Weak identification robust inference in dynamic panel data models*

MAURICE J.G. BUN† and RUTGER W. POLDERMANS‡

this version: 22 January 2015

JEL-code: C13; C23

Keywords: dynamics, endogeneity, GMM, panel data

Abstract

We consider the linear dynamic panel data model with additional endogenous regressors. We analyze the accuracy of various weak identification robust GMM statistics. Although robust to identification strength, we find that they are vulnerable to the use of particular covariance matrix estimators and many moment conditions. Implementations exploiting particular subsets of a limited number of instruments are more or less size correct, while maintaining nontrivial power. Additionally, using uncentered moments in robust covariance matrix estimation substantially improves behavior under the null hypothesis without loss of any power.

1. Introduction

Estimating dynamic panel data models with a small number of time periods induces inference problems such as small sample bias and possibly poor asymptotic approximations. It is well known that the fixed effects least squares estimator is inconsistent for a small number of time periods T and a large cross-section N. This inconsistency is referred to as the Nickell (1981) bias, and is an example of the incidental parameters problem.

It has therefore become common practice to estimate the parameters of dynamic panel data models by the Generalized Method of Moments (GMM). Applying GMM results in

*We like to thank Peter Boswijk, Frank Kleibergen and Frank Windmeijer for helpful comments.

†Tinbergen Institute and Amsterdam School of Economics, University of Amsterdam. E-mail: m.j.g.bun@uva.nl.

‡Tinbergen Institute and Amsterdam School of Economics, University of Amsterdam. E-mail: r.w.poldermans@uva.nl.

consistent estimators and asymptotically valid inference when N is large and T is small (see Arellano and Bond, 1991; Arellano and Bover, 1995; Blundell and Bond, 1998). A main reason for using GMM is that it provides asymptotically efficient inference exploiting a minimal set of statistical assumptions. Furthermore, standard statistical software is available where GMM estimation methods are easily implemented for dynamic panel data analysis.

We consider the class of linear dynamic panel data models where some of the additional regressors are endogenous. Of course one could sometimes argue that these are (weakly) exogenous, or at least conditional on entity and time-specific effects. Nevertheless, especially in policy intervention analysis based on observational panel data, it is more likely that policy variables are not strictly exogenous but simultaneously determined with the outcome variable of interest (Besley and Case, 2000). Even if we are willing to believe that these additional regressors are not simultaneously determined they may be influenced by past values of the outcome variable.

Despite the optimal asymptotic properties of GMM estimators and corresponding Wald test statistics, their behavior in finite samples can be rather peculiar due to the number and weakness of moment conditions. Related to this is the observed sensitivity of conventional two-step GMM inference to important nuisance parameters. A large literature has been devoted to analyze these inherent drawbacks of GMM for dynamic panel data models and to adapt the GMM approach in order to improve quality of inference, see Bun and Sarafidis (2015) for a recent overview.

In dynamic panel data models the number of sequential moment conditions increases rapidly with the number of time periods available. Especially the asymptotically most efficient GMM estimators may suffer from substantial bias and deterioration of the mean squared error by using all these available moment conditions, even having panel data with moderately large N (Ziliak, 1997; Koenker and Machado, 1999; Bun and Kiviet, 2006). In addition, various studies (Binder et al., 2005; Bun and Kiviet, 2006) show that finite sample properties of GMM estimators depend heavily on crucial nuisance parameters like the ratio of the variances of the individual specific effects and idiosyncratic errors. Finally, there is the well known issue of weak instruments (Bound et al., 1995; Staiger and Stock, 1997). In dynamic panel data models lagged levels of endogenous regressors are used as instruments for their current changes or vice versa. When the panel data are persistent, however, such lagged instruments are only weakly correlated with the current endogenous changes. When instruments for dynamic panel data models are weak, GMM inference can perform poorly in finite samples (Blundell and Bond, 1998; Binder et al., 2005; Hahn et al., 2007; Kruiniger, 2009; Bun and Windmeijer, 2010; Bun and Kleibergen, 2013).

In this study we analyze for the dynamic panel data model with additional endogenous regressors to what extent we can overcome the aforementioned problems by exploiting identification-robust GMM inference (Stock and Wright, 2000; Kleibergen, 2005; Newey and Windmeijer, 2009). For the panel AR(1) model Bun and Kleibergen (2014) have shown that identification is possible even in the case of highly persistent panel data. By exploiting robust GMM statistics they also show that the Kleibergen (2005) LM statistic is efficient.

Although robust to weak identification, a potential disadvantage of identification-robust GMM statistics is their sensitivity with respect to the number of moment conditions. For example, Kleibergen and Mavroeidis (2009) find in their Monte Carlo experiments that the size of identification robust GMM tests can increase substantially with the number of instruments. As noted before in dynamic panel data models the number of available moment conditions grows rapidly with the number of time observations. We therefore analyze whether size of weak instrument robust tests is effectively controlled by using particular subsets of the available IVs. We will distinguish subsets based on either lag limits or collapsed instruments (Roodman, 2009). Related to this, we investigate the consequences of exploiting a covariance matrix estimator based on centered moments. We will show that exploiting uncentered moments in covariance matrix estimation is the preferred choice, which result is in contrast with earlier literature on this topic (Hall, 2000). Furthermore, we will investigate how discriminatory power is affected by these simple fixes.

In this study we limit ourselves to GMM, but likelihood based methods that correct for the incidental parameters problem are a viable alternative to achieve identification (Lancaster, 2002; Hsiao et al. (2002); Binder et al., 2005; Dhaene and Jochmans, 2012). Essentially these methods treat the incidental parameters as fixed in estimation, and therefore are largely invariant to these nuisance parameters. In comparison with the single equation GMM approach a limitation is that they impose exogeneity restrictions on the covariates. Endogeneity can be taken into account, but only by specifying a panel VAR model as discussed by Hayakawa and Pesaran (2014) and Juodis (2013).

The rest of the paper is organized as follows. In Section 2 we describe the model and the moment conditions. In Section 3 we list the various GMM statistics and briefly discuss their relative merits. In Section 4 we analyze the practical implementation of robust GMM statistics, with special attention to covariance matrix estimators and instrument proliferation issues. Section 5 reports simulation results from our Monte Carlo study. Section 6 concludes.

2. Model and moment conditions

Suppose the relation between the dependent variable y_{it} and a single regressor x_{it} can be modeled by the following dynamic specification ($i = 1, ...N$; $t = 1, ...T$):

$$y_{it} = \gamma y_{i,t-1} + \beta x_{it} + \eta_i + \varepsilon_{it}, \tag{2.1}$$

where η_i denotes unobserved time-invariant heterogeneity and ε_{it} is the idiosyncratic error component.[1] We assume that y_{i0} and x_{i0} are observed. The dynamic panel data model in (2.1), often denoted as the model in levels (LEV), permits the distinction between the long run, or equilibrium, relationship and the short-run dynamics. Note that x_{it} could also be a vector, containing both contemporaneous and lagged values of explanatory variables. As a result the restricted autoregressive distributed lag model above is easily extended to commonly used specifications in applied research by allowing for additional strictly/weakly exogenous and/or endogenous regressors.

The random individual specific effects are denoted by η_i, and are typically included to account for unobserved heterogeneity. We treat η_i as fixed in estimation, thereby allowing for possible nonzero correlation between error components and included regressors. Moreover, there are other important endogeneity issues in specification (2.1) because possibly $E(\varepsilon_{it}|x_{it}) \neq 0$. In other words, besides the potential correlation between the included regressors and the individual specific effect, the additional covariate x_{it} also can have a nonzero contemporaneous correlation with the idiosyncratic error ε_{it}. We are interested in the situation when the latter occurs and therefore consider models where the idiosyncratic errors obey the following conditional moment condition

$$E\left[\varepsilon_{it}|w_i^{t-1}, \eta_i\right] = 0, \quad t = 2, ...T, \tag{2.2}$$

where[2] $w_{it} = (y_{it}, x_{it})'$ and $w_i^s = (w_{i0}', w_{i1}', ..., w_{is}')'$.

Based on assumption (2.2) for the model in first differences (DIF)

$$\Delta y_{it} = \gamma \Delta y_{i,t-1} + \beta \Delta x_{it} + \Delta \varepsilon_{it}, \tag{2.3}$$

the following moment conditions are available

$$E\left[w_i^{t-2} \Delta \varepsilon_{it}\right] = 0, \quad t = 2, ...T. \tag{2.4}$$

That is to say, lagged levels of the endogenous regressors are used as instrumental variables for current changes, where we consider inference based on the GMM estimator introduced by Arellano and Bond (1991) for the DIF model. As suggested by Ahn and Schmidt (1995), assumption (2.2) implicitly yields $T-2$ additional nonlinear moment conditions, i.e. $(t = 3, ...T)$

$$E\left[(\eta_i + \varepsilon_{it})\Delta \varepsilon_{i,t-1}\right] = 0. \tag{2.5}$$

These moment conditions exploit the lack of serial correlation in ε_{it}.[3] In other words, if

[1] Time-specific effects can also be included explicitly or controlled for by cross-sectional demeaning of the data prior to estimation.

[2] Note that (2.2) does not exclude endogeneity due to correlation between x_{it} and η_i. Likewise, instantaneous and lagged feedback from y to x is allowed for.

[3] Compared with (2.4) the moment conditions in (2.5) make explicit use of $Cov(\eta_i, \varepsilon_{it}) = 0$, which is also implied by assumption (2.2). This is not very restrictive, because in autoregressive models any nonzero correlation between individual effects and idiosyncratic errors tends to vanish over time (Arellano, 2003, p82).

assumption (2.2) holds efficiency gains may occur using (2.5) on top of (2.4). Ahn and Schmidt (1995) show that the GMM estimator using (2.4) and (2.5) is efficient in the class of estimators that make use of second moment information. They also report substantial efficiency gains when comparing asymptotic variances for the covariance stationary AR(1) model. Especially when persistence is high, the additional quadratic moment conditions may still be informative, while the linear moment conditions yield lagged levels to become weak predictors of the first differences. For the panel AR(1) model Bun and Kleibergen (2014) show that the combination of (2.4) and (2.5) always leads to identification for all values of the autoregressive parameter.

Blundell and Bond (1998) propose the use of a different set of additional moment conditions which relies on certain stationarity conditions of the panel data (see Arellano and Bover (1995)). For the model in (2.1) this conditional (mean stationarity) assumption amounts to

$$E\left[\Delta w_{it}|\eta_i\right] = 0, \tag{2.6}$$

implying that the original data in levels have constant correlation over time with the individual specific effects. Assumption (2.6) leads to the following additional moment conditions for the model in levels (2.1)

$$E\left[\Delta w_i^{t-1}(\eta_i + \varepsilon_{it})\right] = 0, \tag{2.7}$$

where $\Delta w_i^s = (\Delta w_{i0}', \Delta w_{i1}', ..., \Delta w_{is}')'$. In words, lagged changes are used as instrumental variables for current levels of the endogenous regressors in the LEV equation. When combining moment conditions (2.4) and (2.7) it should be noted that some are redundant (see e.g. Kiviet, Pleus and Poldermans, 2013, for a proof). Specifically, the non-redundant levels moment conditions are

$$E\left[\Delta w_{i,t-1}(\eta_i + \varepsilon_{it})\right] = 0. \tag{2.8}$$

The complete set of linear moment conditions under assumptions (2.2) and (2.6) leads to the system GMM estimator, which exploits (2.4) and (2.8). Furthermore, using this combination makes the nonlinear moment conditions of Ahn and Schmidt (1995) in (2.5) redundant.

3. GMM estimators and testing procedures

In this Section we introduce the various GMM estimators and corresponding test statistics and briefly discuss their relative merits as well as observed finite sample performance in earlier Monte Carlo studies. Let θ_0 be the unknown K-dimensional parameter vector of interest. Furthermore, let $f_i(\theta)$ be an L-dimensional moment equation with $L > K$ for which we assume that $f_0 = E\left[f_i(\theta_0)\right] = 0$. The moment equation $f_i(\theta)$ can be based on any set of moment conditions discussed in the previous Section.

