UvA ✘✘✘ ECONOMETRICS

Discussion Paper: 2014/09

# Accuracy and efficiency of various GMM inference techniques in dynamic micro panel data models

## Jan F. Kiviet, Milan Pleus and Rutger Poldermans

UvA ✘✘✘ UNIVERSITEIT VAN AMSTERDAM

# Accuracy and efficiency of various GMM inference techniques in dynamic micro panel data models

## Jan F. Kiviet*†, Milan Pleus*‡, Rutger Poldermans*‡

30 December 2014

## Abstract

The performance in finite samples is examined of inference obtained by variants of the Arellano-Bond and the Blundell-Bond GMM estimation techniques for single dynamic panel data models with possibly endogenous regressors and cross-sectional heteroskedasticity. By simulation the effects are examined of using particular instrument strength enhancing reductions and transformations of the matrix of instrumental variables, of less robust implementations of the GMM weighting matrix, and also of corrections to the standard asymptotic variance estimates. We compare the root mean squared errors of the coefficient estimators and also the size of tests on coefficient values and of different implementations of overidentification restriction tests. Also the size and power of tests on the validity of the additional orthogonality conditions exploited by the Blundell-Bond technique are assessed over a pretty wide grid of relevant cases. Surprisingly, particular asymptotically optimal and relatively robust weighting matrices are found to be superior in finite samples to ostensibly more appropriate versions. Most of the variants of tests for overidentification restrictions show serious deficiencies. A recently developed modification of GMM is found to have great potential when the cross-sectional heteroskedasticity is pronounced and the time-series dimension of the sample not too small. Finally all techniques are employed to actual data and lead to some profound insights.

*Amsterdam School of Economics & Tinbergen Institute, University of Amsterdam, Valckenierstraat 65, 1018 XE Amsterdam, The Netherlands; j.f.kiviet@uva.nl, m.pleus@uva.nl, r.w.poldermans@uva.nl
†Emeritus Professor at University of Amsterdam, Visiting Professor at the Division of Economics, School of Humanities and Social Sciences, Nanyang Technological University, 14 Nanyang Drive, Singapore 637332; jfkiviet@ntu.edu.sg

1

# 1 Introduction

One of the major attractions of analyzing panel data rather than single indexed variables is that they allow to cope with the empirically very relevant situation of unobserved heterogeneity correlated with included regressors. Econometric analysis of dynamic relationships on the basis of panel data, where the number of surveyed individuals is relatively large while covering just a few time periods, is very often based on GMM (generalized method of moments). Its reputation is built on its claimed flexibility, generality, ease of use, robustness and efficiency. Widely available standard software enables to estimate models including exogenous, predetermined and endogenous regressors consistently, while allowing for semiparametric approaches regarding the presence of heteroskedasticity and the type of distribution of the disturbances. This software also provides specification checks regarding the adequacy of the internal and external instrumental variables employed and the specific assumptions made regarding (absence of) serial correlation.

Popular are especially the GMM implementations put forward by Arellano and Bond (1991). However, practical problems have often been reported, such as vulnerability due to the abundance of internal instruments, discouraging improvements of 2-step over 1-step GMM findings, poor size control of test statistics, and weakness of instruments especially when the dynamic adjustment process is slow (a root is close to unity). As remedies it has been suggested to reduce the number of instruments by renouncing some valid orthogonality conditions, but also to extend the number of instruments by adopting more orthogonality conditions. Extra orthogonality conditions can be based on certain homoskedasticity or stationarity assumptions or initial value conditions, see Blundell and Bond (1998). By abandoning weak instruments finite sample bias may be reduced, whereas by extending the instrument set with a few strong ones the bias may be further reduced and the efficiency enhanced. Presently, it is not clear yet how practitioners can best make use of these suggestions, because no set of preferred testing tools is yet available that allows to select instruments by assessing both their validity and their strength, and to classify individual regressors accurately as either endogenous, predetermined or strictly exogenous. Therefore it happens often that in applied research models and techniques are selected simply on the basis of the perceived significance and plausibility of their coefficient estimates, whereas it is well known that imposing invalid coefficient restrictions and employing regressors wrongly as instruments will often lead to relatively small estimated standard errors, which then contain misleading information on the actual precision of often seriously biased estimators.

The available studies on the performance of alternative inference techniques for dynamic panel data models have obvious limitations when it comes to advising practitioners on the most effective implementations of estimators and tests under general circumstances. As a rule, they do not consider various empirically relevant issues in conjunction, such as: ($i$) occurrence and the possible endogeneity of regressors additional to the lagged dependent variable, ($ii$) occurrence of individual effect (non-)stationarity of both the lagged dependent variable and other regressors, ($iii$) cross-section and/or time-series heteroskedasticity of the idiosyncratic disturbances, and ($iv$) variation in signal-to-noise ratios and in the relative prominence of individual effects. For example: the simulation results in Arellano and Bover (1995), Hahn and Kuersteiner (2002), Alvarez and Arellano (2003), Hahn et al. (2007), Kiviet (2007), Kruiniger (2008), Okui (2009), Roodman

(2009b), Hayakawa (2009) and Han and Phillips (2013) just concern the panel AR(1) model under homoskedasticity. Although an extra regressor is included in the simulation studies in Arellano and Bond (1991), Kiviet (1995), Bowsher (2002), Hsiao et al. (2002), Bond and Windmeijer (2005), Bun and Carree (2005a,b), Bun and Kiviet (2006), Gouriéroux et al. (2010), Hayakawa (2010), Dhaene and Jochmans (2012), Flannery and Watson Hankins (2013), Everaert (2013) and Kripfganz and Schwarz (2013), this regressor is (weakly-)exogenous and most experiments just concern homoskedastic disturbances and stationarity regarding the impact of individual effects. Blundell et al. (2001) and Bun and Sarafidis (2013) include an endogenous regressor, but their design does not allow to control the degree of simultaneity; moreover, they stick to homoskedasticity. Harris et al. (2009) only examine the effects of neglected endogeneity. Heteroskedasticity is considered in a few simulation experiments in Arellano and Bond (1991) in the model with an exogenous regressor, and just for the panel AR(1) case in Blundell and Bond (1998). Windmeijer (2005) analyzes panel GMM with heteroskedasticity, but without including a lagged dependent variable in the model. Bun and Carree (2006) and Juodis (2013) examine effects of heteroskedasticity in the model with a lagged dependent and a strictly exogenous regressor under stationarity regarding the effects. So, we have to conclude that knowledge is still scarce with respect to the performance of GMM when it is not only needed to cope with genuine simultaneity (which we consider to be the core of econometrics), but also because of occurrence of heteroskedasticity of unknown form. Moreover, many of the simulation studies mentioned above did not systematically explore the effects of relevant nuisance parameter values on the finite sample distortions to asymptotic approximations. Regarding the performance of tests on the validity of instruments worrying results have been obtained in Bowsher (2002) and Roodman (2009b) for homoskedastic models. On the other hand Bun and Sarafidis (2013) report reassuring results, but these just concern models where $T = 3$. Our grid of examined cases will be much wider and have more dimensions. Moreover, we will deliberately explore both feasible and unfeasible versions of estimators and test statistics (in unfeasible versions particular nuisance parameter values are assumed to be known). Therefore we will be able to draw more useful conclusions on what aspects do have major effects on any inference inaccuracies in finite samples.

The data generating process designed here can be simulated for classes of models which may include individual and time effects, a lagged dependent variable regressor and another regressor which may be correlated with these and other individual effects and be either strictly exogenous or jointly dependent with regard to the idiosyncratic disturbances, whereas the latter may show a form of cross-section heteroskedasticity associated with both the individual effects. For a range of relevant parameter values we will verify in moderately large samples the properties of alternative GMM estimators, both 1-step and 2-step, focussing on alternative implementations regarding the weighting matrix and corresponding corrections to variance estimates according to the often practiced approach by Windmeijer (2005). This will include variants of the popular system estimator, which exploit as instruments the first-difference of lagged internal variables for the untransformed model in addition to lagged level internal variables as instruments for the model from which the individual effects have been removed. We will examine cases where the extra instruments are valid, and where they are not, to verify whether particular tests for overidentification restrictions have appropriate size and power, such that with reasonable probabilities valid instruments will be recognized as appropriate

and invalid instruments will be detected and discarded. Moreover, following Kiviet and Feng (2014), we shall investigate a rather novel modification of the traditional GMM implementation which aims at improving the strength of the exploited instruments in the presence of heteroskedasticity. Of course, also the simulation design used here has its limitations. It has only one extra regressor next to the lagged dependent variable, we only consider cross-sectional heteroskedasticity, and all basic random terms have been drawn from the normal distribution. Moreover, the design does not accommodate general forms of cross-sectional dependence between error terms. However, by including individual and time specific effects particular simple forms of cross-sectional dependence are accommodated.

The structure of this study is as follows. In Section 2 we first present the major issues regarding IV and GMM coefficient and variance estimation in linear models and on inference techniques on establishing instrument validity and regarding the coefficient values by standard and by corrected test statistics. Next in Section 3 the generic results of Section 2 are used to discuss in more detail than provided elsewhere the various options for their implementation in linear models for single dynamic simultaneous micro econometric panel data relationships with both individual and time effects and some form of cross-sectional heteroskedasticity. In Section 4 the Monte Carlo design is developed to analyze and compare the performance of alternative often asymptotically equivalent inference methods in finite samples of empirically relevant parametrizations. Section 5 summarizes the simulation results, from which some preferred techniques for use in finite samples of particular models emerge, plus a warning regarding particular types of models that require more refined methods yet to be developed. Next, in Section 6, we examine whether these techniques yield new insights when employed to empirical data on life-cycle labor supply which have been analyzed already extensively in the literature. Finally Section 7 concludes.

# 2    Basic GMM results for linear models

Here we present concisely the major generic results on IV and GMM inference for single indexed data that could either represent time-series or a cross-section. First we define the model and estimators, discuss some of their special properties and consider specific test situations. From these general findings for linear regressions the examined implementations for specific linear dynamic panel data models follow easily in Section 3.

## 2.1    Model and estimators

Let the scalar dependent variable $y_i$ depend linearly on $K$ regressors $x_i$ and an unobserved disturbance term $u_i$, and let there be $L \geq K$ variables $z_i$ (the instruments) that establish orthogonality conditions such that

$$\left. \begin{array}{l} y_i = x_i'\beta_0 + u_i \\ E[z_i(y_i - x_i'\beta_0)] = 0 \end{array} \right\} \ i = 1, ..., n. \tag{2.1}$$

Here $x_i$ and $\beta_0$ are $K \times 1$ vectors, $\beta_0$ containing the true values of the unknown coefficients, and $z_i$ is an $L \times 1$ vector. Applying the analogy principle, the method of moments

(MM) aims to find an estimator for model parameter $\beta$ by solving the $L$ sample moment equations

$$n^{-1} \sum\nolimits_{i=1}^{n} z_i(y_i - x_i'\hat{\beta}) = 0. \tag{2.2}$$

Generally, these have a unique solution only when $L = K$ and then yield

$$\hat{\beta} = \left( \sum\nolimits_{i=1}^{n} z_i x_i' \right)^{-1} \sum\nolimits_{i=1}^{n} z_i y_i, \tag{2.3}$$

provided the inverse exists. For $L > K$ the MM recipe to find a unique estimator is: minimize with respect to $\beta$ the criterion function

$$\sum\nolimits_{j=1}^{n} [(y_j - x_j'\beta)z_j']G \sum\nolimits_{i=1}^{n} [z_i(y_i - x_i'\beta)]$$

for some weighting matrix $G$. It can be shown that the asymptotically optimal choice for $G$ is an expression which has a probability limit that is proportional to the inverse of the asymptotic variance $V$ of

$$n^{-1/2} \sum\nolimits_{i=1}^{n} z_i u_i \xrightarrow{d} N(0, V).$$

When $u_i \sim iid(0, \sigma_u^2)$ an optimal choice for $G$ is proportional to the inverse of $\sum_{i=1}^{n} z_i z_i'$ and the MM estimator is

$$\hat{\beta}_{IV} = [X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}Z'y, \tag{2.4}$$

where $y = (y_1 \cdots y_n)'$, $X = (x_1 \cdots x_N)'$ and $Z = (z_1 \cdots z_N)'$. But, when $E(z_i u_i) = 0$ while $u = (u_1 \cdots u_n)' \sim (0, \sigma_u^2 \Omega)$, where $\Omega$ has full rank and without loss of generality $tr(\Omega) = n$, the optimal choice for $G$ is a matrix proportional to $(Z'\Omega Z)^{-1}$, yielding MM estimator

$$\hat{\beta}_{GMM} = [X'Z(Z'\Omega Z)^{-1}Z'X]^{-1}X'Z(Z'\Omega Z)^{-1}Z'y. \tag{2.5}$$

Note that for $\Omega = I$ the latter formula simplifies to $\hat{\beta}_{IV}$. When $L = K$ both $\hat{\beta}_{GMM}$ and $\hat{\beta}_{IV}$ simplify to (2.3) or $(Z'X)^{-1}Z'y$.

When $\Omega$ is unknown and therefore (2.5) is unfeasible, one should use an informed guess $\Omega^{(0)}$ to obtain the 1-step estimator $\hat{\beta}_{GMM}^{(1)}$, which is sub-optimal when $\Omega^{(0)} \neq \Omega$, though consistent under the assumptions made. Then the residuals

$$\hat{u}^{(1)} = y - X\hat{\beta}_{GMM}^{(1)} \tag{2.6}$$

are consistent for $u$, thus from them it should be possible to obtain an expression $\hat{\Omega}^{(1)}$ such that $\text{plim}\, n^{-1}(Z'\hat{\Omega}^{(1)}Z - Z'\Omega Z) = O$. Substituting in (2.5) yields the 2-step estimator

$$\hat{\beta}_{GMM}^{(2)} = [X'Z(Z'\hat{\Omega}^{(1)}Z)^{-1}Z'X]^{-1}X'Z(Z'\hat{\Omega}^{(1)}Z)^{-1}Z'y, \tag{2.7}$$

which is asymptotically equivalent to $\hat{\beta}_{GMM}$ and thus asymptotically optimal, given the $L$ instruments used.

## 2.2 Some algebraic peculiarities

Defining $P_Z = Z(Z'Z)^{-1}Z'$ and $\hat{X} = P_Z X$ one finds $\hat{\beta}_{IV} = (\hat{X}'\hat{X})^{-1}\hat{X}'y$, which highlights its two-stage least-squares character. Now suppose that $X = (X_1\ X_2)$ and $Z = (Z_1$

$Z_2$) with $Z_2 = X_2$ whereas $X\beta = X_1\beta_1 + X_2\beta_2$, where $\beta_1$ and $\beta_2$ have $K_1$ and $K_2$ elements respectively. Standard results on partitioned regression yields

$$\hat{\beta}_{1,IV} = (\hat{X}_1' M_{\hat{X}_2} \hat{X}_1)^{-1} \hat{X}_1' M_{\hat{X}_2} y = (X_1' P_{M_{X_2} Z_1} X_1)^{-1} X_1' P_{M_{X_2} Z_1} y, \qquad (2.8)$$

which is the IV estimator in the regression of $y$ on just $X_1$ using the $L - K_2$ instruments $M_{X_2} Z_1$. This result is known as partialling out the predetermined regressors $X_2$. It follows from $\hat{X}_2 = P_Z X_2 = X_2$ which yields

$$M_{\hat{X}_2} \hat{X}_1 = M_{X_2} P_Z X_1 = M_{X_2}(P_{X_2} + P_{M_{X_2} Z_1}) X_1 = P_{M_{X_2} Z_1} X_1.$$

A similar result is not straight-forwardly available for GMM because of the following.

Let positive definite matrix $\Omega$ be factorized as follows

$$\Omega^{-1} = \Psi' \Psi, \text{ so } \Omega = \Psi^{-1}(\Psi')^{-1}. \qquad (2.9)$$

Now define

$$y^* = \Psi y, \ X^* = \Psi X, \ Z^\dagger = (\Psi')^{-1} Z, \qquad (2.10)$$

then

$$\begin{aligned}
\hat{\beta}_{GMM} &= [X^{*\prime} Z^\dagger (Z^{\dagger\prime} Z^\dagger)^{-1} Z^{\dagger\prime} X^*]^{-1} X^{*\prime} Z^\dagger (Z^{\dagger\prime} Z^\dagger)^{-1} Z^{\dagger\prime} y^* \\
&= (X^{*\prime} P_{Z^\dagger} X^*)^{-1} X^{*\prime} P_{Z^\dagger} y^*,
\end{aligned}$$

so GMM is equivalent to IV using transformed variables, but where $Z$ has been transformed differently. Therefore, if $X_2$ is such that $X_2^*$ establishes valid instruments in the transformed model $y^* = X^* \beta + u^*$, where $u^* \sim (0, \sigma_u^2 I)$, the regressors $X_2^*$ are not used as instruments in GMM in its IV interpretation. They would, though, if one would deliberately choose $Z_2 = \Omega^{-1} X_2$.

As is well-known and easily verified, linear transformations of the matrix of instruments of the form $Z^\diamond = ZC$, where $C$ is a full rank $L \times L$ matrix, have no effect on $\hat{\beta}_{IV}$ nor on $\hat{\beta}_{GMM}$. However, there is not such invariance when the matrix $Z$ is premultiplied by some transformation matrix, and hence not the columns but the rows of $Z$ are directly affected. It has been shown in Kiviet and Feng (2014) that such transformations, chosen in correspondence with the required transformation of the model when $\Omega \neq I$, may lead to modified GMM estimation achieving higher efficiency levels and better results in finite samples than standard GMM, provided the validity of the transformed instruments is maintained. We will examine here the effects of employing transformation $Z^* = \Psi Z$, which provides the modified GMM estimator

$$\hat{\beta}_{MGMM} = [X' \Omega^{-1} Z (Z' \Omega^{-1} Z)^{-1} Z' \Omega^{-1} X]^{-1} X' \Omega^{-1} Z (Z' \Omega^{-1} Z)^{-1} Z' \Omega^{-1} y. \qquad (2.11)$$

When this can be made feasible, it yields $\hat{\beta}_{MGMM}^{(2)}$.

## 2.3 Particular test procedures

Inference on elements of $\beta_0$ based on $\hat{\beta}_{GMM}^{(2)}$ of (2.7) requires an asymptotic approximation to its distribution. Under correct specification the standard first-order approximation is

$$\hat{\beta}_{GMM}^{(2)} \overset{a}{\sim} N(\beta_0, \hat{\sigma}_u^2 [X'Z(Z'\hat{\Omega}^{(1)}Z)^{-1}Z'X]^{-1}). \qquad (2.12)$$

It allows testing general restrictions by Wald-type tests. For an individual coefficient, say $\beta_{0k} = e'_{K,k}\beta_0$, where $e_{K,k}$ is a $K \times 1$ vector with all elements zero except its $k^{th}$ element $(1 \leq k \leq K)$ which is unity, testing $H_0 : \beta_{0k} = \beta_k^0$ and allowing for one-sided alternatives, amounts to comparing test statistic

$$W_{\beta_k} = (e'_{K,k}\hat{\beta}_{GMM}^{(2)} - \beta_k^0)/\{\hat{\sigma}_u^2 e'_{K,k}[X'Z(Z'\hat{\Omega}^{(1)}Z)^{-1}Z'X]^{-1}e_{K,k}\}^{1/2} \quad (2.13)$$

with the appropriate quantile of the standard normal distribution. Note that this test statistic is actually an asymptotic t-test; in finite samples the type I error probability may deviate from the chosen nominal level, also depending on whether $\hat{\sigma}_u^2$ has been obtained from 1-step or from 2-step residuals and any employed loss of degrees of freedom corrections. In fact it has been observed that the consistent estimator of the variance of two-step GMM estimators $\widehat{Var}(\hat{\beta}_{GMM}^{(2)})$ given in (2.12) often underestimates the finite sample variance, because in its derivation the randomness of $\hat{\Omega}^{(1)}$ is not taken into account. Windmeijer (2005) provides a corrected formula $\widehat{Varc}(\hat{\beta}_{GMM}^{(2)})$, see Appendix A, which can be used in the corrected t-test

$$W_{\beta_k}^c = (e'_{K,k}\hat{\beta}_{GMM}^{(2)} - \beta_k^0)/\{e'_{K,k}\widehat{Varc}(\hat{\beta}_{GMM}^{(2)})e_{K,k}\}^{1/2}. \quad (2.14)$$

When $L > K$ the overidentification restrictions can be tested by the Sargan-Hansen statistic

$$J_Z^{(1)} = n(\hat{u}^{(1)\prime}Z(Z'\hat{\Omega}^{(1)}Z)^{-1}Z'\hat{u}^{(1)})/(\hat{u}^{(1)\prime}\hat{u}^{(1)}), \quad (2.15)$$

which under correct specification and valid instruments $Z$ is distributed as $\chi_{L-K}^2$ asymptotically. Because $\hat{\Omega}^{(1)}$ is based on $\hat{u}^{(1)}$, which relates to a consistent estimator, but not to an asymptotically optimal estimator, it seems better to perform at least one further iteration and use

$$J_Z^{(2)} = n(\hat{u}^{(2)\prime}Z(Z'\hat{\Omega}^{(2)}Z)^{-1}Z'\hat{u}^{(2)})/(\hat{u}^{(2)\prime}\hat{u}^{(2)}), \quad (2.16)$$

where $\hat{u}^{(2)} = y - X\hat{\beta}_{GMM}^{(2)}$ is used to construct $\hat{\Omega}^{(2)}$. When $Z = (Z_m \; Z_a)$, where $Z_m$ is an $n \times L_m$ matrix with $L_m \geq K$ containing the instruments whose validity seems very likely, then, under the maintained hypothesis $E(Z'_m u) = 0$, one can test the validity of the $L - L_m$ additional instruments $Z_a$ by the incremental test statistic

$$JI_{Z_a}^{(2)} = n[\hat{u}^{(2)\prime}Z(Z'\hat{\Omega}^{(2)}Z)^{-1}Z'\hat{u}^{(2)}/(\hat{u}^{(2)\prime}\hat{u}^{(2)}) - \hat{u}_m^{(2)\prime}Z_m(Z'_m\hat{\Omega}_m^{(2)}Z_m)^{-1}Z'_m\hat{u}_m^{(2)}/(\hat{u}_m^{(2)\prime}\hat{u}_m^{(2)})], \quad (2.17)$$

which under correct specification of the model with valid instruments $Z$ is distributed as $\chi_{L-L_m}^2$ asymptotically. Of course, $\hat{u}_m^{(2)}$ and $\hat{\Omega}_m^{(2)}$ are obtained by just using the instruments $Z_m$. Note that for $m = K$ we have $JI_{Z_a}^{(2)} = J_Z^{(2)}$ because in that case $Z'_m\hat{u}_m^{(2)} = 0$. Hence, when $m = K$, explicit specification of component $Z_m$ is meaningless.

In the simulations we will also examine unfeasible versions of the above test statistics, which exploit information that is usually not available in practice. This will produce evidence on what elements of the feasible asymptotic tests may cause any inaccuracies in finite samples. So, next to (2.13), (2.16) and (2.17) we will also examine

$$
\begin{aligned}
W_{\beta_k}^{(u)} &= (e'_{K,k}\hat{\beta}_{GMM} - \beta_k^0)/\{\sigma_u^2 e'_{K,k}[X'Z(Z'\Omega Z)^{-1}Z'X]^{-1}e_{K,k}\}^{1/2}, & (2.18) \\
J_Z^{(u)} &= \hat{u}'Z(Z'\Omega Z)^{-1}Z'\hat{u}/\sigma_u^2, \text{ where } \hat{u} = y - X\hat{\beta}_{GMM}, & (2.19) \\
JI_{Z_a}^{(u)} &= [\hat{u}'Z(Z'\Omega Z)^{-1}Z'\hat{u} - \hat{u}'_m Z_m(Z'_m\Omega Z_m)^{-1}Z'_m\hat{u}_m]/\sigma_u^2. & (2.20)
\end{aligned}
$$

Similar feasible and unfeasible implementations of t-tests and Sargan-Hansen tests for MGMM based estimators follow straight-forwardly.

# 3 Implementations for dynamic micro panel models

## 3.1 Model and assumptions

We consider the balanced linear dynamic panel data model $(i = 1, ..., N; t = 1, ..., T)$

$$y_{it} = x'_{it}\beta + w'_{it}\gamma + v'_{it}\delta + \tau_t + \eta_i + \varepsilon_{it}, \tag{3.1}$$

where $x_{it}$ contains $K_x \geq 0$ strictly exogenous regressors (excluding fixed time effects), $w_{it}$ are $K_w \geq 0$ predetermined regressors (probably including lags of the dependent variable and other variables affected by lagged feedback from $y_{it}$ or just from $\varepsilon_{it}$), $v_{it}$ are $K_v \geq 0$ endogenous regressors (affected by instantaneous feedback from $y_{it}$ and therefore jointly dependent with $y_{it}$), the $\tau_t$ are random or fixed time effects, the $\eta_i$ are random individual specific effects (most likely correlated with many of the regressors) such that

$$\eta_i \sim iid(0, \sigma_\eta^2), \tag{3.2}$$

whereas

$$\left.\begin{array}{l} E(\varepsilon_{it}) = 0 \\ E(\varepsilon_{it}^2) = \sigma_{it}^2 \\ E(\varepsilon_{it}\varepsilon_{js}) = 0 \\ E(\eta_i\varepsilon_{jt}) = 0 \end{array}\right\} \forall i, j, t \neq s. \tag{3.3}$$

The classification of the regressors implies

$$\left.\begin{array}{l} E(x_{it}\varepsilon_{is}) = 0 \\ E(w_{it}\varepsilon_{i,t+l}) = 0 \\ E(v_{it}\varepsilon_{i,t+1+l}) = 0 \end{array}\right\} \forall i, t, s, l \geq 0. \tag{3.4}$$

For the sake of simplicity we assume that all regressors are time varying and that the vectors $x_{it}$, $w_{it}$ or $v_{it}$ are defined for $t = 1, ..., T$. However, their elements may contain observations prior to $t = 1$ for regressors that are actually the $l^{th}$ order lag of a current variable. Only these lagged regressors are observed from $t = 1 - l$ onwards. This means that all regressors, be it current variables or lags of them, have exactly $T$ observations. So, any unbalancedness problems have been defined away; moreover, no internal instrumental variables can be constructed involving observations prior to those included in $x_{i1}$, $w_{i1}$ or $v_{i1}$.

Stacking the $T$ time-series observations of (3.1) the equation in levels can be written

$$y_i = X_i\beta + W_i\gamma + V_i\delta + \tau + \eta_i\iota_T + \varepsilon_i, \tag{3.5}$$

where $y_i = (y_{i1} \cdots y_{iT})'$, $X_i = (x_{i1} \cdots x_{iT})'$, $W_i = (w_{i1} \cdots w_{iT})'$, $V_i = (v_{i1} \cdots v_{iT})'$, $\tau = (\tau_1 \cdots \tau_T)'$ and $\iota_T$ is the $T \times 1$ vector with all its elements equal to unity. Defining the $i$-th $T \times K$ block of the regressor matrix $R_i = (X_i\ W_i\ V_i\ I_T)$, the $K \times 1$ coefficient vector $\alpha = (\beta'\ \gamma'\ \delta'\ \tau')'$ and the compound disturbances as $u_i = \eta_i\iota_T + \varepsilon_i$, the model can compactly be expressed as

$$y_i = R_i\alpha + u_i. \tag{3.6}$$

We do allow $K_x = 0$, $K_w = 0$, $K_v = 0$, but not all three at the same time, so

$$K = K_x + K_w + K_v + T > T. \tag{3.7}$$

The parameter vector $\tau$ could be void too, or have all its elements equal though non-zero. However, it seems better to allow for a general $\tau$, for the same reasons as an ordinary cross-section regression normally should include an intercept, namely to help underpin the assumption that both the idiosyncratic disturbances and the individual effects have expectation zero.

 We focus on micro panels, where the number of time-series observations $T$ is usually very small, possibly a one digit number, and the number of cross-section units $N$ is large, usually at least several hundreds. Therefore asymptotic approximations will be for $N \to \infty$ and $T$ finite.

## 3.2  Removing individual effects by first differencing

First we consider estimating the model by GMM following the approach propounded by Arellano and Bond (1991), see also Holtz-Eakin *et al.* (1988). This estimates the equations ($i = 1, ..., N$; $t = 2, ..., T$)

$$\begin{aligned}\Delta y_{it} &= \Delta x'_{it}\beta + \Delta w'_{it}\gamma + \Delta v'_{it}\delta + \Delta\tau_t + \Delta\varepsilon_{it} \\ &= \Delta r'_{it}\tilde{\alpha} + \Delta\varepsilon_{it},\end{aligned} \tag{3.8}$$

where $\tilde{\alpha} = (\beta' \ \gamma' \ \delta' \ \tilde{\tau}')'$ and $\Delta r_{it} = (\Delta x'_{it} \ \Delta w'_{it} \ \Delta v'_{it} \ e'_{T-1,t})'$ are $(K-1) \times 1$ vectors. Here $\tilde{\tau}$ is the $(T-1) \times 1$ vector $(\Delta\tau_2 \cdots \Delta\tau_T)'$. Stacking the $T-1$ time-series observations of (3.8) yields

$$\tilde{y}_i = \tilde{R}_i\tilde{\alpha} + \tilde{\varepsilon}_i, \tag{3.9}$$

where $\tilde{y}'_i = (\Delta y_{i2} \cdots \Delta y_{iT})$, $\tilde{R}'_i = (\Delta r_{i2} \cdots \Delta r_{iT})$ and $\tilde{\varepsilon}'_i = (\Delta\varepsilon_{i2} \cdots \Delta\varepsilon_{iT})$.

 To construct a full column rank matrix of instrumental variables $Z = (Z'_1 \cdots Z'_N)'$, which expresses as many linearly independent orthogonality conditions as possible for (3.9), while restricting ourselves to internal variables, *i.e.* variables occurring in (3.1), we define the following vectors

$$x_i^{T'} = (x'_{i1} \cdots x'_{iT}), \ w_i^{t'} = (w'_{i1} \cdots w'_{it}), \ v_i^{t'} = (v'_{i1} \cdots v'_{it}). \tag{3.10}$$

Without making this explicit in the notation it should be understood that these three vectors only contain unique elements. Hence, if vector $x_{is}$ (or $w_{is}$) contains for $1 < s \le T$ a particular value and also its lag (which is not possible for $v_{it}$), then this lag should be taken out since it already appears in $x_{i,s-1}$. Matrix $Z_i$ is of order $(T-1) \times L$ and consists of four blocks (though some of these may be void)

$$Z_i = (I_{T-1} \ Z_i^x \ Z_i^w \ Z_i^v). \tag{3.11}$$

The first block is associated with the fundamental moment conditions $E(\tilde{\varepsilon}_i) = 0$ and could therefore form part of $Z_i$ even if one imposes $\tilde{\tau} = 0$. For the other blocks we have

$$Z_i^x = I_{T-1} \otimes x_i^{T'}, \ Z_i^w = \begin{pmatrix} w_i^{1'} & 0' & 0' \\ O & \ddots & O \\ 0' & 0' & w_i^{T-1'} \end{pmatrix}, \ Z_i^v = \begin{pmatrix} 0' & 0' & 0' \\ v_i^{1'} & 0' & 0' \\ O & \ddots & O \\ 0' & 0' & v_i^{T-2'} \end{pmatrix}. \tag{3.12}$$

The maximum possible order of $Z_i^x$ is $(T-1) \times K_x T(T-1)$, of $Z_i^w$ it is $(T-1) \times K_w T(T-1)/2$ and $Z_i^v$ is $(T-1) \times K_v(T-1)(T-2)/2$, thus

$$L \leq (T-1)\{T[K_x + (K_w + K_v)/2] - K_v + 1\}, \tag{3.13}$$

whereas MM estimation requires $L \geq K-1$. It follows from (3.3) and (3.4) that $E(Z_i'\tilde{\varepsilon}_i) = 0$ indeed. In actual estimation one may use a subset of these instruments by taking the linear transformation $Z_i^* = Z_i C$, where $C$ is an $L \times L^*$ matrix (with all its elements often being either zero, one or minus one) of rank $L^* < L$, provided $L^* \geq K$. In the above we have implicitly assumed that the variables are such that $Z = (Z_1' \cdots Z_N')'$ will have full column rank, so another necessary condition is $N(T-1) \geq L$. Of course, it is not required that individual blocks $Z_i$ have full column rank.

