM estimator", in Baltagi B. (ed.), *Advances in Econometrics*, vol. 15, *Nonstationary panels, panel cointegration, and dynamic panels*, JAI Elsevier Sceince, Amsterdam.

[9] **Bond, S. (2002)** "Dynamic panel data models: a guide to micro data methods and practice", *CEMMAP*, Working paper 09/02.

[10] 208b) **Bond S., C. Bowsher F. Windmeijer (2001)** "Criterion-based inference for GMM in autoregressive panel data models", *Economics Letters*, 73, 379-388.

[11] **Bond, S. and F. Windmeijer (2002)** "Finite sample inference for GMM estimators in linear panel data model", *CEMMAP*, Working paper 04/02.

[12] **Bontempi M.E. and J. Mairesse (2008)** "Intangible capital and productivity: an exploration on a panel of Italian manufacturing firms", *NBER*, Working paper n. 14108.

[13] **Bound, J., Jaeger, D.A. and Baker, R.M. (1995)** "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak", *Journal of the American Statistica Association*, vol. 90, pp. 443-450.

[14] **Bowsher, C. G. (2002)** "On testing overidentifying restrictions in dynamic panel data models", *Economics Letters*, vol. 77, pp. 211-220.

[15] **Bun M.J.G. and F. Windmeijer (2010)** "The weak instrument problem of the system GMM estimator in dynamic panal data models", *Econometrics Journal*, vol. 13, pp. 95-126.

[16] **D'Alessio G. (1989)** "Multistep principal components analysis in the study of panel data", in Coppi R. and S. Bolasco (eds.), *multiway Data Analysis*, pp. 375-381, Elsevier Science Publishers B.V. (North-Holland).

[17] **Doran H.E. and Schmidt P. (2006)** "GMM estimators with improved finite sample properties using principla components of the weighting matrix, with an application to the dynamic panel data model", *Journal of econometrics*, Vol. 133, pp. 387-409.

[18] **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.

[19] **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. 540ï¿½554.

[20] **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.

[21] **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.

[22] **Griliches Z. (1998)** "R&D and productivity. The econometric evidence", The University of Chicago Press.

[23] **Griliches Z. and J.A. Hausman (1986)** "Errors in variables in panel data", *Journal of Econometrics*, vol. 31, pp. 93-118.

[24] **Griliches Z. and Mairesse J. (1998)** "Production Functions: the Search for Identification" in S. Strom (ed.), *Essays in Honour of Ragnar Frisch*, Econometric Society Monograph Series, Cambridge University Press, Cambridge, pp. 169-203.

[25] **Groen, J.J.J. and G. Kapetanios (2009)** "Parsimonious estimation with many instruments", *Federal Reserve Bank of New York*, Staff Report n. 386.

[26] **Hall A.R. and F.P.M. Peixe (2003)** "A consistent method for the selection of relevant instruments", *Econometric Reviews*, Vol. 22, pp. 269-287.

[27] **Han C. and P.C.B. Phillips (2006)** "GMM with many moment conditions", *Econometrica*, Vol. 74, pp. 147-192.

[28] **Hayakawa, K. (2009)** "On the effect of mean-nonstationarity in dynamic panel data model", *Journal of Econometrics*, Vol. 153, pp. 133-135.

[29] **Hayakawa, K. and Nagata S. (2012)** "On the Behavior of the GMM Estimator in Persistent Dynamic Panel Data Models with Unrestricted Initial Conditions", (March 1, 2012). Available at SSRN: http://ssrn.com/abstract=2009015 or http://dx.doi.org/10.2139/ssrn.2009015.

[30] **Holtz-Eakin D., Newey W. and Rosen H. (1998)** "Estimating vector autoregressions with panel data", *Econometrica*, Vol. 56, pp. 1371-1395.

[31] **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.

[32] **Kloek, T., and L.B.M. Mennes (1960)** "Simultaneous equations estimation based on principal components of predetermined variables", *Econometrica*, Vol. 28, pp. 45-61.

[33] **Mairesse J. and B.H. Hall (1995)** "Exploring the relationship between R&D and productivity in French manufactirung firms", *Journal of Econometrics*, vol. 65, pp. 263-293.