We define the different GMM estimators as minimizers of the random objective function $Q_N(\theta)$. For the 1-step and 2-step GMM estimator the objective function reads $(i = 1, ..., N)$

$$Q_N(\theta) = f_N(\theta)' W_N^{-1} f_N(\theta)/2, \tag{3.1}$$

where $f_N(\theta) = \frac{1}{N}\sum_{i=1}^N f_i(\theta)$. We impose that the weight matrix W_N converges in probability to W, where W is positive definite (p.d.). Let $q_N(\theta) = \frac{1}{N}\sum_{i=1}^N q_i(\theta)$ with $q_i(\theta) = \partial f_i(\theta)/\partial\theta'$. Furthermore, let $q_0 = E\left[q_i(\theta_0)\right]$. Under regularity conditions

$$\sqrt{N} f_N(\theta_0) \to_d N(0, V),$$

with $V = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N E\left[f_i(\theta_0) f_i(\theta_0)'\right]$. Consequently, for the GMM estimator $\hat{\theta}$ we know that

$$\sqrt{N}(\hat{\theta} - \theta_0) \to_d N(0, \Sigma),$$

where

$$\Sigma = \left[q_0' W^{-1} q_0\right]^{-1} \left[q_0' W^{-1} V W^{-1} q_0\right] \left[q_0' W^{-1} q_0\right]^{-1}. \tag{3.2}$$

For a consistent 1-step GMM estimator $\hat{\theta}^{(1)}$ that uses a known weight matrix W_N, the estimator of Σ equals

$$\Sigma(\hat{\theta}^{(1)}) = \frac{1}{N}\left[q_N(\hat{\theta}^{(1)})' W_N^{-1} q_N(\hat{\theta}^{(1)})\right]^{-1} \left[q_N(\hat{\theta}^{(1)})' W_N^{-1} W_N(\hat{\theta}^{(1)}) W_N^{-1} q_N(\hat{\theta}^{(1)})\right] \left[q_N(\hat{\theta}^{(1)})' W_N^{-1} q_N(\hat{\theta}^{(1)})\right]^{-1},$$

which exploits the Eicker-White covariance matrix

$$W_N(\hat{\theta}^{(1)}) = \frac{1}{N}\sum_{i=1}^N f_i(\hat{\theta}^{(1)}) f_i(\hat{\theta}^{(1)})'. \tag{3.3}$$

For the optimal 2-step GMM estimator $\hat{\theta}^{(2)}$, the conventional estimator of Σ equals

$$\Sigma(\hat{\theta}^{(2)}) = \frac{1}{N}\left[q_N(\hat{\theta}^{(2)})' W_N^{-1}(\hat{\theta}^{(1)}) q_N(\hat{\theta}^{(2)})\right]^{-1}.$$

In the objective function for the CUE, the weight matrix is allowed to vary with θ, i.e.

$$Q_N^{\text{cue}}(\theta) = f_N(\theta)' \left[\frac{1}{N}\sum_{i=1}^N f_i(\theta) f_i(\theta)'\right] f_N(\theta)/2. \tag{3.4}$$

The CUE, denoted as $\hat{\theta}^{\text{cue}}$, solves the minimization problem and its asymptotic distribution is similar to that of the optimal 2-step GMM estimator $\hat{\theta}^{(2)}$ (see Windmeijer, 2006).

We are interested in testing r restrictions of the form $a(\theta_0) = 0$. In many dynamic panel data applications it is common practice to do inference exploiting Wald statistics based on either the 1-step or optimal 2-step GMM estimator. Let $A(\theta) = \partial a(\theta)/\partial\theta'$, where we impose $A_0 = A(\theta_0)$ has full rank, then the Wald type test statistic in its feasible form equals

$$a(\hat{\theta})' \left[A(\hat{\theta})\Sigma(\hat{\theta})A(\hat{\theta})'\right]^{-1} a(\hat{\theta}), \tag{3.5}$$

which is asymptotically $\chi^2(r)$ under the null hypothesis. Note that $\hat{\theta}$ denotes either the consistent 1-step ($\hat{\theta}^{(1)}$) or optimal 2-step ($\hat{\theta}^{(2)}$) GMM estimator, where for $\hat{\theta}^{(2)}$ in addition the correction of Windmeijer (2005) is used[4].

The feasible LM test for the restricted optimal 2-step GMM, denoted as $\tilde{\theta}^{(2)}$,can be expressed as[5]

$$Nf_N(\tilde{\theta}^{(2)})'W_N^{-1}(\tilde{\theta}^{(1)})q_N(\tilde{\theta}^{(2)})'\left[q_N(\tilde{\theta}^{(2)})'W_N^{-1}(\tilde{\theta}^{(1)})q_N(\tilde{\theta}^{(2)})\right]^{-1}q_N(\tilde{\theta}^{(2)})W_N^{-1}(\tilde{\theta}^{(1)})f_N(\tilde{\theta}^{(2)}),$$

with $\tilde{\theta}^{(1)}$ the restricted 1-step GMM estimator. Similar to the Wald test, this test statistic is asymptotically $\chi^2(r)$ under the null hypothesis. As mentioned in Bond and Windmeijer (2005), when the restrictions are linear and all the moment conditions are linear in the parameters, an alternative and numerical equivalent expression for the test statistic is the Wald type form

$$a(\hat{\theta}^{(\tilde{2})})'\left[A(\hat{\theta}^{(\tilde{2})})\Sigma(\hat{\theta}^{(\tilde{2})})A(\hat{\theta}^{(\tilde{2})})'\right]^{-1}a(\hat{\theta}^{(\tilde{2})}),$$

where $\hat{\theta}^{(\tilde{2})}$ is the optimal unrestricted 2-step GMM estimator, using a weight matrix based on restricted 1-step GMM residuals, i.e. $W_N^{-1}(\tilde{\theta}^{(1)})$. Bond and Windmeijer (2005) mention that this approach is potentially useful for simple subset tests of the form $H_0 : \theta_j = c$.

Bond and Windmeijer (2005) consider the conventional asymptotic Wald test, its bootstrapped versions, the LM test and three criterion-based tests. In case of strong identification, the LM-test has the best size properties. Furthermore, the Wald test based on the optimal 2-step estimator can suffer from severe overrejection in finite samples because the asymptotic variance can be considerably downward biased. The variance correction of Windmeijer (2005) yields more accurate inference. Also the resulting corrected Wald test may have better power properties than the 1-step Wald test.

In the Monte Carlo study of Hansen et al. (1996) it has been shown that hypothesis testing based on the CUE can be more reliable than based on optimal 2-step GMM. The computation of the Wald statistic for the CUE is rather straightforward. Although under conventional asymptotics the asymptotic distribution of $\hat{\theta}^{\text{cue}}$ and $\hat{\theta}^{(2)}$ is equivalent, the former estimates the asymptotic variance differently, i.e.

$$\Sigma(\hat{\theta}^{\text{cue}}) = \frac{1}{N}\left[D(\hat{\theta}^{\text{cue}})'W_N^{-1}(\hat{\theta}^{\text{cue}})D(\hat{\theta}^{\text{cue}})\right]^{-1},$$

where column j of $D(\hat{\theta}^{\text{cue}})$ is equal to

$$D_{\cdot,j}(\hat{\theta}^{\text{cue}}) = \frac{1}{N}\sum_{i=1}^{N}\partial f_i(\hat{\theta}^{\text{cue}})/\partial\theta_j - \frac{1}{N}\sum_{i=1}^{N}\partial f_i(\hat{\theta}^{\text{cue}})/\partial\theta_j f_i(\hat{\theta}^{\text{cue}})'W_N^{-1}(\hat{\theta}^{\text{cue}})f_N(\hat{\theta}^{\text{cue}}).$$

The matrix $D(\hat{\theta}^{\text{cue}})$ is the estimator of the (expected) Jacobian based on the CUE and differs from the conventional GMM estimator of the (expected) Jacobian, which equals

[4] For the full correction see Windmeijer (2005).

[5] See Newey and West (1987b) and Bond and Windmeijer (2005).

$q_N(\hat{\theta})$[6]. Kleibergen (2005) shows that $D(\hat{\theta}^{\text{cue}})$ is asymptotically independent of the average moment vector if the Central Limit Theorem (CLT) can be invoked and this property allows for inference without imposing the full rank assumption. However, similar to the issue for the optimal 2-step GMM estimator, the feasible asymptotic standard errors for the CUE tend to be too small as well. The underlying reason differs, that is to say, for the CUE the issue results from the variability of the Jacobian, while for the optimal 2-step GMM the variability in the weight matrix leads underestimation of the asymptotic variance. Newey and Windmeijer (2009) propose a corrected version

$$\Sigma_c(\hat{\theta}^{\text{cue}}) = H(\hat{\theta}^{\text{cue}})^{-1}\Sigma^{-1}(\hat{\theta}^{\text{cue}})H(\hat{\theta}^{\text{cue}})^{-1},$$

where $H(\hat{\theta}^{\text{cue}})^{-1} = \partial^2 Q_N^{\text{cue}}(\hat{\theta}^{\text{cue}})/\partial\theta\partial\theta'$. Newey and Windmeijer (2009, p. 697) note that, although their new variance estimator is not robust to a finite number of weak moment conditions, it is easy to apply in many situations including the presence of multiple endogenous regressors. Their simulation results for a static panel data model with a predetermined regressor indicate good size properties of the corrected Wald test. Hayakawa and Pesaran (2012), however, document size distortions. Furthermore, Kleibergen (2011) mentions that the asymptotic distribution of the Wald statistic based on the CUE is sensitive to weak instruments, because estimated parameter values are used in the covariance matrix estimator.

Conventional Wald tests based on two-step GMM perform poorly when identification is weak. Furthermore, the variance correction of the CUE by Newey and Windmeijer (2009) is also not entirely identification-robust since it relies on many weak moments asymptotics[7]. We therefore consider inference based on identification-robust testing procedures of Kleibergen (2005) and the GMM extension of the Anderson-Rubin (1949) statistic by Stock and Wright (2000), hereafter denoted as KLM and AR respectively. Like the Newey and Windmeijer (2009) Wald test these testing procedures are also based on the CUE.

For the identification-robust KLM and AR statistic we replace the null hypothesis for testing r restrictions of the general form $a(\theta_0) = 0$ by the simple null $H_0 : \theta = \theta_0$. Then the KLM reads

$$KLM(\theta_0) = N\left(\partial Q_N^{\text{cue}}(\theta_0)/\partial\theta\right)' \left[D(\theta_0)'W_N^{-1}(\theta_0)D(\theta_0)\right]^{-1} \left(\partial Q_N^{\text{cue}}(\theta_0)/\partial\theta\right), \tag{3.6}$$

and the AR statistic is equal to

$$AR(\theta_0) = 2NQ_N^{\text{cue}}(\theta_0). \tag{3.7}$$

The KLM statistic is asymptotically $\chi^2(K)$ under the null, whereas the AR statistic is asymptotically $\chi^2(L)$. Note there is no estimation required in computing both test statistics

[6]Under conventional asymptotics, the probability limit of $D(\hat{\theta}^{\text{cue}})$ is equal to q_0. In these circumstances it is not uncommon to use $\Sigma(\hat{\theta}^{\text{cue}}) = \left[q_N(\hat{\theta}^{\text{cue}})'W_N^{-1}(\hat{\theta}^{\text{cue}})q_N(\hat{\theta}^{\text{cue}})\right]^{-1}$ (see Newey and Windmeijer (2006)).

[7]For the exact conditions see Newey and Windmeijer (2009).

under the simple null hypothesis. Furthermore, the KLM is expected to have better power properties since the degrees of freedom is equal to the number of parameters instead of the number of instruments. Finally, Caner (2010) shows that the KLM and AR statistic are robust to a mixture of weak, nearly-weak and strong identification[8], provided the null hypothesis is of the form $H_0 : \theta = \theta_0$. Such a mixture of identification strength among the unknown parameters of interest seems very likely in applied work.

Instead of full vector null hypotheses, the practitioner is often interested in tests on subsets of the unknown parameter vector. The AR and KLM statistics can be easily adapted to accomodate testing of subset null hypotheses. Let $\theta = (\gamma', \beta')'$ with $\gamma : K_\gamma \times 1$, $\beta : K_\beta \times 1$ and $K_\gamma + K_\beta = K$, then Stock and Wright (2000) and Kleibergen (2005) show that the null hypothesis $H_0 : \beta = \beta_0$ can be tested with the same test statistics as in (3.7) and (3.6), but now evaluated in $(\tilde{\gamma}(\beta_0)', \beta_0')'$ with $\tilde{\gamma}(\beta_0)$ the CUE for γ given that $\beta = \beta_0$. Given that the unrestricted parameters in γ are strongly identified the resulting subset AR and KLM statistics are distributed as $\chi^2(L - K_\gamma)$ and $\chi^2(K_\beta)$ respectively (Stock and Wright, 2000); Kleibergen, 2005). In case of weak identification of γ, Kleibergen and Mavroeidis (2009) show that these distributions provide an upper bound for the limiting distributions of AR and KLM subset statistics. In other words, using critical values from these distributions lead to conservative tests in case of weak identification.[9]

Summarizing, there is a trade-off between identification-robust and non-robust testing procedures in terms of testable null hypotheses. By expressing the r restrictions in the null hypothesis $H_0 : a(\theta_0) = 0$ for the non-robust testing procedures, linear and/or nonlinear restrictions are allowed for on either a subset of or the full parameter vector. As mentioned in Caner (2010), this flexibility is not applicable to the KLM and AR testing procedures, which is rather restrictive from an empirical point of view.[10] Perhaps the corrected Wald test of Newey and Windmeijer (2009) provides a middle of the road approach.