Despite its undesirable effect on the asymptotic variance of method of moments estimators, reducing the number of instruments may improve estimation precision, because it may at the same time mitigate estimation bias in finite samples, especially when weak instruments are being removed. So, instead of including the block $I_{T-1}$ in $Z_i$ one could – when the model has no time-effects – replace it by $I_{T-1}\iota_{T-1} = \iota_{T-1}$. Regarding $Z_i^w$ and $Z_i^v$ two alternative instrument reduction methods have been suggested, namely omitting long lags (see Bowsher 2002, Windmijer 2005 and Bun and Kiviet 2006) and collapsing (see Roodman 2009b, but also suggested in Anderson and Hsiao 1981). Both are employed in Ziliak (1997); these two methods can also be combined.

Omitting long lags could be achieved by reducing $Z_i^w$ to, for instance,

$$\begin{pmatrix} w_{i1}' & 0' & 0' & 0' & 0' & & 0' & 0' \\ 0' & w_{i1}' & w_{i2}' & 0' & 0' & & 0' & 0' \\ 0' & 0' & 0' & w_{i2}' & w_{i3}' & & 0' & 0' \\ \vdots & \vdots & \vdots & O' & O' & \ddots & O' & O' \\ 0' & 0' & 0' & 0' & 0' & & w_{i,T-2}' & w_{i,T-1}' \end{pmatrix} \tag{3.14}$$

and similar for $Z_i^v$. The collapsed versions of $Z_i^w$ and of $Z_i^v$ can be denoted as

$$Z_i^{*w} = \begin{pmatrix} w_{i1}' & 0' & \cdots & 0' \\ w_{i2}' & w_{i1}' & & \vdots \\ \vdots & & \ddots & O \\ w_{i,T-1}' & w_{i,T-2}' & \cdots & w_{i,1}' \end{pmatrix}, \quad Z_i^{*v} = \begin{pmatrix} 0' & 0' & \cdots & 0' \\ v_{i1}' & 0' & & 0' \\ v_{i2}' & v_{i1}' & & \vdots \\ \vdots & & \ddots & O \\ v_{i,T-2}' & w_{i,T-3}' & \cdots & v_{i,1}' \end{pmatrix}. \tag{3.15}$$

Collapsing can be combined with omitting long lags, if one removes all the columns of $Z_i^{*w}$ and $Z_i^{*v}$ which have at least a certain number of zero elements (say 1 or 2 or more) in their top rows. In corresponding ways, the column space of $Z_i^x$ can be reduced by including in $Z_i$ either a limited number of lags and leads, or the collapsed matrix

$$Z_i^{*x} = \begin{pmatrix} x_{i2}' & x_{i1}' & 0' & \cdots & 0' \\ x_{i3}' & x_{i2}' & x_{i1}' & & \vdots \\ \vdots & \vdots & & \ddots & O \\ x_{i,T}' & x_{i,T-1}' & x_{i,T-2}' & \cdots & x_{i,1}' \end{pmatrix}, \tag{3.16}$$

or just its first two or three columns – or what is often done in practice – simply the difference between the first two columns, the $K_x$ regressors $(\Delta x_{i2}, ..., \Delta x_{iT})'$.

In the simulations we will examine the following specific forms of instrument matrix reduction of the case where all instruments associated with valid linear moment restrictions are being used. The latter case we label as A (all); the reductions are labelled C (standard collapsing), L0, L1, L2, L3 (which primarily restrict the lag length), and C0, C1, C2, C3 (which combine the two reduction principles). In all the reductions we replace $I_{T-1}$ by $\iota_{T-1}$ when the model does not include time-effects. Regarding $Z_i^x$, $Z_i^w$ and $Z_i^v$ different types of reductions can be taken, which we will distinguish by using for example the characterization: $A^v$, $L2^w$, $C1^x$ etc. The particular reductions that we will consider are defined as indicated in Table 1.

**Table 1.** Definition of labels for particular instrument matrix reductions

| $A^x$ : $Z_i^x$ | $A^w$ : $Z_i^w$ | $A^v$ : $Z_i^v$ |
|---|---|---|
| $L0^x : diag(x_{i2}', ..., x_{iT}')$ | $L0^w : diag(w_{i1}', ..., w_{i,T-1}')$ | $L0^v : [0, diag(v_{i1}, ..., v_{i,T-2})]'$ |
| $L1^x : diag(\Delta x_{i2}', ..., \Delta x_{iT}')$ | $L1^w : diag(w_{i1}', \Delta w_{i2}', ..., \Delta w_{i,T-1}')$ | $L1^v : [0, diag(v_{i1}, \Delta v_{i2}, ..., \Delta v_{i,T-2})]'$ |
| $L2^x : diag(x_{i1}', ..., x_{i,T-1}')$, $L0^x$ | $L2^w : [0, diag(w_{i1}, ..., w_{i,T-2})]'$, $L0^w$ | $L2^v : [0, 0, diag(v_{i1}, ..., v_{i,T-3})]'$, $L0^v$ |
| $L3^x : [0, diag(x_{i1}, ..., x_{i,T-2})]'$, $L2^x$ | $L3^w : [0, 0, diag(w_{i1}, ..., w_{i,T-3})]'$, $L2^w$ | $L3^v : [0, 0, 0, diag(v_{i1}, ..., v_{i,T-4})]'$, $L2^v$ |
| $C^x$ : $Z_i^{*x}$ | $C^w$ : $Z_i^{*w}$ | $C^v$ : $Z_i^{*v}$ |
| $C0^x : (x_{i2}, ..., x_{iT})'$ | $C0^w : (w_{i1}, ..., w_{i,T-1})'$ | $C0^v : (0, v_{i1}, ..., v_{i,T-2})'$ |
| $C1^x : (\Delta x_{i2}, ..., \Delta x_{iT})'$ | $C1^w : (0, \Delta w_{i2}, ..., \Delta w_{i,T-1})'$ | $C1^v : (0, 0, \Delta v_{i2}, ..., \Delta v_{i,T-2})'$ |
| $C2^x : C0^x, (x_{i1}, ..., x_{i,T-1})'$ | $C2^w : C0^w, (0, w_{i1}, ..., w_{i,T-2})'$ | $C2^v : C0^v, (0, 0, v_{i1}, ..., v_{i,T-3})'$ |
| $C3^x : C2^x, (0, x_{i1}, ..., x_{i,T-2})'$ | $C3^w : C2^w, (0, 0, w_{i1}, ..., w_{i,T-3})'$ | $C3^v : C2^v, (0, 0, 0, v_{i1}, ..., v_{i,T-4})'$ |

Note that for all three types of regressors L2, like L1, uses one extra lag compared to L0, but does not impose the first-difference restrictions characterizing L1. We skipped a similar intermediary case between L2 and L3. Self-evidently $L2^x$ can also be represented by combining $diag(x_{i1}', ..., x_{i,T-1}')$ with $L1^x$, and similar for $L2^w$ and $L2^v$. The reductions C0 and C1, which yield just one instrument per regressor, consitute generalizations of the classic instruments suggested by Anderson and Hsiao (1981). These may lead to just identified models where the number of instruments equals the number of regressors and the occurrence of the non-existence of moments problem. To avoid that, and also because we suppose that in general some degree of overidentification will have advantages regarding both estimation precision and the opportunity to test model adequacy, we will restrict ourselves to the popular $C1^x$ and the reductions C and C3 as far as collapsing is concerned. In C3 just the first three columns of the matrices in (3.15) and (3.16) are being used as instruments.

### 3.2.1 Alternative weighting matrices

We assumed in (3.3) that the $\varepsilon_{it}$ are serially and cross-sectionally uncorrelated but may be heteroskedastic. Let us define the matrices

$$\Omega_i = \begin{pmatrix} \sigma_{i1}^2 & 0' & 0 \\ 0 & \ddots & 0 \\ 0 & 0' & \sigma_{iT}^2 \end{pmatrix} \text{ and } D = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & -1 & 1 \end{pmatrix}, \quad (3.17)$$

where $\varepsilon_i \sim (0, \Omega_i)$ and $D$ is $(T-1) \times T$, so $\tilde{\varepsilon}_i = D\varepsilon_i \sim (0, D\Omega_i D')$. Under standard regularity we have

$$N^{-1/2} \sum_{i=1}^N Z_i'\tilde{\varepsilon}_i \xrightarrow{d} N(0, \text{plim } N^{-1} \sum_{i=1}^N Z_i'D\Omega_i D'Z_i). \quad (3.18)$$

Hence, the optimal GMM estimator of $\tilde{\alpha}$ of (3.9) should use a weighting matrix such that its inverse has probability limit proportional to plim $N^{-1} \sum_{i=1}^{N} Z_i' D\Omega_i D' Z_i$. This can be achieved by starting with a consistent 1-step GMM estimator $\widehat{\alpha}^{(1)}$, which, assuming incorrectly $\Omega_i = \sigma_\varepsilon^2 I$, uses the weighting matrix

$$G^{(0)} = \left( \sum_{i=1}^{N} Z_i' DD' Z_i \right)^{-1}, \tag{3.19}$$

and then yields, what in the simulations we will call the AB1 estimator,

$$\widehat{\alpha}^{(1)} = \left[ \left( \sum_{i=1}^{N} \tilde{R}_i' Z_i \right) G^{(0)} \left( \sum_{i=1}^{N} Z_i' \tilde{R}_i \right) \right]^{-1} \left( \sum_{i=1}^{N} \tilde{R}_i' Z_i \right) G^{(0)} \left( \sum_{i=1}^{N} Z_i' \tilde{y}_i \right). \tag{3.20}$$

Next, in a second step, the consistent 1-step residuals $\widehat{\tilde{\varepsilon}}_i^{(1)} = \tilde{y}_i - \tilde{R}_i \widehat{\alpha}^{(1)}$ can be used to construct the asymptotically optimal weighting matrix

$$\hat{G}_a^{(1)} = \left( \sum_{i=1}^{N} Z_i' \widehat{\tilde{\varepsilon}}_i^{(1)} \widehat{\tilde{\varepsilon}}_i^{(1)\prime} Z_i \right)^{-1}. \tag{3.21}$$

However, an alternative is using

$$\hat{G}_b^{(1)} = \left( \sum_{i=1}^{N} Z_i' \hat{H}_i^{(1)b} Z_i \right)^{-1}, \tag{3.22}$$

where $\hat{H}_i^{(1)b}$ is the band matrix

$$\hat{H}_i^{(1)b} = \begin{pmatrix} \widehat{\tilde{\varepsilon}}_{i2}^{(1)}\widehat{\tilde{\varepsilon}}_{i2}^{(1)} & \widehat{\tilde{\varepsilon}}_{i2}^{(1)}\widehat{\tilde{\varepsilon}}_{i3}^{(1)} & 0 & \cdots & & 0 \\ \widehat{\tilde{\varepsilon}}_{i3}^{(1)}\widehat{\tilde{\varepsilon}}_{i2}^{(1)} & \widehat{\tilde{\varepsilon}}_{i3}^{(1)}\widehat{\tilde{\varepsilon}}_{i3}^{(1)} & \widehat{\tilde{\varepsilon}}_{i3}^{(1)}\widehat{\tilde{\varepsilon}}_{i4}^{(1)} & & & 0 \\ 0 & \widehat{\tilde{\varepsilon}}_{i4}^{(1)}\widehat{\tilde{\varepsilon}}_{i3}^{(1)} & \widehat{\tilde{\varepsilon}}_{i4}^{(1)}\widehat{\tilde{\varepsilon}}_{i4}^{(1)} & \ddots & & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & & & \ddots & \widehat{\tilde{\varepsilon}}_{i,T-1}^{(1)}\widehat{\tilde{\varepsilon}}_{i,T-1}^{(1)} & \widehat{\tilde{\varepsilon}}_{i,T-1}^{(1)}\widehat{\tilde{\varepsilon}}_{iT}^{(1)} \\ 0 & 0 & 0 & \cdots & \widehat{\tilde{\varepsilon}}_{iT}^{(1)}\widehat{\tilde{\varepsilon}}_{i,T-1}^{(1)} & \widehat{\tilde{\varepsilon}}_{iT}^{(1)}\widehat{\tilde{\varepsilon}}_{iT}^{(1)} \end{pmatrix}. \tag{3.23}$$

Note that both $(N\hat{G}_a^{(1)})^{-1}$ and $(N\hat{G}_b^{(1)})^{-1}$ have a probability limit equal to the limiting variance of (3.18). The latter is less robust, but may converge faster when $\Omega_i$ is diagonal indeed. On the other hand (3.22) may not be positive definite, whereas (3.21) is.

For the special case $\Omega_i = \sigma_{\varepsilon,i}^2 I$ of cross-section heteroskedasticity but time-series homoskedasticity, one could use

$$\hat{G}_c^{(1)} = \left( \sum_{i=1}^{N} Z_i' \hat{H}_i^{(1)c} Z_i \right)^{-1}, \tag{3.24}$$

with

$$\hat{H}_i^{(1)c} = \hat{\sigma}_{\varepsilon,i}^{2(1)} H = \hat{\sigma}_{\varepsilon,i}^{2(1)} \begin{pmatrix} 2 & -1 & 0 & \cdots & & 0 \\ -1 & 2 & -1 & & & 0 \\ 0 & -1 & 2 & \ddots & & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & & & \ddots & 2 & -1 \\ 0 & 0 & 0 & \cdots & -1 & 2 \end{pmatrix}, \tag{3.25}$$

where $H = DD'$ and

$$\hat{\sigma}_{\varepsilon,i}^{2(1)} = \widehat{\tilde{\varepsilon}}_i^{(1)\prime} H^{-1} \widehat{\tilde{\varepsilon}}_i^{(1)} / (T-1), \tag{3.26}$$

although this estimator is not consistent for $T$ finite.

All these different weighting matrices can be used to calculate alternative $\widehat{\alpha}_{(j)}^{(2)}$ estimators for $j \in \{a, b, c\}$, to be labelled AB2a, AB2b and AB2c, according to

$$\widehat{\alpha}_{(j)}^{(2)} = \left[ \left( \sum_{i=1}^N \tilde{R}_i' Z_i \right) \hat{G}_j^{(1)} \left( \sum_{i=1}^N Z_i' \tilde{R}_i \right) \right]^{-1} \left( \sum_{i=1}^N \tilde{R}_i' Z_i \right) \hat{G}_j^{(1)} \left( \sum_{i=1}^N Z_i' \tilde{y}_i \right). \tag{3.27}$$

When the employed weighting matrix is asymptotically optimal indeed, the first-order asymptotic approximation to the variance of $\widehat{\alpha}_{(j)}^{(2)}$ is given by the inverse of the matrix in square brackets. From this (corrected) t-tests are easily obtained, see Section 3.4. Matching implementations of Sargan-Hansen statistics follow easily too, see Section 3.5.

## 3.3  Respecting the equation in levels as well

Since $\tau_t = \tau_1 + \sum_{s=2}^t \tilde{\tau}_s$ for $t \geq 2$, defining the $T \times (T-1)$ matrix

$$Q = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & & \\ 1 & 1 & \ddots & \\ \vdots & \vdots & \ddots & 0 \\ 1 & 1 & \cdots & 1 \end{pmatrix}, \tag{3.28}$$

we may write $\tau = \tau_1 \iota_T + Q\tilde{\tau}$. Therefore, we can rewrite model (3.6) as

$$
\begin{aligned}
y_i &= X_i\beta + W_i\gamma + V_i\delta + Q\tilde{\tau} + \tau_1\iota_T + u_i \\
&= \bar{R}_i\tilde{\alpha} + \tau_1\iota_T + u_i \\
&= \bar{R}_i^*\ddot{\alpha} + u_i,
\end{aligned}
\tag{3.29}
$$

where $\bar{R}_i = (X_i \ W_i \ V_i \ Q)$, $\bar{R}_i^* = (\bar{R}_i \ \iota_T)$ and $K \times 1$ vector $\ddot{\alpha} = (\tilde{\alpha}' \ \tau_1)'$, with coefficient $\tau_1$ now being an overall intercept.

A valid instrument for model (3.29) is the (as a regressor possibly redundant) intercept, because this embodies the single orthogonality condition $E[\sum_{t=1}^T (\eta_i + \varepsilon_{it})] = E[\sum_{t=1}^T u_{it}] = 0 \ (\forall i)$, which is implied by the $T+1$ assumptions $E(\eta_i) = 0$ and $E(\varepsilon_{it}) = 0$ (for $t = 1, ..., T$) made in (3.2) and (3.3). These $T+1$ assumptions can also be represented (through linear transformation) by $(i)$ $E(\eta_i) = 0$, $(ii)$ $E(\Delta\varepsilon_{it}) = 0$ (for $t = 2, ..., T$) and $(iii)$ $E(\sum_{t=1}^T u_{it}) = 0$. Because we cannot express $\eta_i$ exclusively in observed variables and unknown parameters it is impossible to convert $(i)$ into a separate sample orthogonality condition. The $T - 1$ orthogonality conditions $(ii)$ are already employed by Arellano-Bond estimation, through including $I_{T-1}$ in $Z_i$ of (3.11) for the equation in first differences. Orthogonality condition $(iii)$, which is in terms of the level disturbance, can be exploited by including the column $\iota_T$ in the $i$-th block of an instrument matrix for level equation (3.29).

Combining the $T-1$ difference equations and the $T$ level equations in a system yields

$$\ddot{y}_i = \ddot{R}_i\ddot{\alpha} + \ddot{u}_i, \tag{3.30}$$

for each individual $i$, where $\ddot{y}_i = (\tilde{y}_i' \ y_i')'$, $\ddot{R}_i = (\tilde{R}_i^{*\prime} \ \bar{R}_i^{*\prime})'$, with $\tilde{R}_i^* = (\tilde{R}_i \ 0)$, so it is extended by an extra column of zeros (to annihilate coefficient $\tau_1$ in the equation in first differences), and $\ddot{u}_i = (\tilde{\varepsilon}_i' \ u_i')'$. We find that $E(\tilde{\varepsilon}_i u_i') = E(D\varepsilon_i \varepsilon_i') = D\Omega_i$ and $E(u_i u_i') = E[(\eta_i \iota_T + \varepsilon_i)(\eta_i \iota_T + \varepsilon_i)'] = \sigma_\eta^2 \iota_T \iota_T' + \Omega_i$, so

$$E(\ddot{u}_i \ddot{u}_i') = \begin{pmatrix} D\Omega_i D' & D\Omega_i \\ \Omega_i D' & \Omega_i + \sigma_\eta^2 \iota_T \iota_T' \end{pmatrix}. \tag{3.31}$$

Model (3.30) can be estimated by MM using the $N(2T-1) \times (L+1)$ matrix of instruments with blocks

$$\ddot{Z}_i = \begin{pmatrix} Z_i & 0 \\ O & \iota_T \end{pmatrix}, \tag{3.32}$$

provided $L + 1 \geq K$ and $N(2T-1) \geq L + 1$. Since both $\ddot{R}_i$ and $\ddot{Z}_i$ contain a column $(0', \iota_T')'$, and due to the occurrence of the $O$-block in $\ddot{Z}_i$, by a minor generalization of result (2.8) the IV estimator of $\ddot{\alpha}$ obtained by using instrument blocks $\ddot{Z}_i$ in (3.30) will be equivalent regarding $\tilde{\alpha}$ with the IV estimator of equation (3.9) using instruments with blocks $Z_i$. That the same holds here for GMM under cross-sectional heteroskedasticity when using optimal instruments is due to the very special shape of $\ddot{Z}_i$ and is proved in Appendix B. Hence, there is no reason to estimate the system, just in order to exploit the extra valid instrument $(0', \iota_T')'$.

### 3.3.1 Effect stationarity

More internal instruments can be found for the equation in levels when some of the regressors are known to be uncorrelated with the individual effects. We will assume that after first differencing some of the explanatory variables may be uncorrelated with $\eta_i$. Let $r_{it}^\blacklozenge = (x_{it}^{\blacklozenge\prime} \ w_{it}^{\blacklozenge\prime} \ v_{it}^{\blacklozenge\prime})'$ contain the $K^\blacklozenge = K_x^\blacklozenge + K_w^\blacklozenge + K_v^\blacklozenge$ unique elements of $r_{it}$ which are effect stationary, by which we mean that $E(r_{it}\eta_i)$ is time-invariant, so that

$$E(\Delta r_{it}^\blacklozenge \eta_i) = 0, \ \forall i, \ t = 2, ..., T.$$

This implies that for the equation in levels the orthogonality conditions

$$\left. \begin{array}{l} E[\Delta x_{it}^\blacklozenge(\eta_i + \varepsilon_{is})] = 0 \\ E[\Delta w_{it}^\blacklozenge(\eta_i + \varepsilon_{i,t+l})] = 0 \\ E[\Delta v_{it}^\blacklozenge(\eta_i + \varepsilon_{i,t+1+l})] = 0 \end{array} \right\} \forall i, t > 1, s \geq 1, l \geq 0 \tag{3.33}$$

hold. When $w_{it}^\blacklozenge$ includes $y_{i,t-1}$, then apparently $y_{it}$ is effect stationary so that the adopted model (3.1) suggests that all regressors in $r_{it}$ must be effect stationary, resulting in $K^\blacklozenge = K$.

Like for the $T-1$ conditions $E(\Delta\varepsilon_{it}) = 0$ discussed below (3.6), many of the conditions (3.33) are already implied by the orthogonality conditions $E(Z_i'\tilde{\varepsilon}_i) = 0$ for the equation in first-differences. In Appendix C we demonstrate that a matrix $\tilde{Z}_i^s$ of instruments can be designed for the equation in levels (3.6) just containing instruments additional to those already exploited by $E(Z_i'\tilde{\varepsilon}_i) = 0$, whilst $E[\tilde{Z}_i^{s\prime}(\eta_i\iota_T + \varepsilon_i)] = 0$. This is the $T \times L^\blacklozenge$ matrix

$$\tilde{Z}_i^s = (\iota_T \ \tilde{Z}_i^x \ \tilde{Z}_i^w \ \tilde{Z}_i^v), \tag{3.34}$$

where $L^{\blacklozenge} = 1 + K^{\blacklozenge}(T-1) - K_v^{\blacklozenge}$, with

$$
\tilde{Z}_i^x = \begin{pmatrix} 0' & 0' & \cdots & 0' \\ \Delta x_{i2}^{\blacklozenge\prime} & 0' & \cdots & 0' \\ 0' & \Delta x_{i3}^{\blacklozenge\prime} & & 0' \\ \vdots & & \ddots & \\ 0' & 0' & & \Delta x_{iT}^{\blacklozenge\prime} \end{pmatrix}, \quad \tilde{Z}_i^w = \begin{pmatrix} 0' & 0' & \cdots & 0' \\ \Delta w_{i2}^{\blacklozenge\prime} & 0' & \cdots & 0' \\ 0' & \Delta w_{i3}^{\blacklozenge\prime} & & 0' \\ \vdots & & \ddots & \\ 0' & 0' & & \Delta w_{iT}^{\blacklozenge\prime} \end{pmatrix},
$$

$$
\tilde{Z}_i^v = \begin{pmatrix} 0' & \cdots & 0' \\ 0' & \cdots & 0' \\ \Delta v_{i2}^{\blacklozenge\prime} & & 0' \\ & \ddots & \\ 0' & & \Delta v_{i,T-1}^{\blacklozenge\prime} \end{pmatrix}.
$$

Under effect stationarity of the $K^{\blacklozenge}$ variables (3.33) the system (3.30) can be estimated while exploiting the matrix of instruments

$$
\ddot{Z}_i^s = \begin{pmatrix} Z_i & O \\ O & \tilde{Z}_i^s \end{pmatrix}. \tag{3.35}
$$

If one decides to collapse the instruments included in $Z_i$, it seems reasonable to collapse $\tilde{Z}_i^s$ as well and replace it by

$$
\ddot{Z}_i^{*s} = \begin{pmatrix} 1 & 0' & 0' & 0' \\ 1 & \Delta x_{i2}^{\blacklozenge\prime} & \Delta w_{i2}^{\blacklozenge\prime} & 0' \\ 1 & \Delta x_{i3}^{\blacklozenge\prime} & \Delta w_{i3}^{\blacklozenge\prime} & \Delta v_{i2}^{\blacklozenge\prime} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \Delta x_{iT}^{\blacklozenge\prime} & \Delta w_{iT}^{\blacklozenge\prime} & \Delta v_{i,T-1}^{\blacklozenge\prime} \end{pmatrix}. \tag{3.36}
$$

Note that $\ddot{Z}_i^{*s}$ has $L^{\blacklozenge} = 1 + K^{\blacklozenge}$ columns.

### 3.3.2   Alternative weighting matrices under effect stationarity

For the above system we have

$$
N^{-1/2} \sum_{i=1}^{N} \ddot{Z}_i^{s\prime} \ddot{u}_i \xrightarrow{d} N(0, \operatorname{plim} N^{-1} \sum_{i=1}^{N} \Phi_i), \tag{3.37}
$$

with

$$
\Phi_i = \begin{pmatrix} Z_i' D \Omega_i D' Z_i & Z_i' D \Omega_i \tilde{Z}_i^s \\ \tilde{Z}_i^{s\prime} \Omega_i D' Z_i & \tilde{Z}_i^{s\prime}(\Omega_i + \sigma_\eta^2 \iota_T \iota_T') \tilde{Z}_i^s \end{pmatrix}. \tag{3.38}
$$

Hence a feasible initial weighting matrix is given by

$$
S^{(0)}(q) = \left( \sum_{i=1}^{N} \Phi_i^{(0)}(q) \right)^{-1}, \tag{3.39}
$$

where

$$
\Phi_i^{(0)}(q) = \begin{pmatrix} Z_i' D D' Z_i & Z_i' D \tilde{Z}_i^s \\ \tilde{Z}_i^{s\prime} D' Z_i & \tilde{Z}_i^{s\prime}(I + q \iota_T \iota_T') \tilde{Z}_i^s \end{pmatrix},
$$

with $q$ some nonnegative real value. Weighting matrix $S^{(0)}(q)$ would be optimal if $\Omega_i = \sigma_\varepsilon^2 I$ with $q = \sigma_\eta^2/\sigma_\varepsilon^2$. The for any $q$ consistent 1-step GMM system estimator BB1(q) is given by

$$\widehat{\alpha}^{(1)}(q) = \left[\left(\sum_{i=1}^{N} \ddot{R}_i' \ddot{Z}_i^s\right) S^{(0)}(q) \left(\sum_{i=1}^{N} \ddot{Z}_i^{s\prime} \ddot{R}_i\right)\right]^{-1} \left(\sum_{i=1}^{N} \ddot{R}_i' \ddot{Z}_i^s\right) S^{(0)}(q) \left(\sum_{i=1}^{N} \ddot{Z}_i^{s\prime} \ddot{y}_i\right). \tag{3.40}$$

Next, in a second step, the consistent 1-step residuals $\widehat{\ddot{u}}_i^{(1)} = \ddot{y}_i - \ddot{R}_i \widehat{\alpha}^{(1)}(q)$ can be used to construct the asymptotically optimal weighting matrix

$$\hat{S}_a^{(1)} = \left(\sum_{i=1}^{N} \ddot{Z}_i^{s\prime} \widehat{\ddot{u}}_i^{(1)} \widehat{\ddot{u}}_i^{(1)\prime} \ddot{Z}_i^s\right)^{-1}, \tag{3.41}$$

where $\widehat{\ddot{u}}_i^{(1)} = (\widehat{\ddot{\varepsilon}}_i^{s(1)\prime} \ \hat{u}_i^{s(1)\prime})'$ with $\widehat{\ddot{\varepsilon}}_i^{s(1)} = \tilde{y}_i - \tilde{R}_i^* \widehat{\alpha}^{(1)}(q)$ and $\hat{u}_i^{s(1)} = y_i - \bar{R}_i^* \widehat{\alpha}^{(1)}(q)$. However, several alternatives are possible. Consider weighting matrix

$$\hat{S}_b^{(1)} = \left[\sum_{i=1}^{N} \left( \begin{matrix} Z_i' \hat{H}_i^{s(1)} Z_i & Z_i' \hat{D}_i^{s(1)} \tilde{Z}_i^s \\ \tilde{Z}_i^{s\prime} \hat{D}_i^{s(1)\prime} Z_i & \tilde{Z}_i^{s\prime} \hat{u}_i^{s(1)} \hat{u}_i^{s(1)\prime} \tilde{Z}_i^s \end{matrix} \right) \right]^{-1}, \tag{3.42}$$

where $\hat{H}_i^{s(1)}$ is self-evidently like $\hat{H}_i^{(1)b}$ but on the basis of the residuals $\widehat{\ddot{\varepsilon}}_i^{s(1)}$, and

$$\hat{D}_i^{s(1)} = \left( \begin{matrix} \widehat{\ddot{\varepsilon}}_{i2}^{s(1)} \hat{u}_{i1}^{s(1)} & \widehat{\ddot{\varepsilon}}_{i2}^{s(1)} \hat{u}_{i2}^{s(1)} & 0 & \cdots & 0 & 0 \\ 0 & \widehat{\ddot{\varepsilon}}_{i3}^{s(1)} \hat{u}_{i2}^{s(1)} & \widehat{\ddot{\varepsilon}}_{i3}^{s(1)} \hat{u}_{i3}^{s(1)} & & & 0 \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & & & \ddots & \widehat{\ddot{\varepsilon}}_{i,T-1}^{s(1)} \hat{u}_{i,T-1}^{s(1)} & \widehat{\ddot{\varepsilon}}_{i,T-1}^{s(1)} \hat{u}_{iT}^{s(1)} \\ 0 & 0 & 0 & \cdots & 0 & \widehat{\ddot{\varepsilon}}_{iT}^{s(1)} \hat{u}_{iT}^{s(1)} \end{matrix} \right). \tag{3.43}$$

For the special case $\sigma_\varepsilon^2 \Omega_i = \sigma_{\varepsilon,i}^2 I$ of cross-section heteroskedasticity and time-series homoskedasticity one can use the weighting matrix

$$\hat{S}_c^{(1)} = \left[\sum_{i=1}^{N} \hat{\sigma}_{\varepsilon,i}^{2,s(1)} \left( \begin{matrix} Z_i' H Z_i & Z_i' D \tilde{Z}_i^s \\ \tilde{Z}_i^{s\prime} D' Z_i & \tilde{Z}_i^{s\prime} [I + (\hat{\sigma}_\eta^{2(1)}/\hat{\sigma}_{\varepsilon,i}^{2,s(1)}) \iota_T \iota_T'] \tilde{Z}_i^s \end{matrix} \right) \right]^{-1}, \tag{3.44}$$

where

$$\hat{\sigma}_{\varepsilon,i}^{2,s(1)} = \widehat{\ddot{\varepsilon}}_i^{s(1)\prime} H^{-1} \widehat{\ddot{\varepsilon}}_i^{s(1)}/(T-1), \tag{3.45}$$

$$\hat{\sigma}_\eta^{2,s(1)} = \sum_{i=1}^{N} [(\iota_T' \hat{u}_i^{s(1)})^2 - \hat{u}_i^{s(1)\prime} \hat{u}_i^{s(1)}]/[NT(T-1)]. \tag{3.46}$$

Note that, unlike the former, the latter is consistent, because for $t \neq s$

$$\begin{aligned} \text{plim } N^{-1} \sum_{i=1}^{N} \hat{u}_{it}^{s(1)} \hat{u}_{is}^{s(1)} &= \lim N^{-1} \sum_{i=1}^{N} E(u_{it} u_{is}) \\ &= \lim N^{-1} \sum_{i=1}^{N} E(\eta_i + \varepsilon_{it})(\eta_i + \varepsilon_{is}) = \sigma_\eta^2. \end{aligned}$$

The three alternatives yield 2-step system estimators $\widehat{\alpha}_{(j)}^{(2)}$ for $j \in \{a, b, c\}$, labelled BB2a, BB2b and BB2c, where

$$\widehat{\alpha}_{(j)}^{(2)} = \left[ \left( \sum_{i=1}^{N} \ddot{R}_i' \ddot{Z}_i^s \right) \hat{S}_j^{(1)} \left( \sum_{i=1}^{N} \ddot{Z}_i^{s\prime} \ddot{R}_i \right) \right]^{-1} \left( \sum_{i=1}^{N} \ddot{R}_i' \ddot{Z}_i^s \right) \hat{S}_j^{(1)} \left( \sum_{i=1}^{N} \ddot{Z}_i^{s\prime} \ddot{y}_i \right),$$
(3.47)

and the inverse matrix expression can be used again to estimate the variance of $\widehat{\alpha}_{(j)}^{(2)}$ if all employed moment conditions are valid.