[34] **Mairesse J. and B.H. Hall (1996)** "Estimating the productivity of research and development in French and US manufacturing firms: an exploration of simultaneity issues with GMM methods", in K. Wagner and Bart Van Ark (eds.) *International Productivity Differences and Their Explanations*, Elsevier Science, pp. 285-315.

[35] **Mairesse J. and M. Sassenou (1991)** "R&D and productivity: a survey of econometric studies at the firm level", *NBER*, Working paper n. 3666.

[36] **Mammi I., (2011)** "Essays in GMM estimation of dynamic panel data models", *PhD dissertation*, IMT Institute for Advanced Studies, Lucca, Italy.

[37] **Mehrhoff, J., (2009)** "A solution to the problem of too many instruments in dynamic panel data GMM", *Discussion paper n. 1/2009*, Deutsche Bundesbank.

[38] **Kloek, T., and L.B.M. Mennes (1960)** "Simultaneous equations estimation based on principal components of predetermined variables", *Econometrica*, Vol. 28, pp. 45-61.

[39] **Nickell, S. (1981)** "Biases in Dynamic Models with Fixed Effects", *Econometrica*, Vol. 49, pp. 1417-26.

[40] **Roodman, D. (2009)** "A Note on the theme of too many instruments", *Oxford Bulletin of Economics and Statistics*, vol. 71, pp. 135-158.

[41] **Satchachai P. and Schmidt P. (2008)** "GMM with more moment conditions than observations", *Economics Letters*, vol. 99, pp. 252-255.

[42] **Stock, J.H. and M.W. Watson, (1998)** "Diffusion indexes", *NBER*, Working Paper 6702.

[43] **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.

[44] **Stock, J.H. and M.W. Watson, (2002b)** "Macroeconomic forecasting using diffusion indexes", *Journal of Business and Economic Statistics*, vol. 20, pp. 147-162.

[45] **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*.

[46] **Windmeijer F. (2005)** "A finite sample correction for the variance of linear efficient two-step GMM estimators", *Journal of Econometric*, vol. 126, pp. 25-51.

[47] **Ziliak J. P. (1997)** "Efficient estimation with panel data when instruments are predetermined: an empirical comparison of moment-condition estimators", *Journal of Business and Economic Statistics*, vol. 15, pp. 419-431.

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

$$\mathbf{pc}_k = \mathbf{u}'_k \mathbf{x} \tag{25}$$

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:

$$\mathbf{u_k}'\mathbf{u_k} = 1 \tag{26}$$

$$\mathbf{u_k}'\mathbf{u_j} = 0 \text{ for } i \neq j. \tag{27}$$

$\mathbf{pc}_1 = \mathbf{u}'_1\mathbf{x}$ 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 $\mathbf{C}$:

$$\mathbf{C} = \mathbf{V}\Lambda\mathbf{V}' = \sum_{i=1}^{p} \lambda_i \mathbf{v}_i \mathbf{v}'_i \tag{28}$$

where $\mathbf{V}$ is the matrix consisting of the eigenvectors (principal components) of $\mathbf{C}$, Λ 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 (26) and (27), that is if $\mathbf{u}_k$ is such to have unit length, the variance of the k^{th} principal component, $\text{var}(\mathbf{pc}_k)$, 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:

$$\sum_{k=1}^{p} \lambda_k = \text{tr}(\mathbf{C}). \tag{29}$$

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

$$P_k = \frac{\lambda_k}{\text{tr}(\mathbf{C})}. \tag{30}$$

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

$$\mathbf{S} = \mathbf{XV} \tag{31}$$

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:

$$\mathbf{X} = \mathbf{V}'\mathbf{S} \tag{32}$$

where $\mathbf{V}$ and $\mathbf{S}$ 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 $\mathbf{C}$ 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:

$$\frac{\sum_{k=1}^{q} \lambda_k}{\text{tr}(\mathbf{C})}. \tag{33}$$

Accordingly, in the matrix $\mathbf{V}$ only q eigenvectors will be retained and the scores will be computed form the reduced $\mathbf{V}$ 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.