4. Implementing robust GMM inference in practice

In the previous Section we discussed various GMM estimators and corresponding testing procedures. We briefly discussed the statistical properties of each procedure, which give rise to some potential issues for the applied researcher. In this Section we discuss in more detail the practical implementation of identification robust KLM and AR statistics. In Section 4.1 we discuss the issue of using centered moment conditions in covariance matrix estimation. In Section 4.2 we investigate various ways of exploiting only a subset of the

[8] For the precise definitions of the different types of identification strength see Caner (2010).

[9] Recently, Guggenberger et al. (2012) show in the linear IV model with homoskedastic errors that for the AR statistic inference on β is still reliable without imposing the strong identification assumption on γ. However, the KLM statistic may suffer from asymptotic size distortions in some cases.

[10] For example, it is not possible to test the long-run effect $\beta/(1-\gamma)$ of x_{it} on y_{it} in our model.

available moment conditions.

4.1. Covariance estimation and centering

For general values of θ the standard Eicker-White estimator of the covariance matrix of the sample moments is:

$$W_N^u(\theta) = \frac{1}{N}\sum_{i=1}^N f_i(\theta) f_i(\theta)', \tag{4.1}$$

where we have added the superscript u to indicate that uncentered moment conditions have been used. In calculating the two-step GMM estimator the expression in (4.1) is evaluated in the one-step GMM estimate $\hat{\theta}^{(1)}$ resulting in (3.3). Analyzing the Sargan-Hansen overidentifying restrictions test, Hall (2000) notes that this covariance matrix estimator is consistent only under the null hypothesis that the population moment conditions hold. In case of an incorrect model specification, however, this consistency is lost. In those cases some of the moment conditions are not valid, and test power may fall as a result of using an inconsistent one-step GMM coefficient estimate. Hall (2000) therefore proposes[11] the use of sample moments in mean deviation form in the variance estimator, i.e.

$$W_N^c(\theta) = \frac{1}{N}\sum_{i=1}^N \left[f_i(\theta) - f_N(\theta)\right]\left[f_i(\theta) - f_N(\theta)\right]'. \tag{4.2}$$

Using centered sample moments in computation of the variance estimator yields consistency under both the null and the alternative hypothesis. Applying centered sample moments in HACC estimators Hall (2000) shows a power gain in the Sargan-Hansen test statistic.

Calculating coefficient restrictions tests based on weak identification robust GMM statistics a similar issue occurs regarding the use of uncentered or centered moment conditions. While Stock and Wright (2000) and Kleibergen (2005) exploit an Eicker-White covariance matrix estimator based on centered moment conditions, Newey and Windmeijer (2009) instead use uncentered moments.

One can expect a similar power gain when applying a centered Eicker-White estimator to coefficient restrictions tests based on any (robust or nonrobust) GMM statistic, even in case of a correct model specification. The centered Eicker-White covariance matrix estimator in (4.2) can be expressed as:

$$\begin{aligned} W_N^c(\theta) &= \frac{1}{N}\sum_{i=1}^N f_i(\theta) f_i(\theta)' - f_N(\theta) f_N(\theta)' \\ &= W_N^u(\theta) - f_N(\theta) f_N(\theta)', \end{aligned} \tag{4.3}$$

where $W_N^u(\theta)$ is the covariance matrix estimator based on uncentered moments. Therefore, we can write:

[11] See also Andrews (1999).

$$
\begin{aligned}
W_N^c(\theta)^{-1} &= \left(W_N^u(\theta) - f_N(\theta) f_N(\theta)'\right)^{-1} \\
&= W_N^u(\theta)^{-1} + W_N^u(\theta)^{-1} f_N(\theta) \left(\frac{1}{1 - f_N(\theta)' W_N^u(\theta)^{-1} f_N(\theta)} \right) f_N(\theta)' W_N^u(\theta)^{-1}.
\end{aligned} \tag{4.4}
$$

Considering the 'centered' GMM-AR statistic, we have:

$$
\begin{aligned}
N f_N(\theta)' W_N^c(\theta)^{-1} f_N(\theta) &= N f_N(\theta)' W_N^u(\theta)^{-1} f_N(\theta) \\
&\quad + N f_N(\theta)' W_N^u(\theta)^{-1} f_N(\theta) \left(\frac{1}{1 - f_N(\theta)' W_N^u(\theta)^{-1} f_N(\theta)} \right) f_N(\theta)' W_N^u(\theta)^{-1} f_N(\theta) \\
&= N f_N(\theta)' W_N^u(\theta)^{-1} f_N(\theta) \left(\frac{1}{1 - f_N(\theta)' W_N^u(\theta)^{-1} f_N(\theta)} \right),
\end{aligned} \tag{4.5}
$$

which shows that it is always larger than the 'uncentered' GMM-AR statistic. Under the alternative hypothesis, we have that $f_N(\theta)$ converges to a nonzero value, hence power of the 'centered' GMM-AR statistic will be larger.

Although the use of centered sample moments renders GMM test statistics to have more power in finite samples, the actual rejection frequency under the null hypothesis is potentially adversely affected. Under the null hypothesis, $f_N(\theta_0)$ converges to a zero value, hence asymptotically centered or uncentered implementations are equivalent. Nevertheless, in finite samples the difference may not be negligible. Exploiting Theorem 1 of Kleibergen (2011) we have the following asymptotic expansion around the true value θ_0 of the GMM-AR statistic:

$$
\begin{aligned}
AR(\theta_0) &= N f_N(\theta_0)' V_{ff}(\theta_0)^{-1} f_N(\theta_0) \\
&\quad - N f_N(\theta_0)' V_{ff}(\theta_0)^{-1} \left(\hat{V}_{ff}(\theta_0) - V_{ff}(\theta_0) \right) V_{ff}(\theta_0)^{-1} f_N(\theta_0) \\
&\quad + O_p\left(N^{-2}\right),
\end{aligned} \tag{4.6}
$$

with $\hat{V}_{ff}$ a covariance matrix estimator. The first and second order terms in (4.6) are $O_p(1)$ and $O_p\left(N^{-1}\right)$ respectively. The leading term in (4.6) converges in distribution to a χ^2_L distribution. The second order term involves the estimator $\hat{V}_{ff}$. Substituting $W_N^c(\theta_0)$ or $W_N^u(\theta)$ for $\hat{V}_{ff}$ results in centered and uncentered GMM-AR statistics (labeled $AR^c(\theta_0)$ and $AR^u(\theta_0)$ respectively). The relation between these two GMM-AR statistics is:

$$
AR^c(\theta_0) = AR^u(\theta_0) + \frac{1}{N} \left(N f_N(\theta_0)' V_{ff}(\theta_0)^{-1} f_N(\theta_0) \right)^2 + O_p\left(N^{-2}\right). \tag{4.7}
$$

Hence, centering leads to the following additional higher-order term:

$$
\frac{1}{N} \left(N f_N(\theta_0)' V_{ff}(\theta_0)^{-1} f_N(\theta_0) \right)^2, \tag{4.8}
$$

which is of order $O_p(N^{-1})$. The term between parentheses in (4.8) is asymptotically χ^2_L distributed. In case of a finite number of instruments the difference is $O_p(N^{-1})$, hence negligible small. However, in case of many instruments, i.e. L large, the term (4.8) becomes of importance. To illustrate this, note that for $Q_N(\theta_0) = Nf_N(\theta_0)'V_{ff}(\theta_0)^{-1}f_N(\theta_0)$ we have the following approximation:

$$\frac{Q_N(\theta_0) - L}{\sqrt{2L}} \underset{L\to\infty}{\longrightarrow} N(0,1),$$

hence $Q_N(\theta_0) \overset{a}{\sim} N(L, 2L)$ and

$$\left(\frac{Q_N(\theta_0) - L}{\sqrt{2L}}\right)^2 \underset{L\to\infty}{\longrightarrow} \chi^2_1.$$

This implies that

$$E\left[\frac{1}{N}\left(Nf_N(\theta_0)'V_{ff}(\theta_0)^{-1}f_N(\theta_0)\right)^2\right] \approx \frac{L^2+2L}{N}.$$

Compared with the expectation of the first-order asymptotic distribution, i.e. L from χ^2_L, this number is small when T is small. For example, when $N = 100$, $T = 4$ and considering DIF moments, we have $L = 6$ for each endogenous regressor and $\frac{L^2+2L}{N} = 0.48$, which is less than 10%. However, when $T = 10$, we already have $L = 45$ and $\frac{L^2+2L}{N} = 21.15$, which is almost 50% of L. This will certainly lead to a bias in hypothesis testing using centered moments, but can be easily avoided exploiting an uncentered covariance matrix estimator.

Although exploiting centering may lead to some power gain, size distortions may occur as well. Some indication of the net effect of these opposing consequences are given by Allen (2007), who analyzes by Monte Carlo simulation size and power of the Sargan-Hansen test. He shows that size adjusted power of a centered test is not significantly greater than the standard, uncentered Sargan-Hansen test. This means that if the test using centered moments has larger power than the standard approach exploiting uncentered moments, it is only because the former test suffers from size distortions.

Summarizing, once adjusted for their size distortion tests based on centered moments are not an improvement over the standard, uncentered tests. Regarding the GMM-AR statistic it can easily be seen that size-corrected power is even identical. According to (4.5) we can write:

$$AR^c(\theta) = AR^u(\theta)\left(\frac{1}{1-\frac{AR^u(\theta)}{N}}\right),$$

hence we see that $AR^c(\theta)$ is a monotonic increasing function of $AR^u(\theta)$. Therefore, size-corrected power functions of both tests are identical. In our Monte Carlo study we will further investigate the finite sample behaviour of coefficient restrictions tests based on these and other GMM statistics based on centered and uncentered covariance matrix estimators.

4.2. Abundance of moment conditions

In dynamic panel data models the number of available moment conditions increases rapidly with the number of time observations T. For example, the number of DIF moment conditions as defined in (2.4) is of order $O(T^2)$. The use of an abundance of moment conditions leads to more bias in GMM coefficient estimators (Ziliak 1997, Koenker and Machado, 1999, Alvarez and Arellano, 2003; Bun and Kiviet, 2006). For example, Bun and Kiviet (2006) show analytically that the order of magnitude of coefficient bias of several one-step GMM estimators depends on L. For example, regarding the DIF GMM estimator they show that exploiting all $L = O(T^2)$ moment conditions leads to a bias of order $O(N^{-1})$, while using only a subset of $O(T)$ or $O(1)$ number of instruments reduces the bias to $O(N^{-1}T^{-1})$.

Finite sample bias in one- and two-step GMM coefficient estimators will carry over to bias in corresponding Wald statistics. The continuously updating GMM estimator (Hansen et al., 1996) has been shown in Monte Carlo studies to have less finite sample bias than the two-step GMM estimator. Newey and Smith (2004) proof analytically that the CUE has better second-order bias properties. Nevertheless, it has be shown by Guggenberger (2005) that the CUE may suffer from a moment problem. That is to say, he shows in his simulation study, among other things, that the median bias of the CUE is lower than that of the 2SLS estimator in many cases, whereas the CUE is substantially more dispersed[12].

A main advantage of the CUE is that it provides identification robust GMM inference (Stock and Wright, 2000; Kleibergen, 2005). A possible disadvantage[13] of the GMM-AR and KLM statistics is their lack of robustness with respect to the number of moment conditions. As noted before the number of available moment conditions grows rapidly with the number of time observations. Hence, especially the power of the GMM-AR statistic might be negatively affected even when the number of time observations in the panel data is moderate. The results of Kleibergen (2005), Newey and Windmeijer (2009) and Bun and Kleibergen (2014) show that the KLM statistic provides a more powerful test than the GMM-AR statistic.

But a more serious issue is lack of size control. Although the limiting distributions of identification robust GMM statistics do not alter under a many instrument limit sequence (see e.g. Newey and Windmeijer, 2009), the dimensionality of the moment equations and their derivatives is an issue with respect to the involved covariance matrix estimators. For example, Kleibergen and Mavroeidis (2009) find in their Monte Carlo experiments that the size of identification robust GMM tests can increase substantially with the number of instruments. Especially the GMM-AR test is vulnerable to the number of moment condi-

[12] Hansen et al. (1996) also recognize this tendency when introducing the CUE.