## 3.4 Coefficient restriction tests

Simple Student-type coefficient test statistics can be obtained from 1-step and 2-step AB and BB estimation for the different weighting matrices considered. The 1-step estimators can be used in combination with a robust variance estimate (which takes possible heteroskedasticity into account). The 2-step estimators can be used in combination with the standard or a corrected variance estimate.[1]

When testing particular coefficient values, the relevant element of AB1 estimator $\widehat{\alpha}^{(1)}$ given in (3.20) should under homoskedasticity be scaled by the corresponding diagonal element of the standard expression for its estimated variance given by

$$\widehat{Var}(\widehat{\alpha}^{(1)}) = N^{-1} \sum_{i=1}^{N} \hat{\sigma}_{\varepsilon,i}^{2(1)} \Psi, \text{ with } \Psi = \left[ \left( \sum_{i=1}^{N} \tilde{R}_i' Z_i \right) G^{(0)} \left( \sum_{i=1}^{N} Z_i' \tilde{R}_i \right) \right]^{-1},$$
(3.48)

where $\hat{\sigma}_{\varepsilon,i}^{2(1)}$ is given in (3.26). This test will be indicated as AB1. Its robust version under cross-sectional heteroskedasticity uses for $j \in \{a, b, c\}$

$$\widehat{Var}_{(j)}(\widehat{\alpha}^{(1)}) = \Psi \left( \sum_{i=1}^{N} \tilde{R}_i' Z_i \right) G^{(0)} \left[ \hat{G}_j^{(1)} \right]^{-1} G^{(0)} \left( \sum_{i=1}^{N} Z_i' \tilde{R}_i \right) \Psi,$$
(3.49)

which will be indicated as either AB1aR, AB1bR or AB1cR. However, under heteroskedasticity the estimators $\widehat{\alpha}_{(j)}^{(2)}$ given in (3.47) are more efficient. The standard estimator for their variance is

$$\widehat{Var}(\widehat{\alpha}_{(j)}^{(2)}) = \left[ \left( \sum_{i=1}^{N} \tilde{R}_i' Z_i \right) \hat{G}_j^{(1)} \left( \sum_{i=1}^{N} Z_i' \tilde{R}_i \right) \right]^{-1},$$
(3.50)

which yields test statistics AB2a etc. The corrected version $\widehat{Varc}(\widehat{\alpha}_{(j)}^{(2)})$ requires derivation for $k = 1, ..., K-1$ of the actual implementation of matrix $\partial \Omega(\beta)/\partial \beta_k$ of Appendix A which is here $N(T-1) \times N(T-1)$. We denote its $i$-th block as $\partial \tilde{\Omega}_{(j)i}(\tilde{\alpha})/\partial \tilde{\alpha}_k$. For the a-type weighting matrix[2] the relevant $T-1 \times T-1$ matrix $\partial \tilde{\varepsilon}_i \tilde{\varepsilon}_i'/\partial \tilde{\alpha}_k$ with $\tilde{\varepsilon}_i = \tilde{y}_i - \tilde{R}_i \tilde{\alpha}$, is $-(\tilde{\varepsilon}_i \tilde{R}_{ik}' + \tilde{R}_{ik} \tilde{\varepsilon}_i')$, where $\tilde{R}_{ik}$ denotes the $k$-th column of $\tilde{R}_i$. For weighting matrix b it simplifies to the matrix consisting of the main diagonal and the two first sub-diagonals of $-(\tilde{\varepsilon}_i \tilde{R}_{ik}' + \tilde{R}_{ik} \tilde{\varepsilon}_i')$ with all other elements zero. And $\partial \tilde{\Omega}_{(c)i}(\tilde{\alpha})/\partial \tilde{\alpha}_k = -2[\tilde{\varepsilon}_i' H^{-1} \tilde{R}_{ik}/(T-1)]H$. So, we find

$$\widehat{Varc}(\widehat{\alpha}_{(j)}^{(2)}) = \widehat{Var}(\widehat{\alpha}_{(j)}^{(2)}) + \hat{F}_{(j)} \widehat{Var}(\widehat{\alpha}_{(j)}^{(2)}) + \widehat{Var}(\widehat{\alpha}_{(j)}^{(2)}) \hat{F}_{(j)} + \hat{F}_{(j)} \widehat{Var}_{(j)}(\widehat{\alpha}^{(1)}) \hat{F}_{(j)}', \quad (3.51)$$

---

[1]Many authors and the Stata program confusingly address the corrected 2-step variance as robust.
[2]This is the only variant considered in Windmeijer (2005).

with the $k$-th column of $\hat{F}_{(j)}$ given by

$$\hat{F}_{(j)\cdot k} = -\widehat{Var}(\widehat{\tilde{\alpha}}_{(j)}^{(2)}) \left(\sum_{i=1}^{N} \tilde{R}_i' Z_i\right) \hat{G}_j^{(1)} \left(\sum_{i=1}^{N} Z_i' \left.\frac{\partial \tilde{\Omega}_{(j)i}(\tilde{\alpha})}{\partial \tilde{\alpha}_k}\right|_{\widehat{\tilde{\alpha}}^{(1)}} Z_i\right) \hat{G}_j^{(1)} \left(\sum_{i=1}^{N} Z_i' \widehat{\tilde{\varepsilon}}_i^{(2)}\right). \tag{3.52}$$

This gives rise to tests indicated as AB2aW etc.

All above expressions become a bit more complex when considering BB estimation of the $K$ coefficients $\ddot{\alpha}$. The suboptimal 1-step BB estimator (3.40) of $\ddot{\alpha}$ should not be used for testing, unless in combination with a robust variance estimator, which is

$$\widehat{Var}_{(j)}(\widehat{\ddot{\alpha}}^{(1)}) = \Phi \left(\sum_{i=1}^{N} \ddot{R}_i' \ddot{Z}_i^s\right) S^{(0)}(q) \left[\hat{S}_j^{(1)}\right]^{-1} S^{(0)}(q) \left(\sum_{i=1}^{N} \ddot{Z}_i^{s\prime} \ddot{R}_i\right) \Phi, \tag{3.53}$$

where $\Phi = \left[\left(\sum_{i=1}^{N} \ddot{R}_i' \ddot{Z}_i^s\right) S^{(0)}(q) \left(\sum_{i=1}^{N} \ddot{Z}_i^{s\prime} \ddot{R}_i\right)\right]^{-1}$. This yields BB1aR, BB1bR and BB1cR. It seems better of course to use the efficient estimator $\widehat{\ddot{\alpha}}_{(j)}^{(2)}$ of (3.47). The standard expression for its estimated variance is

$$\widehat{Var}(\widehat{\ddot{\alpha}}_{(j)}^{(2)}) = \left[\left(\sum_{i=1}^{N} \ddot{R}_i' \ddot{Z}_i^s\right) \hat{S}_j^{(1)} \left(\sum_{i=1}^{N} \ddot{Z}_i^{s\prime} \ddot{R}_i\right)\right]^{-1}, \tag{3.54}$$

leading to tests BB2a etc. Their corrected versions (BB2aW etc.) can be obtained by a formula similar to (3.51) upon changing $\tilde{\alpha}$ in $\ddot{\alpha}$ and $\hat{F}_{(j)\cdot k}$ of (3.52) in

$$\hat{F}_{(j)\cdot k} = -\widehat{Var}(\widehat{\ddot{\alpha}}_{(j)}^{(2)}) \left(\sum_{i=1}^{N} \ddot{R}_i' \ddot{Z}_i^s\right) \hat{S}_j^{(1)} \left(\sum_{i=1}^{N} \ddot{Z}_i^{s\prime} \left.\frac{\partial \ddot{\Omega}_{(j)i}(\ddot{\alpha})}{\partial \ddot{\alpha}_k}\right|_{\widehat{\ddot{\alpha}}^{(1)}} \ddot{Z}_i^s\right) \hat{S}_j^{(1)} \left(\sum_{i=1}^{N} \ddot{Z}_i^{s\prime} \widehat{\tilde{\varepsilon}}_i^{(2)}\right), \tag{3.55}$$

where the block of $\partial \ddot{\Omega}_{(j)i}(\ddot{\alpha})/\partial \ddot{\alpha}_k$ corresponding to the equation in first differences is similar as for AB, but with an extra column and row of zeros for the intercept. The block corresponding to the equation in levels we took for weighting matrices a and b equal to $\partial u_i u_i'/\partial \ddot{\alpha}_k = -(u_i \bar{R}_{ik}^{*\prime} + \bar{R}_{ik}^* u_i')$, and for type c

$$\partial \{\tilde{\varepsilon}_i' H^{-1} \tilde{\varepsilon}_i / (T-1) + \sum_{i=1}^{N} [(\iota_T' u_i)^2 - u_i' u_i]/[NT(T-1)]\} I_T / \partial \ddot{\alpha}_k,$$

for which the first term yields $-2[(\tilde{\varepsilon}_i' H^{-1} \tilde{R}_{ik})/(T-1)] I_T$, and the second gives

$$-2\{\sum_{i=1}^{N} (\iota_T' u_i \bar{R}_{ik}^{*\prime} \iota_T - \bar{R}_{ik}^{*\prime} u_i)/[NT(T-1)]\} I_T.$$

For the nondiagonal upperblock of $\partial \ddot{\Omega}_{(j)i}(\ddot{\alpha})/\partial \ddot{\alpha}_k$ we took in cases a and b $\partial \tilde{\varepsilon}_i u_i'/\partial \tilde{\alpha}_k = -(\tilde{\varepsilon}_i \bar{R}_{ik}^{*\prime} + \tilde{R}_{ik} u_i')$ and for the derivative with respect to the intercept $-\tilde{\varepsilon}_i \iota_T'$. In case c it is $-2[(\tilde{\varepsilon}_i' H^{-1} \tilde{R}_{ik})/(T-1)] D$ and a zero matrix for the derivative with respect to the intercept.

## 3.5 Tests of overidentification restrictions

Using Arellano-Bond and Blundell-Bond type estimation, many options exist with respect to testing the overidentification restrictions. These options differ in the residuals and weighting matrices being employed. After 1-step Arellano-Bond estimation, see (3.20) and (3.26), we have test statistic

$$JAB^{(1,0)} = \left(\sum_{i=1}^{N} \widehat{\widetilde{\varepsilon}}_i^{(1)\prime} Z_i\right) \left(\sum_{i=1}^{N} Z_i' H Z_i\right)^{-1} \left(\sum_{i=1}^{N} Z_i'\widehat{\widetilde{\varepsilon}}_i^{(1)}\right) / \left(N^{-1}\sum_{i=1}^{N} \hat{\sigma}_i^{2(1)}\right),$$
$$(3.56)$$

which is only valid in case of homoskedasticity. Alternatively, after 1-step estimation the heteroskedasticity-robust test statistics

$$JAB_j^{(1,1)} = \left(\sum_{i=1}^{N} \widehat{\widetilde{\varepsilon}}_i^{(1)\prime} Z_i\right) \hat{G}_j^{(1)} \left(\sum_{i=1}^{N} Z_i'\widehat{\widetilde{\varepsilon}}_i^{(1)}\right), \ j \in \{a,b,c\} \qquad (3.57)$$

may be used, where $\hat{G}_j^{(1)}$ is given in (3.21), (3.22) and (3.24). Given the type of weighting matrix being used, one obtains 2-step residuals $\widehat{\widetilde{\varepsilon}}_{i(j)}^{(2)} = \tilde{y}_i - \tilde{R}_i^*\widehat{\tilde{\alpha}}_{(j)}^{(2)}$, and from these overidentification restrictions test statistics can be obtained, which may differ depending on whether the weighting matrix is now obtained still from 1-step or already from 2-step residuals.[3] This leads to

$$JAB_j^{(2,h)} = \left(\sum_{i=1}^{N} \widehat{\widetilde{\varepsilon}}_{a,i}^{(2)\prime} Z_i\right) \hat{G}_j^{(h)} \left(\sum_{i=1}^{N} Z_i'\widehat{\widetilde{\varepsilon}}_{a,i}^{(2)}\right), \ \text{for } h \in \{1,2\} \qquad (3.58)$$

where the 2-step weighting matrices are either $\hat{G}_a^{(2)} = \left(\sum_{i=1}^{N} Z_i'\widehat{\widetilde{\varepsilon}}_{i(a)}^{(2)}\widehat{\widetilde{\varepsilon}}_{i(a)}^{(2)\prime} Z_i\right)^{-1}$, $\hat{G}_b^{(2)} = \left(\sum_{i=1}^{N} Z_i' \hat{H}_i^{(2)b} Z_i\right)^{-1}$ or $\hat{G}_c^{(2)} = \left(\sum_{i=1}^{N} \hat{\sigma}_{i(c)}^{2(2)} Z_i' H Z_i\right)^{-1}$, and $\hat{H}_i^{(2)b}$ is like $\hat{H}_i^{(1)b}$ of (3.23), though using $\widehat{\widetilde{\varepsilon}}_{i(b)}^{(2)}$ instead of $\widehat{\widetilde{\varepsilon}}_i^{(1)}$; furthermore $\hat{\sigma}_{i(c)}^{2(2)} = \widehat{\widetilde{\varepsilon}}_{i(c)}^{(2)\prime} H^{-1}\widehat{\widetilde{\varepsilon}}_{i(c)}^{(2)}/(T-1)$.

Exploiting effect stationarity of a subset of the regressors by estimating the Blundell-Bond system leads to the 1-step test statistics

$$JBB^{(1,0)} \ = \ \left(\sum_{i=1}^{N} \widehat{\ddot{u}}_i^{(1)\prime} \ddot{Z}_i^s\right) S^{(0)}(\hat{\sigma}_\eta^{2,s(1)}/\hat{\sigma}_\varepsilon^{2,s(1)}) \left(\sum_{i=1}^{N} \ddot{Z}_i^{s\prime}\widehat{\ddot{u}}_i^{(1)}\right)/\hat{\sigma}_\varepsilon^{2,s(1)}, \quad (3.59)$$

$$JBB_j^{(1,1)} \ = \ \left(\sum_{i=1}^{N} \widehat{\ddot{u}}_i^{(1)\prime} \ddot{Z}_i^s\right) \hat{S}_j^{(1)} \left(\sum_{i=1}^{N} \ddot{Z}_i^{s\prime}\widehat{\ddot{u}}_i^{(1)}\right), \ j \in \{a,b,c\} \qquad (3.60)$$

where $\hat{\sigma}_\varepsilon^{2,s(1)} = \sum_{i=1}^{N} \hat{\sigma}_{\varepsilon,i}^{2,s(1)}/N$ and $S^{(0)}(\cdot)$ and $\hat{S}_j^{(1)}$ can be found in (3.39), (3.41), (3.42) and (3.44). Defining the various 2-step residuals and variance estimators as $\widehat{\ddot{u}}_{i(j)}^{(2)} = \ddot{y}_i - \ddot{R}_i\widehat{\ddot{\alpha}}_{(j)}^{(2)} = (\widehat{\tilde{\varepsilon}}_{i(j)}^{s(2)\prime} \ \hat{u}_{i(j)}^{s(2)\prime})'$ and $\hat{\sigma}_{\varepsilon,i(j)}^{2,s(2)}$ and $\hat{\sigma}_{\eta(j)}^{2(2)}$ similar to (3.45) and (3.46) though obtained from the appropriate two-step residuals $\widehat{\tilde{\varepsilon}}_{i(j)}^{s(2)} = \tilde{y}_i - \tilde{R}_i^*\widehat{\tilde{\alpha}}_{(j)}^{(2)}$ and $\hat{u}_{i(j)}^{s(2)} = y_i - \bar{R}_i\widehat{\alpha}_{(j)}^{(2)}$, the statistics to be used after 2-step estimation are

$$JBB_j^{(2,h)} = \left(\sum_{i=1}^{N} \widehat{\ddot{u}}_{i(j)}^{(2)\prime} \ddot{Z}_i^s\right) \hat{S}_j^{(h)} \left(\sum_{i=1}^{N} \ddot{Z}_i^{s\prime}\widehat{\ddot{u}}_{i(j)}^{(2)}\right), \qquad (3.61)$$

where $\hat{S}_a^{(2)}$ and $\hat{S}_b^{(2)}$ are like $\hat{S}_a^{(1)}$ and $\hat{S}_b^{(1)}$, except that they use $\widehat{\ddot{u}}_{i(a)}^{(2)}$ and $\widehat{\ddot{u}}_{i(b)}^{(2)}$ instead of $\widehat{\ddot{u}}_i^{(1)}$. With respect to $\hat{S}_c^{(2)}$ one can use

---

[3]Package xtabond2 for Stata always reports $JAB^{(1,0)}$ after Arellano-Bond estimation, which is inappropriate when there is heteroskedasticity. After requesting for robust standard errors in 1-step estimation it presents also $JABa^{(2,1)}$ instead of $JABa^{(1,1)}$. Requesting 2-step estimation also presents both $JAB^{(1,0)}$ and $JABa^{(2,1)}$. Blundell-Bond estimation yields $JBB^{(1,0)}$ and $JBB^{(2,1)}$, although a version of $JBB^{(1,0)}$ is reported that does not use weighting matrix $S^{(0)}(\hat{\sigma}_\eta^{2,s(1)}/\hat{\sigma}_\varepsilon^{2,s(1)})$, but $S^{(0)}(0)$, which is only valid under homoskedasticity and $\sigma_\eta^2 = 0$.

$$\hat{S}_c^{(2)} = \left[ \sum_{i=1}^{N} \hat{\sigma}_{\varepsilon,i(j)}^{2,s(2)} \left( \begin{array}{cc} Z_i'HZ_i & Z_i'D\tilde{Z}_i^s \\ \tilde{Z}_i^{s\prime}D'Z_i & \tilde{Z}_i^{s\prime}[I + (\hat{\sigma}_{\eta(c)}^{2(2)}/\hat{\sigma}_{\varepsilon,i(c)}^{2,s(2)})\iota_T\iota_T']\tilde{Z}_i^s \end{array} \right) \right]^{-1}.$$

Under their respective null hypotheses the tests based on Arellano-Bond estimation follow asymptotically $\chi^2$ distributions with $L - K + 1$ degrees of freedom, whereas the tests based on Blundell-Bond estimates have $L + L^{\blacklozenge} - K$ degrees of freedom. Self-evidently tests on the effect stationarity related orthogonality conditions are given by

$$\begin{aligned} JES^{(1,0)} &= JBB^{(1,0)} - JAB^{(1,0)}, & (3.62) \\ JES_j^{(l,h)} &= JBB_j^{(l,h)} - JAB_j^{(l,h)}, \ l \le h \in \{1,2\}, \ j \in \{a,b,c\} & (3.63) \end{aligned}$$

and should be compared with a $\chi^2$ critical value for $L^{\blacklozenge} - 1$ degrees of freedom.[4]

## 3.6 Modified GMM

In the special case that panel model (3.5) has cross-sectional heteroskedasticity and no time-series heteroskedasticity, hence

$$\sigma_\varepsilon^2 \Omega_i = \sigma_{\varepsilon,i}^2 I, \ \text{with} \ \sum_{i=1}^{N} \sigma_{\varepsilon,i}^2 = \sigma_\varepsilon^2 N, \tag{3.64}$$

we can easily employ MGMM estimator (2.11). However, because $H^{-1}$ is not a lower-triangular matrix, not all instruments $\sigma_{\varepsilon,i}^{-2}H^{-1}Z_i$ would be valid for the equation in first-differences. This problem can be avoided by using, instead of first-differencing, the forward orthogonal deviation (FOD) transformation for removing the individual effects. Let

$$B = \begin{bmatrix} \frac{T-1}{T} & 0 & 0 & \cdots & 0 \\ 0 & \frac{T-2}{T-1} & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \frac{2}{3} & 0 \\ 0 & 0 & \cdots & 0 & \frac{1}{2} \end{bmatrix}^{1/2} \begin{bmatrix} 1 & -\frac{1}{T-1} & \cdots & \cdots & -\frac{1}{T-1} & -\frac{1}{T-1} \\ 0 & 1 & -\frac{1}{T-2} & \cdots & -\frac{1}{T-2} & -\frac{1}{T-2} \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \ddots & 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & \cdots & 0 & 1 & -1 \end{bmatrix}, \tag{3.65}$$

and $\check{\varepsilon}_i = B\varepsilon_i$. Then $Bu_i = \check{\varepsilon}_i \sim (0, \sigma_{\varepsilon,i}^2 I_{T-1})$ provided (3.64) holds, whereas $E(Z_i'\check{\varepsilon}_i) = 0$ for $Z_i$ given by (3.11). Hence, estimating model

$$\check{y}_i = \check{R}_i \tilde{\alpha} + \check{\varepsilon}_i, \tag{3.66}$$

where $\check{y}_i = By_i$ and $\check{R}_i = BR_i$, by GMM, but using an instrument matrix with components $\sigma_{\varepsilon,i}^{-2}Z_i$ yields the unfeasible MABu estimator for the model with cross-sectional heteroskedasticity, which is

$$\begin{aligned} \widehat{\tilde{\alpha}}_{MABu} &= \left[ \left( \sum_{i=1}^{N} \sigma_{\varepsilon,i}^{-2}\check{R}_i'Z_i \right) \left( \sum_{i=1}^{N} \sigma_{\varepsilon,i}^{-2}Z_i'Z_i \right)^{-1} \left( \sum_{i=1}^{N} \sigma_{\varepsilon,i}^{-2}Z_i'\check{R}_i \right) \right]^{-1} \times \\ & \left( \sum_{i=1}^{N} \sigma_{\varepsilon,i}^{-2}\check{R}_i'Z_i \right) \left( \sum_{i=1}^{N} \sigma_{\varepsilon,i}^{-2}Z_i'Z_i \right)^{-1} \left( \sum_{i=1}^{N} \sigma_{\varepsilon,i}^{-2}Z_i'\check{y}_i \right). \quad (3.67) \end{aligned}$$

---

[4]By specifying instruments in separate groups, xtabond2 presents for each separate group the corresponding incremental $J$ test. However, not the version as defined in (2.17), but an asymptotically equivalent one as suggested in Hayashi (2000, p.220) which will never be negative.

Note that the exploited moment conditions are here $E(\sigma_{\varepsilon,i}^{-2} Z_i' \check{\varepsilon}_i) = \sigma_{\varepsilon,i}^{-2} E(Z_i' \check{\varepsilon}_i) = 0$. For $\sigma_{\varepsilon,i}^2 > 0$ these are intrinsically equivalent with $E(Z_i' \check{\varepsilon}_i) = 0$, but they induce the use of a different set of instruments yielding a different estimator. That it is most likely that the unfeasible standard AB estimator ABu, which uses instruments $\sigma_{\varepsilon,i} Z_i$ for regressors $\sigma_{\varepsilon,i}^{-1} \check{R}_i$, will generally exploit weaker instruments than MABu, which uses $\sigma_{\varepsilon,i}^{-1} Z_i$ for regressors $\sigma_{\varepsilon,i}^{-1} \check{R}_i$, should be intuitively obvious.

To convert this into a feasible procedure, one could initially assume that all $\sigma_{\varepsilon,i}^2$ are equal. Then the first-step MGMM estimator $\widehat{\alpha}_{MAB}^{(1)}$ is numerically equivalent to AB1 of (3.20), provided all instruments are being used.[5] Next, exploiting (3.26), the feasible 2-step MAB2 estimator can be obtained by

$$
\widehat{\alpha}_{MAB2}^{(2)} = \left[ \left( \sum_{i=1}^{N} \check{R}_i' Z_i / \hat{\sigma}_{\varepsilon,i}^{2(1)} \right) \left( \sum_{i=1}^{N} Z_i' Z_i / \hat{\sigma}_{\varepsilon,i}^{2(1)} \right)^{-1} \left( \sum_{i=1}^{N} Z_i' \check{R}_i / \hat{\sigma}_{\varepsilon,i}^{2(1)} \right) \right]^{-1} \times
$$
$$
\left( \sum_{i=1}^{N} \check{R}_i' Z_i / \hat{\sigma}_{\varepsilon,i}^{2(1)} \right) \left( \sum_{i=1}^{N} Z_i' Z_i / \hat{\sigma}_{\varepsilon,i}^{2(1)} \right)^{-1} \left( \sum_{i=1}^{N} Z_i' \check{y}_i / \hat{\sigma}_{\varepsilon,i}^{2(1)} \right). \quad (3.68)
$$

Modifying the system estimator is more problematic, primarily because the inverse of the matrix $Var(u_i) = \Sigma_i = \sigma_{\varepsilon,i}^2 I_T + \sigma_\eta^2 \iota_T \iota_T'$, which is $\Sigma_i^{-1} = \sigma_{\varepsilon,i}^{-2} [I_T + (T + \sigma_{\varepsilon,i}^2 / \sigma_\eta^2)^{-1} \iota_T \iota_T']$, is nondiagonal. So, although $E(\tilde{Z}_i^{s\prime} u_i) = 0$, surely $E(\tilde{Z}_i^{s\prime} \Sigma_i^{-1} u_i) \neq 0$. However, as an unfeasible modified system estimator we can combine estimation of the model for $\check{y}_i$ using instruments $\sigma_{\varepsilon,i}^{-2} Z_i$ with estimation of the model for $y_i$ using instruments $(\sigma_{\varepsilon,i}^2 + \sigma_\eta^2)^{-1} \tilde{Z}_i^s$. So, the system is then given by the model

$$
\dddot{y}_i = \dddot{R}_i \ddot{\alpha} + \dddot{u}_i, \quad (3.69)
$$

where $\dddot{y}_i = (\check{y}_i' \ y_i')'$, $\dddot{R}_i = (\check{R}_i^{*\prime} \ \bar{R}_i^{*\prime})'$, with $\check{R}_i^* = (\check{R}_i \ 0)$, and $\dddot{u}_i = (\check{\varepsilon}_i' \ u_i')'$.

For the 1-step estimator we could again choose some nonnegative value $q$ and calculate the 1-step estimator BB1 given in (3.40) in order to find residuals and obtain the estimators $\hat{\sigma}_{\varepsilon,i}^{2,s(1)}$ and $\hat{\sigma}_\eta^{2,s(1)}$ of (3.45) and (3.46). Building on $E(\dddot{u}_i \dddot{u}_i')$ and instrument matrix block $\dddot{Z}_i$, given by

$$
E(\dddot{u}_i \dddot{u}_i') = \sigma_{\varepsilon,i}^2 \begin{pmatrix} I_{T-1} & B \\ B' & I_T + (\sigma_\eta^2 / \sigma_{\varepsilon,i}^2) \iota_T \iota_T' \end{pmatrix} \text{ and } \dddot{Z}_i = \begin{pmatrix} \hat{\sigma}_{\varepsilon,i}^{-2,s(1)} Z_i & O \\ O & (\hat{\sigma}_{\varepsilon,i}^{2,s(1)} + \hat{\sigma}_\eta^{2,s(1)})^{-1} \tilde{Z}_i^s \end{pmatrix},
$$

one obtains weighting matrix

$$
\hat{S}_c^{B(1)} = \left[ \sum_{i=1}^{N} \begin{pmatrix} \hat{\sigma}_{\varepsilon,i}^{-2,s(1)} Z_i' Z_i & (\hat{\sigma}_{\varepsilon,i}^{2,s(1)} + \hat{\sigma}_\eta^{2,s(1)})^{-1} Z_i' B \tilde{Z}_i^s \\ (\hat{\sigma}_{\varepsilon,i}^{2,s(1)} + \hat{\sigma}_\eta^{2,s(1)})^{-1} \tilde{Z}_i^{s\prime} B' Z_i & (\hat{\sigma}_{\varepsilon,i}^{2,s(1)} + \hat{\sigma}_\eta^{2,s(1)})^{-2} \tilde{Z}_i^{s\prime} [\hat{\sigma}_{\varepsilon,i}^{2,s(1)} I + \hat{\sigma}_\eta^{2,s(1)} \iota_T \iota_T'] \tilde{Z}_i^s \end{pmatrix} \right]^{-1},
$$
$$
(3.70)
$$

which can be exploited in the feasible 2-step MGMM system estimator MBB2

$$
\widehat{\alpha}_{MBB}^{(2)} = \left[ \left( \sum_{i=1}^{N} \dddot{R}_i' \dddot{Z}_i \right) \hat{S}_c^{B(1)} \left( \sum_{i=1}^{N} \dddot{Z}_i' \dddot{R}_i \right) \right]^{-1} \left( \sum_{i=1}^{N} \dddot{R}_i' \dddot{Z}_i \right) \hat{S}_c^{B(1)} \left( \sum_{i=1}^{N} \dddot{Z}_i' \dddot{y}_i \right). \quad (3.71)
$$

For both $\widehat{\alpha}_{MAB}^{(2)}$ and $\widehat{\alpha}_{MBB}^{(2)}$ relevant t-test and Sargan-Hansen test statistics can be constructed. Regarding the latter we will just examine

$$
JMAB = \left( \sum_{i=1}^{N} \widehat{\check{\varepsilon}}_i^{(2)\prime} Z_i / \hat{\sigma}_{\varepsilon,i}^{2,s(1)} \right) \left( \sum_{i=1}^{N} Z_i' Z_i / \hat{\sigma}_{\varepsilon,i}^{2,s(1)} \right)^{-1} \left( \sum_{i=1}^{N} Z_i' \widehat{\check{\varepsilon}}_i^{(2)} / \hat{\sigma}_{\varepsilon,i}^{2,s(1)} \right), \quad (3.72)
$$

---

[5]Proved in Arellano and Bover (1995).

where $\widehat{\widetilde{\varepsilon}}_i^{(2)} = \breve{y}_i - \breve{R}_i \widehat{\widetilde{\alpha}}_{MGMMch}^{(2)}$, and

$$JMBB = \left( \sum_{i=1}^N \widehat{\ddot{u}}_i^{(2)\prime} \dddot{Z}_i^s \right) \hat{S}_c^{B(1)} \left( \sum_{i=1}^N \dddot{Z}_i^{s\prime} \widehat{\ddot{u}}_i^{(2)} \right), \qquad (3.73)$$

with $\widehat{\ddot{u}}_i^{(2)} = \dddot{y}_i - \dddot{R}_i \widehat{\ddot{\alpha}}_{MBB}^{(2)} = (\widehat{\widetilde{\varepsilon}}_i^{s(2)\prime} \ \hat{u}_i^{s(2)\prime})'$. Under their respective null hypotheses these follow asymptotically $\chi^2$ distributions with $L - K + 1$ and $L + L^\blacklozenge - K$ degrees of freedom. Self-evidently, the test on the effect stationarity related orthogonality conditions is given by

$$JESM = JMBB - JMAB. \qquad (3.74)$$

# 4  Simulation design

We will examine the stable dynamic simultaneous heteroskedastic DGP ($i = 1, ..., N$, $t = 1, ..., T$)

$$y_{it} = \alpha_y + \gamma y_{i,t-1} + \beta x_{it} + \sigma_\eta \eta_i^\circ + \sigma_\varepsilon \omega_i^{1/2} \varepsilon_{it}^\circ \qquad (|\gamma| < 1). \qquad (4.1)$$

Here $\beta$ has just one element relating to the for each $i$ stable autoregressive regressor

$$x_{it} = \alpha_x + \xi x_{i,t-1} + \pi_\eta \eta_i^\circ + \pi_\lambda \lambda_i^\circ + \sigma_v \omega_i^{1/2} v_{it}^\circ \qquad (|\xi| < 1), \qquad (4.2)$$

where

$$v_{it}^\circ = \rho_{v\varepsilon} \varepsilon_{it}^\circ + (1 - \rho_{v\varepsilon}^2)^{1/2} \zeta_{it}^\circ. \qquad (4.3)$$

All random drawings $\eta_i^\circ$, $\varepsilon_{it}^\circ$, $\lambda_i^\circ$, $\zeta_{it}^\circ$ are $IID(0,1)$ and mutually independent. Parameter $\rho_{v\varepsilon}$ indicates the correlation between the disturbances $\varepsilon_{it} = \sigma_\varepsilon \omega_i^{1/2} \varepsilon_{it}^\circ$ and $v_{it} = \sigma_v \omega_i^{1/2} v_{it}^\circ$. These may have different variances, but have similar cross-sectional heteroskedasticity pattern, determined by the fixed values $\omega_1, ..., \omega_N$. How we did generate these $\omega_i$ and the start-up values $x_{i,0}$ and $y_{i,0}$ and chose relevant numerical values for the other eleven parameters will be discussed extensively below.