[13] Another disadvantage of identification robust statistics is that exhaustive grid search is needed to calculate confidence sets. Additionally, regarding subset hypotheses they need to be adapted into subset statistics.

tions, but also the KLM statistic is not entirely robust in this direction. The Edgeworth expansions of Kleibergen (2011) provide a quantitative statement of this problem. Higher order terms for GMM-AR depend on L^2, while those for KLM depend on L only. In panel data models, however, using all available moment conditions $L = O(T^2)$, which is already substantially large when T is still a one-digit number.

A simple solution to the instrument proliferation issue in panel data models (see e.g. Roodman, 2009) is to consider only recent lags as instruments (lag limits), or to combine moment conditions from different periods (collapsed instruments).[14] As an example of this consider the DIF moment conditions in (2.4). It is common practice in applied research not to use all instruments when the number of time periods gets larger. Taking lag limits such that only the nearest lag is included as instrument for each time period results in:

$$E\left[w_{i,t-2}\Delta\varepsilon_{it}\right] = 0, \quad t = 2, ...T. \tag{4.9}$$

Hence, for each endogenous regressor in w_{it} we only exploit the $T-2$ nearest lagged instruments out of $\frac{1}{2}T(T-1)$ available moment conditions. Alternatively, Roodman (2009) proposes to collapse the set of moment conditions in (2.4) as:

$$E\left[\sum_{s=2}^{t} w_{i,s-2}\Delta\varepsilon_{it}\right] = 0, \quad t = 2, ...T. \tag{4.10}$$

Collapsing implies that for each of the $T-2$ estimation periods we only exploit the sum of the available moment conditions. These transformed moments therefore also include information of further lagged instruments.[15]

From an asymptotic point of view more moment conditions are always beneficial in terms of efficiency, but we have seen that this may lead to anomalous results regarding size distortions in finite samples. A priori it is difficult to judge whether taking lag limits or exploiting a collapsed instrument set is more advantageous. In the following we provide a partial answer to this by considering the power of tests based on the full set of available DIF moment conditions (2.4), the nearest lagged instruments (4.9) only and the collapsed moment conditions (4.10).

Comparing test power we confine ourselves to the case of strong identification. We compare power by exploiting a classical Pitman drift $\theta = \theta_0 + \frac{\mu}{\sqrt{N}}$. In this case Newey and West (1987) show that two-step GMM Wald and LM tests of the null hypothesis $H_0 : a(\theta_0) = 0$ converge to a noncentral chi-squared distribution with r degrees of freedom and noncentrality parameter equal to $\mu' A' \left[A\Sigma A'\right]^{-1} A'\mu$. Because the two-step GMM estimator

[14] Alternatively, bootstrap procedures (Kleibergen, 2011) can be exploited to obtain size corrections.

[15] Following Anderson and Hsiao (1982) and Bun and Kiviet (2006) the instrument set can be made even more collapsed or sparse by exploiting only $E\left[\sum_{t=2}^{T} w_{i,t-2}\Delta\varepsilon_{it}\right] = 0$ (and possibly $E\left[\sum_{t=3}^{T} w_{i,t-3}\Delta\varepsilon_{it}\right] = 0$ to achieve some degree of overidentification).

and CUE are asymptotically equivalent in case of strong identification, the CUE Wald and CUE LM statistics will have identical noncentrality parameter.

In general test power is depending mainly on the asymptotic variance Σ of the corresponding coefficient estimator. Therefore, by comparing asymptotic variances of GMM coefficient estimators based on either lag limits or collapsed instruments we can answer the question which of the two approaches will result in largest power. We further simplify our analysis by considering the pure AR(1) model. For the Arellano and Bond (1991) moment conditions and $T = 3$, we derive in the Appendix the asymptotic variances of the GMM estimators using all, nearest lag and collapsed instruments (labeled Σ_f, Σ_l and Σ_c respectively). Following Hayakawa and Nagata (2012) we specify the initial condition for the AR(1) model as:

$$y_{i0} = \delta \frac{\eta_i}{1-\gamma} + \varepsilon_{i0}, \qquad Var(\varepsilon_{i0}) = \frac{\lambda \sigma_\varepsilon^2}{1-\gamma^2}, \tag{4.11}$$

with $\delta \neq 0$ and $\lambda > 0$. The additional parameters δ and λ model the deviations from covariance stationarity, which case is equal to $\delta = \lambda = 1$.

We can numerically compare the resulting asymptotic variances Σ_f, Σ_l and Σ_c as a function of the autoregressive parameter γ, the variance ratio $vr = \sigma_\eta^2/\sigma_\varepsilon^2$, δ and λ. Without loss of generalization we normalize $\sigma_\varepsilon^2 = 1$ and calculate Σ_f, Σ_l and Σ_c for particular values $\{\gamma, vr, \delta, \lambda\}$.

In case of covariance stationarity we choose $\delta = \lambda = 1$ and $0 \leq \gamma \leq 0.9$ and $0 \leq vr \leq 5$. We find the following:

$$\Sigma_f \leq \Sigma_c \leq \Sigma_l, \qquad \forall \gamma, vr, \tag{4.12}$$

$$\Sigma_f = \Sigma_l = \Sigma_c \text{ if } \gamma + vr = 1. \tag{4.13}$$

These results show that, in case of covariance stationarity, using nearest lags never leads to an efficiency gain compared with collapsed instrument sets, while obviously using all moment conditions is efficient. Furthermore, for small variance ratios, i.e. $0 < vr \leq 1$, there is always a value for γ resulting in identical asymptotic variances for all three estimators.

In absence of covariance stationarity, however, there is no clear ranking between taking lag limits or collapsing. Maintaining mean stationarity ($\delta = 1$), but allowing $0.2 \leq \lambda \leq 5$ produces $0.95 < \frac{\Sigma_l}{\Sigma_c} < 1.21$. In most cases collapsing is slightly more efficient than taking nearest lags. Maintaining $\lambda = 1$, but allowing a deviation from mean stationarity $-1 \leq \delta \leq 3$ produces $0.04 < \frac{\Sigma_l}{\Sigma_c} < 2.30$. This implies large efficiency gains for lag limits over collapsing for particular parameter configurations.

Summarizing, under covariance stationarity collapsing leads to more powerful tests than taking nearest lags. This is also often the case when there is no covariance stationarity, but only mean stationarity. Under mean stationarity the practitioner will most likely use the system GMM approach of Blundell and Bond (1998) and our results suggest that to economize on the number of moment conditions one can better collapse in most cases. However, under mean-nonstationarity no clear ranking exist, and actually in some cases taking lag limits can have large efficiency gains compared with collapsing instruments. As the practitioner doesn't know the magnitude of the deviation from mean stationarity, it is not clear whether in applied research one should take nearest lag instruments or collapsed instruments in this case. Finally, we note that these results have been derived under strong identification. Under weak identification asymptotic theory is non-standard, and a different ranking may result. In the next Section we will investigate by simulation what in this case the more powerful implementation is.

5. Monte Carlo study

In this Section we analyze by simulation the impact of number, weakness and type of moment conditions on the finite sample properties of a wide variety of GMM statistics. We first investigate size and power of both Wald and LM statistics for the AR(1) model without any additional regressors. Reason for this is that it is computational burdensome to conduct a power comparison for the model with multiple endogenous regressors. In particular we analyze whether not centering the covariance matrix estimator or economizing on the number of moment conditions is effective in controlling size. Furthermore, we quantify the consequences for power of these simple fixes as discussed in the previous Section. Having determined in the AR(1) model the optimal choice, we next analyze size and power in the model with an additional regressor.

5.1. AR(1) model

The data generating process for y_{it} is given by:

$$y_{it} = \gamma_0 y_{i,t-1} + \eta_i + \varepsilon_{it}, \tag{5.1}$$

and we specify for the initial condition:

$$y_{i0} = \frac{\eta_i}{1-\gamma_0} + \varepsilon_{i0}. \tag{5.2}$$

Furthermore, we choose $\sigma_\varepsilon^2 = \sigma_\eta^2 = 1$. We calculate rejection frequencies both under null and alternative hypotheses. We specify $\gamma_0 = \{0.5, 0.6, 0.7, 0.8, 0.9, 0.99\}$ and test $H_0 : \gamma = \{0.5, 0.99\}$ at a nominal 5% significance level. Furthermore, we choose $T = \{4, 9\}$ and $N = \{100, 250\}$. Each experiment has 10000 replications. We limit ourselves to the finite sample performance of LM, KLM and GMM-AR tests and do not exhaustively analyze Wald statistics here. We implement these tests for DIF, AS and SYS moment conditions. Regarding the DIF part of the AS and SYS moment conditions we exploit all available instruments as in (2.4) or the nearest lags subset as defined in (4.9). We label the the full set and subset of instruments using nearest lags as I and II respectively. We also considered collapsed instruments as in (4.10), but the size of those tests is very close to that of nearest lags. Furthermore, because we assumed mean stationarity the power of tests based on collapsed moment conditions will be bounded by full set and nearest lag results. Finally, we denote with C and U robust covariance matrix estimation by centered and uncentered moments respectively.

The parameter configurations cover 4 important cases, mainly determined by the values for γ_0 and T:

1. a finite number of strong moment conditions (both T and γ_0 small);

2. a finite number of weak moment conditions (T small and γ_0 close to one);

3. many moment conditions (T large and γ_0 small);

4. many weak moment conditions (T large and γ_0 close to one).

Table 1 shows actual rejection frequencies for the LM, KLM and GMM-AR tests under $H_0 : \gamma = \gamma_0$. As long as we have a small number of strong moment conditions ($T = 4$ and $\gamma_0 = 0.5$), i.e. columns (1) and (2), size distortions are small for all 3 tests. Increasing γ_0 to 0.99, i.e. columns (5) and (6), leads to weak identification issues for the DIF LM test. Somewhat remarkable is that AS and SYS LM tests still show rather satisfactory performance. The performance of KLM and GMM-AR tests is not affected by this change in identification strength, as expected. This shows the robustness to identification strength, i.e. the limiting null distributions of KLM and GMM-AR statistics are robust to instrument quality.

Increasing T to 9, but keeping $\gamma_0 = 0.5$ (columns (3) and (4)), leads to instruments proliferation and a many moment conditions problem for all tests, especially when $N = 100$. Not surprisingly, especially the GMM-AR test is vulnerable to the use of many moment conditions. Increasing the cross-sectional dimension from $N = 100$ to $N = 250$ mitigates size distortions, as expected. Finally, increasing both T and γ_0 (columns (7) and (8)) leads to similar size distortions.

It is also apparent from Table 1 that reducing the set of moment conditions or exploiting an uncentered covariance matrix estimator are both very effective in controlling the size of the various tests. The rows labeled II and U show much less size distortions than the implementations exploiting all instruments and centered moments. Compare, for example, centered and uncentered SYS KLM tests. When $\gamma_0 = 0.5$ the SYS KLM rejection frequency of a centered test increases from 0.086 to 0.325 when T grows from 4 to 9 periods. For the SYS GMM-AR statistic the rejection frequency when $T = 9$ is 0.900 implying a huge size distortion compared with only 0.117 when $T = 4$. In contrast, the same statistics using an uncentered covariance matrix estimator have rejection frequencies very close to the nominal

0.05 level. Similar reductions can be seen when applying statistics with centered covariance matriz estimators, but reducing the number of moment conditions to nearest lags only.

The next question is whether exploiting uncentered moments or only a subset of IVs leads to a loss of power. This is investigated by varying γ_0, while keeping the value for γ fixed under H_0. Table 2 shows size and power for testing $\gamma = 0.5$ and varying γ_0 between 0 and 0.99 when $T = 9$ and $N = 250$. Note that due to the size distortions of original tests we only compare size corrected power. All DIF statistics show that the power of the statistics for testing $H_0 : \gamma = 0.5$ reduces when γ_0 gets close to one, while this is not the case for the AS and SYS statistics. In other words, there is for DIF an identification problem near the unit root, while AS and SYS moment conditions always identify θ. This issue is extensively discussed in Bun and Kleibergen (2014). They show that the moment equation and its derivative contains a divergent component when γ_0 goes to one. The AS and SYS moment conditions, however, contain a part not depending on this divergent component and this leads to nontrivial power when γ_0 is near the unit root.

Using a subset of moment conditions does lead to a loss of power, especially for alternatives close to H_0. Between LM, KLM and GMM-AR statistics there is no clear pattern of which statistic is most vulnerable, but in relative terms power loss seems largest for tests based on the AS moment conditions. For example, testing $H_0 : \gamma = 0.5$ when $\gamma_0 = 0.6$, the loss in power for the AS KLM statistic is $0.820 - 0.423 = 0.397$. Overall the KLM statistic performs best in terms of power irrespective of the number of moment conditions used, which corroborates the findings of Bun and Kleibergen (2014).

Using uncentered moments in covariance matrix estimation does not lead to a loss of power at all. As discussed earlier, remarkable is that for the GMM-AR statistic size-corrected power of centered and uncentered statistics is even identical. This implies that if statistics using uncentered covariance matrix estimators have lower power than statistics using centered moments, it is only because the latter suffers from size distortions.

Summarizing, we can achieve size control for the robust GMM statistics by exploiting either an uncentered covariance matrix estimator or reducing the nunber of moment conditions. As is apparent from Table 2 the former implementation generally leads to more powerful tests than the latter. Using all moment conditions in combination with an uncentered covariance matrix estimator seems therefore the preferred choice when implementing robust GMM statistics in practice. In the next Monte Carlo design we will investigate for the dynamic model with an additional regressor what the quality is of the various testing procedures.

5.2. ARX model

We adopt the Monte Carlo design of Kiviet, Pleus and Poldermans (2014), henceforth denoted as KPP. The data generating process for y_{it} is given by ($i = 1, ..., N$, $t = 1, ..., T$):

$$y_{it} = \alpha_y + \gamma y_{i,t-1} + \beta x_{it} + \sigma_\eta \eta_i^\circ + \sigma_\varepsilon \omega_i^{1/2} \varepsilon_{it}^\circ \qquad (|\gamma| < 1), \tag{5.3}$$

while the data generating process for x_{it} is equal to

$$\begin{aligned} x_{it} &= \alpha_x + \xi x_{i,t-1} + \pi_\eta \eta_i^\circ + \pi_\lambda \lambda_i^\circ + \sigma_v \omega_i^{1/2} v_{it}^\circ, \text{ where} \\ v_{it}^\circ &= \rho_{v\varepsilon} \varepsilon_{it}^\circ + (1 - \rho_{v\varepsilon}^2)^{1/2} \zeta_{it}^\circ, \end{aligned}$$

with $|\xi| < 1$ and $|\rho_{v\varepsilon}| < 1$. All random drawings η_i°, ε_{it}°, λ_i°, ζ_{it}° are $IID(0, 1)$ and mutually independent. Parameter $\rho_{v\varepsilon}$ indicates the correlation between the cross-sectionally heteroskedastic disturbances $\varepsilon_{it} = \sigma_\varepsilon \omega_i^{1/2} \varepsilon_{it}^\circ$ and $v_{it} = \sigma_v \omega_i^{1/2} v_{it}^\circ$, where the fixed values $\omega_i^{1/2}$ parametrize cross-section heteroskedasticity with $\sum_{i=1}^N \omega_i = N$. The cross-sectional heteroskedasticity follows a lognormal pattern

$$\omega_i = e^{h_i(\theta)}, \text{ with } h_i(\theta) = -\theta^2/2 + \theta[\kappa^{1/2} \eta_i^\circ + (1 - \kappa)^{1/2} \lambda_i^\circ] \sim NID(-\theta^2/2, \theta^2).$$

In our simulation study we will generate pre-sample periods for the processes for x_{it} and y_{it} by taking $x_{is} = 0$ and $y_{is} = 0$ for $s = -50$. Next, we generate x_{it} and y_{it} for the indices $t = s + 1, ..., T$. We will discard the first 50 observations for estimation[16], where we assume that both series achieved their stationary tracks.

Without loss of generality we may chose $\sigma_\varepsilon = 1$ and $\alpha_y = \alpha_x = 0$. Furthermore, we allow for pretty serious cross-sectional heteroskedasticity by choosing $\theta = 1$ and $\kappa = 0.5$. The values of the parameters β, σ_η, π_η, π_λ, σ_v and $\rho_{v\varepsilon}$ are chosen according to the six design parameters introduced[17] in KPP, i.e. $\bar{V}_x$, EVF_x, IEF_x^η, $\bar{\rho}_{x\varepsilon}$, DEN_y^η and SNR. The first three parameters control particular variance components of the long-run stationary path of the process for x_{it}, whereas $\bar{\rho}_{x\varepsilon}$ fixes the average simultaneity. The last two parameters DEN_y^η and SNR characterize the long-run stationary path of the process for y_{it}. That is to say, the former implicitly fixes the variance ratio between the individual effect directly in y_{it} and the idiosyncratic error, whereas the latter, although not well-defined when $\bar{\rho}_{x\varepsilon} \neq 0$, corresponds to the signal-noise ratio.

In particular we are interested in the empirical relevant case where the series x_{it} is correlated with the idiosyncratic error ε_{it}, without the need to specify a VAR model as suggested for likelihood-based inference procedures. Moreover, we consider different instruments sets used for estimation and examine two different DGPs to discriminate between identification

[16] As a result, the number of time observations for each individual i corresponds to $T + 1$.

[17] The Monte Carlo design of Kiviet, Pleus and Poldermans (2014) is described in full detail in section 4 of their study.

strength. To do so, the approach is less clear-cut than in the previously discussed AR(1) model, yet similarities remain. In DGP A, which represents stronger identification, we chose moderate values for the AR parameters, i.e. $\xi = \gamma = 0.5$. Furthemore, the variance of the individual effects in both the series x_{it} and y_{it} is relatively small by fixing $EVF_x = 0.3$ and $DEN_y^{\eta} = 1$. Finally, the degree of simultaneity is mild ($\bar{\rho}_{x\varepsilon} = 0.15$) and the signal-noise ratio is reasonable ($SNR = 5$). The identification strength is reduced considerably in DGP B by adjusting these parameters accordingly. That is to say, we reduce the speed of convergence by fixing $\xi = \gamma = 0.8$ and increase the impact of the individual effects by setting $EVF_x = 0.6$ and $DEN_y^{\eta} = 4$. Besides, the prescence of moderate simultaneity ($\bar{\rho}_{x\varepsilon} = 0.30$) and a smaller signal-noise ratio ($SNR = 3$) yield further loss of identication strength.

Regaring the different instrument sets exploited in estimation we proceed as follows. Initially, all the available DIF moment conditions in (2.4) and all available LEV moment conditions in (2.7) are exploited. Furthermore, the identity matrix I_{T-1} is included in the instrument matrix for the DIF equation, whereas a constant is included in the instrument matrix for the LEV equation. This approach is in accordance with the fundamental moment condition in (2.2). Second, the number of DIF moment conditions in (2.4) is reduced by using only the first available lag of both regressors, yet maintaining the identity matrix I_{T-1}. Also, all available LEV moment conditions are used. Both instrument sets are denoted as set I and set II respectively. Subsequently, both instrument sets are collapsed for the equation in differences and the equation in levels , which we refer to as set I^{coll} and set II^{coll}. For example[18], when $T = 3$ the number of instruments in sets I, II, I^{coll} and II^{coll} correspond to 8, 6, 5 and 3 respectively for the DIF equation, whereas they correspond to 5, 5, 3 and 3 for the LEV equation.

For the both DGPs A and B considered, which involve sample size $N \in \{200, 400\}$ and $T \in \{3, 7\}$, we used the very same realizations of the underlying standardized random components η_i°, λ_i°, ε_{it}° and ζ_{it}° over the respective 2500 replications that we performed. At this stage, all these components have been drawn from the standard normal distribution. To speed-up the convergence of our simulation results, in each replication we have modified the N drawings η_i° and λ_i° such, that they have sample mean zero, sample variance 1 and sample correlation zero. This rescaling is achieved by replacing the N draws η_i° first by $[\eta_i^{\circ} - N^{-1}\sum_{i=1}^{N}\eta_i^{\circ}]$ and next by $\eta_i^{\circ}/[N^{-1}\sum_{i=1}^{N}(\eta_i^{\circ})^2]^{1/2}$, and by replacing the λ_i° by the residuals obtained after regressing λ_i° on η_i° and an intercept, and next scaling them by taking $\lambda_i^{\circ}/[N^{-1}\sum_{i=1}^{N}(\lambda_i^{\circ})^2]^{1/2}$. In addition, we have rescaled in each replication the ω_i by dividing them by $N^{-1}\sum_{i=1}^{N}\omega_i$, so that the resulting ω_i have sum N as they should in order to avoid that presence of heteroskedasticity is conflated with larger or smaller average disturbance variance.

We report rejection frequencies under the joint null hypothesis $H_0 : \gamma = \gamma_0, \beta = \beta_0$.

[18]The intrument sets I, II, I^{coll} and II^{coll} are given in the appendix for $T = 3$.

We consider statistics based on either DIF or SYS moment conditions (labeled AB and BB respectively). We report reject frequencies for CUE Wald, CUE Wald with variance estimator of Newey and Windmeijer (2009), LM, Kleibergen (2005) LM and AR (Stock and Wright, 2000) statistics.

5.2.1. DGP A

Table 3 shows actual rejection probabilities ($\alpha = 0.05$) for Wald CUE, (K)LM and AR tests under $H_0 : \gamma = \gamma_0$, $\beta = \beta_0$ for estimation based on both AB and BB, each exploiting four different instrument sets.

Although DGP A represents strong identification, the rejection frequency of the Wald CUE (W^{cue}) shows severe overrejection for instrument set I, especially for T large and/or BB. Reducing the instruments according to set II or set I^{c} mitigates the overrejection, although the actual rejection probabilities depart from the nominal level considerably. The Wald CUE that uses the variance correction ($W_{\text{c}}^{\text{cue}}$) proposed by Newey and Windmeijer (2009) remedies the tendency of overrejection in all cases. Nevertheless, inference based on this approach is unreliable in finite samples, even for $N = 400$. As expected, the differences between the rejection frequency of W^{cue} and $W_{\text{c}}^{\text{cue}}$ are largest when many instruments are used, such as in instrument sets I and II when T is large. Remarkably, the rejection frequency of $W_{\text{c}}^{\text{cue}}$ is largest when all instruments are used for estimation. Exploiting instrument set II^{c} yields actual rejection frequency closest to the nominal level, where the differences between the corrected and uncorrected version are small. Interestingly, for $N = 400$ the actual rejection frequency of $W_{\text{c}}^{\text{cue}}$ and of the Wald test based on optimal 2-step using the Windmeijer (2005) correction (not reported here) is almost similar.

Turning to the LM statistic we note that its performance is vulnerable to the number of instruments as well, although for AB the performance seems reasonable for $N = 400$ and instrument set I^{c}. Clearly, the degree of overrejection for the LM statistic is much smaller compared to the overrejection of CUE Wald.

The performance of the KLM statistic is superior for AB, where the rejection frequency is invariant with respect to the number of instruments. This superior performance seems to occur for BB as well when $N = 400$ and the number of instruments become not too large. The AR statistic shows a general tendency of small underrejection. Overrejection may occurs for BB when T is large and for instrument set I^{c}. For $N = 200$ and $T = 3$ the performance of the AR statistic for BB seems very reasonable. Interestingly, when instrument set II^{c} is used for AB or BB and N is large, the rejection frequency of all statistics are close to the nominal level. Nevertheless, the use of this instrument set is expected to yield considerable power loss.

5.2.2. DGP B

The results for DGP B are reported in Table 4. When testing the full parameter vector, the rejection frequency of W^{cue} clearly shows that it suffers from the weak identification in DGP B. Although the CUE has more desirable bias properties in finite samples relative to optimal 2-step, the Wald approach yields unreliable inference. Of course, the proposed correction by Newey and Windmeijer (2009) yields no improvement, since it is not robust to weak identification. If we compare the rejection frequency of the LM statistic of AB to BB, we conclude that the weak identification in DGP B manifests itself especially for the DIF equation.

The KLM statistic shows rejection frequency closest to the nominal level when $N = 400$, except for BB when $T = 7$ and either instrument set I or II is used for estimation. The AR statistic underrejects somewhat in general. The results in Table 4 show that weak identification robust inference is of importance in DGP B.

6. Preliminary conclusions

We have analyzed the finite sample properties of various GMM statistics, robust and non-robust to weak identification, for the dynamic panel data model with additional endogenous regressors. In particular, we have analyzed the effectiveness of some simple solutions proposed in the dynamic panel data literature to reduce the instrument count. Furthermore, we consider both centered and uncentered moments in constructing covariance matrix estimators.

Our theoretical results show that exploiting centered moment conditions in covariance matrix estimation can lead to considerable size distortions. Furthermore, we have shown that power of GMM tests based on an uncentered Eicker-White covariance matrix estimator is not affected at all. In other words, if centering the covariance matrix estimator does lead to more test power, it is only because the resulting test is size distorted.