Note that in this DGP $x_{it}$ is either strictly exogenous ($\rho_{v\varepsilon} = 0$) or otherwise endogenous[6]; the only weakly exogenous regressor is $y_{i,t-1}$. Regressor $x_{it}$ may be affected contemporaneously by two independent individual specific effects when $\pi_\eta \neq 0$ and $\pi_\lambda \neq 0$, but also with delays if $\xi \neq 0$. The dependent variable $y_{it}$ may be affected contemporaneously by the (standardized) individual effect $\eta_i^\circ$, both directly and indirectly; directly if $\sigma_\eta \neq 0$, and indirectly via $x_{it}$ when $\beta \pi_\eta \neq 0$. However, $\eta_i^\circ$ will also have delayed effects on $y_{it}$, when $\gamma \neq 0$ or $\xi \beta \pi_\eta \neq 0$, and so has $\lambda_i^\circ$ when $\xi \beta \pi_\lambda \neq 0$.

For the cross-sectional heteroskedasticity we follow an approach similar to Kiviet and Feng (2014). It is determined by both $\eta_i^\circ$ and $\lambda_i^\circ$, the two standardized individual effects, and is thus associated with the regressors $x_{it}$ and $y_{i,t-1}$. It follows a lognormal pattern when both $\eta_i^\circ$ and $\delta_i^\circ$ are standard normal, because we take

$$\omega_i = e^{h_i(\theta)}, \text{ with } h_i(\theta) = -\theta^2/2 + \theta[\kappa^{1/2} \eta_i^\circ + (1-\kappa)^{1/2} \lambda_i^\circ] \sim NID(-\theta^2/2, \theta^2), \quad (4.4)$$

where $0 \leq \kappa \leq 1$. This establishes a lognormal distribution with $E(\omega_i) = 1$ and $Var(\omega_i) = e^{\theta^2} - 1$. So, for $\theta = 0$ the $\varepsilon_{it}$ and $v_{it}$ are homoskedastic. The seriousness

---

[6]If we would strictly follow the notation of the earlier sections the coefficient $\beta$ should actually be called $\delta$ when $\rho_{v\varepsilon} \neq 0$.

of the heteroskedasticity increases with the absolute value of $\theta$. Obviously, for $\kappa = 0$ the error components $\eta_i$ and $\varepsilon_{it}$ are independent (like $\eta_i^\circ$ and $\varepsilon_{it}^\circ$), but not for $0 < \kappa \leq 1$. From $h_i(\theta)/2 \sim NID(-\theta^2/4, \theta^2/4)$ it follows that $\omega_i^{1/2} = e^{h_i(\theta)/2}$ is lognormally distributed too, with $E(\omega_i^{1/2}) = e^{-\theta^2/8}$. Table 2 presents some quantiles of the distributions of $\omega_i$ and $\omega_i^{1/2}$ (taken as the positive square root of $\omega_i$) and the expectation of the latter in order to disclose the effects of parameter $\theta$. It shows that $\theta \geq 1$ implies pretty serious heteroskedasticity, whereas it may be qualified mild when $\theta \leq 0.3$, say.

**Table 2.** Heteroskedasticity quantiles and $E(\omega_i^{1/2})$

| $\lvert\theta\rvert$ | $\omega_i$ | | | $\omega_i^{1/2}$ | | | |
|---|---|---|---|---|---|---|---|
| | $q_{0.01}$ | $q_{0.5}$ | $q_{0.99}$ | $q_{0.01}$ | $q_{0.5}$ | $q_{0.99}$ | $E(\omega_i^{1/2})$ |
| .1 | .789 | .995 | 1.256 | .888 | .998 | 1.121 | .999 |
| .3 | .476 | .956 | 1.921 | .690 | .978 | 1.386 | .999 |
| .5 | .276 | .883 | 2.824 | .525 | .939 | 1.681 | .969 |
| .7 | .154 | .783 | 3.989 | .391 | .885 | 1.997 | .941 |
| 1.0 | .059 | .607 | 6.211 | .243 | .779 | 2.492 | .882 |
| 1.3 | .021 | .430 | 8.840 | .145 | .655 | 2.973 | .810 |
| 1.6 | .007 | .278 | 11.498 | .082 | .527 | 3.391 | .726 |
| 2.0 | .001 | .135 | 14.192 | .036 | .368 | 3.767 | .607 |

Without loss of generality we may chose $\sigma_\varepsilon = 1$ and $\alpha_y = \alpha_x = 0$. Note that (4.1) implicitly specifies $\tilde{\tau} = 0$. All simulation results refer to estimators where these $T - 1$ restrictions have been imposed (there are no time effects), but $\alpha_y = \alpha_x = 0$ have not been imposed. Hence, when estimating the model in levels $\iota_T$ is one of the regressors. Moreover, we may always include $I_{T-1}$ in $Z_i$ and $\iota_T$ in $\tilde{Z}_i^s$ in order to exploit the fundamental moment conditions $E(\tilde{\varepsilon}_{it}) = 0$ (for $t = 2, ..., T$) and $E[\sum_{t=1}^{T}(\eta_i + \varepsilon_{it})] = 0$ for $i = 1, ..., N$.

Apart from values for $\theta$ and $\kappa$, we have to make choices on relevant values for eight more parameters. We could choose $\gamma \in \{0.2, 0.5, 0.8\}$, which covers a broad range of adjustment processes for dynamic behavioral relationships, and $\xi \in \{0.5, 0.8, 0.95\}$ to include less and more smooth $x_{it}$ processes. Next, interesting values should be given to the remaining six parameters, namely $\beta$, $\sigma_\eta$, $\pi_\eta$, $\pi_\lambda$, $\sigma_v$ and $\rho_{v\varepsilon}$. We will do this by choosing relevant values for six alternative more meaningful notions, which are all functions of some of the eight DGP parameters and allow to establish relevant numerical values for them.

The first three notions will be based on (ratios of) particular variance components of the long-run stationary path of the process for $x_{it}$. Using lag-operator notation and assuming that $v_{it'}^\circ$ (and $\varepsilon_{it'}^\circ$) exist for $t' = -\infty, ...., 0, 1, ..., T$, we find[7] that the long-run path for $x_{it}$ consists of three mutually independent components, namely

$$x_{it} = \frac{\pi_\eta}{1 - \xi}\eta_i^\circ + \frac{\pi_\lambda}{1 - \xi}\lambda_i^\circ + \frac{\sigma_v \omega_i^{1/2}}{1 - \xi L}v_{it}^\circ. \tag{4.5}$$

The third component, the accumulated contributions of $v_{it}^\circ$, is a stationary AR(1) process with variance $\sigma_v^2 \omega_i/(1 - \xi^2)$. Approximating $N^{-1}\sum_{i=1}^{N}\omega_i$ by 1, the average variance is

---

[7]Note that $(1 - \xi L)^{-1} = 1 + \xi L + \xi^2 L^2 + ...$ and therefore $(1 - \xi L)^{-1}\eta_i^\circ = \sum_{j=0}^{\infty}\xi^j\eta_i^\circ = \eta_i^\circ/(1 - \xi)$.

$\sigma_v^2/(1 - \xi^2)$. The other two components have variances $\pi_\eta^2/(1 - \xi)^2$ and $\pi_\lambda^2/(1 - \xi)^2$ respectively, so the average long-run variance of $x_{it}$ equals

$$\bar{V}_x = \frac{\pi_\eta^2 + \pi_\lambda^2}{(1 - \xi)^2} + \frac{\sigma_v^2}{1 - \xi^2}. \tag{4.6}$$

A first characterization of the $x_{it}$ series can be obtained by setting $\bar{V}_x = 1$. This is an innocuous normalization, because $\beta$ is still a free parameter. As a second characterization of the $x_{it}$ series, we can choose what we call the (average) effects variance fraction of $x_{it}$, given by

$$EVF_x = \frac{\pi_\eta^2 + \pi_\lambda^2}{(1 - \xi)^2}/\bar{V}_x, \tag{4.7}$$

with $0 \leq EVF_x \leq 1$, for which we could take, say, $EVF_x \in \{0, 0.3, 0.6\}$. To balance the two individual effect variances we define for the case $EVF_x > 0$ what we call the individual effect fraction of $\eta_i^\circ$ in $x_{it}$ given by

$$IEF_x^\eta = \pi_\eta^2/(\pi_\eta^2 + \pi_\lambda^2). \tag{4.8}$$

So $IEF_x^\eta$, with $0 \leq IEF_x^\eta \leq 1$, expresses the fraction due to $\pi_\lambda \eta_i^\circ$ of the (long-run) variance of $x_{it}$ stemming from the two individual effects. We could take, say, $IEF_x^\eta \in \{0, 0.3, 0.6\}$.

From these three characterizations we obtain

$$\pi_\lambda = (1 - \xi)[(1 - IEF_x^\eta)EVF_x]^{1/2}, \tag{4.9}$$
$$\pi_\eta = (1 - \xi)[IEF_x^\eta EVF_x]^{1/2}, \tag{4.10}$$
$$\sigma_v = [(1 - \xi^2)(1 - EVF_x)]^{1/2}. \tag{4.11}$$

For all three we will only consider the nonnegative root, because changing the sign would have no effects on the characteristics of $x_{it}$, as we will generate the series $\eta_i^\circ$, $\varepsilon_{it}^\circ$, $\lambda_i^\circ$ and $\zeta_{it}^\circ$ from symmetric distributions. The above choices regarding the $x_{it}$ process have the following implications for the average correlations between $x_{it}$ and its two constituting effects:

$$\bar{\rho}_{x\eta} = \pi_\eta/(1 - \xi) = [(1 - IEF_x^\eta)EVF_x]^{1/2}, \tag{4.12}$$
$$\bar{\rho}_{x\lambda} = \pi_\lambda/(1 - \xi) = [IEF_x^\eta EVF_x]^{1/2}. \tag{4.13}$$

Now the $x_{it}$ series can be generated upon choosing a value for $\rho_{v\varepsilon}$. This we obtain from $E(x_{it}\varepsilon_{it}) = \sigma_\varepsilon \sigma_v \rho_{v\varepsilon}\omega_i$, which on average is $\sigma_v \rho_{v\varepsilon}$. Hence, fixing the average simultaneity[8] to $\bar{\rho}_{x\varepsilon}$, we should choose

$$\rho_{v\varepsilon} = \bar{\rho}_{x\varepsilon}/\sigma_v. \tag{4.14}$$

In order that both correlations are smaller than 1 in absolute value an admissibility restriction has to be satisfied, namely $\bar{\rho}_{x\varepsilon}^2 \leq \sigma_v^2$, giving

$$\bar{\rho}_{x\varepsilon}^2 \leq (1 - \xi^2)(1 - EVF_x). \tag{4.15}$$

---

[8]Such control is not exercised in the simulation designs of Blundell et al. (2001) and Bun and Sarafidis (2013). They do consider simultaneity, but its magnitude has not been mentioned and it is not kept constant over different designs.

When choosing $EVF_x = 0.6$ and $\xi = 0.8$ we should have $|\bar{\rho}_{x\varepsilon}| \leq 0.379$. That we should not exclude negative values of $\bar{\rho}_{x\varepsilon}$ will become obvious in due course. For the moment it seems interesting to examine, say, $\bar{\rho}_{x\varepsilon} \in \{-0.3, 0, 0.3\}$.

The remaining choices concern $\beta$ and $\sigma_\eta$ which both directly affect the DGP for $y_{it}$. Substituting (4.5) and (4.3) in (4.1) we find that the long-run stationary path for $y_{it}$ entails four mutually independent components, since

$$
\begin{aligned}
y_{it} &= \frac{\beta}{1 - \gamma L} x_{it} + \frac{\sigma_\eta}{1 - \gamma} \eta_i^\circ + \frac{\sigma_\varepsilon \omega_i^{1/2}}{1 - \gamma L} \varepsilon_{it}^\circ \\
&= \frac{\beta \pi_\eta + (1 - \xi)\sigma_\eta}{(1 - \gamma)(1 - \xi)} \eta_i^\circ + \frac{\beta \pi_\lambda}{(1 - \gamma)(1 - \xi)} \lambda_i^\circ \\
&\quad + \frac{\beta \sigma_v \omega_i^{1/2}}{(1 - \gamma L)(1 - \xi L)} v_{it}^\circ + \frac{\sigma_\varepsilon \omega_i^{1/2}}{1 - \gamma L} \varepsilon_{it}^\circ \\
&= \frac{\beta \pi_\eta + (1 - \xi)\sigma_\eta}{(1 - \gamma)(1 - \xi)} \eta_i^\circ + \frac{\beta \pi_\lambda}{(1 - \gamma)(1 - \xi)} \lambda_i^\circ \\
&\quad \frac{\beta \sigma_v (1 - \rho_{v\varepsilon}^2)^{1/2} \omega_i^{1/2}}{(1 - \gamma L)(1 - \xi L)} \zeta_{it}^\circ + \frac{[\beta \rho_{v\varepsilon} \sigma_v + (1 - \xi L)\sigma_\varepsilon]\omega_i^{1/2}}{(1 - \gamma L)(1 - \xi L)} \varepsilon_{it}^\circ
\end{aligned}
\tag{4.16}
$$

The third term of the final expression constitutes for each $i$ an AR(2) process and the fourth one an ARMA(2,1) process. The variance of each of the four terms is given (derivations in the Appendix) by

$$
\begin{aligned}
V_\eta &= \frac{[\beta \pi_\eta + (1 - \xi)\sigma_\eta]^2}{(1 - \gamma)^2(1 - \xi)^2} \\
V_\lambda &= \frac{\beta^2 \pi_\lambda^2}{(1 - \gamma)^2(1 - \xi)^2} \\
V_\zeta(i) &= \frac{\beta^2 \sigma_v^2 (1 - \rho_{v\varepsilon}^2)(1 + \gamma \xi)}{(1 - \gamma^2)(1 - \xi^2)(1 - \gamma \xi)} \omega_i \\
V_\varepsilon(i) &= \frac{[(1 + \beta \rho_{v\varepsilon} \sigma_v)^2 + \xi^2](1 + \gamma \xi) - 2\xi(1 + \beta \rho_{v\varepsilon} \sigma_v)(\gamma + \xi)}{(1 - \gamma^2)(1 - \xi^2)(1 - \gamma \xi)} \omega_i.
\end{aligned}
$$

Averaging the last two over all $i$ yields $\bar{V}_\zeta$ and $\bar{V}_\varepsilon$. For the average long-run variance of $y_{it}$ we then can evaluate

$$
\bar{V}_y = V_\eta + V_\lambda + \bar{V}_\zeta + \bar{V}_\varepsilon.
\tag{4.17}
$$

When choosing fixed values for ratios involving these components to obtain values for $\beta$ and $\sigma_\eta$ we will run into the problem of multiple solutions. On the other hand, the four components of (4.17) have particular invariance properties regarding the signs of $\beta$, $\sigma_\eta$ and $\rho_{v\varepsilon}$, since changing the sign of all three yields exactly the same value of $\bar{V}_y$. We coped with this as follows. Although we note that $V_\eta$ does depend on $\beta \pi_\eta$, we set $\sigma_\eta$ simply by fixing the direct cumulated effect impact of $\eta_i^\circ$ on $y_{it}$ relative to the current noise $\sigma_\varepsilon = 1$. This is

$$
DEN_y^\eta = \sigma_\eta / (1 - \gamma).
\tag{4.18}
$$

Because the direct and indirect (via $x_{it}$) effects from $\eta_i^\circ$ may have opposite signs, $DEN_y^\eta$ could be given negative values too, but we restricted ourselves to $DEN_y^\eta \in \{1, 4\}$,

yielding

$$\sigma_\eta = (1 - \gamma)DEN_y^\eta. \tag{4.19}$$

Finally we fix a signal-noise ratio, which gives a value for $\beta$. Because under simultaneity the noise and current signal conflate, we focus on the case where $\rho_{x\varepsilon} = 0$. Then we have

$$\bar{V}_\zeta = \frac{\beta^2 \sigma_v^2 (1 + \gamma\xi)}{(1 - \gamma^2)(1 - \xi^2)(1 - \gamma\xi)}, \quad \bar{V}_\varepsilon = \frac{1}{1 - \gamma^2}.$$

Leaving the variance due to the effects aside, the average signal variance is $\bar{V}_\zeta + \bar{V}_\varepsilon - 1$, because the current average noise variance is unity. Hence, we may define a signal-noise ratio as

$$SNR = \bar{V}_\zeta + \bar{V}_\varepsilon - 1 = \frac{\beta^2 (1 - EVR_x)(1 + \gamma\xi)}{(1 - \gamma^2)(1 - \gamma\xi)} + \frac{\gamma^2}{(1 - \gamma^2)}, \tag{4.20}$$

where we have substituted (4.11). For this we may choose, say, $SNR \in \{2, 3, 5\}$, in order to find

$$\beta = \left[ \frac{1 - \gamma\xi}{1 + \gamma\xi} \frac{SNR - \gamma^2(SNR + 1)}{1 - EVR_x} \right]^{1/2}. \tag{4.21}$$

Note that here another admissibility restriction crops up, namely

$$\gamma^2 \leq SNR/(SNR + 1). \tag{4.22}$$

However, for $|\gamma| \leq 0.8$ this is satisfied for $SNR \geq 1.78$. From (4.21) we only examined the positive root.

Instead of fixing SNR another approach would be to fix the total multiplier

$$TM = \beta/(1 - \gamma), \tag{4.23}$$

which would directly lead to a value for $\beta$, given $\gamma$. However, different $TM$ values will then lead to different $SNR$ values, because

$$SNR = TM^2(1 - EVF_x)\frac{(1 - \gamma)(1 + \gamma\xi)}{(1 + \gamma)(1 - \gamma\xi)} + \frac{\gamma^2}{(1 - \gamma^2)}. \tag{4.24}$$

At this stage it is hard to say what would yield more useful information from the Monte Carlo, fixing $TM$ or $SNR$. Keeping both constant for different $\gamma$ and some other characteristics of this DGP is out of the question. We chose to fix $SNR = 3$. which yields $TM$ values in the range 1.5-1.8. When comparing with results for $TM = 1$ we did not note substantial principle differences.

For all different design parameter combinations considered, which involve sample size $N = 200$ and $T \in \{3, 6, 9\}$, we used the very same realizations of the underlying standardized random components $\eta_i^\circ$, $\lambda_i^\circ$, $\varepsilon_{it}^\circ$ and $\zeta_{it}^\circ$ over the respective 10000 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 $\eta_i^\circ$ and $\lambda_i^\circ$ such, that they have sample mean zero, sample variance 1 and sample correlation zero. This rescaling is achieved by replacing the $N$ draws $\eta_i^\circ$ 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 $\lambda_i^\circ$ by the residuals obtained after regressing $\lambda_i^\circ$ on $\eta_i^\circ$ 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 $\omega_i$ by dividing them by $N^{-1} \sum_{i=1}^{N} \omega_i$, so that the resulting $\omega_i$ have sum $N$ as they should in order to avoid that presence of heteroskedasticity is conflated with larger or smaller average disturbance variance.

In the simulation experiments we will start-up the processes for $x_{it}$ and $y_{it}$ at pre-sample period $s < 0$ by taking $x_{is} = 0$ and $y_{i,s} = 0$ and next generate $x_{it}$ and $y_{it}$ for the indices $t = s + 1, ..., T$. The data with time-indices $s, ..., -1$ will be discarded when estimating the model. We suppose that for $s = -50$ both series will be on their stationary track from $t = 0$ onwards. When taking $s = -1$ or $-2$ the initial values $y_{i0}$ and $x_{i1}$ will be such that effect-stationarity has not yet been achieved. Due to the fixed zero startups (which are equal to the unconditional expectations) the (cross-)autocorrelations of the $x_{it}$ and $y_{it}$ series have a very peculiar start then too, so our results regarding effect nonstationarity will certainly not be fully general, but in a way for $s$ close to zero we mimic the situation that a process only started very recently.

Another simple way to mimic a situation in which lagged first-differenced variables are invalid instruments for the model in levels can be designed as follows. Equations (4.5) and (4.16) highlight that in the long-run $\Delta x_{it}$ and $\Delta y_{it}$ are uncorrelated with the effects $\eta_i^\circ$ and $\lambda_i^\circ$. This can be undermined by perturbing $x_{i0}$ and $y_{i0}$ as obtained from $s = -50$ in such a way that we add to them the values

$$(\phi - 1)\left[\frac{\pi_\eta}{1 - \xi}\eta_i^\circ + \frac{\pi_\lambda}{1 - \xi}\lambda_i^\circ\right] \text{ and } (\phi - 1)\left[\frac{\beta\pi_\eta + (1 - \xi)\sigma_\eta}{(1 - \gamma)(1 - \xi)}\eta_i^\circ + \frac{\beta\pi_\lambda}{(1 - \gamma)(1 - \xi)}\lambda_i^\circ\right]$$
(4.25)

respectively. Note that for $\phi = 1$ effect stationarity is maintained, whereas for $0 \leq \phi < 1$ the dependence of $x_{i0}$ and $y_{i0}$ on the effects is mitigated in comparison to the stationary track (upon maintaining stationarity regarding $\varepsilon_{it}^\circ$ and $\zeta_{it}^\circ$), whereas for $\phi > 1$ this dependence is inflated. Note that this is a straight-forward generalization of the approach followed in Kiviet (2007) for the panel AR(1) model.

# 5   Simulation results

To limit the number of tables we proceed as follows. Often we will first produce results on unfeasible implementations of the various inference techniques in relatively simple DGPs. These exploit the true values of $\omega_1, ..., \omega_N$, $\sigma_\varepsilon^2$ and $\sigma_\eta^2$ instead of their estimates. Although this information is generally not available in practice, only when such unfeasible techniques behave reasonably well in finite samples it seems useful to examine in more detail the performance of feasible implementations. Results for the unfeasible Arellano Bond (1991) and Blundell Bond (1998) GMM estimators are denoted as ABu and BBu respectively. As indicated before, their feasible counterparts are denoted as AB1 and BB1 for the 1-step (which under homoskedasticity are equivalent to their unfeasible counterparts) and AB2 and BB2 for the 2-step estimators. For 2-step estimators the lower case letters a, b or c are used (as in for instance AB2c) to indicate which type of weighting matrix has been exploited, as discussed in sections 3.2.1 and 3.3.2. For corresponding MGMM implementations these acronyms are preceded by the letter M. Under homoskedasticity their unfeasible implementation has been omitted when this is equivalent to GMM. In BB estimation we have always used $q = 1$.

First we will discuss the results for DGPs in which the initial conditions are such that BB estimation will be consistent and more efficient than AB, and subsequently

in a separate subsection the situation where BB is inconsistent is examined. Within these two subsections (5.1 and 5.2) we will in separate subsections examine different parameter value combinations for the DGP. We will start by presenting results for a reference parametrization (indicated P0) which has been chosen such that the model has in fact four parameters less, by choosing $\bar{\rho}_{x\varepsilon} = 0$ ($x_{it}$ is strictly exogenous), $EVF_x = 0$ (hence $\pi_\lambda = \pi_\eta = 0$, so $x_{it}$ is neither correlated with $\lambda_i$ nor with $\eta_i$) and $\kappa = 0$ (any cross-sectional heteroskedasticity is just related with $\lambda_i$). These choices (implying that any heteroskedasticity will be unrelated to the regressor $x_{it}$) may (hopefully) lead to results where little difference between unfeasible and feasible estimation will be found and where test sizes are relatively close to the nominal level of 5%. Next we will discuss the effects of settings (to be labelled P1, P2 etc.) which deviate from this reference parametrization P0 in one or more aspects regarding the various correlations and variance fractions and ratios. In P0 the relationship for $y_{it}$ will be characterized by $DEN_y^\eta = 1$ (the impact on $y_{it}$ of the individual effect $\eta_i$ and of the idiosyncratic disturbance $\varepsilon_{it}$ have equal variance). The two remaining parameters are fixed over all cases examined (including P0). The $x_{it}$ series has autoregressive coefficient $\xi = 0.8$ and regarding $y_{it}$ we take $SNR = 3$ (excluding the impacts of the individual effects, the variance of the explanatory part of $y_{it}$ is three times as large as $\sigma_\varepsilon^2$).

In section 3.2 we already indicated that we will examine implementations of GMM where all internal instruments associated with linear moment conditions will be employed (A), but also particular reductions based either on collapsing (C) or omitting long lags (L3, etc.), or a combination (C3, etc.). On top of this we will also distinguish situations that may lead to reductions of the instruments that are being used, because the regressor $x_{it}$ in model (4.1), which will either be strictly exogenous or endogenous with respect to $\varepsilon_{it}$, might be rightly or wrongly treated as either strictly exogenous, or as predetermined (weakly exogenous), or as endogenous. These three distinct situations will be indicated by the letters X, W and E respectively. So, in parametrization P0, where $x_{it}$ is strictly exogenous, the instruments used by either A, C or, say, L2, are not the same under the situations X, W and E. This is hopefully clarified in the next paragraph.

Since we assume that for estimation just the observations $y_{i0}, ..., y_{iT}$ and $x_{i1}, ..., x_{iT}$ are available, the number of internal instruments that are used under XA (all instruments, $x_{it}$ treated as strictly exogenous) for estimation of the equation in first differences is: $T - 1$ (time-dummies) $+ T(T-1)/2$ (lags of $y_{it}$) $+ (T-1)T$ (lags and leads of $x_{it}$). This yields $\{11, 50, 116\}$ instruments for $T = \{3, 6, 9\}$. Under WA this is $\{8, 35, 80\}$ and under EA $\{6, 30, 72\}$. From section 3.3.1 it follows that for BB estimation this number of instruments increases with 1 (intercept) $+ T - 1$ (when $y_{i,t-1}$ is supposed to be effect stationary) $+ T - 1$ (when $x_{i,t}$ is supposed to be effect stationary) $-1$ (when $x_{it}$ is treated as endogenous). This implies for $T = \{3, 6, 9\}$ a total of $\{5, 11, 17\}$ extra instruments under XA and WA, and of $\{4, 10, 16\}$ under EA, whereas these extra instruments will be valid in section 5.1 and invalid in section 5.2.

In the tables to follow we always examine the three values $T \in \{3, 6, 9\}$, while $N = 200$ (as in the classic Arellano and Bond, 1991 study) and the three values $\gamma \in \{0.2, 0.5, 0.8\}$ for the dynamic adjustment coefficient. This is done for both $\theta = 0$ (homoskedasticity) and $\theta = 1$ (substantial cross-sectional heteroskedasticity). Tables have a code which starts by the design parametrization, followed by the character u or f, indicating whether the table contains unfeasible or feasible results. Because of the many feasible variants not all results can be combined in just one table. Therefore, the f is

followed by c, t or J, where c indicates that the table just contains results on coefficient estimates, which are estimated bias, standard deviation (Stdv) and RMSE (root mean squared error; below often loosely addressed as precision); t refers to estimates of the actual rejection probabilities of tests on true coefficient values; and J indicates that the table only contains results on Sargan-Hansen tests. Next, after a bar (-), the earlier discussed code for how regressor $x_{it}$ is actually treated when selecting the instruments is given, followed by the type of instrument reduction.

## 5.1 DGPs under effect stationarity

Here we focus on the case where BB is consistent and more efficient than AB ($s = -50$, $\phi = 1$).

### 5.1.1 Results for the reference parametrization P0

Table 3, with code P0u-XA, gives results for unfeasible GMM coefficient estimators, unfeasible single coefficient tests, and for unfeasible Sargan-Hansen tests for the reference parametrization P0 when $x_{it}$ is (correctly) treated as strictly exogenous and all available instruments are being used. Table 4 (P0fc-XA) presents a selection of feasible counterparts regarding the coefficient estimators. Under homoskedasticity we see that for $\hat{\gamma}_{ABu} = \hat{\gamma}_{AB1}$ its bias (which is negative), stdv and thus its RMSE increase with $\gamma$ and decrease with $T$, whereas the bias of $\hat{\beta}_{ABu} = \hat{\beta}_{AB1}$ is moderate and its RMSE, although decreasing in $T$, is almost invariant with respect to $\beta$. The BBu coefficient estimates are superior indeed, the more so for larger $\gamma$ values (as is already well-known), but less so for $\beta$. As already conjectured in section 3.6 under cross-sectional heteroskedasticity both ABu and BBu are substantially less precise than under homoskedasticity. However, modifying the instruments under cross-sectional heteroskedasticity as is done by MABu and MBBu yields considerable improvements in performance both in terms of bias and RMSE. In fact, the precision of the unfeasible modified estimators under heteroskedasticity comes very close to their counterparts under homoskedasticity.

The simulation results for feasible estimation do not contain the b variant of the weighting matrix[9] because it is so bad, whereas both the a and c variants yield RMSE values very close to their unfeasible counterparts, under homoskedasticity as well as heteroskedasticity. Although the best unfeasible results under heteroskedasticity are obtained by MBBu, this does not fully carry over to MBB, because for $T$ small and also for moderate $T$ and large $\gamma$, BB2c performs much better. The performance of MAB and AB2c is rather similar, whereas we established that their unfeasible variants differ a lot when $\gamma$ is large. Apparently, the modified estimators can be much more vulnerable when the variances of the error components, $\sigma_{\varepsilon,i}^2$ and $\sigma_\eta^2$, are unknown, probably because their estimates have to be inverted in (3.68) and (3.70).