Furthermore, reducing the number of instruments, by considering lag limits or collapsed instrument matrices, are both quite effective in controlling the size of GMM tests. Regarding power, however, there may be substantially losses especially for local alternatives when using a reduced number of instruments. Therefore, using uncentered moments in robust covariance matrix estimation in combination with all moment conditions and an identification robust GMM statistic results in a viable test procedure.

Our Monte Carlo simulations corroborate these theoretical findings. We find that particular Wald statistics are size distorted in case of many and/or weak moment conditions. Weak instrument robust statistics using an uncentered covariance matrix estimator are size correct, while maintaining nontrivial power.

References

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

Alvarez, J. and M. Arellano (2003). The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica* 71, 1121-1159.

Anderson, T.W. and C. Hsiao (1981), Estimation of dynamic models with error components, *Journal of the American Statistical Association* 76, 598-606.

Anderson, T.W. and C. Hsiao (1982), Formulation and estimation of dynamic models using panel data, *Journal of Econometrics* 18, 47-82.

Anderson, T. W., and Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics* 20, 46–63.

Andrews, D.W.K. (1999). Consistent moment selection procedures for generalized method of moments estimation. *Econometrica* 67, 543-564.

Arellano, M. (2003). Panel Data Econometrics. *Oxford University Press.*

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58, 277-298.

Arellano, M. and O. Bover (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* 68, 29-51.

Besley, T. and A. Case (2000). Unnatural experiments? Estimating the incidence of endogenous policies. *The Economic Journal* 110, F672-F694.

Binder, M., Hsiao, C. and M. H. Pesaran (2005). Estimation and inference in short panel vector autoregressions with unit roots and cointegration. *Econometric Theory* 21, 795–837.

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

Blundell, R. and S. Bond (2000). GMM Estimation with persistent panel data: an application to production functions. *Econometric Reviews* 19, 321-340.

Blundell, R., Bond, S. and F. Windmeijer (2001). Estimation in dynamic panel data models: Improving on the performance of the standard GMM estimator. In: B.H. Baltagi, T.B. Fomby, R. Carter Hill (eds.), Nonstationary Panels, Panel Cointegration, and Dynamic Panels. Advances in Econometrics, Volume 15, Emerald Group Publishing Limited, 53-91.

Bond, S. and F. Windmeijer (2005). Reliable inference for GMM estimators? Finite sample properties of alternative test procedures in linear panel data models. *Econometric Reviews* 24, 1-37.

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

Bun, M.J.G. and M.A. Carree (2005). Bias-corrected estimation in dynamic panel data models. *Journal of Business & Economic Statistics* 23, 200-210.

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

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

Bun, M.J.G. and F. Kleibergen (2013).Identification and inference in moments based analysis of linear dynamic panel data models. *UvA-Econometrics working paper* 2013/07, University of Amsterdam.

Bun, M.J.G. and V. Sarafidis (2015). Dynamic panel data models. Chapter 3 in The Oxford Handbook of Panel Data. *Oxford University Press.*

Dhaene, G. and K. Jochmans (2012). An adjusted profile likelihood for non-stationary panel data models with fixed effects. Working paper, KU Leuven.

Doornik, J.A., Arellano, M. and S. Bond (2006). Panel data estimation using DPD for Ox. mimeo, University of Oxford.

Hall, A.R. (2000). Covariance matrix estimation and the power of the overidentifying restrictions test. *Econometrica* 68, 1517-1527.

Han, C. and P.C.B. Phillips (2010). GMM estimation for dynamic panels with fixed effects and strong instruments at unity. *Econometric Theory* 26, 119-151.

Hansen, L.P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029-1054.

Hansen, L. P., Heaton, J., and Yaron, A. (1996). Finite sample properties of some alternative GMM Eetimators. *Journal of Business & Economic Statistics* 14, 262–280.

Hayakawa, K. and S. Nagata. On the behavior of the GMM estimator in persistent dynamic panel data models with unrestricted initial conditions. *Working Paper*, Hiroshima University.

Hayakawa, K. and M.H. Pesaran (2014). Robust standard errors in transformed likelihood estimation of dynamic panel data models with cross-sectional heteroskedasticity. CWPE 1224, University of Cambridge.

Hsiao, C., Pesaran, M.H. and A.K. Tahmiscioglu (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics* 109,107–150.

Holtz-Eakin, D., Newey, W. and H.S. Rosen (1988). Estimating vector autoregressions with panel data. *Econometrica*, 56, 1371–1395.

Juodis, A. (2013). First difference transformation in panel VAR models: robustness, estimation and inference. Mimeo, *UvA-Econometrics working paper* 2013/06, University of

Amsterdam.

Kiviet, J.F. (1995). On bias, inconsistency and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics* 1995, 68, 53-78.

Kiviet, J.F., Pleus, M. and R. Poldermans (2013). Accuracy and efficiency of various GMM inference techniques in dynamic micro panel data models. Mimeo, University of Amsterdam.

Kleibergen, F. (2005). Testing parameters in GMM without assuming that they are identified. *Econometrica* 73, 1103-1124.

Kleibergen, F. and S. Mavroeidis (2009). Weak instrument robust tests in GMM and the new Keynesian Phillips curve. *Journal of Business and Economic Statistics* 27, 293-311.

Kleibergen, F. (2011). Improved accuracy of weak instrument robust GMM statistics through bootstrap and Edgeworth approximations, working paper, Brown University.

Koenker, R. and J.A.F. Machado (1999). GMM inference when the number of moment conditions is large. Journal of Econometrics 93, 327-344.

Kruiniger, H. (2009). GMM estimation and inference in dynamic panel data models with persistent data. *Econometric Theory* 25, 1348-1391.

Lancaster, T. (2002), Orthogonal parameters and panel data, *Review of Economic Studies* 69, 647-666.

Newey, W.K. and R.J. Smith (2004). Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators. *Econometrica* 72, 219-255.

Newey, W.K. and F. Windmeijer (2009). Generalized method of moments with many weak moment conditions. *Econometrica* 77, 687-719.

Newey, W.K. and K.D. West (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review* 28, 777-787.

Nickell, S. (1981). Biases in Dynamic Models with Fixed Effects. *Econometrica* 49, 1417-1426.

Roodman, D. (2009). A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics* 71, 135-158.

Staiger, D. and J.H. Stock (1997). Instrumental Variables regression with weak instruments. *Econometrica* 65, 557-586.

Stock, J.H. and J.H. Wright (2000). GMM with weak identification. *Econometrica* 68, 1055-1096.

Stock, J.H., Wright, J.H. and M. Yogo (2002). A survey of weak instruments and weak identification in Generalized Method of Moments, *Journal of Business & Economic Statistics*, 518-529.

Windmeijer, F. (2005), A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics* 126, 25-517.

Ziliak, J.P. (1997). Efficient estimation with panel data when instruments are predetermined: An empirical comparison of moment-condition estimators. *Journal of Business & Economic Statistics* 15, 419-431.

Table 1: actual rejection frequencies under H_0 for the AR(1) model

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
T	4	4	9	9	4	4	9	9
N	100	250	100	250	100	250	100	250
γ	0.50	0.50	0.50	0.50	0.99	0.99	0.99	0.99
DIF LM I U	0.060	0.055	0.070	0.058	0.254	0.259	0.562	0.714
DIF LM I C	0.079	0.063	0.268	0.109	0.298	0.279	0.811	0.796
DIF LM II C	0.067	0.063	0.073	0.057	0.110	0.095	0.199	0.196
AS LM I U	0.053	0.058	0.059	0.053	0.060	0.059	0.066	0.056
AS LM I C	0.083	0.066	0.300	0.123	0.085	0.068	0.326	0.103
AS LM II C	0.069	0.057	0.099	0.067	0.064	0.050	0.099	0.063
SYS LM I U	0.047	0.053	0.063	0.052	0.067	0.067	0.069	0.055
SYS LM I C	0.074	0.063	0.313	0.119	0.105	0.075	0.338	0.114
SYS LM II C	0.060	0.053	0.098	0.071	0.077	0.071	0.111	0.080
DIF KLM I U	0.057	0.056	0.063	0.046	0.042	0.056	0.066	0.063
DIF KLM I C	0.074	0.062	0.230	0.094	0.059	0.063	0.268	0.108
DIF KLM II C	0.058	0.059	0.067	0.052	0.051	0.060	0.068	0.056
AS KLM I U	0.049	0.050	0.057	0.060	0.051	0.059	0.061	0.060
AS KLM I C	0.075	0.060	0.317	0.127	0.075	0.071	0.307	0.121
AS KLM II C	0.063	0.058	0.089	0.064	0.069	0.059	0.104	0.066
SYS KLM I U	0.055	0.057	0.065	0.061	0.054	0.043	0.063	0.062
SYS KLM I C	0.086	0.068	0.325	0.130	0.075	0.053	0.322	0.120
SYS KLM II C	0.063	0.059	0.092	0.066	0.063	0.050	0.097	0.061
DIF AR I U	0.044	0.051	0.032	0.047	0.049	0.052	0.039	0.055
DIF AR I C	0.086	0.063	0.705	0.245	0.091	0.063	0.709	0.246
DIF AR II C	0.059	0.053	0.085	0.059	0.062	0.054	0.082	0.055
AS AR I U	0.050	0.060	0.033	0.056	0.047	0.057	0.033	0.056
AS AR I C	0.110	0.079	0.878	0.333	0.115	0.076	0.876	0.337
AS AR II C	0.077	0.067	0.168	0.080	0.078	0.058	0.165	0.083
SYS AR I U	0.048	0.057	0.032	0.054	0.053	0.058	0.032	0.062
SYS AR I C	0.117	0.079	0.900	0.347	0.121	0.079	0.893	0.345
SYS AR II C	0.074	0.061	0.188	0.096	0.084	0.061	0.178	0.090

Note: $\sigma_\eta^2 = \sigma_\varepsilon^2 = 1$, 2000 replications, nominal level is 0.05.
(U) C means covariance matrix estimation based on (un)centered moments.
I and II refer to using all moments and only a subset of moments respectively.

Table 2: size-corrected power, $H_0 : \gamma = 0.5, T = 9, N = 250$

γ_0	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99
DIF LM I U	1.000	1.000	1.000	1.000	0.763	0.050	0.324	0.794	0.904	0.633	0.096
DIF LM I C	1.000	1.000	1.000	1.000	0.759	0.050	0.327	0.805	0.916	0.647	0.093
DIF LM II C	1.000	1.000	1.000	0.998	0.644	0.050	0.331	0.655	0.504	0.083	0.049
AS LM I U	1.000	1.000	1.000	1.000	0.893	0.050	0.797	0.998	1.000	0.999	0.385
AS LM I C	1.000	1.000	1.000	1.000	0.893	0.050	0.794	0.999	1.000	1.000	0.544
AS LM II C	1.000	1.000	1.000	0.998	0.610	0.050	0.361	0.822	0.952	0.945	0.353
SYS LM I U	1.000	1.000	1.000	1.000	0.898	0.050	0.818	0.997	0.997	0.916	0.193
SYS LM I C	1.000	1.000	1.000	1.000	0.898	0.050	0.826	0.998	0.999	0.960	0.341
SYS LM II C	1.000	1.000	1.000	0.998	0.674	0.050	0.622	0.974	0.993	0.963	0.326
DIF KLM I U	1.000	1.000	1.000	0.999	0.677	0.050	0.523	0.908	0.955	0.797	0.074
DIF KLM I C	1.000	1.000	1.000	0.999	0.682	0.050	0.526	0.911	0.961	0.811	0.073
DIF KLM II C	1.000	1.000	1.000	0.996	0.603	0.050	0.407	0.732	0.618	0.164	0.043
AS KLM I U	1.000	1.000	1.000	1.000	0.879	0.050	0.817	0.999	1.000	1.000	1.000
AS KLM I C	1.000	1.000	1.000	1.000	0.886	0.050	0.820	0.999	1.000	1.000	1.000
AS KLM II C	1.000	1.000	1.000	0.993	0.556	0.050	0.423	0.882	0.989	0.999	1.000
SYS KLM I U	1.000	1.000	1.000	1.000	0.885	0.050	0.825	0.999	1.000	1.000	1.000
SYS KLM I C	1.000	1.000	1.000	1.000	0.884	0.050	0.829	1.000	1.000	1.000	1.000
SYS KLM II C	1.000	1.000	1.000	0.998	0.712	0.050	0.591	0.985	1.000	1.000	1.000
DIF AR I U	1.000	1.000	1.000	0.859	0.188	0.050	0.128	0.357	0.560	0.386	0.050
DIF AR I C	1.000	1.000	1.000	0.859	0.188	0.050	0.128	0.357	0.560	0.386	0.050
DIF AR II C	1.000	1.000	1.000	0.953	0.291	0.050	0.183	0.408	0.332	0.092	0.047
AS AR I U	1.000	1.000	1.000	0.985	0.334	0.050	0.232	0.831	0.997	1.000	1.000
AS AR I C	1.000	1.000	1.000	0.985	0.334	0.050	0.232	0.831	0.997	1.000	1.000
AS AR II C	1.000	1.000	1.000	0.879	0.200	0.050	0.149	0.511	0.855	0.978	0.999
SYS AR I U	1.000	1.000	1.000	0.989	0.343	0.050	0.246	0.844	0.998	1.000	1.000
SYS AR I C	1.000	1.000	1.000	0.989	0.343	0.050	0.246	0.844	0.998	1.000	1.000
SYS AR II C	1.000	1.000	1.000	0.934	0.272	0.050	0.212	0.793	0.990	1.000	1.000

Note: $\sigma_\eta^2 = \sigma_\varepsilon^2 = 1$, 2000 replications, nominal level is 0.05.
(U) C means covariance matrix estimation based on (un)centered moments.
I and II refer to using all moments and only a subset of moments respectively.

Table 3. DGP A*: Rejection probabilities ($\alpha = 0.05$) Wald, (K)LM test, AR for $H_0 : \gamma = \gamma_0, \beta = \beta_0$

			AB						BB				
$N = 200$	Set	L	W^{cue}	W_c^{cue}	LM	KLM	AR	L	W^{cue}	W_c^{cue}	LM	KLM	AR
$T = 3$	I	8	0.154	0.109	0.064	0.042	0.028	13	0.289	0.173	0.076	0.062	0.041
	II	6	0.108	0.081	0.060	0.042	0.034	11	0.254	0.158	0.060	0.048	0.046
	I^{coll}	5	0.098	0.084	0.054	0.040	0.035	8	0.166	0.122	0.056	0.048	0.042
	II^{coll}	3	0.062	0.061	0.048	0.043	0.040	6	0.129	0.106	0.046	0.040	0.041
$T = 7$	I	48	0.452	0.206	0.096	0.044	0.018	61	0.679	0.414	0.118	0.143	0.138
	II	18	0.190	0.098	0.068	0.044	0.030	31	0.466	0.252	0.100	0.076	0.074
	I^{coll}	13	0.136	0.086	0.050	0.044	0.024	16	0.201	0.127	0.077	0.063	0.055
	II^{coll}	3	0.057	0.056	0.037	0.039	0.043	6	0.105	0.091	0.057	0.054	0.038
$N = 400$	Set	L	W^{cue}	W_c^{cue}	LM	KLM	AR	L	W^{cue}	W_c^{cue}	LM	KLM	AR
$T = 3$	I	8	0.110	0.070	0.059	0.046	0.040	13	0.193	0.108	0.065	0.056	0.041
	II	6	0.085	0.062	0.050	0.044	0.043	11	0.166	0.095	0.060	0.047	0.037
	I^{coll}	5	0.079	0.064	0.052	0.046	0.044	8	0.115	0.078	0.060	0.054	0.036
	II^{coll}	3	0.060	0.055	0.046	0.043	0.052	6	0.100	0.074	0.051	0.044	0.034
$T = 7$	I	48	0.244	0.092	0.097	0.039	0.031	61	0.379	0.176	0.107	0.091	0.102
	II	18	0.128	0.066	0.061	0.044	0.034	31	0.255	0.114	0.086	0.056	0.051
	I^{coll}	13	0.083	0.060	0.046	0.043	0.029	16	0.121	0.081	0.068	0.056	0.035
	II^{coll}	3	0.057	0.054	0.039	0.039	0.048	6	0.082	0.068	0.052	0.048	0.033

*$R = 2500$ simulation replications. Design parameter values: $EVF_x = 0.3$, $IEF_x = 0.3$, $DEN_y = 1.0$, $\bar{\rho}_{x\varepsilon} = 0.15$, $SNR = 5$, $\xi = \gamma = 0.5$, $\sigma_\varepsilon = 1$, $\kappa = 0.5$, $\theta = 1.0$. These yield DGP parameter values: $\pi_\lambda = 0.23$, $\pi_\eta = 0.15$, $\sigma_v = 0.72$, $\sigma_\eta = 0.50$, $\rho_{v\varepsilon} = 0.21$, $\beta = 1.73$.

Table4. DGP B*: Rejection probabilities ($\alpha = 0.05$) Wald, (K)LM test, AR for $H_0 : \gamma = \gamma_0, \beta = \beta_0$

			AB						BB				
$N = 200$	Set	L	W^{cue}	W_c^{cue}	LM	KLM	AR	L	W^{cue}	W_c^{cue}	LM	KLM	AR
$T = 3$	I	8	0.438	0.424	0.339	0.041	0.032	13	0.310	0.290	0.170	0.054	0.047
	II	6	0.368	0.364	0.208	0.043	0.044	11	0.282	0.276	0.138	0.041	0.047
	I^{coll}	5	0.333	0.339	0.147	0.040	0.036	8	0.245	0.237	0.099	0.046	0.042
	II^{coll}	3	0.154	0.164	0.059	0.041	0.041	6	0.190	0.194	0.053	0.037	0.033
$T = 7$	I	48	0.544	0.454	0.833	0.044	0.018	61	0.760	0.657	0.362	0.197	0.143
	II	18	0.279	0.252	0.464	0.045	0.038	31	0.562	0.438	0.249	0.080	0.072
	I^{coll}	13	0.199	0.177	0.266	0.054	0.034	16	0.319	0.278	0.164	0.072	0.055
	II^{coll}	3	0.093	0.087	0.047	0.047	0.050	6	0.187	0.184	0.051	0.052	0.047
$N = 400$	Set	L	W^{cue}	W_c^{cue}	LM	KLM	AR	L	W^{cue}	W_c^{cue}	LM	KLM	AR
$T = 3$	I	8	0.305	0.268	0.284	0.050	0.039	13	0.225	0.204	0.144	0.053	0.040
	II	6	0.254	0.246	0.174	0.047	0.044	11	0.210	0.204	0.117	0.050	0.045
	I^{coll}	5	0.222	0.209	0.115	0.044	0.037	8	0.165	0.150	0.099	0.044	0.031
	II^{coll}	3	0.123	0.126	0.055	0.044	0.048	6	0.135	0.134	0.056	0.039	0.036
$T = 7$	I	48	0.238	0.153	0.856	0.044	0.030	61	0.518	0.386	0.395	0.128	0.117
	II	18	0.108	0.079	0.342	0.042	0.032	31	0.360	0.246	0.213	0.066	0.054
	I^{coll}	13	0.099	0.070	0.178	0.042	0.037	16	0.165	0.132	0.132	0.056	0.042
	II^{coll}	3	0.070	0.065	0.038	0.035	0.043	6	0.104	0.100	0.050	0.046	0.026

*$R = 2500$ simulation replications. Design parameter values: $EVF_x = 0.6$, $IEF_x = 0.3$, $DEN_y = 4.0$, $\bar{\rho}_{x\varepsilon} = 0.30$, $SNR = 3$, $\xi = \gamma = 0.8$, $\sigma_\varepsilon = 1$, $\kappa = 0.5$, $\theta = 1.0$. These yield DGP parameter values: $\pi_\lambda = 0.13$, $\pi_\eta = 0.08$, $\sigma_v = 0.38$, $\sigma_\eta = 0.80$, $\rho_{v\varepsilon} = 0.79$, $\beta = 0.49$.

Appendix

Comparing asymptotic variances We analyze the first-order autoregressive panel data model

$$y_{it} = \gamma y_{i,t-1} + \eta_i + \varepsilon_{it}, \qquad i = 1, ..., N; t = 1, ..., T, \tag{6.1}$$

where we observe y_{i0}. Assume

$$E\left[\varepsilon_{it}|y_i^{t-1}, \eta_i\right] = 0, \tag{6.2}$$

$$E\left[\varepsilon_{it}^2|y_i^{t-1}, \eta_i\right] = \sigma_\varepsilon^2, \tag{6.3}$$

with $y_i^{t-s} = (y_{i0}, ..., y_{it-s})$. Furthermore, we assume:

$$E\left[y_{i0}|\eta_i\right] = \frac{\delta\eta_i}{1-\gamma}, \tag{6.4}$$

$$Var\left[y_{i0}|\eta_i\right] = \frac{\lambda\sigma_\varepsilon^2}{1-\gamma^2}, \tag{6.5}$$

so that $\delta = 1$ results in mean stationarity and $\delta = \lambda = 1$ implies covariance stationarity. Under assumption (6.2) the following $\frac{1}{2}T(T-1)$ linear moment conditions (labelled DIF) hold for each individual i

$$E\left[y_i^{t-2}(\Delta y_{it} - \gamma\Delta y_{it-1})\right] = 0, \qquad t = 2, ..., T. \tag{6.6}$$

We will analyze GMM estimators of γ making use of the DIF moment conditions. In general the estimators can be expressed as

$$\widehat{\gamma}_j = \frac{\Delta y_{-1}' Z \left(Z'\Omega Z\right)^{-1} Z'\Delta y}{\Delta y_{-1}' Z \left(Z'\Omega Z\right)^{-1} Z'\Delta y_{-1}}, \tag{6.7}$$

where Z is the instrument matrix and $\Omega = I_N \otimes \Omega_T$ with $\Omega_T = E\left[\Delta\varepsilon_i \left(\Delta\varepsilon_i\right)'\right]$. Usual asymptotic reasoning for fixed T yields the normal limiting distribution

$$\sqrt{N}\left(\widehat{\gamma} - \gamma\right) \xrightarrow{d} \mathcal{N}[0,\ \Sigma], \tag{6.8}$$

with

$$\begin{aligned}\Sigma &= \left(\operatorname*{plim}_{N\to\infty} \Delta y_{-1}' Z \left(\operatorname*{plim}_{N\to\infty} Z'\Omega Z\right)^{-1} \operatorname*{plim}_{N\to\infty} Z'\Delta y_{-1}\right)^{-1} \\ &= \left(E\left[\Delta y_{i,-1}' Z_i\right] \left(E\left[Z_i'\Delta\varepsilon_i \left(\Delta\varepsilon_i\right)' Z_i\right]\right)^{-1} E\left[Z_i'\Delta y_{i,-1}\right]\right)^{-1}.\end{aligned} \tag{6.9}$$

The instrument matrix for the DIF moments is:

$$Z_{fi} = \begin{pmatrix} y_{i,0} & 0 & 0 & & \cdots & 0 & \cdots & 0 \\ 0 & y_{i,0} & y_{i,1} & & & \vdots & & \vdots \\ \vdots & & \vdots & & \ddots & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & y_{i,0} & \cdots & y_{i,T-2} \end{pmatrix}, \tag{6.10}$$

We consider two different subsets of the DIF moments, i.e. collapsed and lag limits. The lag limits instrument matrix reads:

$$Z_{li} = \begin{pmatrix} y_{i,0} & 0\dots0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0\dots0 & y_{i,T-2} \end{pmatrix}, \tag{6.11}$$

hence a subset of only $(T-1) = O(T)$ instruments is used only. Alternatively, Roodman (2009) proposes to collapse the instrument matrix as

$$Z_{ci} = \begin{pmatrix} y_{i0} & 0 & \cdots & 0 \\ y_{i1} & y_{i0} & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ y_{i,T-2} & y_{i,T-3} & \cdots & y_{i0} \end{pmatrix}. \tag{6.12}$$