From the type I error estimates for unfeasible single coefficient tests in Table 3 we see that the standard test procedures work pretty well for all techniques regarding $\beta$, but with respect to $\gamma$ ABu fails for larger $\gamma$. This gets even worse under heteroskedasticity, but less so for MABu. For BBu and MBBu the results are reasonable. Here the test seems to benefit from the small bias of BBu. For the feasible variants we find in Table 5

---

[9]The b variant differs from a only for $T > 3$ and may then be not positive definite. For $T = 6, 9$ it proved to be so bad for both AB and BB that we discarded it completely from the presented tables.

## Table 3. P0u-XA*

Design parameters header: $\bar{\rho}_{x\varepsilon}=0.0$

### Unfeasible coefficient estimators

| | L (AB) | L (BB) | $\gamma$ | θ=0 ABu Bias | Stdv | RMSE | θ=0 BBu Bias | Stdv | RMSE | θ=1 ABu Bias | Stdv | RMSE | θ=1 BBu Bias | Stdv | RMSE | θ=1 MABu Bias | Stdv | RMSE | θ=1 MBBu Bias | Stdv | RMSE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| T=3 | 11 | 16 | 0.20 | -0.012 | 0.058 | 0.060 | -0.001 | 0.049 | 0.049 | -0.022 | 0.080 | 0.083 | -0.003 | 0.067 | 0.067 | -0.015 | 0.066 | 0.068 | -0.002 | 0.055 | 0.055 |
| | | | 0.50 | -0.022 | 0.076 | 0.080 | -0.003 | 0.055 | 0.055 | -0.042 | 0.105 | 0.113 | -0.007 | 0.075 | 0.075 | -0.029 | 0.087 | 0.091 | -0.003 | 0.061 | 0.061 |
| | | | 0.80 | -0.077 | 0.134 | 0.155 | -0.009 | 0.067 | 0.068 | -0.144 | 0.182 | 0.232 | -0.018 | 0.096 | 0.097 | -0.096 | 0.150 | 0.178 | -0.006 | 0.071 | 0.071 |
| T=6 | 50 | 61 | 0.20 | -0.009 | 0.029 | 0.030 | 0.000 | 0.026 | 0.026 | -0.017 | 0.040 | 0.044 | 0.001 | 0.036 | 0.036 | -0.010 | 0.030 | 0.032 | -0.002 | 0.029 | 0.029 |
| | | | 0.50 | -0.017 | 0.034 | 0.038 | 0.000 | 0.028 | 0.028 | -0.030 | 0.046 | 0.055 | -0.000 | 0.038 | 0.038 | -0.020 | 0.037 | 0.041 | -0.000 | 0.031 | 0.031 |
| | | | 0.80 | -0.054 | 0.052 | 0.075 | -0.002 | 0.032 | 0.032 | -0.094 | 0.070 | 0.117 | -0.005 | 0.043 | 0.043 | -0.065 | 0.057 | 0.087 | 0.001 | 0.034 | 0.034 |
| T=9 | 116 | 133 | 0.20 | -0.008 | 0.021 | 0.023 | 0.001 | 0.020 | 0.020 | -0.015 | 0.029 | 0.032 | 0.001 | 0.027 | 0.027 | -0.009 | 0.022 | 0.024 | -0.002 | 0.021 | 0.021 |
| | | | 0.50 | -0.014 | 0.024 | 0.027 | 0.001 | 0.020 | 0.020 | -0.024 | 0.031 | 0.040 | 0.002 | 0.027 | 0.027 | -0.016 | 0.025 | 0.029 | -0.001 | 0.022 | 0.022 |
| | | | 0.80 | -0.041 | 0.033 | 0.053 | -0.000 | 0.022 | 0.022 | -0.069 | 0.043 | 0.081 | -0.001 | 0.028 | 0.028 | -0.049 | 0.036 | 0.061 | 0.003 | 0.023 | 0.023 |

| | L (AB) | L (BB) | $\beta$ | θ=0 ABu Bias | Stdv | RMSE | θ=0 BBu Bias | Stdv | RMSE | θ=1 ABu Bias | Stdv | RMSE | θ=1 BBu Bias | Stdv | RMSE | θ=1 MABu Bias | Stdv | RMSE | θ=1 MBBu Bias | Stdv | RMSE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| T=3 | 11 | 16 | 1.43 | 0.003 | 0.100 | 0.100 | 0.001 | 0.096 | 0.096 | 0.004 | 0.148 | 0.148 | 0.005 | 0.136 | 0.136 | 0.004 | 0.100 | 0.100 | 0.001 | 0.097 | 0.097 |
| | | | 0.93 | 0.002 | 0.099 | 0.099 | 0.003 | 0.092 | 0.092 | 0.002 | 0.146 | 0.146 | 0.009 | 0.131 | 0.131 | 0.002 | 0.099 | 0.099 | 0.002 | 0.094 | 0.094 |
| | | | 0.31 | -0.002 | 0.097 | 0.097 | 0.006 | 0.092 | 0.093 | -0.005 | 0.142 | 0.142 | 0.012 | 0.132 | 0.133 | -0.003 | 0.097 | 0.097 | 0.004 | 0.093 | 0.093 |
| T=6 | 50 | 61 | 1.43 | 0.006 | 0.054 | 0.055 | -0.000 | 0.053 | 0.053 | 0.011 | 0.078 | 0.078 | -0.000 | 0.074 | 0.074 | 0.007 | 0.055 | 0.055 | 0.001 | 0.054 | 0.054 |
| | | | 0.93 | 0.007 | 0.053 | 0.053 | -0.000 | 0.051 | 0.051 | 0.012 | 0.075 | 0.076 | 0.001 | 0.070 | 0.070 | 0.008 | 0.053 | 0.054 | 0.000 | 0.052 | 0.052 |
| | | | 0.31 | 0.004 | 0.051 | 0.051 | 0.002 | 0.048 | 0.048 | 0.006 | 0.073 | 0.073 | 0.004 | 0.066 | 0.066 | 0.005 | 0.051 | 0.051 | 0.000 | 0.048 | 0.048 |
| T=9 | 116 | 133 | 1.43 | 0.007 | 0.040 | 0.041 | -0.001 | 0.039 | 0.039 | 0.012 | 0.056 | 0.057 | -0.001 | 0.054 | 0.054 | 0.008 | 0.040 | 0.041 | 0.001 | 0.040 | 0.040 |
| | | | 0.93 | 0.009 | 0.039 | 0.040 | -0.001 | 0.037 | 0.037 | 0.014 | 0.054 | 0.056 | -0.001 | 0.051 | 0.051 | 0.010 | 0.039 | 0.040 | 0.001 | 0.038 | 0.038 |
| | | | 0.31 | 0.006 | 0.037 | 0.037 | 0.001 | 0.034 | 0.034 | 0.010 | 0.051 | 0.052 | 0.002 | 0.047 | 0.047 | 0.008 | 0.037 | 0.037 | -0.000 | 0.035 | 0.035 |

### Unfeasible t-test: actual significance level

| | L (AB) | L (BB) | $\gamma$ | θ=0 ABu | θ=0 BBu | $\beta$ | θ=1 ABu | θ=1 BBu | θ=1 MABu | θ=1 MBBu |
|---|---|---|---|---|---|---|---|---|---|---|
| T=3 | 11 | 16 | 0.20 | 0.058 | 0.051 | 1.43 | 0.060 | 0.051 | 0.062 | 0.049 |
| | | | 0.50 | 0.061 | 0.053 | 0.93 | 0.066 | 0.053 | 0.067 | 0.048 |
| | | | 0.80 | 0.089 | 0.056 | 0.31 | 0.123 | 0.058 | 0.099 | 0.049 |
| T=6 | 50 | 61 | 0.20 | 0.059 | 0.041 | 1.43 | 0.071 | 0.048 | 0.061 | 0.049 |
| | | | 0.50 | 0.074 | 0.044 | 0.93 | 0.099 | 0.053 | 0.079 | 0.049 |
| | | | 0.80 | 0.172 | 0.052 | 0.31 | 0.267 | 0.058 | 0.197 | 0.049 |
| T=9 | 116 | 133 | 0.20 | 0.071 | 0.043 | 1.43 | 0.082 | 0.048 | 0.072 | 0.048 |
| | | | 0.50 | 0.095 | 0.047 | 0.93 | 0.127 | 0.049 | 0.101 | 0.049 |
| | | | 0.80 | 0.246 | 0.053 | 0.31 | 0.377 | 0.053 | 0.281 | 0.050 |

### Unfeasible Sargan-Hansen test: rejection probability

| | df (AB) | df (BB) | df (Inc) | $\gamma$ | θ=0 JABu | θ=0 JBBu | θ=0 JESu | $\beta$ | θ=1 JMABu | θ=1 JMMBu | θ=1 JESMu |
|---|---|---|---|---|---|---|---|---|---|---|---|
| T=3 | 9 | 13 | 4 | 0.20 | 0.049 | 0.047 | 0.048 | 1.43 | 0.050 | 0.050 | 0.049 |
| | | | | 0.50 | 0.047 | 0.049 | 0.048 | 0.93 | 0.049 | 0.051 | 0.051 |
| | | | | 0.80 | 0.033 | 0.051 | 0.063 | 0.31 | 0.038 | 0.050 | 0.067 |
| T=6 | 48 | 58 | 10 | 0.20 | 0.048 | 0.048 | 0.048 | 1.43 | 0.045 | 0.048 | 0.050 |
| | | | | 0.50 | 0.043 | 0.045 | 0.049 | 0.93 | 0.042 | 0.047 | 0.052 |
| | | | | 0.80 | 0.030 | 0.043 | 0.075 | 0.31 | 0.035 | 0.043 | 0.077 |
| T=9 | 114 | 130 | 16 | 0.20 | 0.049 | 0.053 | 0.048 | 1.43 | 0.048 | 0.052 | 0.051 |
| | | | | 0.50 | 0.048 | 0.051 | 0.053 | 0.93 | 0.045 | 0.051 | 0.055 |
| | | | | 0.80 | 0.036 | 0.049 | 0.086 | 0.31 | 0.037 | 0.049 | 0.095 |

*R = 10000 simulation replications. Design parameter values: $N=200$, $SNR=3$, $DEN_y=1.0$, $EVF_x=0.0$, $\bar{\rho}_{x\varepsilon}=0.0$, $\xi=0.8$, $\kappa=0.00$, $\sigma_\varepsilon=1$, $q=1$, $\phi=1.0$.
These yield the DGP parameter values: $\pi_\lambda=0.00$, $\pi_\eta=0.00$, $\sigma_v=0.60$, $\sigma_\eta=1.0*(1-\gamma)$, $\rho_{v\varepsilon}=0.0$ (and $\bar{\rho}_{\varepsilon\eta}=0.00$, $\bar{\rho}_{\varepsilon\lambda}=0.00$).

30

**Table 4.** P0fc-XA*

Feasible coefficient estimators for Arellano-Bond

ρ̄xε = 0.0. Estimator groups AB1, AB2a, AB2c (θ = 0); AB2a, AB2c, ABI, MAB (θ = 1).

| T | L | γ | AB1 Bias | AB1 Stdv | AB1 RMSE | AB2a(θ0) Bias | AB2a(θ0) Stdv | AB2a(θ0) RMSE | AB2c(θ0) Bias | AB2c(θ0) Stdv | AB2c(θ0) RMSE | AB2a(θ1) Bias | AB2a(θ1) Stdv | AB2a(θ1) RMSE | AB2c(θ1) Bias | AB2c(θ1) Stdv | AB2c(θ1) RMSE | ABI Bias | ABI Stdv | ABI RMSE | MAB Bias | MAB Stdv | MAB RMSE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 11 | 0.20 | -0.012 | 0.058 | 0.060 | -0.011 | 0.061 | 0.062 | -0.012 | 0.059 | 0.061 | -0.019 | 0.081 | 0.083 | -0.022 | 0.081 | 0.084 | -0.023 | 0.086 | 0.089 | -0.022 | 0.083 | 0.086 |
| | | 0.50 | -0.022 | 0.076 | 0.080 | -0.022 | 0.079 | 0.082 | -0.022 | 0.078 | 0.081 | -0.036 | 0.106 | 0.112 | -0.040 | 0.106 | 0.114 | -0.044 | 0.112 | 0.121 | -0.042 | 0.109 | 0.117 |
| | | 0.80 | -0.077 | 0.134 | 0.155 | -0.077 | 0.141 | 0.161 | -0.075 | 0.137 | 0.156 | -0.132 | 0.188 | 0.230 | -0.139 | 0.184 | 0.231 | -0.146 | 0.194 | 0.243 | -0.143 | 0.190 | 0.238 |
| 6 | 50 | 0.20 | -0.009 | 0.029 | 0.030 | -0.009 | 0.032 | 0.033 | -0.009 | 0.029 | 0.031 | -0.016 | 0.041 | 0.043 | -0.017 | 0.040 | 0.044 | -0.019 | 0.045 | 0.049 | -0.014 | 0.036 | 0.039 |
| | | 0.50 | -0.017 | 0.034 | 0.038 | -0.017 | 0.038 | 0.041 | -0.017 | 0.034 | 0.038 | -0.028 | 0.047 | 0.055 | -0.030 | 0.047 | 0.055 | -0.035 | 0.052 | 0.063 | -0.026 | 0.043 | 0.050 |
| | | 0.80 | -0.054 | 0.052 | 0.075 | -0.055 | 0.059 | 0.081 | -0.053 | 0.053 | 0.075 | -0.091 | 0.074 | 0.118 | -0.094 | 0.071 | 0.117 | -0.105 | 0.078 | 0.131 | -0.087 | 0.068 | 0.110 |
| 9 | 116 | 0.20 | -0.008 | 0.021 | 0.023 | -0.008 | 0.024 | 0.025 | -0.008 | 0.021 | 0.023 | -0.015 | 0.031 | 0.035 | -0.014 | 0.029 | 0.032 | -0.017 | 0.033 | 0.037 | -0.011 | 0.024 | 0.027 |
| | | 0.50 | -0.014 | 0.024 | 0.027 | -0.014 | 0.026 | 0.030 | -0.014 | 0.024 | 0.028 | -0.026 | 0.034 | 0.043 | -0.024 | 0.032 | 0.040 | -0.028 | 0.036 | 0.046 | -0.019 | 0.028 | 0.034 |
| | | 0.80 | -0.041 | 0.033 | 0.053 | -0.042 | 0.037 | 0.056 | -0.041 | 0.033 | 0.053 | -0.074 | 0.048 | 0.088 | -0.070 | 0.043 | 0.082 | -0.078 | 0.050 | 0.093 | -0.061 | 0.040 | 0.073 |

| T | L | β | AB1 Bias | AB1 Stdv | AB1 RMSE | AB2a(θ0) Bias | AB2a(θ0) Stdv | AB2a(θ0) RMSE | AB2c(θ0) Bias | AB2c(θ0) Stdv | AB2c(θ0) RMSE | AB2a(θ1) Bias | AB2a(θ1) Stdv | AB2a(θ1) RMSE | AB2c(θ1) Bias | AB2c(θ1) Stdv | AB2c(θ1) RMSE | ABI Bias | ABI Stdv | ABI RMSE | MAB Bias | MAB Stdv | MAB RMSE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 11 | 1.43 | 0.003 | 0.100 | 0.100 | 0.004 | 0.103 | 0.103 | 0.003 | 0.101 | 0.101 | 0.005 | 0.145 | 0.145 | 0.005 | 0.147 | 0.147 | 0.006 | 0.158 | 0.158 | 0.005 | 0.148 | 0.148 |
| | | 0.93 | 0.002 | 0.099 | 0.099 | 0.003 | 0.103 | 0.103 | 0.002 | 0.100 | 0.100 | 0.003 | 0.144 | 0.144 | 0.002 | 0.146 | 0.146 | 0.004 | 0.157 | 0.157 | 0.003 | 0.147 | 0.147 |
| | | 0.31 | -0.002 | 0.097 | 0.097 | -0.002 | 0.101 | 0.101 | -0.002 | 0.099 | 0.099 | -0.005 | 0.141 | 0.141 | -0.005 | 0.143 | 0.143 | -0.004 | 0.152 | 0.152 | -0.004 | 0.142 | 0.142 |
| 6 | 50 | 1.43 | 0.006 | 0.054 | 0.055 | 0.006 | 0.060 | 0.061 | 0.006 | 0.055 | 0.055 | 0.010 | 0.077 | 0.077 | 0.011 | 0.078 | 0.079 | 0.013 | 0.087 | 0.088 | 0.009 | 0.067 | 0.067 |
| | | 0.93 | 0.007 | 0.053 | 0.053 | 0.007 | 0.059 | 0.059 | 0.007 | 0.053 | 0.054 | 0.011 | 0.074 | 0.075 | 0.012 | 0.076 | 0.077 | 0.014 | 0.085 | 0.086 | 0.010 | 0.065 | 0.066 |
| | | 0.31 | 0.004 | 0.051 | 0.051 | 0.004 | 0.057 | 0.057 | 0.004 | 0.052 | 0.052 | 0.005 | 0.072 | 0.072 | 0.006 | 0.073 | 0.074 | 0.007 | 0.082 | 0.082 | 0.005 | 0.063 | 0.063 |
| 9 | 116 | 1.43 | 0.007 | 0.040 | 0.041 | 0.008 | 0.045 | 0.045 | 0.007 | 0.041 | 0.041 | 0.013 | 0.060 | 0.061 | 0.012 | 0.056 | 0.058 | 0.014 | 0.065 | 0.066 | 0.009 | 0.046 | 0.047 |
| | | 0.93 | 0.009 | 0.039 | 0.040 | 0.009 | 0.043 | 0.044 | 0.009 | 0.039 | 0.040 | 0.016 | 0.058 | 0.060 | 0.014 | 0.054 | 0.056 | 0.017 | 0.062 | 0.064 | 0.012 | 0.044 | 0.046 |
| | | 0.31 | 0.006 | 0.037 | 0.037 | 0.007 | 0.041 | 0.041 | 0.006 | 0.037 | 0.037 | 0.011 | 0.055 | 0.056 | 0.010 | 0.051 | 0.052 | 0.012 | 0.059 | 0.060 | 0.009 | 0.042 | 0.043 |

Feasible coefficient estimators for Blundell-Bond

ρ̄xε = 0.0. Estimator groups BB1, BB2a, BB2c (θ = 0); BB2a, BB2c, BBI, MBB (θ = 1).

| T | L | γ | BB1 Bias | BB1 Stdv | BB1 RMSE | BB2a(θ0) Bias | BB2a(θ0) Stdv | BB2a(θ0) RMSE | BB2c(θ0) Bias | BB2c(θ0) Stdv | BB2c(θ0) RMSE | BB2a(θ1) Bias | BB2a(θ1) Stdv | BB2a(θ1) RMSE | BB2c(θ1) Bias | BB2c(θ1) Stdv | BB2c(θ1) RMSE | BBI Bias | BBI Stdv | BBI RMSE | MBB Bias | MBB Stdv | MBB RMSE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 16 | 0.20 | -0.003 | 0.049 | 0.050 | -0.000 | 0.051 | 0.051 | -0.000 | 0.051 | 0.051 | -0.004 | 0.067 | 0.068 | -0.002 | 0.070 | 0.070 | -0.010 | 0.073 | 0.074 | -0.003 | 0.073 | 0.073 |
| | | 0.50 | -0.009 | 0.057 | 0.058 | -0.004 | 0.058 | 0.058 | -0.003 | 0.057 | 0.057 | -0.010 | 0.078 | 0.078 | -0.005 | 0.079 | 0.079 | -0.021 | 0.085 | 0.087 | -0.006 | 0.443 | 0.443 |
| | | 0.80 | -0.029 | 0.075 | 0.081 | -0.013 | 0.076 | 0.077 | -0.010 | 0.073 | 0.074 | -0.031 | 0.108 | 0.112 | -0.014 | 0.107 | 0.108 | -0.055 | 0.111 | 0.124 | 0.023 | 2.456 | 2.456 |
| 6 | 61 | 0.20 | -0.002 | 0.026 | 0.026 | -0.001 | 0.028 | 0.028 | 0.001 | 0.027 | 0.027 | -0.006 | 0.037 | 0.037 | 0.000 | 0.038 | 0.038 | -0.010 | 0.041 | 0.042 | -0.003 | 0.034 | 0.034 |
| | | 0.50 | -0.008 | 0.029 | 0.030 | -0.003 | 0.032 | 0.032 | 0.000 | 0.030 | 0.030 | -0.013 | 0.040 | 0.043 | -0.001 | 0.040 | 0.040 | -0.021 | 0.045 | 0.050 | -0.003 | 0.039 | 0.039 |
| | | 0.80 | -0.029 | 0.038 | 0.048 | -0.014 | 0.039 | 0.041 | -0.005 | 0.035 | 0.035 | -0.042 | 0.052 | 0.067 | -0.009 | 0.047 | 0.048 | -0.058 | 0.056 | 0.081 | 0.009 | 0.254 | 0.254 |
| 9 | 133 | 0.20 | -0.002 | 0.020 | 0.020 | -0.001 | 0.021 | 0.021 | 0.001 | 0.020 | 0.020 | -0.008 | 0.030 | 0.031 | 0.001 | 0.028 | 0.028 | -0.009 | 0.031 | 0.033 | -0.003 | 0.024 | 0.024 |
| | | 0.50 | -0.008 | 0.021 | 0.022 | -0.006 | 0.022 | 0.023 | 0.001 | 0.021 | 0.021 | -0.017 | 0.031 | 0.036 | 0.001 | 0.029 | 0.029 | -0.019 | 0.033 | 0.038 | -0.004 | 0.026 | 0.027 |
| | | 0.80 | -0.027 | 0.026 | 0.038 | -0.021 | 0.027 | 0.034 | -0.003 | 0.024 | 0.024 | -0.049 | 0.038 | 0.062 | -0.006 | 0.031 | 0.032 | -0.053 | 0.040 | 0.066 | -0.004 | 0.037 | 0.037 |

| T | L | β | BB1 Bias | BB1 Stdv | BB1 RMSE | BB2a(θ0) Bias | BB2a(θ0) Stdv | BB2a(θ0) RMSE | BB2c(θ0) Bias | BB2c(θ0) Stdv | BB2c(θ0) RMSE | BB2a(θ1) Bias | BB2a(θ1) Stdv | BB2a(θ1) RMSE | BB2c(θ1) Bias | BB2c(θ1) Stdv | BB2c(θ1) RMSE | BBI Bias | BBI Stdv | BBI RMSE | MBB Bias | MBB Stdv | MBB RMSE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 16 | 1.43 | 0.002 | 0.096 | 0.096 | 0.002 | 0.101 | 0.101 | 0.002 | 0.097 | 0.097 | 0.004 | 0.138 | 0.138 | 0.008 | 0.137 | 0.137 | 0.007 | 0.149 | 0.149 | 0.009 | 0.139 | 0.139 |
| | | 0.93 | 0.004 | 0.094 | 0.094 | 0.004 | 0.097 | 0.097 | 0.005 | 0.093 | 0.093 | 0.008 | 0.133 | 0.133 | 0.014 | 0.132 | 0.133 | 0.009 | 0.146 | 0.146 | 0.019 | 0.269 | 0.270 |
| | | 0.31 | 0.005 | 0.094 | 0.094 | 0.007 | 0.098 | 0.098 | 0.007 | 0.094 | 0.094 | 0.010 | 0.134 | 0.134 | 0.015 | 0.135 | 0.135 | 0.010 | 0.145 | 0.146 | 0.035 | 0.965 | 0.966 |
| 6 | 61 | 1.43 | 0.002 | 0.053 | 0.053 | 0.001 | 0.059 | 0.059 | -0.000 | 0.054 | 0.054 | 0.004 | 0.075 | 0.075 | 0.003 | 0.075 | 0.075 | 0.007 | 0.085 | 0.085 | 0.003 | 0.065 | 0.065 |
| | | 0.93 | 0.004 | 0.052 | 0.052 | 0.002 | 0.057 | 0.057 | 0.000 | 0.051 | 0.051 | 0.007 | 0.072 | 0.073 | 0.003 | 0.072 | 0.072 | 0.011 | 0.082 | 0.083 | 0.005 | 0.063 | 0.063 |
| | | 0.31 | 0.004 | 0.050 | 0.050 | 0.004 | 0.054 | 0.054 | 0.003 | 0.048 | 0.048 | 0.007 | 0.070 | 0.070 | 0.006 | 0.067 | 0.068 | 0.009 | 0.079 | 0.079 | 0.007 | 0.271 | 0.272 |
| 9 | 133 | 1.43 | 0.002 | 0.039 | 0.039 | 0.002 | 0.043 | 0.043 | -0.001 | 0.040 | 0.040 | 0.007 | 0.060 | 0.060 | -0.001 | 0.055 | 0.055 | 0.008 | 0.064 | 0.064 | 0.003 | 0.046 | 0.046 |
| | | 0.93 | 0.005 | 0.038 | 0.038 | 0.004 | 0.041 | 0.042 | -0.000 | 0.038 | 0.038 | 0.011 | 0.057 | 0.058 | 0.000 | 0.052 | 0.052 | 0.012 | 0.061 | 0.062 | 0.003 | 0.044 | 0.044 |
| | | 0.31 | 0.006 | 0.036 | 0.036 | 0.005 | 0.039 | 0.039 | 0.002 | 0.035 | 0.035 | 0.010 | 0.054 | 0.055 | 0.004 | 0.048 | 0.048 | 0.011 | 0.057 | 0.058 | 0.004 | 0.042 | 0.042 |

* R = 10000 simulation replications. Design parameter values: N = 200, SNR = 3, DEN_y = 1.0, EVF_x = 0.0, ρ̄_xε = 0.0, ξ = 0.8, κ = 0.00, σ_ε = 1, q = 1, φ = 1.0. These yield the DGP parameter values: π_λ = 0.00, π_η = 0.00, σ_υ = 0.60, σ_η = 1.0 * (1 − γ), ρ_υε = 0.0 (and ρ̄_xη = 0.00, ρ̄_xλ = 0.00).

31

**Table 5.** P0ft-XA*

Feasible t-test Arellano-Bond: actual significance level

| $\bar{\rho}_{x\varepsilon}=0.0$ | | | $\theta=0$ | | | | | | | $\theta=1$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | L | $\gamma$ | AB1 | AB1aR | AB1cR | AB2a | AB2aW | AB2c | AB2cW | AB1 | AB1aR | AB1cR | AB2a | AB2aW | AB2c | AB2cW | MAB |
| $T=3$ | 11 | 0.20 | 0.062 | 0.064 | 0.057 | 0.084 | 0.062 | 0.065 | 0.059 | 0.213 | 0.088 | 0.070 | 0.134 | 0.074 | 0.082 | 0.067 | 0.556 |
| | | 0.50 | 0.072 | 0.072 | 0.065 | 0.093 | 0.071 | 0.073 | 0.069 | 0.237 | 0.104 | 0.086 | 0.150 | 0.083 | 0.101 | 0.083 | 0.584 |
| | | 0.80 | 0.120 | 0.119 | 0.114 | 0.144 | 0.104 | 0.121 | 0.114 | 0.348 | 0.190 | 0.172 | 0.248 | 0.146 | 0.192 | 0.166 | 0.697 |
| $T=6$ | 50 | 0.20 | 0.061 | 0.063 | 0.052 | 0.159 | 0.061 | 0.058 | 0.053 | 0.246 | 0.091 | 0.067 | 0.354 | 0.077 | 0.078 | 0.066 | 0.201 |
| | | 0.50 | 0.078 | 0.078 | 0.071 | 0.182 | 0.071 | 0.076 | 0.071 | 0.299 | 0.123 | 0.098 | 0.395 | 0.096 | 0.108 | 0.091 | 0.241 |
| | | 0.80 | 0.191 | 0.183 | 0.176 | 0.317 | 0.142 | 0.182 | 0.174 | 0.547 | 0.329 | 0.285 | 0.617 | 0.234 | 0.299 | 0.267 | 0.465 |
| $T=9$ | 116 | 0.20 | 0.073 | 0.074 | 0.062 | 0.324 | 0.070 | 0.069 | 0.064 | 0.272 | 0.101 | 0.075 | 0.689 | 0.095 | 0.087 | 0.075 | 0.154 |
| | | 0.50 | 0.098 | 0.096 | 0.090 | 0.358 | 0.084 | 0.096 | 0.090 | 0.344 | 0.149 | 0.117 | 0.728 | 0.139 | 0.130 | 0.115 | 0.198 |
| | | 0.80 | 0.255 | 0.242 | 0.239 | 0.552 | 0.192 | 0.246 | 0.235 | 0.653 | 0.411 | 0.367 | 0.893 | 0.376 | 0.396 | 0.366 | 0.460 |
| | L | $\beta$ | AB1 | AB1aR | AB1cR | AB2a | AB2aW | AB2c | AB2cW | AB1 | AB1aR | AB1cR | AB2a | AB2aW | AB2c | AB2cW | MAB |
| $T=3$ | 11 | 1.43 | 0.052 | 0.052 | 0.051 | 0.073 | 0.053 | 0.057 | 0.052 | 0.216 | 0.070 | 0.065 | 0.117 | 0.062 | 0.071 | 0.056 | 0.586 |
| | | 0.93 | 0.051 | 0.052 | 0.051 | 0.073 | 0.053 | 0.057 | 0.051 | 0.215 | 0.073 | 0.065 | 0.118 | 0.062 | 0.072 | 0.055 | 0.582 |
| | | 0.31 | 0.050 | 0.051 | 0.050 | 0.072 | 0.052 | 0.055 | 0.050 | 0.214 | 0.068 | 0.062 | 0.116 | 0.061 | 0.072 | 0.057 | 0.581 |
| $T=6$ | 50 | 1.43 | 0.051 | 0.055 | 0.050 | 0.145 | 0.055 | 0.056 | 0.051 | 0.222 | 0.068 | 0.059 | 0.315 | 0.060 | 0.068 | 0.057 | 0.210 |
| | | 0.93 | 0.054 | 0.054 | 0.052 | 0.145 | 0.055 | 0.058 | 0.053 | 0.225 | 0.069 | 0.061 | 0.315 | 0.062 | 0.069 | 0.058 | 0.216 |
| | | 0.31 | 0.054 | 0.054 | 0.054 | 0.144 | 0.052 | 0.059 | 0.054 | 0.232 | 0.066 | 0.064 | 0.317 | 0.058 | 0.070 | 0.058 | 0.229 |
| $T=9$ | 116 | 1.43 | 0.050 | 0.052 | 0.048 | 0.296 | 0.057 | 0.054 | 0.051 | 0.232 | 0.068 | 0.057 | 0.652 | 0.067 | 0.067 | 0.055 | 0.138 |
| | | 0.93 | 0.053 | 0.055 | 0.052 | 0.298 | 0.057 | 0.058 | 0.054 | 0.241 | 0.072 | 0.064 | 0.657 | 0.071 | 0.070 | 0.059 | 0.145 |
| | | 0.31 | 0.057 | 0.056 | 0.058 | 0.297 | 0.057 | 0.061 | 0.057 | 0.241 | 0.069 | 0.069 | 0.654 | 0.066 | 0.071 | 0.061 | 0.153 |

Feasible t-test Blundell-Bond: actual significance level

| $\bar{\rho}_{x\varepsilon}=0.0$ | | | $\theta=0$ | | | | | | | $\theta=1$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | L | $\gamma$ | BB1 | BB1aR | BB1cR | BB2a | BB2aW | BB2c | BB2cW | BB1 | BB1aR | BB1cR | BB2a | BB2aW | BB2c | BB2cW | MBB |
| $T=3$ | 16 | 0.20 | 0.051 | 0.056 | 0.041 | 0.093 | 0.059 | 0.058 | 0.050 | 0.196 | 0.076 | 0.051 | 0.158 | 0.064 | 0.073 | 0.058 | 0.494 |
| | | 0.50 | 0.056 | 0.057 | 0.046 | 0.098 | 0.057 | 0.062 | 0.053 | 0.209 | 0.079 | 0.056 | 0.165 | 0.067 | 0.080 | 0.063 | 0.502 |
| | | 0.80 | 0.066 | 0.065 | 0.048 | 0.103 | 0.057 | 0.057 | 0.040 | 0.252 | 0.100 | 0.071 | 0.208 | 0.073 | 0.087 | 0.060 | 0.559 |
| $T=6$ | 61 | 0.20 | 0.042 | 0.052 | 0.034 | 0.178 | 0.056 | 0.045 | 0.041 | 0.206 | 0.074 | 0.048 | 0.389 | 0.067 | 0.059 | 0.051 | 0.160 |
| | | 0.50 | 0.053 | 0.060 | 0.044 | 0.184 | 0.054 | 0.052 | 0.044 | 0.248 | 0.092 | 0.064 | 0.413 | 0.073 | 0.069 | 0.055 | 0.167 |
| | | 0.80 | 0.116 | 0.122 | 0.100 | 0.217 | 0.061 | 0.064 | 0.042 | 0.424 | 0.218 | 0.161 | 0.562 | 0.129 | 0.080 | 0.050 | 0.255 |
| $T=9$ | 133 | 0.20 | 0.049 | 0.056 | 0.041 | 0.364 | 0.054 | 0.050 | 0.046 | 0.221 | 0.076 | 0.051 | 0.735 | 0.072 | 0.057 | 0.049 | 0.119 |
| | | 0.50 | 0.063 | 0.070 | 0.056 | 0.381 | 0.058 | 0.056 | 0.050 | 0.279 | 0.108 | 0.078 | 0.767 | 0.101 | 0.063 | 0.054 | 0.121 |
| | | 0.80 | 0.173 | 0.176 | 0.156 | 0.506 | 0.109 | 0.066 | 0.043 | 0.552 | 0.313 | 0.249 | 0.894 | 0.282 | 0.074 | 0.046 | 0.166 |
| | L | $\beta$ | BB1 | BB1aR | BB1cR | BB2a | BB2aW | BB2c | BB2cW | BB1 | BB1aR | BB1cR | BB2a | BB2aW | BB2c | BB2cW | MBB |
| $T=3$ | 16 | 1.43 | 0.051 | 0.054 | 0.052 | 0.083 | 0.056 | 0.058 | 0.058 | 0.210 | 0.070 | 0.061 | 0.147 | 0.064 | 0.073 | 0.068 | 0.571 |
| | | 0.93 | 0.050 | 0.053 | 0.051 | 0.084 | 0.055 | 0.057 | 0.057 | 0.214 | 0.070 | 0.061 | 0.148 | 0.064 | 0.073 | 0.068 | 0.565 |
| | | 0.31 | 0.048 | 0.051 | 0.049 | 0.087 | 0.056 | 0.059 | 0.059 | 0.217 | 0.072 | 0.062 | 0.155 | 0.065 | 0.078 | 0.072 | 0.616 |
| $T=6$ | 61 | 1.43 | 0.048 | 0.052 | 0.047 | 0.166 | 0.055 | 0.053 | 0.054 | 0.216 | 0.069 | 0.053 | 0.371 | 0.059 | 0.064 | 0.067 | 0.200 |
| | | 0.93 | 0.051 | 0.055 | 0.050 | 0.166 | 0.053 | 0.054 | 0.057 | 0.223 | 0.069 | 0.058 | 0.376 | 0.063 | 0.064 | 0.068 | 0.205 |
| | | 0.31 | 0.053 | 0.055 | 0.053 | 0.169 | 0.055 | 0.053 | 0.057 | 0.229 | 0.068 | 0.060 | 0.388 | 0.064 | 0.065 | 0.068 | 0.228 |
| $T=9$ | 133 | 1.43 | 0.047 | 0.051 | 0.046 | 0.342 | 0.052 | 0.051 | 0.053 | 0.220 | 0.067 | 0.054 | 0.717 | 0.064 | 0.060 | 0.066 | 0.128 |
| | | 0.93 | 0.050 | 0.052 | 0.049 | 0.349 | 0.053 | 0.053 | 0.058 | 0.232 | 0.070 | 0.060 | 0.718 | 0.067 | 0.061 | 0.069 | 0.132 |
| | | 0.31 | 0.055 | 0.055 | 0.055 | 0.356 | 0.056 | 0.056 | 0.060 | 0.234 | 0.070 | 0.066 | 0.730 | 0.066 | 0.064 | 0.068 | 0.146 |

*$R = 10000$ simulation replications. Design parameter values: $N = 200$, $SNR = 3$, $DEN_y = 1.0$, $EVF_x = 0.0$, $\bar{\rho}_{x\varepsilon} = 0.0$, $\xi = 0.8$, $\kappa = 0.00$, $\sigma_\varepsilon = 1$, $q = 1$, $\phi = 1.0$. These yield the DGP parameter values: $\pi_\lambda = 0.00$, $\pi_\eta = 0.00$, $\sigma_v = 0.60$, $\sigma_\eta = 1.0*(1-\gamma)$, $\rho_{v\varepsilon} = 0.0$ (and $\bar{\rho}_{x\eta} = 0.00$, $\bar{\rho}_{x\lambda} = 0.00$).

(P0ft-XA) that under homoskedasticity AB1 has reasonable actual significance level for $\beta$, but for $\gamma$ only when it is small. The same holds for AB2c. Under heteroskedasticity $\sigma^2_{\varepsilon,i}$ has to be estimated and then AB2c overrejects, especially for $\gamma$ or $T$ large, but only mildly so for tests on $\beta$. Both AB2a and MAB overreject enormously. Employing the Windmeijer correction mitigates the overrejection probabilities in many cases, but not in all. AB2cW has appropriate size for tests on $\beta$, but for tests on $\gamma$ the size increases both with $\gamma$ and with $T$ from 7% to 37% over the grid examined. Since the test based on ABu shows a similar pattern, it is self-evident that a correction which takes into account the randomness of AB1 cannot be fully effective. Oddly enough the Windmeijer correction is occasionally more effective for the heavily oversized AB2a than for the less oversized AB2c. Under homoskedasticity both BB2c and BB2cW behave very reasonable, both for tests on $\beta$ and on $\gamma$. Under heteroskedasticity BB2cW is still very reasonable, but all other implementations fail in some instances, especially for tests on $\gamma$ when $\gamma$ or $T$ are large.

Regarding the unfeasible $J$ tests Table 3 shows reasonable size properties under homoskedasticity, especially for $JBBu$, but less so for the incremental test on effect

stationarity when $\gamma$ is large. Under heteroskedasticity this problem is more serious, though less so for the unfeasible modified procedure. Heteroskedasticity and $\gamma$ large lead to underrejection of the $JABu$ test, especially when $T$ is large too. Turning now to some of the many variants of feasible $J$ tests, of which some[10] are presented in Table 6 (P0fJ-XA), we first focus on $JAB$. Under homoskedasticity $JAB^{(1,0)}$ behaves reasonable, though when applied when $\theta = 1$ it rejects with high probability (thus detecting heteroskedasticity instead of instrument invalidity, probably due to underestimation of the variance of the still valid moment conditions). Of the $JAB^{(1,1)}$ tests, both for $\theta = 0$ and $\theta = 1$, the c variant severely underrejects when $T = 9$ (when there is an abundance of instruments), but less so than the a version. Such severe underrejection had already been noted by Bowsher (2002). An almost similar pattern we note for $JAB^{(2,1)}$ and $JAB^{(2,2)}$. Test $JMAB^{(2,1)}$ overrejects severely for $T = 3$ and underrejects otherwise, whereas $JMAB^{(2,2)}$ is fine for $T = 3$ but underrejects for larger values of $T$. Turning now to feasible $JBB$ tests we find that $JBB^{(1,0)}$ underrejects when $\theta = 0$ and, like $JAB^{(1,0)}$, rejects with high probability when $\theta = 1$. Both the a and c variants of test $JBB^{(1,1)}$, like $JAB^{(1,1)}$, have rejection probabilities that are not invariant with respect to $T$, $\gamma$ and $\theta$. The c variants seem the least vulnerable, and therefore also yield an almost reasonable incremental test $JES^{(1,1)}$, although it underrejects when $\theta = 0$ and overrejects when $\theta = 1$ for $\gamma = 0.8$. For $JBB^{(2,1)}$ and $JBB^{(2,2)}$ too the c variant has rejection probabilities which vary the least with $T$, $\gamma$ and $\theta$, but they are systematically below the nominal significance level, which is also the case for the resulting incremental tests. Oddly enough, the incremental tests resulting from the a variants have type I error probabilities reasonably close to 5%, despite the serious underrejection of both the $JAB$ and $JBB$ tests from which they result.

When treating regressor $x_{it}$ as predetermined (P0-WA), whereas it is strictly exoge-

**Table 6.** P0fJ-XA*

Feasible Sargan-Hansen test: rejection probability

| $\bar{\rho}_{x\varepsilon}=0.0$ | df | | | | $\theta = 0$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | AB | BB | Inc | $\gamma$ | $JAB_a^{(2,1)}$ | $JBB_a^{(2,1)}$ | $JES_a^{(2,1)}$ | $JAB_c^{(2,1)}$ | $JBB_c^{(2,1)}$ | $JES_c^{(2,1)}$ | $JMAB$ | $JMBB$ | $JESM$ |
| $T = 3$ | 9 | 13 | 4 | 0.20 | 0.047 | 0.050 | 0.053 | 0.043 | 0.033 | 0.030 | 0.244 | 0.304 | 0.306 |
| | | | | 0.50 | 0.050 | 0.051 | 0.052 | 0.044 | 0.034 | 0.025 | 0.246 | 0.328 | 0.327 |
| | | | | 0.80 | 0.061 | 0.055 | 0.047 | 0.055 | 0.035 | 0.022 | 0.251 | 0.369 | 0.365 |
| $T = 6$ | 48 | 58 | 10 | 0.20 | 0.034 | 0.038 | 0.068 | 0.026 | 0.025 | 0.030 | 0.032 | 0.386 | 0.439 |
| | | | | 0.50 | 0.037 | 0.039 | 0.062 | 0.027 | 0.023 | 0.023 | 0.033 | 0.391 | 0.442 |
| | | | | 0.80 | 0.046 | 0.042 | 0.056 | 0.031 | 0.022 | 0.013 | 0.039 | 0.405 | 0.453 |
| $T = 9$ | 114 | 130 | 16 | 0.20 | 0.007 | 0.002 | 0.056 | 0.021 | 0.023 | 0.033 | 0.022 | 0.409 | 0.466 |
| | | | | 0.50 | 0.007 | 0.002 | 0.053 | 0.021 | 0.021 | 0.026 | 0.022 | 0.411 | 0.467 |
| | | | | 0.80 | 0.009 | 0.002 | 0.048 | 0.025 | 0.019 | 0.014 | 0.026 | 0.416 | 0.471 |

| $\bar{\rho}_{x\varepsilon}=0.0$ | df | | | | $\theta = 1$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | AB | BB | Inc | $\gamma$ | $JAB_a^{(2,1)}$ | $JBB_a^{(2,1)}$ | $JES_a^{(2,1)}$ | $JAB_c^{(2,1)}$ | $JBB_c^{(2,1)}$ | $JES_c^{(2,1)}$ | $JMAB$ | $JMBB$ | $JESM$ |
| $T = 3$ | 9 | 13 | 4 | 0.20 | 0.036 | 0.035 | 0.047 | 0.037 | 0.034 | 0.041 | 0.278 | 0.569 | 0.557 |
| | | | | 0.50 | 0.041 | 0.035 | 0.045 | 0.042 | 0.038 | 0.042 | 0.280 | 0.593 | 0.580 |
| | | | | 0.80 | 0.057 | 0.041 | 0.045 | 0.061 | 0.047 | 0.048 | 0.300 | 0.604 | 0.591 |
| $T = 6$ | 48 | 58 | 10 | 0.20 | 0.016 | 0.015 | 0.054 | 0.020 | 0.022 | 0.028 | 0.037 | 0.727 | 0.754 |
| | | | | 0.50 | 0.018 | 0.016 | 0.051 | 0.023 | 0.020 | 0.027 | 0.036 | 0.730 | 0.756 |
| | | | | 0.80 | 0.024 | 0.017 | 0.048 | 0.033 | 0.028 | 0.028 | 0.042 | 0.737 | 0.760 |
| $T = 9$ | 114 | 130 | 16 | 0.20 | 0.001 | 0.000 | 0.046 | 0.015 | 0.017 | 0.032 | 0.024 | 0.764 | 0.788 |
| | | | | 0.50 | 0.001 | 0.000 | 0.044 | 0.017 | 0.015 | 0.025 | 0.023 | 0.766 | 0.788 |
| | | | | 0.80 | 0.001 | 0.000 | 0.038 | 0.025 | 0.019 | 0.020 | 0.027 | 0.770 | 0.791 |

*$R = 10000$ simulation replications. Design parameter values: $N = 200$, $SNR = 3$, $DEN_y = 1.0$, $EVF_x = 0.0$, $\bar{\rho}_{x\varepsilon} = 0.0$, $\xi = 0.8$, $\kappa = 0.00$, $\sigma_\varepsilon = 1$, $q = 1$, $\phi = 1.0$. These yield the DGP parameter values: $\pi_\lambda = 0.00$, $\pi_\eta = 0.00$, $\sigma_v = 0.60$, $\sigma_\eta = 1.0 * (1 - \gamma)$, $\rho_{v\varepsilon} = 0.0$ (and $\bar{\rho}_{x\eta} = 0.00$, $\bar{\rho}_{x\lambda} = 0.00$).

---

[10]A full set of Monte Carlo results can be obtained from the authors upon request. Below various results will be discussed in the text without referring to a table presented here, simply because we did not find it worthwhile to include it in the paper respecting reasonable space limitations.

nous, fewer instruments are being used. Since the ones that are now abstained from are most probably the strongest ones regarding $\Delta x_{it}$, it is no surprise that in the simulation results (not tabulated here) we note that especially the standard deviation of the $\beta$ coefficient suffers. Also the rejection probabilities of the various tests differ slightly between implementations WA and XA, but not in a very systematic way as it seems. When treating $x_{it}$ as endogenous (P0-EA) the precision of the estimators gets worse, with again no striking effects on the performance of test procedures under their respective null hypotheses. Upon comparing for P0 the instrument set A (and set C) with the one where $A^x$ ($C^x$) is replaced by $C1^x$ it has been found that the in practice quite popular choice $C1^x$ yields often slightly less efficient estimates for $\beta$, but much less efficient estimates for $\gamma$.

When $x_{it}$ is again treated as strictly exogenous, but the number of instruments is reduced by collapsing the instruments stemming from both $y_{it}$ and $x_{it}$, then we note from Table 7 (P0fc-XC) a mixed picture regarding the coefficient estimates. Although any substantial bias always reduces by collapsing, standard errors always increase at the same time, leading either to an increase or a decrease in RMSE. Decreases occur for the AB estimators of $\gamma$, especially when $\gamma$ is large; for $\beta$ just increases occur. A noteworthy reduction in RMSE does show up for BB2a when $\gamma$ is large, $T = 9$ and $\theta = 1$, but then the RMSE of BB2c using all instruments is in fact smaller. However, Table 8 (P0ft-XC) shows that collapsing is certainly found to be very beneficial for the type I error probability of coefficient tests, especially in cases where collapsing yields substantially reduced coefficient bias. The AB tests benefit a lot from collapsing, especially the c variant, leaving only little room for further improvement by employing the Windmeijer correction. After collapsing AB1 works well under homoskedasticity, and also under heteroskedasticity provided robust standard errors are being used, where the c version is clearly superior to the a version. AB2c has appropriate type I error probabilities, except for testing $\gamma$ when it is 0.8 at $T = 3$ and $\theta = 1$ (which is not repaired by a Windmeijer correction either), and is for most cases superior to AB2aW. After collapsing BB2a shows overrejection which is not completely repaired by BB2aW when $\theta = 1$. BB2c and BB2cW generally show lower rejection probabilities, with occasionally some underrejection. Tests based on MAB and MBB still heavily overreject. Table 9 (P0fJ-XC) shows that by collapsing $JAB$ and $JBB$ tests suffer much less from underrejection when $T$ is larger than 3. However, both the a and c versions of the $J^{(2,1)}$ and $J^{(2,2)}$ tests usually still underreject, mostly by about 1 or 2 percentage points. Good performance is shown by $JES_a^{(2,1)}$ and $JES_c^{(2,1)}$.

When $x_{it}$ is still correctly treated as strictly exogenous but for the level instruments just a few lags or first differences are being used (XL0 ... XL3) for both $y_{it}$ and $x_{it}$ then we find the following. Regarding feasible AB and BB estimation collapsing (XC) always gives smaller RMSE values than XL0 and XL1 (which is much worse than XL0), but this is not the case for XL2 and XL3. Whereas XC yields smaller bias, XL2 and XL3 often reach smaller Stdv and RMSE. Especially regarding $\beta$ XL3 performs better than XL2. Probably due to the smaller bias of XC it is more successful in mitigating size problems of coefficient tests than XL0 through XL3. The effects on $J$ tests is less clear-cut. Combining collapsing with restricting the lag length we find that XC2 and XC3 are in some aspects slightly worse but in others occasionally better than XC for P0. We also examined the hybrid instrumentation which seems popular amongst practitioners where $C^w$ is combined with $L1^x$. Especially for $\gamma$ this leads to loss of estimator precision

**Table 7.** P0fc-XC*

Feasible coefficient estimators for Arellano-Bond

| $\bar{\rho}_{x\varepsilon}=0.0$ | | | θ=0 AB1 Bias | AB1 Stdv | AB1 RMSE | AB2a Bias | AB2a Stdv | AB2a RMSE | AB2c Bias | AB2c Stdv | AB2c RMSE | θ=1 AB1 Bias | AB1 Stdv | AB1 RMSE | AB2a Bias | AB2a Stdv | AB2a RMSE | AB2c Bias | AB2c Stdv | AB2c RMSE | MAB Bias | MAB Stdv | MAB RMSE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | L | γ | | | | | | | | | | | | | | | | | | | | | |
| T=3 | 6 | 0.20 | -0.005 | 0.061 | 0.061 | -0.006 | 0.062 | 0.062 | -0.007 | 0.061 | 0.062 | -0.010 | 0.090 | 0.090 | -0.010 | 0.086 | 0.086 | -0.013 | 0.087 | 0.088 | -0.010 | 0.086 | 0.086 |
| | | 0.50 | -0.010 | 0.080 | 0.080 | -0.010 | 0.081 | 0.082 | -0.012 | 0.080 | 0.081 | -0.019 | 0.118 | 0.120 | -0.018 | 0.113 | 0.114 | -0.022 | 0.114 | 0.116 | -0.018 | 0.113 | 0.115 |
| | | 0.80 | -0.033 | 0.144 | 0.148 | -0.033 | 0.146 | 0.150 | -0.037 | 0.144 | 0.149 | -0.065 | 0.217 | 0.227 | -0.062 | 0.207 | 0.216 | -0.074 | 0.206 | 0.219 | -0.065 | 0.210 | 0.220 |
| T=6 | 12 | 0.20 | -0.003 | 0.032 | 0.032 | -0.003 | 0.033 | 0.033 | -0.003 | 0.032 | 0.032 | -0.006 | 0.049 | 0.050 | -0.004 | 0.046 | 0.046 | -0.006 | 0.047 | 0.048 | -0.004 | 0.039 | 0.040 |
| | | 0.50 | -0.005 | 0.038 | 0.038 | -0.005 | 0.040 | 0.040 | -0.005 | 0.038 | 0.039 | -0.011 | 0.059 | 0.060 | -0.007 | 0.054 | 0.055 | -0.010 | 0.056 | 0.057 | -0.007 | 0.047 | 0.047 |
| | | 0.80 | -0.015 | 0.061 | 0.063 | -0.015 | 0.063 | 0.065 | -0.016 | 0.061 | 0.063 | -0.035 | 0.094 | 0.100 | -0.026 | 0.088 | 0.092 | -0.033 | 0.089 | 0.095 | -0.024 | 0.074 | 0.078 |
| T=9 | 18 | 0.20 | -0.002 | 0.023 | 0.024 | -0.002 | 0.025 | 0.025 | -0.002 | 0.024 | 0.025 | -0.005 | 0.037 | 0.037 | -0.003 | 0.033 | 0.033 | -0.004 | 0.035 | 0.035 | -0.002 | 0.027 | 0.027 |
| | | 0.50 | -0.004 | 0.027 | 0.028 | -0.003 | 0.029 | 0.029 | -0.004 | 0.027 | 0.029 | -0.009 | 0.042 | 0.043 | -0.006 | 0.038 | 0.038 | -0.007 | 0.040 | 0.041 | -0.004 | 0.031 | 0.031 |
| | | 0.80 | -0.010 | 0.041 | 0.042 | -0.010 | 0.043 | 0.044 | -0.011 | 0.041 | 0.044 | -0.024 | 0.061 | 0.066 | -0.017 | 0.056 | 0.059 | -0.021 | 0.058 | 0.062 | -0.014 | 0.045 | 0.047 |
| | L | β | | | | | | | | | | | | | | | | | | | | | |
| T=3 | 6 | 1.43 | 0.001 | 0.113 | 0.113 | 0.001 | 0.115 | 0.115 | 0.000 | 0.114 | 0.114 | 0.004 | 0.181 | 0.181 | 0.003 | 0.172 | 0.172 | 0.002 | 0.174 | 0.174 | 0.004 | 0.159 | 0.159 |
| | | 0.93 | 0.001 | 0.113 | 0.113 | 0.001 | 0.115 | 0.115 | 0.000 | 0.114 | 0.114 | 0.003 | 0.180 | 0.180 | 0.002 | 0.172 | 0.172 | 0.002 | 0.174 | 0.174 | 0.004 | 0.159 | 0.159 |
| | | 0.31 | -0.000 | 0.113 | 0.113 | -0.001 | 0.115 | 0.115 | -0.001 | 0.114 | 0.114 | 0.001 | 0.179 | 0.179 | -0.001 | 0.171 | 0.171 | -0.001 | 0.173 | 0.173 | 0.000 | 0.158 | 0.158 |
| T=6 | 12 | 1.43 | 0.001 | 0.068 | 0.068 | 0.000 | 0.071 | 0.071 | 0.001 | 0.069 | 0.069 | 0.004 | 0.110 | 0.111 | 0.000 | 0.102 | 0.102 | 0.003 | 0.106 | 0.106 | 0.002 | 0.079 | 0.079 |
| | | 0.93 | 0.001 | 0.068 | 0.068 | 0.000 | 0.070 | 0.070 | 0.001 | 0.068 | 0.068 | 0.004 | 0.109 | 0.109 | 0.000 | 0.101 | 0.101 | 0.003 | 0.105 | 0.105 | 0.002 | 0.077 | 0.077 |
| | | 0.31 | 0.000 | 0.068 | 0.068 | -0.000 | 0.070 | 0.070 | 0.000 | 0.068 | 0.068 | 0.001 | 0.109 | 0.109 | -0.001 | 0.101 | 0.101 | 0.000 | 0.105 | 0.105 | 0.001 | 0.076 | 0.076 |
| T=9 | 18 | 1.43 | 0.001 | 0.052 | 0.052 | -0.000 | 0.055 | 0.055 | 0.001 | 0.052 | 0.052 | 0.003 | 0.084 | 0.084 | 0.000 | 0.075 | 0.075 | 0.002 | 0.080 | 0.080 | 0.001 | 0.056 | 0.056 |
| | | 0.93 | 0.001 | 0.052 | 0.052 | -0.000 | 0.054 | 0.054 | 0.001 | 0.052 | 0.052 | 0.003 | 0.083 | 0.083 | 0.001 | 0.074 | 0.074 | 0.002 | 0.079 | 0.079 | 0.002 | 0.055 | 0.055 |
| | | 0.31 | -0.000 | 0.051 | 0.051 | -0.001 | 0.053 | 0.053 | -0.000 | 0.051 | 0.051 | 0.001 | 0.082 | 0.082 | -0.001 | 0.073 | 0.073 | 0.000 | 0.078 | 0.078 | 0.000 | 0.053 | 0.053 |

Feasible coefficient estimators for Blundell-Bond

| $\bar{\rho}_{x\varepsilon}=0.0$ | | | θ=0 BB1 Bias | BB1 Stdv | BB1 RMSE | BB2a Bias | BB2a Stdv | BB2a RMSE | BB2c Bias | BB2c Stdv | BB2c RMSE | θ=1 BB1 Bias | BB1 Stdv | BB1 RMSE | BB2a Bias | BB2a Stdv | BB2a RMSE | BB2c Bias | BB2c Stdv | BB2c RMSE | MBB Bias | MBB Stdv | MBB RMSE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | L | γ | | | | | | | | | | | | | | | | | | | | | |
| T=3 | 9 | 0.20 | -0.001 | 0.053 | 0.053 | -0.001 | 0.054 | 0.054 | -0.002 | 0.054 | 0.054 | -0.005 | 0.078 | 0.078 | -0.002 | 0.074 | 0.074 | -0.004 | 0.075 | 0.075 | 0.001 | 0.076 | 0.076 |
| | | 0.50 | -0.005 | 0.062 | 0.062 | -0.004 | 0.062 | 0.062 | -0.005 | 0.061 | 0.061 | -0.011 | 0.092 | 0.093 | -0.007 | 0.086 | 0.086 | -0.008 | 0.087 | 0.088 | 0.008 | 0.439 | 0.439 |
| | | 0.80 | -0.019 | 0.085 | 0.087 | -0.013 | 0.084 | 0.085 | -0.015 | 0.084 | 0.085 | -0.037 | 0.129 | 0.134 | -0.024 | 0.123 | 0.126 | -0.025 | 0.126 | 0.128 | 0.086 | 2.228 | 2.229 |
| T=6 | 15 | 0.20 | -0.001 | 0.029 | 0.029 | -0.000 | 0.030 | 0.030 | -0.001 | 0.029 | 0.029 | -0.004 | 0.046 | 0.046 | -0.001 | 0.042 | 0.042 | -0.002 | 0.044 | 0.044 | -0.001 | 0.036 | 0.036 |
| | | 0.50 | -0.003 | 0.033 | 0.033 | -0.001 | 0.034 | 0.034 | -0.002 | 0.033 | 0.033 | -0.007 | 0.052 | 0.052 | -0.002 | 0.047 | 0.047 | -0.003 | 0.048 | 0.049 | 0.001 | 0.041 | 0.036 |
| | | 0.80 | -0.010 | 0.046 | 0.047 | -0.005 | 0.045 | 0.045 | -0.007 | 0.044 | 0.044 | -0.021 | 0.071 | 0.074 | -0.010 | 0.063 | 0.063 | -0.011 | 0.065 | 0.066 | 0.006 | 1.180 | 1.180 |
| T=9 | 21 | 0.20 | -0.001 | 0.022 | 0.022 | -0.000 | 0.023 | 0.023 | -0.001 | 0.022 | 0.022 | -0.003 | 0.035 | 0.035 | -0.001 | 0.031 | 0.031 | -0.001 | 0.033 | 0.033 | -0.001 | 0.026 | 0.026 |
| | | 0.50 | -0.002 | 0.025 | 0.025 | -0.001 | 0.025 | 0.025 | -0.002 | 0.024 | 0.024 | -0.006 | 0.038 | 0.039 | -0.002 | 0.034 | 0.034 | -0.003 | 0.036 | 0.036 | -0.001 | 0.028 | 0.028 |
| | | 0.80 | -0.007 | 0.033 | 0.034 | -0.003 | 0.033 | 0.033 | -0.005 | 0.032 | 0.032 | -0.016 | 0.051 | 0.053 | -0.007 | 0.044 | 0.045 | -0.007 | 0.047 | 0.047 | 0.003 | 0.042 | 0.042 |
| | L | β | | | | | | | | | | | | | | | | | | | | | |
| T=3 | 9 | 1.43 | 0.000 | 0.109 | 0.109 | -0.000 | 0.111 | 0.111 | -0.000 | 0.109 | 0.109 | 0.004 | 0.170 | 0.170 | 0.001 | 0.162 | 0.162 | -0.000 | 0.163 | 0.163 | 0.007 | 0.157 | 0.157 |
| | | 0.93 | 0.001 | 0.107 | 0.107 | 0.001 | 0.108 | 0.108 | 0.002 | 0.107 | 0.107 | 0.006 | 0.168 | 0.168 | 0.004 | 0.158 | 0.158 | 0.002 | 0.160 | 0.160 | 0.020 | 0.391 | 0.391 |
| | | 0.31 | 0.003 | 0.110 | 0.110 | 0.004 | 0.111 | 0.111 | 0.004 | 0.110 | 0.110 | 0.008 | 0.172 | 0.172 | 0.007 | 0.164 | 0.164 | 0.004 | 0.165 | 0.166 | 0.044 | 0.633 | 0.634 |
| T=6 | 15 | 1.43 | 0.001 | 0.067 | 0.067 | 0.001 | 0.069 | 0.069 | 0.001 | 0.067 | 0.067 | 0.004 | 0.107 | 0.107 | 0.002 | 0.098 | 0.098 | 0.004 | 0.100 | 0.100 | 0.005 | 0.078 | 0.078 |
| | | 0.93 | 0.001 | 0.066 | 0.066 | 0.001 | 0.068 | 0.068 | 0.001 | 0.065 | 0.065 | 0.005 | 0.105 | 0.105 | 0.003 | 0.095 | 0.096 | 0.006 | 0.100 | 0.100 | 0.013 | 0.076 | 0.076 |
| | | 0.31 | 0.001 | 0.066 | 0.066 | 0.002 | 0.068 | 0.068 | 0.002 | 0.066 | 0.066 | 0.004 | 0.106 | 0.106 | 0.003 | 0.097 | 0.097 | 0.005 | 0.101 | 0.101 | 0.013 | 0.212 | 0.213 |
| T=9 | 21 | 1.43 | 0.001 | 0.051 | 0.051 | 0.000 | 0.054 | 0.054 | 0.001 | 0.051 | 0.051 | 0.003 | 0.082 | 0.082 | 0.002 | 0.073 | 0.073 | 0.003 | 0.078 | 0.078 | 0.001 | 0.056 | 0.056 |
| | | 0.93 | 0.001 | 0.050 | 0.050 | 0.001 | 0.052 | 0.052 | 0.001 | 0.050 | 0.050 | 0.004 | 0.080 | 0.080 | 0.003 | 0.071 | 0.071 | 0.005 | 0.075 | 0.075 | 0.002 | 0.054 | 0.054 |
| | | 0.31 | 0.001 | 0.050 | 0.050 | 0.001 | 0.052 | 0.052 | 0.001 | 0.050 | 0.050 | 0.003 | 0.079 | 0.080 | 0.003 | 0.070 | 0.070 | 0.004 | 0.075 | 0.075 | 0.004 | 0.074 | 0.074 |

*$R = 10000$ simulation replications. Design parameter values: $N = 200$, $SNR = 3$, $DEN_y = 1.0$, $EVF_x = 0.0$, $\bar{\rho}_{x\varepsilon} = 0.0$, $\xi = 0.8$, $\kappa = 0.00$, $\sigma_\varepsilon = 1$, $q = 1$, $\phi = 1.0$. These yield the DGP parameter values: $\pi_\lambda = 0.00$, $\pi_\eta = 0.00$, $\sigma_v = 0.60$, $\sigma_\eta = 1.0 * (1 - \gamma)$, $\rho_{v\varepsilon} = 0.0$ (and $\bar{\rho}_{x\eta} = 0.00$, $\bar{\rho}_{x\lambda} = 0.00$).