The differences between the two implementations are seen from $T = 3$ onwards. When $T = 3$ we have

$$E\left[Z_{li}'\Delta y_{i,-1}\right] = E\begin{pmatrix} y_{i0} & 0 \\ 0 & y_{i1} \end{pmatrix}\begin{pmatrix} \Delta y_{i1} \\ \Delta y_{i2} \end{pmatrix} = \begin{pmatrix} E\left[y_{i0}\Delta y_{i1}\right] \\ E\left[y_{i1}\Delta y_{i2}\right] \end{pmatrix}. \tag{6.13}$$

$$\begin{aligned} E\left[Z_{li}'\Delta\varepsilon_i\left(\Delta\varepsilon_i\right)' Z_{li}\right] &= E\begin{pmatrix} y_{i0} & 0 \\ 0 & y_{i1} \end{pmatrix}\begin{pmatrix} \left(\Delta\varepsilon_{i2}\right)^2 & \Delta\varepsilon_{i2}\Delta\varepsilon_{i3} \\ \Delta\varepsilon_{i2}\Delta\varepsilon_{i3} & \left(\Delta\varepsilon_{i3}\right)^2 \end{pmatrix}\begin{pmatrix} y_{i0} & 0 \\ 0 & y_{i1} \end{pmatrix} \\ &= E\begin{pmatrix} y_{i0}^2\left(\Delta\varepsilon_{i2}\right)^2 & y_{i0}y_{i1}\Delta\varepsilon_{i2}\Delta\varepsilon_{i3} \\ y_{i0}y_{i1}\Delta\varepsilon_{i2}\Delta\varepsilon_{i3} & y_{i1}^2\left(\Delta\varepsilon_{i3}\right)^2 \end{pmatrix} \\ &= \sigma_\varepsilon^2\begin{pmatrix} 2E\left[y_{i0}^2\right] & -E\left[y_{i0}y_{i1}\right] \\ -E\left[y_{i0}y_{i1}\right] & 2E\left[y_{i1}^2\right] \end{pmatrix}, \end{aligned} \tag{6.14}$$

where the last line follows from the fact that (see also Kiviet, 2007):

$$\begin{aligned} E\left[y_{i,t-2}^2\left(\Delta\varepsilon_{it}\right)^2\right] &= E\left[E_{t-2}\left[y_{i,t-2}^2\left(\Delta\varepsilon_{it}\right)^2\right]\right] \\ &= E\left[y_{i,t-2}^2 E_{t-2}\left[\left(\Delta\varepsilon_{it}\right)^2\right]\right] \\ &= 2\sigma_\varepsilon^2 E\left[y_{i,t-2}^2\right], \end{aligned} \tag{6.15}$$

$$\begin{aligned} E\left[y_{i,t-3}y_{i,t-2}\Delta\varepsilon_{i,t-1}\Delta\varepsilon_{it}\right] &= E\left[y_{i,t-3}y_{i,t-2}\varepsilon_{i,t-1}\Delta\varepsilon_{it}\right] - E\left[y_{i,t-3}y_{i,t-2}\varepsilon_{i,t-2}\Delta\varepsilon_{it}\right] \\ &= E\left[E_{t-2}\left[y_{i,t-3}y_{i,t-2}\varepsilon_{i,t-1}\Delta\varepsilon_{it}\right]\right] - E\left[E_{t-2}\left[y_{i,t-3}y_{i,t-2}\varepsilon_{i,t-2}\Delta\varepsilon_{it}\right]\right] \\ &= E\left[y_{i,t-3}y_{i,t-2}E_{t-2}\left[\varepsilon_{i,t-1}\Delta\varepsilon_{it}\right]\right] - E\left[y_{i,t-3}y_{i,t-2}\varepsilon_{i,t-2}E_{t-2}\left[\Delta\varepsilon_{it}\right]\right] \\ &= -\sigma_\varepsilon^2 E\left[y_{i,t-3}y_{i,t-2}\right], \end{aligned} \tag{6.16}$$

In general, we have (Hayakawa and Nagata, 2012):

$$y_{it} = (1 - (1-\delta)\gamma^t)\frac{\eta_i}{1-\gamma} + \sum_{j=0}^{t-1} \gamma^j \varepsilon_{i,t-j} + \gamma^t \varepsilon_{i0}, \tag{6.17}$$

hence we can write for $s \leq t$:

$$E\left[y_{is}, y_{it}\right] = (1-(1-\delta)\gamma^s)(1-(1-\delta)\gamma^t)\frac{\sigma_\eta^2}{(1-\gamma)^2} + (\gamma^{t-s}(1-(1-\lambda)\gamma^{2s})\frac{\sigma_\varepsilon^2}{1-\gamma^2}. \tag{6.18}$$

Substituting for appropriate chosen s and t this expression into the various elements of Σ we can express the asymptotic variance as a function of $\gamma, \sigma_\eta^2, \sigma_\varepsilon^2, \delta$ and λ.

As an example, assume covariance stationarity ($\delta = \lambda = 1$). We can write in this special case:

$$E\left[y_{it}^2\right] = \frac{\sigma_\eta^2}{(1-\gamma)^2} + \frac{\sigma_\varepsilon^2}{1-\gamma^2}, \tag{6.19}$$

$$\begin{aligned} E\left[y_{it}y_{i,t-1}\right] &= \gamma E\left[y_{i,t-1}^2\right] + E\left[\eta_i y_{i,t-1}\right] + E\left[\varepsilon_{it} y_{i,t-1}\right] \\ &= \gamma\left(\frac{\sigma_\eta^2}{(1-\gamma)^2} + \frac{\sigma_\varepsilon^2}{1-\gamma^2}\right) + \frac{\sigma_\eta^2}{1-\gamma} + 0 \\ &= \frac{\sigma_\eta^2}{(1-\gamma)^2} + \frac{\gamma\sigma_\varepsilon^2}{1-\gamma^2} \end{aligned} \tag{6.20}$$

$$\begin{aligned} E\left[y_{i,t-1}\Delta y_{it}\right] &= E\left[y_{i,t-1}y_{it}\right] - E\left[y_{i,t-1}^2\right] \\ &= (\gamma-1)E\left[y_{i,t-1}^2\right] + E\left[\eta_i y_{i,t-1}\right] + E\left[\varepsilon_{it} y_{i,t-1}\right] \\ &= (\gamma-1)\left(\frac{\sigma_\eta^2}{(1-\gamma)^2} + \frac{\sigma_\varepsilon^2}{1-\gamma^2}\right) + \frac{\sigma_\eta^2}{1-\gamma} + 0 \\ &= -\frac{\sigma_\varepsilon^2}{1+\gamma}. \end{aligned} \tag{6.21}$$

Collecting terms we find for the asymptotic variance of the GMM estimator using lag limits the following for $T = 3$:

$$\begin{aligned} \Sigma_l &= \left(\left(-\frac{\sigma_\varepsilon^2}{1+\gamma}\right)^2 \frac{1}{\sigma_\varepsilon^2}\begin{pmatrix} 1 & 1 \end{pmatrix}\begin{pmatrix} 2\left(\frac{\sigma_\eta^2}{(1-\gamma)^2} + \frac{\sigma_\varepsilon^2}{1-\gamma^2}\right) & -\left(\frac{\sigma_\eta^2}{(1-\gamma)^2} + \frac{\gamma\sigma_\varepsilon^2}{1-\gamma^2}\right) \\ -\left(\frac{\sigma_\eta^2}{(1-\gamma)^2} + \frac{\gamma\sigma_\varepsilon^2}{1-\gamma^2}\right) & 2\left(\frac{\sigma_\eta^2}{(1-\gamma)^2} + \frac{\sigma_\varepsilon^2}{1-\gamma^2}\right) \end{pmatrix}^{-1}\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right)^{-1} \\ &= \frac{\gamma+1}{2\gamma^2\sigma_\varepsilon^2 - 4\gamma\sigma_\varepsilon^2 + 2\sigma_\varepsilon^2}\left(\gamma^2\sigma_\varepsilon^2 - 3\gamma\sigma_\varepsilon^2 + \gamma\sigma_\eta^2 + 2\sigma_\varepsilon^2 + \sigma_\eta^2\right) \end{aligned} \tag{6.22}$$

Regarding the collapsed implementation we find:

$$E\left[Z_{ci}'\Delta y_{i,-1}\right] = E\begin{pmatrix} y_{i0} & y_{i1} \\ 0 & y_{i0} \end{pmatrix}\begin{pmatrix} \Delta y_{i1} \\ \Delta y_{i2} \end{pmatrix} = \begin{pmatrix} E\left[y_{i0}\Delta y_{i1}\right] + E\left[y_{i1}\Delta y_{i2}\right] \\ E\left[y_{i0}\Delta y_{i2}\right] \end{pmatrix}. \tag{6.23}$$

$$E\left[Z_{ci}'\Delta\varepsilon_i\left(\Delta\varepsilon_i\right)'Z_{ci}\right] = E\begin{pmatrix} y_{i0} & y_{i1} \\ 0 & y_{i0}\end{pmatrix}\begin{pmatrix}(\Delta\varepsilon_{i2})^2 & \Delta\varepsilon_{i2}\Delta\varepsilon_{i3} \\ \Delta\varepsilon_{i2}\Delta\varepsilon_{i3} & (\Delta\varepsilon_{i3})^2\end{pmatrix}\begin{pmatrix} y_{i0} & 0 \\ y_{i1} & y_{i0}\end{pmatrix}$$
$$= \sigma_\varepsilon^2\begin{pmatrix} 2E\left[y_{i0}^2\right] - 2E\left[y_{i0}y_{i1}\right] + 2E\left[y_{i1}^2\right] & -E\left[y_{i0}^2\right] + 2E\left[y_{i0}y_{i1}\right] \\ -E\left[y_{i0}^2\right] + 2E\left[y_{i0}y_{i1}\right] & 2E\left[y_{i0}^2\right]\end{pmatrix}. \tag{6.24}$$

Under covariance stationarity, we can furthermore derive that:

$$E\left[y_{i,t-2}\Delta y_{it}\right] = \gamma E\left[y_{i,t-2}\Delta y_{i,t-1}\right] + E\left[y_{i,t-2}\Delta\varepsilon_{it}\right]$$
$$= -\frac{\gamma\sigma_\varepsilon^2}{1+\gamma}. \tag{6.25}$$

Collecting terms we find for the asymptotic variance of the GMM estimator using collapsed instruments the following expression for $T = 3$:

$$\Sigma_c = \left(\left(-\frac{\sigma_\varepsilon^2}{1+\gamma}\right)^2\frac{1}{\sigma_\varepsilon^2}\begin{pmatrix} 2 & \gamma\end{pmatrix}\begin{pmatrix}\left(\frac{2\sigma_\eta^2}{(1-\gamma)^2} + \frac{2(2-\gamma)\sigma_\varepsilon^2}{1-\gamma^2}\right) & \left(\frac{\sigma_\eta^2}{(1-\gamma)^2} + \frac{(2\gamma-1)\sigma_\varepsilon^2}{1-\gamma^2}\right) \\ \frac{\sigma_\eta^2}{(1-\gamma)^2} + \frac{(2\gamma-1)\sigma_\varepsilon^2}{1-\gamma^2} & 2\left(\frac{\sigma_\eta^2}{(1-\gamma)^2} + \frac{\sigma_\varepsilon^2}{1-\gamma^2}\right)\end{pmatrix}^{-1}\begin{pmatrix}2 \\ \gamma\end{pmatrix}\right)^{-1}. \tag{6.26}$$

In a similar way the asymptotic variance of the GMM estimator using the full set of DIF moments can be expressed as function of γ, σ_η^2 and σ_ε^2.

Simulation cases Case I

$$Z_{si} = \begin{pmatrix} Z_{di} & O \\ O & Z_{li}\end{pmatrix}, \; Z_{di} = \begin{bmatrix} I_3 & Z_{di}^w\end{bmatrix},$$

$$Z_{di}^w = \begin{pmatrix} w_i^{0\prime} & 0' \\ 0' & w_i^{1\prime}\end{pmatrix},$$

$$Z_{li} = \begin{pmatrix} 1 & 0' & 0' \\ 1 & \triangle w_{i1}' & 0' \\ 1 & 0' & \triangle w_{i2}'\end{pmatrix}.$$

Case II

$$Z_{si} = \begin{pmatrix} Z_{di} & O \\ O & Z_{li}\end{pmatrix}, \; Z_{di} = \begin{bmatrix} I_3 & Z_{di}^w\end{bmatrix},$$

$$Z_{di}^w = \begin{pmatrix} w_{i0}' & 0' \\ 0' & w_{i1}'\end{pmatrix},$$

$$Z_{li} = \begin{pmatrix} 1 & 0' & 0' \\ 1 & \triangle w_{i1}' & 0' \\ 1 & 0' & \triangle w_{i2}'\end{pmatrix}.$$

Case III

$$Z_{si} = \begin{pmatrix} Z_{di} & O \\ O & Z_{li} \end{pmatrix}, \ Z_{di} = \begin{bmatrix} \iota_3 & Z_{di}^w \end{bmatrix},$$

$$Z_{di}^w = \begin{pmatrix} w_{i0}' & 0' \\ w_{i1}' & w_{i0}' \end{pmatrix},$$

$$Z_{li} = \begin{pmatrix} 1 & 0' \\ 1 & \triangle w_{i1}' \\ 1 & \triangle w_{i2}' \end{pmatrix}.$$

Case IV

$$Z_{si} = \begin{pmatrix} Z_{di} & O \\ O & Z_{li} \end{pmatrix}, \ Z_{di} = \begin{bmatrix} \iota_3 & Z_{di}^w \end{bmatrix},$$

$$Z_{di}^w = \begin{pmatrix} w_{i1}' \\ w_{i2}' \end{pmatrix},$$

$$Z_{li} = \begin{pmatrix} 1 & 0' \\ 1 & \triangle w_{i1}' \\ 1 & \triangle w_{i2}' \end{pmatrix}.$$