**Table 8.** P0ft-XC*

Feasible t-test Arellano-Bond: actual significance level

| $\bar{\rho}_{x\varepsilon} = 0.0$ | | | $\theta = 0$ | | | | | | | $\theta = 1$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $L$ | $\gamma$ | AB1 | AB1aR | AB1cR | AB2a | AB2aW | AB2c | AB2cW | AB1 | AB1aR | AB1cR | AB2a | AB2aW | AB2c | AB2cW | MAB |
| $T = 3$ | 6 | 0.20 | 0.055 | 0.059 | 0.044 | 0.068 | 0.058 | 0.047 | 0.045 | 0.192 | 0.073 | 0.051 | 0.097 | 0.071 | 0.059 | 0.055 | 0.502 |
| | | 0.50 | 0.058 | 0.060 | 0.047 | 0.069 | 0.059 | 0.051 | 0.050 | 0.198 | 0.079 | 0.058 | 0.100 | 0.072 | 0.066 | 0.061 | 0.511 |
| | | 0.80 | 0.071 | 0.072 | 0.064 | 0.079 | 0.067 | 0.066 | 0.064 | 0.230 | 0.110 | 0.091 | 0.132 | 0.094 | 0.102 | 0.095 | 0.556 |
| $T = 6$ | 12 | 0.20 | 0.046 | 0.050 | 0.036 | 0.068 | 0.052 | 0.041 | 0.039 | 0.208 | 0.070 | 0.047 | 0.120 | 0.062 | 0.048 | 0.046 | 0.157 |
| | | 0.50 | 0.049 | 0.051 | 0.039 | 0.070 | 0.051 | 0.043 | 0.042 | 0.206 | 0.069 | 0.049 | 0.121 | 0.061 | 0.050 | 0.048 | 0.147 |
| | | 0.80 | 0.061 | 0.062 | 0.053 | 0.081 | 0.058 | 0.055 | 0.053 | 0.239 | 0.090 | 0.067 | 0.139 | 0.071 | 0.072 | 0.069 | 0.151 |
| $T = 9$ | 18 | 0.20 | 0.053 | 0.056 | 0.045 | 0.079 | 0.056 | 0.046 | 0.046 | 0.216 | 0.070 | 0.047 | 0.148 | 0.064 | 0.051 | 0.048 | 0.111 |
| | | 0.50 | 0.052 | 0.055 | 0.045 | 0.081 | 0.053 | 0.047 | 0.047 | 0.216 | 0.071 | 0.049 | 0.144 | 0.064 | 0.054 | 0.053 | 0.108 |
| | | 0.80 | 0.059 | 0.058 | 0.051 | 0.088 | 0.056 | 0.052 | 0.052 | 0.233 | 0.090 | 0.068 | 0.165 | 0.068 | 0.068 | 0.066 | 0.107 |
| | $L$ | $\beta$ | AB1 | AB1aR | AB1cR | AB2a | AB2aW | AB2c | AB2cW | AB1 | AB1aR | AB1cR | AB2a | AB2aW | AB2c | AB2cW | MAB |
| $T = 3$ | 6 | 1.43 | 0.051 | 0.055 | 0.053 | 0.061 | 0.054 | 0.054 | 0.051 | 0.212 | 0.069 | 0.063 | 0.091 | 0.068 | 0.068 | 0.061 | 0.586 |
| | | 0.93 | 0.051 | 0.055 | 0.053 | 0.061 | 0.054 | 0.054 | 0.051 | 0.211 | 0.066 | 0.061 | 0.089 | 0.067 | 0.066 | 0.059 | 0.584 |
| | | 0.31 | 0.050 | 0.053 | 0.050 | 0.059 | 0.052 | 0.052 | 0.050 | 0.206 | 0.066 | 0.059 | 0.089 | 0.064 | 0.061 | 0.054 | 0.572 |
| $T = 6$ | 12 | 1.43 | 0.047 | 0.050 | 0.048 | 0.065 | 0.051 | 0.050 | 0.049 | 0.215 | 0.064 | 0.054 | 0.114 | 0.062 | 0.058 | 0.055 | 0.208 |
| | | 0.93 | 0.048 | 0.050 | 0.048 | 0.065 | 0.052 | 0.051 | 0.050 | 0.217 | 0.065 | 0.054 | 0.114 | 0.062 | 0.056 | 0.055 | 0.206 |
| | | 0.31 | 0.049 | 0.052 | 0.049 | 0.068 | 0.051 | 0.051 | 0.050 | 0.214 | 0.065 | 0.052 | 0.116 | 0.064 | 0.055 | 0.053 | 0.209 |
| $T = 9$ | 18 | 1.43 | 0.051 | 0.052 | 0.050 | 0.078 | 0.054 | 0.052 | 0.051 | 0.221 | 0.064 | 0.051 | 0.135 | 0.058 | 0.055 | 0.054 | 0.138 |
| | | 0.93 | 0.052 | 0.052 | 0.052 | 0.078 | 0.053 | 0.053 | 0.052 | 0.222 | 0.063 | 0.052 | 0.138 | 0.059 | 0.054 | 0.053 | 0.137 |
| | | 0.31 | 0.051 | 0.053 | 0.050 | 0.076 | 0.052 | 0.053 | 0.053 | 0.219 | 0.063 | 0.052 | 0.139 | 0.057 | 0.053 | 0.051 | 0.139 |

Feasible t-test Blundell-Bond: actual significance level

| $\bar{\rho}_{x\varepsilon} = 0.0$ | | | $\theta = 0$ | | | | | | | $\theta = 1$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $L$ | $\gamma$ | BB1 | BB1aR | BB1cR | BB2a | BB2aW | BB2c | BB2cW | BB1 | BB1aR | BB1cR | BB2a | BB2aW | BB2c | BB2cW | MBB |
| $T = 3$ | 9 | 0.20 | 0.054 | 0.054 | 0.041 | 0.072 | 0.057 | 0.047 | 0.043 | 0.196 | 0.070 | 0.046 | 0.110 | 0.066 | 0.058 | 0.050 | 0.472 |
| | | 0.50 | 0.055 | 0.055 | 0.037 | 0.072 | 0.057 | 0.047 | 0.041 | 0.201 | 0.071 | 0.046 | 0.109 | 0.067 | 0.064 | 0.054 | 0.480 |
| | | 0.80 | 0.057 | 0.058 | 0.034 | 0.076 | 0.056 | 0.041 | 0.034 | 0.219 | 0.079 | 0.046 | 0.131 | 0.071 | 0.064 | 0.053 | 0.533 |
| $T = 6$ | 15 | 0.20 | 0.047 | 0.051 | 0.034 | 0.070 | 0.050 | 0.038 | 0.036 | 0.203 | 0.066 | 0.043 | 0.129 | 0.060 | 0.048 | 0.046 | 0.143 |
| | | 0.50 | 0.046 | 0.050 | 0.034 | 0.069 | 0.048 | 0.038 | 0.037 | 0.203 | 0.067 | 0.043 | 0.127 | 0.058 | 0.048 | 0.045 | 0.142 |
| | | 0.80 | 0.054 | 0.056 | 0.038 | 0.073 | 0.049 | 0.038 | 0.035 | 0.225 | 0.081 | 0.050 | 0.135 | 0.062 | 0.049 | 0.045 | 0.198 |
| $T = 9$ | 21 | 0.20 | 0.049 | 0.053 | 0.040 | 0.081 | 0.054 | 0.043 | 0.042 | 0.208 | 0.067 | 0.047 | 0.153 | 0.063 | 0.050 | 0.048 | 0.102 |
| | | 0.50 | 0.052 | 0.056 | 0.043 | 0.079 | 0.051 | 0.044 | 0.043 | 0.209 | 0.069 | 0.047 | 0.155 | 0.059 | 0.049 | 0.047 | 0.101 |
| | | 0.80 | 0.052 | 0.056 | 0.042 | 0.081 | 0.051 | 0.041 | 0.040 | 0.225 | 0.078 | 0.054 | 0.160 | 0.058 | 0.049 | 0.045 | 0.108 |
| | $L$ | $\beta$ | BB1 | BB1aR | BB1cR | BB2a | BB2aW | BB2c | BB2cW | BB1 | BB1aR | BB1cR | BB2a | BB2aW | BB2c | BB2cW | MBB |
| $T = 3$ | 9 | 1.43 | 0.050 | 0.053 | 0.050 | 0.065 | 0.054 | 0.053 | 0.053 | 0.210 | 0.068 | 0.062 | 0.104 | 0.066 | 0.069 | 0.068 | 0.581 |
| | | 0.93 | 0.050 | 0.053 | 0.048 | 0.064 | 0.053 | 0.051 | 0.051 | 0.213 | 0.068 | 0.061 | 0.105 | 0.068 | 0.068 | 0.067 | 0.586 |
| | | 0.31 | 0.051 | 0.054 | 0.050 | 0.066 | 0.055 | 0.053 | 0.051 | 0.215 | 0.070 | 0.062 | 0.111 | 0.073 | 0.072 | 0.072 | 0.639 |
| $T = 6$ | 15 | 1.43 | 0.048 | 0.050 | 0.048 | 0.071 | 0.052 | 0.050 | 0.051 | 0.214 | 0.065 | 0.056 | 0.124 | 0.063 | 0.062 | 0.063 | 0.213 |
| | | 0.93 | 0.049 | 0.050 | 0.049 | 0.071 | 0.054 | 0.050 | 0.051 | 0.216 | 0.066 | 0.056 | 0.125 | 0.064 | 0.062 | 0.063 | 0.212 |
| | | 0.31 | 0.050 | 0.051 | 0.050 | 0.073 | 0.053 | 0.051 | 0.052 | 0.217 | 0.068 | 0.054 | 0.128 | 0.065 | 0.060 | 0.063 | 0.241 |
| $T = 9$ | 21 | 1.43 | 0.050 | 0.052 | 0.050 | 0.084 | 0.055 | 0.051 | 0.051 | 0.220 | 0.065 | 0.052 | 0.147 | 0.059 | 0.056 | 0.059 | 0.140 |
| | | 0.93 | 0.050 | 0.052 | 0.050 | 0.083 | 0.053 | 0.051 | 0.051 | 0.219 | 0.063 | 0.052 | 0.149 | 0.059 | 0.055 | 0.059 | 0.138 |
| | | 0.31 | 0.051 | 0.053 | 0.050 | 0.080 | 0.054 | 0.051 | 0.051 | 0.221 | 0.064 | 0.053 | 0.150 | 0.057 | 0.056 | 0.055 | 0.148 |

*$R = 10000$ simulation replications. Design parameter values: $N = 200$, $SNR = 3$, $DEN_y = 1.0$, $EVF_x = 0.0$, $\bar{\rho}_{x\varepsilon} = 0.0$, $\xi = 0.8$, $\kappa = 0.00$, $\sigma_\varepsilon = 1$, $q = 1$, $\phi = 1.0$. These yield the DGP parameter values: $\pi_\lambda = 0.00$, $\pi_\eta = 0.00$, $\sigma_v = 0.60$, $\sigma_\eta = 1.0 * (1 - \gamma)$, $\rho_{v\varepsilon} = 0.0$ (and $\bar{\rho}_{x\eta} = 0.00$, $\bar{\rho}_{x\lambda} = 0.00$).

without any other clear advantages, so it does not outperform the XC results for P0. From examining P0-WC (and P0-EC) we find that in comparison to P0-WA (P0-EA) there is often some increase in RMSE, but the size control of especially the t-tests is much better.

Summarizing the results for P0 on feasible estimators and tests we note that when choosing between different possible instrument sets a trade off has to be made between estimator precision and test size control. For both some form of reduction of the instrument set is often but not always beneficial. Not one single method seems superior irrespective of the actual values of $\gamma$, $\beta$ and $T$. Using all instruments is not necessarily a bad choice; also XC, XL3 and XC3 often work well. To mitigate estimator bias and foster test size control while not sacrificing too much estimator precision using collapsing (C) for all regressors seems a reasonable compromise, as far as P0 is concerned. Coefficient and $J$ tests based on the feasible modified estimator behave so poorly, that in the remainder we no longer mention its results.

**Table 9.** P0fJ-XC*

Feasible Sargan-Hansen test: rejection probability

| $\bar{\rho}_{x\varepsilon}=0.0$ | df | | | | $\theta = 0$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | AB | BB | Inc | $\gamma$ | $JAB_a^{(2,1)}$ | $JBB_a^{(2,1)}$ | $JES_a^{(2,1)}$ | $JAB_c^{(2,1)}$ | $JBB_c^{(2,1)}$ | $JES_c^{(2,1)}$ | $JMAB$ | $JMBB$ | $JESM$ |
| $T=3$ | 4 | 6 | 2 | 0.20 | 0.044 | 0.047 | 0.047 | 0.045 | 0.043 | 0.040 | 0.268 | 0.254 | 0.250 |
| | | | | 0.50 | 0.045 | 0.045 | 0.045 | 0.045 | 0.039 | 0.033 | 0.262 | 0.278 | 0.272 |
| | | | | 0.80 | 0.048 | 0.046 | 0.044 | 0.046 | 0.040 | 0.032 | 0.268 | 0.356 | 0.347 |
| $T=6$ | 10 | 12 | 2 | 0.20 | 0.045 | 0.045 | 0.052 | 0.040 | 0.040 | 0.048 | 0.049 | 0.294 | 0.371 |
| | | | | 0.50 | 0.048 | 0.046 | 0.047 | 0.040 | 0.039 | 0.042 | 0.048 | 0.301 | 0.376 |
| | | | | 0.80 | 0.050 | 0.049 | 0.047 | 0.043 | 0.041 | 0.040 | 0.051 | 0.331 | 0.407 |
| $T=9$ | 16 | 18 | 2 | 0.20 | 0.041 | 0.042 | 0.051 | 0.037 | 0.038 | 0.048 | 0.042 | 0.299 | 0.400 |
| | | | | 0.50 | 0.042 | 0.042 | 0.052 | 0.038 | 0.040 | 0.047 | 0.037 | 0.305 | 0.405 |
| | | | | 0.80 | 0.046 | 0.044 | 0.050 | 0.039 | 0.039 | 0.047 | 0.042 | 0.312 | 0.411 |

| $\bar{\rho}_{x\varepsilon}=0.0$ | df | | | | $\theta = 1$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | AB | BB | Inc | $\gamma$ | $JAB_a^{(2,1)}$ | $JBB_a^{(2,1)}$ | $JES_a^{(2,1)}$ | $JAB_c^{(2,1)}$ | $JBB_c^{(2,1)}$ | $JES_c^{(2,1)}$ | $JMAB$ | $JMBB$ | $JESM$ |
| $T=3$ | 4 | 6 | 2 | 0.20 | 0.037 | 0.037 | 0.045 | 0.042 | 0.040 | 0.044 | 0.300 | 0.491 | 0.471 |
| | | | | 0.50 | 0.040 | 0.037 | 0.042 | 0.045 | 0.042 | 0.042 | 0.298 | 0.515 | 0.501 |
| | | | | 0.80 | 0.048 | 0.040 | 0.041 | 0.054 | 0.048 | 0.046 | 0.308 | 0.538 | 0.522 |
| $T=6$ | 10 | 12 | 2 | 0.20 | 0.032 | 0.030 | 0.048 | 0.036 | 0.035 | 0.042 | 0.054 | 0.655 | 0.699 |
| | | | | 0.50 | 0.031 | 0.029 | 0.046 | 0.037 | 0.037 | 0.044 | 0.053 | 0.661 | 0.703 |
| | | | | 0.80 | 0.037 | 0.035 | 0.047 | 0.040 | 0.040 | 0.054 | 0.056 | 0.672 | 0.712 |
| $T=9$ | 16 | 18 | 2 | 0.20 | 0.027 | 0.027 | 0.047 | 0.033 | 0.034 | 0.050 | 0.046 | 0.701 | 0.745 |
| | | | | 0.50 | 0.029 | 0.028 | 0.046 | 0.034 | 0.032 | 0.049 | 0.044 | 0.705 | 0.746 |
| | | | | 0.80 | 0.032 | 0.030 | 0.051 | 0.035 | 0.036 | 0.054 | 0.046 | 0.710 | 0.748 |

*$R = 10000$ simulation replications. Design parameter values: $N = 200$, $SNR = 3$, $DEN_y = 1.0$, $EVF_x = 0.0$, $\bar{\rho}_{x\varepsilon} = 0.0$, $\xi = 0.8$, $\kappa = 0.00$, $\sigma_\varepsilon = 1$, $q = 1$, $\phi = 1.0$. These yield the DGP parameter values: $\pi_\lambda = 0.00$, $\pi_\eta = 0.00$, $\sigma_v = 0.60$, $\sigma_\eta = 1.0 * (1 - \gamma)$, $\rho_{v\varepsilon} = 0.0$ (and $\bar{\rho}_{x\eta} = 0.00$, $\bar{\rho}_{x\lambda} = 0.00$).

### 5.1.2 Results for alternative parametrizations

Next we examine a series of alternative parametrizations where each time we just change one of the parameter values of one of the already examined cases. In P1 we increase $DEN_y^\eta$ from 1 to 4 (hence, substantially increasing the relative variance of the individual effects). We note that for P1-XA (not tabulated here) all estimators regarding $\gamma$ are more biased and dispersed than for P0-XA, but there is little or no effect on the $\beta$ estimates. For both $T$ and $\gamma$ large this leads to serious overrejection for the unfeasible coefficient tests regarding $\gamma$, in particular for ABu. Self-evidently, this carries over to the feasible tests and, although a Windmeijer correction has a mitigating effect, the overrejection remains often serious for both AB and BB based tests. Tests on $\beta$ based on AB behave reasonable, apart from not robustified AB1 and AB2a. For the latter a Windmeijer correction proves reasonably effective. When exploiting the effect stationarity the BB2c implementation seems preferable. The unfeasible $J$ tests show a similar though slightly more extreme pattern as for P0-XA. Among the feasible tests both serious underrejection and overrejection occurs. As far as the incremental tests concerns $JES_c^{(2,2)}$ behaves remarkably well.

In Tables 10, 11 and 12 (P1f$j$-XC, $j$ = c,t,J) we find that collapsing leads again to reduced bias, slightly deteriorated precision though improved size control (here all unfeasible tests behave reasonably well). All feasible AB1R and AB2W tests have reasonable size control, apart from tests on $\gamma$ when $T$ is small and $\gamma$ large. These give actual significance levels close to 10%. BB2cW seems slightly better than BB2aW. The 1-step $J$ tests show some serious overrejection whereas the 2-step $J$ tests behave quite satisfactorily. For C3 reasonably similar results are obtained, but those for L3 are generally slightly less attractive.

In P2 we increase $EVF_x$ from 0 to 0.6, upon having again $IEF_x^\eta = 0$ (hence, $x_{it}$ is still uncorrelated with effect $\eta_i$ though correlated with effect $\lambda_i$, which determines any heteroskedasticity). This leads to increased $\beta$ values. Results for P2-XA show larger

**Table 10.** P1fc-XC*

Feasible coefficient estimators for Arellano-Bond

| | | | θ=0 AB1 | | | AB2a | | | AB2c | | | θ=1 AB1 | | | AB2a | | | AB2c | | | MAB | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ρ̄_xε = 0.0 | L | γ | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE |
| T = 3 | 6 | 0.20 | -0.011 | 0.096 | 0.097 | -0.012 | 0.098 | 0.099 | -0.012 | 0.097 | 0.098 | -0.017 | 0.134 | 0.135 | -0.020 | 0.127 | 0.129 | -0.021 | 0.126 | 0.128 | -0.017 | 0.132 | 0.134 |
| | | 0.50 | -0.023 | 0.126 | 0.128 | -0.023 | 0.129 | 0.131 | -0.023 | 0.126 | 0.128 | -0.036 | 0.175 | 0.179 | -0.038 | 0.167 | 0.171 | -0.041 | 0.164 | 0.169 | -0.035 | 0.172 | 0.176 |
| | | 0.80 | -0.079 | 0.223 | 0.236 | -0.082 | 0.230 | 0.244 | -0.081 | 0.223 | 0.238 | -0.131 | 0.303 | 0.330 | -0.134 | 0.298 | 0.326 | -0.135 | 0.284 | 0.315 | -0.136 | 0.298 | 0.328 |
| T = 6 | 12 | 0.20 | -0.005 | 0.042 | 0.042 | -0.005 | 0.043 | 0.043 | -0.005 | 0.042 | 0.042 | -0.009 | 0.062 | 0.063 | -0.007 | 0.058 | 0.058 | -0.008 | 0.058 | 0.059 | -0.006 | 0.052 | 0.053 |
| | | 0.50 | -0.009 | 0.052 | 0.053 | -0.009 | 0.054 | 0.055 | -0.009 | 0.052 | 0.053 | -0.017 | 0.077 | 0.079 | -0.014 | 0.072 | 0.074 | -0.016 | 0.072 | 0.074 | -0.011 | 0.065 | 0.066 |
| | | 0.80 | -0.033 | 0.090 | 0.095 | -0.034 | 0.094 | 0.100 | -0.033 | 0.090 | 0.096 | -0.061 | 0.126 | 0.139 | -0.053 | 0.121 | 0.133 | -0.056 | 0.117 | 0.130 | -0.049 | 0.113 | 0.123 |
| T = 9 | 18 | 0.20 | -0.003 | 0.029 | 0.029 | -0.003 | 0.030 | 0.030 | -0.003 | 0.029 | 0.029 | -0.006 | 0.044 | 0.044 | -0.005 | 0.040 | 0.040 | -0.005 | 0.041 | 0.042 | -0.004 | 0.035 | 0.035 |
| | | 0.50 | -0.006 | 0.035 | 0.035 | -0.005 | 0.036 | 0.037 | -0.006 | 0.035 | 0.035 | -0.012 | 0.052 | 0.053 | -0.009 | 0.048 | 0.049 | -0.010 | 0.049 | 0.050 | -0.007 | 0.041 | 0.042 |
| | | 0.80 | -0.020 | 0.057 | 0.061 | -0.020 | 0.061 | 0.064 | -0.020 | 0.057 | 0.061 | -0.040 | 0.081 | 0.090 | -0.032 | 0.077 | 0.084 | -0.035 | 0.076 | 0.083 | -0.028 | 0.070 | 0.075 |
| | L | β | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE |
| T = 3 | 6 | 1.43 | 0.001 | 0.114 | 0.114 | 0.001 | 0.115 | 0.115 | 0.001 | 0.114 | 0.114 | 0.005 | 0.182 | 0.182 | 0.002 | 0.174 | 0.174 | 0.003 | 0.176 | 0.176 | 0.007 | 0.161 | 0.161 |
| | | 0.93 | -0.000 | 0.114 | 0.114 | -0.000 | 0.115 | 0.115 | -0.000 | 0.115 | 0.115 | 0.003 | 0.181 | 0.181 | 0.000 | 0.171 | 0.174 | 0.001 | 0.175 | 0.175 | 0.006 | 0.159 | 0.159 |
| | | 0.31 | -0.003 | 0.113 | 0.113 | -0.003 | 0.114 | 0.114 | -0.003 | 0.113 | 0.113 | -0.002 | 0.177 | 0.177 | -0.006 | 0.171 | 0.171 | -0.004 | 0.172 | 0.172 | -0.002 | 0.157 | 0.157 |
| T = 6 | 12 | 1.43 | 0.002 | 0.068 | 0.068 | 0.001 | 0.071 | 0.071 | 0.002 | 0.069 | 0.069 | 0.004 | 0.110 | 0.110 | 0.002 | 0.103 | 0.103 | 0.004 | 0.106 | 0.106 | 0.003 | 0.080 | 0.080 |
| | | 0.93 | 0.002 | 0.068 | 0.068 | 0.001 | 0.070 | 0.070 | 0.002 | 0.068 | 0.068 | 0.004 | 0.109 | 0.109 | 0.001 | 0.102 | 0.102 | 0.004 | 0.105 | 0.105 | 0.004 | 0.078 | 0.078 |
| | | 0.31 | -0.001 | 0.068 | 0.068 | -0.001 | 0.070 | 0.070 | -0.001 | 0.068 | 0.068 | 0.000 | 0.108 | 0.108 | -0.002 | 0.102 | 0.102 | -0.001 | 0.105 | 0.105 | 0.002 | 0.076 | 0.076 |
| T = 9 | 18 | 1.43 | 0.001 | 0.052 | 0.052 | 0.000 | 0.055 | 0.055 | 0.001 | 0.052 | 0.052 | 0.003 | 0.084 | 0.084 | 0.001 | 0.075 | 0.075 | 0.003 | 0.079 | 0.079 | 0.002 | 0.057 | 0.057 |
| | | 0.93 | 0.001 | 0.051 | 0.051 | 0.000 | 0.054 | 0.054 | 0.001 | 0.052 | 0.052 | 0.004 | 0.083 | 0.083 | 0.001 | 0.075 | 0.075 | 0.003 | 0.079 | 0.079 | 0.003 | 0.053 | 0.055 |
| | | 0.31 | -0.001 | 0.051 | 0.051 | -0.001 | 0.053 | 0.053 | -0.001 | 0.051 | 0.051 | 0.000 | 0.082 | 0.082 | -0.001 | 0.074 | 0.074 | -0.000 | 0.078 | 0.078 | 0.001 | 0.053 | 0.053 |

Feasible coefficient estimators for Blundell-Bond

| | | | θ=0 BB1 | | | BB2a | | | BB2c | | | θ=1 BB1 | | | BB2a | | | BB2c | | | MBB | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ρ̄_xε = 0.0 | L | γ | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE |
| T = 3 | 9 | 0.20 | 0.026 | 0.108 | 0.111 | 0.016 | 0.091 | 0.093 | 0.015 | 0.090 | 0.091 | 0.023 | 0.125 | 0.128 | 0.018 | 0.114 | 0.116 | 0.016 | 0.114 | 0.115 | 0.021 | 0.123 | 0.124 |
| | | 0.50 | 0.016 | 0.100 | 0.101 | 0.014 | 0.100 | 0.101 | 0.012 | 0.099 | 0.100 | 0.011 | 0.124 | 0.125 | 0.012 | 0.124 | 0.124 | 0.010 | 0.126 | 0.126 | 0.017 | 0.126 | 0.126 |
| | | 0.80 | -0.008 | 0.103 | 0.103 | -0.005 | 0.113 | 0.113 | -0.005 | 0.115 | 0.115 | -0.022 | 0.142 | 0.144 | -0.015 | 0.148 | 0.149 | -0.016 | 0.153 | 0.154 | 0.027 | 0.297 | 0.298 |
| T = 6 | 15 | 0.20 | 0.009 | 0.048 | 0.049 | 0.002 | 0.040 | 0.040 | 0.001 | 0.039 | 0.039 | 0.007 | 0.061 | 0.062 | 0.004 | 0.053 | 0.054 | 0.002 | 0.055 | 0.055 | 0.002 | 0.049 | 0.049 |
| | | 0.50 | 0.006 | 0.047 | 0.048 | 0.003 | 0.046 | 0.046 | 0.001 | 0.045 | 0.045 | 0.002 | 0.064 | 0.064 | 0.004 | 0.061 | 0.061 | 0.002 | 0.063 | 0.063 | 0.001 | 0.056 | 0.056 |
| | | 0.80 | -0.005 | 0.054 | 0.054 | -0.000 | 0.057 | 0.057 | -0.004 | 0.057 | 0.057 | -0.016 | 0.078 | 0.080 | -0.005 | 0.075 | 0.075 | -0.009 | 0.078 | 0.078 | 0.008 | 0.081 | 0.082 |
| T = 9 | 21 | 0.20 | 0.005 | 0.033 | 0.033 | 0.001 | 0.028 | 0.028 | -0.000 | 0.027 | 0.027 | 0.003 | 0.044 | 0.044 | 0.002 | 0.038 | 0.038 | -0.000 | 0.039 | 0.039 | 0.000 | 0.033 | 0.033 |
| | | 0.50 | 0.003 | 0.033 | 0.033 | 0.002 | 0.032 | 0.032 | -0.000 | 0.031 | 0.031 | -0.001 | 0.046 | 0.046 | 0.002 | 0.042 | 0.042 | -0.001 | 0.044 | 0.044 | -0.001 | 0.037 | 0.037 |
| | | 0.80 | -0.004 | 0.039 | 0.039 | -0.000 | 0.041 | 0.041 | -0.003 | 0.040 | 0.040 | -0.014 | 0.056 | 0.058 | -0.004 | 0.053 | 0.053 | -0.007 | 0.055 | 0.056 | -0.002 | 0.051 | 0.051 |
| | L | β | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE |
| T = 3 | 9 | 1.43 | -0.014 | 0.124 | 0.125 | 0.001 | 0.117 | 0.117 | 0.001 | 0.114 | 0.114 | -0.010 | 0.185 | 0.185 | -0.000 | 0.175 | 0.175 | 0.000 | 0.174 | 0.174 | -0.009 | 0.171 | 0.171 |
| | | 0.93 | -0.008 | 0.115 | 0.115 | -0.001 | 0.115 | 0.115 | -0.001 | 0.113 | 0.113 | -0.003 | 0.176 | 0.176 | -0.000 | 0.170 | 0.170 | 0.000 | 0.170 | 0.170 | -0.001 | 0.163 | 0.163 |
| | | 0.31 | -0.001 | 0.111 | 0.111 | -0.001 | 0.113 | 0.113 | -0.001 | 0.112 | 0.112 | 0.004 | 0.174 | 0.174 | 0.002 | 0.167 | 0.167 | 0.003 | 0.168 | 0.169 | 0.017 | 0.207 | 0.207 |
| T = 6 | 15 | 1.43 | -0.005 | 0.073 | 0.073 | 0.000 | 0.071 | 0.071 | 0.000 | 0.069 | 0.069 | -0.002 | 0.112 | 0.112 | -0.001 | 0.103 | 0.103 | 0.001 | 0.106 | 0.106 | -0.001 | 0.082 | 0.082 |
| | | 0.93 | -0.002 | 0.069 | 0.069 | 0.001 | 0.070 | 0.070 | 0.000 | 0.068 | 0.068 | 0.001 | 0.108 | 0.108 | 0.001 | 0.101 | 0.101 | 0.002 | 0.104 | 0.104 | 0.000 | 0.079 | 0.079 |
| | | 0.31 | 0.000 | 0.067 | 0.067 | 0.001 | 0.069 | 0.069 | 0.000 | 0.067 | 0.067 | 0.003 | 0.106 | 0.107 | 0.002 | 0.100 | 0.100 | 0.003 | 0.103 | 0.103 | 0.004 | 0.077 | 0.077 |
| T = 9 | 21 | 1.43 | -0.003 | 0.056 | 0.056 | -0.001 | 0.055 | 0.055 | -0.000 | 0.052 | 0.052 | -0.001 | 0.086 | 0.086 | -0.000 | 0.076 | 0.076 | 0.001 | 0.081 | 0.081 | -0.000 | 0.058 | 0.058 |
| | | 0.93 | -0.001 | 0.052 | 0.052 | 0.000 | 0.054 | 0.054 | 0.000 | 0.051 | 0.051 | 0.002 | 0.082 | 0.082 | 0.001 | 0.073 | 0.073 | 0.002 | 0.078 | 0.078 | 0.000 | 0.056 | 0.056 |
| | | 0.31 | 0.001 | 0.050 | 0.050 | 0.001 | 0.053 | 0.053 | 0.000 | 0.050 | 0.050 | 0.002 | 0.080 | 0.080 | 0.002 | 0.073 | 0.073 | 0.002 | 0.076 | 0.076 | 0.001 | 0.054 | 0.054 |

*R = 10000 simulation replications. Design parameter values: N = 200, SNR = 3, DEN_y = 4.0, EVF_x = 0.0, ρ̄_xε = 0.0, ξ = 0.8, κ = 0.00, σ_ε = 1, q = 1, φ = 1.0. These yield the DGP parameter values: π_λ = 0.00, π_η = 0.00, σ_v = 0.60, σ_η = 4.0 * (1 − γ), ρ_vε = 0.0 (and ρ̄_xλ = 0.00).

**Table 11.** P1ft-XC*

Feasible t-test Arellano-Bond: actual significance level

| $\bar{\rho}_{x\varepsilon} = 0.0$ | | | $\theta = 0$ | | | | | | | $\theta = 1$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $L$ | $\gamma$ | AB1 | AB1aR | AB1cR | AB2a | AB2aW | AB2c | AB2cW | AB1 | AB1aR | AB1cR | AB2a | AB2aW | AB2c | AB2cW | MAB |
| $T=3$ | 6 | 0.20 | 0.054 | 0.058 | 0.053 | 0.069 | 0.059 | 0.057 | 0.054 | 0.165 | 0.075 | 0.064 | 0.099 | 0.072 | 0.068 | 0.063 | 0.558 |
| | | 0.50 | 0.062 | 0.064 | 0.062 | 0.073 | 0.062 | 0.064 | 0.062 | 0.179 | 0.082 | 0.073 | 0.106 | 0.076 | 0.080 | 0.073 | 0.571 |
| | | 0.80 | 0.095 | 0.092 | 0.096 | 0.106 | 0.086 | 0.099 | 0.097 | 0.236 | 0.134 | 0.130 | 0.165 | 0.111 | 0.137 | 0.130 | 0.623 |
| $T=6$ | 12 | 0.20 | 0.049 | 0.049 | 0.045 | 0.069 | 0.053 | 0.045 | 0.045 | 0.192 | 0.070 | 0.053 | 0.115 | 0.060 | 0.052 | 0.050 | 0.172 |
| | | 0.50 | 0.050 | 0.052 | 0.048 | 0.071 | 0.053 | 0.049 | 0.047 | 0.192 | 0.072 | 0.058 | 0.117 | 0.059 | 0.056 | 0.054 | 0.176 |
| | | 0.80 | 0.072 | 0.065 | 0.068 | 0.094 | 0.064 | 0.070 | 0.069 | 0.215 | 0.099 | 0.086 | 0.155 | 0.076 | 0.088 | 0.085 | 0.188 |
| $T=9$ | 18 | 0.20 | 0.054 | 0.053 | 0.050 | 0.079 | 0.055 | 0.051 | 0.050 | 0.201 | 0.070 | 0.054 | 0.140 | 0.063 | 0.055 | 0.053 | 0.116 |
| | | 0.50 | 0.054 | 0.055 | 0.051 | 0.081 | 0.053 | 0.053 | 0.052 | 0.201 | 0.071 | 0.057 | 0.140 | 0.063 | 0.058 | 0.055 | 0.120 |
| | | 0.80 | 0.069 | 0.065 | 0.066 | 0.100 | 0.060 | 0.066 | 0.065 | 0.225 | 0.095 | 0.084 | 0.174 | 0.067 | 0.078 | 0.076 | 0.132 |
| | $L$ | $\beta$ | AB1 | AB1aR | AB1cR | AB2a | AB2aW | AB2c | AB2cW | AB1 | AB1aR | AB1cR | AB2a | AB2aW | AB2c | AB2cW | MAB |
| $T=3$ | 6 | 1.43 | 0.050 | 0.054 | 0.050 | 0.060 | 0.052 | 0.054 | 0.051 | 0.207 | 0.066 | 0.060 | 0.086 | 0.065 | 0.065 | 0.058 | 0.582 |
| | | 0.93 | 0.049 | 0.054 | 0.051 | 0.060 | 0.052 | 0.054 | 0.052 | 0.204 | 0.064 | 0.057 | 0.086 | 0.065 | 0.061 | 0.055 | 0.576 |
| | | 0.31 | 0.045 | 0.049 | 0.046 | 0.057 | 0.050 | 0.047 | 0.045 | 0.195 | 0.061 | 0.051 | 0.083 | 0.061 | 0.057 | 0.050 | 0.564 |
| $T=6$ | 12 | 1.43 | 0.047 | 0.049 | 0.048 | 0.067 | 0.050 | 0.050 | 0.049 | 0.213 | 0.065 | 0.053 | 0.111 | 0.064 | 0.056 | 0.053 | 0.206 |
| | | 0.93 | 0.049 | 0.051 | 0.049 | 0.065 | 0.051 | 0.051 | 0.050 | 0.215 | 0.066 | 0.053 | 0.112 | 0.062 | 0.057 | 0.053 | 0.207 |
| | | 0.31 | 0.048 | 0.051 | 0.048 | 0.066 | 0.051 | 0.049 | 0.049 | 0.211 | 0.064 | 0.051 | 0.114 | 0.062 | 0.055 | 0.052 | 0.206 |
| $T=9$ | 18 | 1.43 | 0.050 | 0.051 | 0.050 | 0.080 | 0.053 | 0.051 | 0.050 | 0.219 | 0.063 | 0.051 | 0.131 | 0.060 | 0.053 | 0.052 | 0.134 |
| | | 0.93 | 0.051 | 0.052 | 0.052 | 0.078 | 0.052 | 0.052 | 0.052 | 0.219 | 0.063 | 0.051 | 0.131 | 0.057 | 0.054 | 0.052 | 0.136 |
| | | 0.31 | 0.051 | 0.053 | 0.051 | 0.076 | 0.052 | 0.052 | 0.051 | 0.215 | 0.063 | 0.050 | 0.130 | 0.056 | 0.052 | 0.051 | 0.137 |

Feasible t-test Blundell-Bond: actual significance level

| $\bar{\rho}_{x\varepsilon} = 0.0$ | | | $\theta = 0$ | | | | | | | $\theta = 1$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $L$ | $\gamma$ | BB1 | BB1aR | BB1cR | BB2a | BB2aW | BB2c | BB2cW | BB1 | BB1aR | BB1cR | BB2a | BB2aW | BB2c | BB2cW | MBB |
| $T=3$ | 9 | 0.20 | 0.085 | 0.087 | 0.079 | 0.098 | 0.070 | 0.076 | 0.057 | 0.118 | 0.087 | 0.068 | 0.132 | 0.080 | 0.085 | 0.062 | 0.551 |
| | | 0.50 | 0.088 | 0.089 | 0.077 | 0.118 | 0.091 | 0.092 | 0.075 | 0.138 | 0.083 | 0.063 | 0.146 | 0.093 | 0.097 | 0.073 | 0.497 |
| | | 0.80 | 0.062 | 0.064 | 0.039 | 0.120 | 0.091 | 0.089 | 0.064 | 0.144 | 0.063 | 0.034 | 0.141 | 0.088 | 0.088 | 0.065 | 0.502 |
| $T=6$ | 15 | 0.20 | 0.055 | 0.063 | 0.051 | 0.077 | 0.052 | 0.046 | 0.042 | 0.124 | 0.071 | 0.048 | 0.127 | 0.062 | 0.052 | 0.047 | 0.170 |
| | | 0.50 | 0.057 | 0.063 | 0.050 | 0.083 | 0.056 | 0.051 | 0.041 | 0.141 | 0.065 | 0.045 | 0.133 | 0.065 | 0.059 | 0.049 | 0.150 |
| | | 0.80 | 0.049 | 0.053 | 0.037 | 0.092 | 0.063 | 0.058 | 0.044 | 0.161 | 0.065 | 0.038 | 0.138 | 0.069 | 0.058 | 0.045 | 0.181 |
| $T=9$ | 21 | 0.20 | 0.053 | 0.059 | 0.050 | 0.089 | 0.055 | 0.045 | 0.044 | 0.138 | 0.064 | 0.047 | 0.150 | 0.062 | 0.052 | 0.049 | 0.106 |
| | | 0.50 | 0.053 | 0.060 | 0.048 | 0.085 | 0.054 | 0.048 | 0.044 | 0.155 | 0.065 | 0.045 | 0.153 | 0.060 | 0.051 | 0.047 | 0.104 |
| | | 0.80 | 0.050 | 0.055 | 0.040 | 0.092 | 0.058 | 0.052 | 0.043 | 0.175 | 0.068 | 0.042 | 0.157 | 0.061 | 0.053 | 0.043 | 0.106 |
| | $L$ | $\beta$ | BB1 | BB1aR | BB1cR | BB2a | BB2aW | BB2c | BB2cW | BB1 | BB1aR | BB1cR | BB2a | BB2aW | BB2c | BB2cW | MBB |
| $T=3$ | 9 | 1.43 | 0.051 | 0.053 | 0.050 | 0.064 | 0.052 | 0.051 | 0.053 | 0.176 | 0.068 | 0.058 | 0.097 | 0.065 | 0.065 | 0.067 | 0.591 |
| | | 0.93 | 0.049 | 0.052 | 0.049 | 0.065 | 0.054 | 0.052 | 0.055 | 0.191 | 0.066 | 0.056 | 0.097 | 0.066 | 0.066 | 0.069 | 0.580 |
| | | 0.31 | 0.048 | 0.051 | 0.048 | 0.064 | 0.053 | 0.051 | 0.055 | 0.208 | 0.069 | 0.057 | 0.103 | 0.068 | 0.067 | 0.068 | 0.597 |
| $T=6$ | 15 | 1.43 | 0.050 | 0.052 | 0.050 | 0.069 | 0.049 | 0.050 | 0.051 | 0.196 | 0.065 | 0.055 | 0.119 | 0.061 | 0.058 | 0.061 | 0.216 |
| | | 0.93 | 0.049 | 0.052 | 0.048 | 0.071 | 0.051 | 0.050 | 0.053 | 0.205 | 0.064 | 0.053 | 0.119 | 0.062 | 0.059 | 0.061 | 0.218 |
| | | 0.31 | 0.050 | 0.052 | 0.049 | 0.070 | 0.053 | 0.051 | 0.053 | 0.214 | 0.067 | 0.053 | 0.124 | 0.063 | 0.059 | 0.063 | 0.218 |
| $T=9$ | 21 | 1.43 | 0.052 | 0.055 | 0.052 | 0.084 | 0.053 | 0.051 | 0.053 | 0.200 | 0.063 | 0.053 | 0.139 | 0.061 | 0.056 | 0.059 | 0.142 |
| | | 0.93 | 0.051 | 0.054 | 0.051 | 0.081 | 0.051 | 0.053 | 0.054 | 0.207 | 0.064 | 0.054 | 0.141 | 0.061 | 0.055 | 0.059 | 0.141 |
| | | 0.31 | 0.052 | 0.053 | 0.052 | 0.079 | 0.053 | 0.052 | 0.054 | 0.217 | 0.063 | 0.054 | 0.141 | 0.060 | 0.054 | 0.056 | 0.141 |

*$R = 10000$ simulation replications. Design parameter values: $N = 200$, $SNR = 3$, $DEN_y = 4.0$, $EVF_x = 0.0$, $\bar{\rho}_{x\varepsilon} = 0.0$, $\xi = 0.8$, $\kappa = 0.00$, $\sigma_\varepsilon = 1$, $q = 1$, $\phi = 1.0$. These yield the DGP parameter values: $\pi_\lambda = 0.00$, $\pi_\eta = 0.00$, $\sigma_v = 0.60$, $\sigma_\eta = 4.0 * (1 - \gamma)$, $\rho_{v\varepsilon} = 0.0$ (and $\bar{\rho}_{x\eta} = 0.00$, $\bar{\rho}_{x\lambda} = 0.00$).

absolute values for the standard deviations of the $\beta$ estimates than for P0-XA, but they are almost similar in relative terms. The patterns in the rejection probabilities under the respective null hypotheses are hardly affected, and P2-XC shows again improved behavior of the test statistics due to reduced estimator bias, whereas the RMSE values have slightly increased.

In P3 we change $IEF_x^\eta$ from 0 to 0.3, while keeping $EVF_x = 0.6$ (hence, realizing now dependence between regressor $x_{it}$ and the individual effect $\eta_i$). Comparing the results for P3-XA with those for P2-XA (which have the same $\beta$ values) we find that all patterns are pretty similar. Also P3-XC follows the P2-XC picture closely.

P4 differs from P3 because $\kappa = 0.25$, thus now the heteroskedasticity is determined by $\eta_i$ too. This has a noteworthy effect on MBB estimation and a minor effect on JBB (and thus on JES) testing only.

P5 differs from P0 just in having $\bar{\rho}_{x\varepsilon} = 0.3$, so $x_{it}$ is now endogenous with respect to $\varepsilon_{it}$. P5-EA uses all instruments available when correctly taking the endogeneity into account. This leads to very unsatisfactory results. The coefficient estimates of $\gamma$ have serious negative bias, and those for $\beta$ positive bias, whereas the standard deviation

**Table 12.** P1ft-XC*

Feasible Sargan-Hansen test: rejection probability

| $\bar\rho_{x\varepsilon}=0.0$ | df | | | | $\theta=0$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | AB | BB | Inc | $\gamma$ | $JAB_a^{(2,1)}$ | $JBB_a^{(2,1)}$ | $JES_a^{(2,1)}$ | $JAB_c^{(2,1)}$ | $JBB_c^{(2,1)}$ | $JES_c^{(2,1)}$ | $JMAB$ | $JMBB$ | $JESM$ |
| $T=3$ | 4 | 6 | 2 | 0.20 | 0.043 | 0.053 | 0.071 | 0.045 | 0.054 | 0.067 | 0.281 | 0.255 | 0.249 |
| | | | | 0.50 | 0.047 | 0.048 | 0.063 | 0.049 | 0.047 | 0.055 | 0.280 | 0.249 | 0.245 |
| | | | | 0.80 | 0.054 | 0.045 | 0.052 | 0.056 | 0.040 | 0.036 | 0.282 | 0.301 | 0.294 |
| $T=6$ | 10 | 12 | 2 | 0.20 | 0.043 | 0.046 | 0.059 | 0.043 | 0.049 | 0.061 | 0.053 | 0.291 | 0.367 |
| | | | | 0.50 | 0.045 | 0.048 | 0.057 | 0.044 | 0.050 | 0.058 | 0.051 | 0.288 | 0.364 |
| | | | | 0.80 | 0.048 | 0.047 | 0.048 | 0.049 | 0.046 | 0.044 | 0.055 | 0.295 | 0.369 |
| $T=9$ | 16 | 18 | 2 | 0.20 | 0.041 | 0.043 | 0.056 | 0.040 | 0.044 | 0.057 | 0.045 | 0.289 | 0.388 |
| | | | | 0.50 | 0.041 | 0.043 | 0.055 | 0.041 | 0.045 | 0.058 | 0.043 | 0.290 | 0.389 |
| | | | | 0.80 | 0.045 | 0.042 | 0.049 | 0.045 | 0.045 | 0.050 | 0.045 | 0.296 | 0.391 |

| $\bar\rho_{x\varepsilon}=0.0$ | df | | | | $\theta=1$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | AB | BB | Inc | $\gamma$ | $JAB_a^{(2,1)}$ | $JBB_a^{(2,1)}$ | $JES_a^{(2,1)}$ | $JAB_c^{(2,1)}$ | $JBB_c^{(2,1)}$ | $JES_c^{(2,1)}$ | $JMAB$ | $JMBB$ | $JESM$ |
| $T=3$ | 4 | 6 | 2 | 0.20 | 0.039 | 0.043 | 0.061 | 0.043 | 0.047 | 0.054 | 0.305 | 0.432 | 0.418 |
| | | | | 0.50 | 0.042 | 0.043 | 0.055 | 0.048 | 0.042 | 0.046 | 0.313 | 0.438 | 0.425 |
| | | | | 0.80 | 0.053 | 0.036 | 0.042 | 0.059 | 0.038 | 0.036 | 0.321 | 0.493 | 0.475 |
| $T=6$ | 10 | 12 | 2 | 0.20 | 0.035 | 0.036 | 0.055 | 0.039 | 0.041 | 0.053 | 0.057 | 0.594 | 0.639 |
| | | | | 0.50 | 0.034 | 0.035 | 0.052 | 0.039 | 0.042 | 0.054 | 0.056 | 0.597 | 0.644 |
| | | | | 0.80 | 0.043 | 0.035 | 0.044 | 0.045 | 0.042 | 0.044 | 0.064 | 0.613 | 0.661 |
| $T=9$ | 16 | 18 | 2 | 0.20 | 0.029 | 0.029 | 0.056 | 0.037 | 0.039 | 0.054 | 0.048 | 0.646 | 0.696 |
| | | | | 0.50 | 0.029 | 0.028 | 0.053 | 0.037 | 0.039 | 0.054 | 0.048 | 0.647 | 0.697 |
| | | | | 0.80 | 0.031 | 0.026 | 0.049 | 0.040 | 0.039 | 0.046 | 0.053 | 0.658 | 0.703 |

*$R=10000$ simulation replications. Design parameter values: $N=200$, $SNR=3$, $DEN_y=4.0$, $EVF_x=0.0$, $\bar\rho_{x\varepsilon}=0.0$, $\xi=0.8$, $\kappa=0.00$, $\sigma_\varepsilon=1$, $q=1$, $\phi=1.0$. These yield the DGP parameter values: $\pi_\lambda=0.00$, $\pi_\eta=0.00$, $\sigma_v=0.60$, $\sigma_\eta=4.0*(1-\gamma)$, $\rho_{v\varepsilon}=0.0$ (and $\bar\rho_{x\eta}=0.00$, $\bar\rho_{x\lambda}=0.00$).

is slightly larger than for P0-EA, which are substantially larger than for P0-XA. All coefficient tests are very seriously oversized, also after a Windmeijer correction, both for AB and BB. Tests $JABu$ and $JBBu$ show underrejection, whereas the matching $JES$ tests show serious overrejection when $T$ is large, but the feasible 2-step variants are not all that bad. From Tables 13, 14 and 15 (P5f$j$-EC, $j$ = c,t,J) we see that most results which correctly handle the simultaneity of $x_{it}$ are still bad after collapsing, especially for $T$ small (where collapsing can only lead to a minor reduction of the instrument set), although not as bad as those for P5-EA and larger values of $T$. For P5-EC the rejection probabilities of the corrected coefficient tests are usually in the 10-20% range, but those of the 2-step $J$ tests are often close to 5%. Both AB and BB are inconsistent when treating $x_{it}$ either as predetermined or as exogenous. For P5-WA and P5-XA the coefficient bias is almost similar but much more serious than for P5-EA. For the inconsistent estimators the bias does not reduce when collapsing the instruments. Because the inconsistent estimators have a much smaller standard deviation than the consistent estimators practitioners should be warned never to select an estimator simply because of its attractive estimated standard error. The consistency of AB and BB should be tested with the Sargan-Hansen test. In this study we did not examine the particular incremental test which focusses on the validity of the extra instruments when comparing E with W or E with X. Here we just examine the rejection probabilities of the overall overidentification $J$ tests for case P5 using all instruments and can compare the rejection frequencies when treating $x_{it}$ correctly as endogenous, or incorrectly as either predetermined or exogenous.

From the Tables 16, 17 and 18 (P5fJ-$j$A, $j$ = E,W,X) we find that the detection of inconsistency in this way has often a higher probability when the null hypothesis is W than when it is X. The probability generally increases with $T$ and with $\gamma$ and is often better for the c variant than for the a variant and slightly better for BB implementations than for AB implementations, whereas in general heteroskedasticity mitigates the rejection probability. In the situation where all instruments have been collapsed, where we already established that the $J$ tests do have reasonable size control, we find the fol-

## Table 13. P5fc-EC*

**Feasible coefficient estimators for Arellano-Bond**

$\bar\rho_{x\varepsilon}=0.3$

### $\theta = 0$

| T | L | $\gamma$ | ABI Bias | ABI Stdv | ABI RMSE | AB2a Bias | AB2a Stdv | AB2a RMSE | AB2c Bias | AB2c Stdv | AB2c RMSE |
|---|---|---|---|---|---|---|---|---|---|---|---|
| T=3 | 4 | 0.20 | -0.117 | 0.266 | 0.291 | -0.114 | 0.266 | 0.290 | -0.117 | 0.269 | 0.293 |
| | | 0.50 | -0.115 | 0.291 | 0.313 | -0.115 | 0.295 | 0.317 | -0.117 | 0.289 | 0.312 |
| | | 0.80 | -0.082 | 0.300 | 0.311 | -0.081 | 0.304 | 0.315 | -0.083 | 0.298 | 0.310 |
| T=6 | 10 | 0.20 | -0.054 | 0.075 | 0.092 | -0.054 | 0.078 | 0.095 | -0.055 | 0.076 | 0.094 |
| | | 0.50 | -0.053 | 0.081 | 0.097 | -0.054 | 0.085 | 0.101 | -0.052 | 0.082 | 0.097 |
| | | 0.80 | -0.042 | 0.086 | 0.096 | -0.042 | 0.090 | 0.099 | -0.041 | 0.087 | 0.096 |
| T=9 | 16 | 0.20 | -0.037 | 0.048 | 0.061 | -0.037 | 0.051 | 0.063 | -0.038 | 0.049 | 0.062 |
| | | 0.50 | -0.037 | 0.053 | 0.065 | -0.037 | 0.056 | 0.067 | -0.037 | 0.053 | 0.065 |
| | | 0.80 | -0.028 | 0.054 | 0.061 | -0.028 | 0.057 | 0.064 | -0.027 | 0.054 | 0.061 |
| | L | $\beta$ | | | | | | | | | |
| T=3 | 4 | 1.43 | 0.559 | 1.325 | 1.438 | 0.547 | 1.325 | 1.433 | 0.559 | 1.336 | 1.449 |
| | | 0.93 | 0.374 | 0.992 | 1.060 | 0.369 | 1.002 | 1.067 | 0.381 | 0.982 | 1.053 |
| | | 0.31 | 0.170 | 0.687 | 0.708 | 0.165 | 0.690 | 0.710 | 0.177 | 0.677 | 0.700 |
| T=6 | 10 | 1.43 | 0.230 | 0.315 | 0.389 | 0.234 | 0.330 | 0.405 | 0.229 | 0.320 | 0.393 |
| | | 0.93 | 0.168 | 0.257 | 0.307 | 0.173 | 0.268 | 0.319 | 0.161 | 0.260 | 0.306 |
| | | 0.31 | 0.085 | 0.189 | 0.207 | 0.088 | 0.194 | 0.214 | 0.081 | 0.189 | 0.206 |
| T=9 | 16 | 1.43 | 0.147 | 0.186 | 0.237 | 0.150 | 0.196 | 0.246 | 0.148 | 0.189 | 0.240 |
| | | 0.93 | 0.116 | 0.159 | 0.196 | 0.117 | 0.167 | 0.204 | 0.111 | 0.160 | 0.195 |
| | | 0.31 | 0.059 | 0.116 | 0.131 | 0.061 | 0.122 | 0.136 | 0.057 | 0.117 | 0.130 |

### $\theta = 1$

| T | L | $\gamma$ | ABI Bias | ABI Stdv | ABI RMSE | AB2a Bias | AB2a Stdv | AB2a RMSE | AB2c Bias | AB2c Stdv | AB2c RMSE | MAB Bias | MAB Stdv | MAB RMSE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| T=3 | 4 | 0.20 | -0.165 | 0.297 | 0.340 | -0.169 | 0.298 | 0.343 | -0.169 | 0.291 | 0.337 | -0.178 | 0.287 | 0.338 |
| | | 0.50 | -0.191 | 0.379 | 0.424 | -0.197 | 0.372 | 0.421 | -0.195 | 0.360 | 0.409 | -0.204 | 0.357 | 0.412 |
| | | 0.80 | -0.181 | 0.477 | 0.511 | -0.195 | 0.474 | 0.513 | -0.197 | 0.453 | 0.494 | -0.197 | 0.443 | 0.485 |
| T=6 | 10 | 0.20 | -0.097 | 0.101 | 0.140 | -0.090 | 0.101 | 0.135 | -0.099 | 0.099 | 0.141 | -0.098 | 0.100 | 0.140 |
| | | 0.50 | -0.104 | 0.111 | 0.152 | -0.092 | 0.112 | 0.144 | -0.096 | 0.108 | 0.144 | -0.093 | 0.107 | 0.142 |
| | | 0.80 | -0.091 | 0.126 | 0.155 | -0.078 | 0.123 | 0.146 | -0.083 | 0.121 | 0.147 | -0.068 | 0.110 | 0.130 |
| T=9 | 16 | 0.20 | -0.073 | 0.070 | 0.101 | -0.062 | 0.066 | 0.091 | -0.073 | 0.068 | 0.100 | -0.060 | 0.062 | 0.086 |
| | | 0.50 | -0.077 | 0.075 | 0.108 | -0.063 | 0.072 | 0.096 | -0.069 | 0.072 | 0.100 | -0.056 | 0.065 | 0.086 |
| | | 0.80 | -0.064 | 0.080 | 0.102 | -0.050 | 0.076 | 0.091 | -0.055 | 0.076 | 0.094 | -0.039 | 0.063 | 0.074 |
| | L | $\beta$ | | | | | | | | | | | | |
| T=3 | 4 | 1.43 | 0.790 | 1.455 | 1.655 | 0.810 | 1.451 | 1.661 | 0.804 | 1.420 | 1.632 | 0.823 | 1.362 | 1.591 |
| | | 0.93 | 0.613 | 1.322 | 1.457 | 0.638 | 1.286 | 1.436 | 0.644 | 1.242 | 1.399 | 0.631 | 1.195 | 1.351 |
| | | 0.31 | 0.362 | 1.109 | 1.166 | 0.394 | 1.100 | 1.168 | 0.418 | 1.068 | 1.146 | 0.383 | 0.989 | 1.061 |
| T=6 | 10 | 1.43 | 0.409 | 0.432 | 0.595 | 0.399 | 0.437 | 0.591 | 0.420 | 0.423 | 0.596 | 0.381 | 0.395 | 0.548 |
| | | 0.93 | 0.324 | 0.362 | 0.486 | 0.307 | 0.364 | 0.476 | 0.314 | 0.351 | 0.471 | 0.264 | 0.309 | 0.406 |
| | | 0.31 | 0.178 | 0.283 | 0.334 | 0.170 | 0.276 | 0.324 | 0.179 | 0.273 | 0.326 | 0.121 | 0.216 | 0.247 |
| T=9 | 16 | 1.43 | 0.286 | 0.270 | 0.394 | 0.255 | 0.261 | 0.365 | 0.289 | 0.262 | 0.390 | 0.213 | 0.218 | 0.305 |
| | | 0.93 | 0.235 | 0.235 | 0.332 | 0.205 | 0.224 | 0.304 | 0.222 | 0.224 | 0.315 | 0.157 | 0.179 | 0.238 |
| | | 0.31 | 0.129 | 0.179 | 0.221 | 0.112 | 0.167 | 0.201 | 0.123 | 0.171 | 0.211 | 0.070 | 0.121 | 0.140 |

**Feasible coefficient estimators for Blundell-Bond**

$\bar\rho_{x\varepsilon}=0.3$

### $\theta = 0$

| T | L | $\gamma$ | BBI Bias | BBI Stdv | BBI RMSE | BB2a Bias | BB2a Stdv | BB2a RMSE | BB2c Bias | BB2c Stdv | BB2c RMSE |
|---|---|---|---|---|---|---|---|---|---|---|---|
| T=3 | 7 | 0.20 | -0.066 | 0.135 | 0.151 | -0.066 | 0.143 | 0.157 | -0.069 | 0.143 | 0.159 |
| | | 0.50 | -0.075 | 0.140 | 0.159 | -0.070 | 0.149 | 0.164 | -0.071 | 0.148 | 0.164 |
| | | 0.80 | -0.070 | 0.136 | 0.153 | -0.060 | 0.142 | 0.155 | -0.060 | 0.140 | 0.152 |
| T=6 | 13 | 0.20 | -0.034 | 0.058 | 0.067 | -0.030 | 0.061 | 0.068 | -0.035 | 0.060 | 0.069 |
| | | 0.50 | -0.036 | 0.061 | 0.071 | -0.029 | 0.063 | 0.069 | -0.032 | 0.062 | 0.069 |
| | | 0.80 | -0.030 | 0.061 | 0.068 | -0.023 | 0.061 | 0.065 | -0.024 | 0.059 | 0.064 |
| T=9 | 19 | 0.20 | -0.026 | 0.040 | 0.048 | -0.024 | 0.043 | 0.049 | -0.027 | 0.041 | 0.049 |
| | | 0.50 | -0.027 | 0.043 | 0.050 | -0.022 | 0.044 | 0.050 | -0.024 | 0.043 | 0.049 |
| | | 0.80 | -0.021 | 0.042 | 0.047 | -0.016 | 0.043 | 0.045 | -0.017 | 0.041 | 0.045 |
| | L | $\beta$ | | | | | | | | | |
| T=3 | 7 | 1.43 | 0.320 | 0.631 | 0.708 | 0.317 | 0.671 | 0.742 | 0.327 | 0.665 | 0.741 |
| | | 0.93 | 0.250 | 0.474 | 0.536 | 0.231 | 0.498 | 0.549 | 0.234 | 0.491 | 0.544 |
| | | 0.31 | 0.153 | 0.333 | 0.367 | 0.131 | 0.339 | 0.363 | 0.133 | 0.334 | 0.360 |
| T=6 | 13 | 1.43 | 0.150 | 0.236 | 0.280 | 0.140 | 0.246 | 0.283 | 0.147 | 0.243 | 0.284 |
| | | 0.93 | 0.122 | 0.195 | 0.230 | 0.108 | 0.199 | 0.227 | 0.106 | 0.196 | 0.223 |
| | | 0.31 | 0.072 | 0.145 | 0.161 | 0.063 | 0.142 | 0.156 | 0.059 | 0.139 | 0.151 |
| T=9 | 19 | 1.43 | 0.108 | 0.154 | 0.188 | 0.103 | 0.162 | 0.192 | 0.107 | 0.157 | 0.190 |
| | | 0.93 | 0.091 | 0.132 | 0.160 | 0.082 | 0.138 | 0.160 | 0.081 | 0.132 | 0.155 |
| | | 0.31 | 0.052 | 0.099 | 0.112 | 0.048 | 0.100 | 0.111 | 0.045 | 0.096 | 0.106 |

### $\theta = 1$

| T | L | $\gamma$ | BBI Bias | BBI Stdv | BBI RMSE | BB2a Bias | BB2a Stdv | BB2a RMSE | BB2c Bias | BB2c Stdv | BB2c RMSE | MBB Bias | MBB Stdv | MBB RMSE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| T=3 | 7 | 0.20 | -0.106 | 0.169 | 0.199 | -0.102 | 0.179 | 0.206 | -0.108 | 0.174 | 0.205 | -0.089 | 0.187 | 0.207 |
| | | 0.50 | -0.126 | 0.187 | 0.225 | -0.117 | 0.195 | 0.228 | -0.120 | 0.193 | 0.227 | -0.105 | 0.193 | 0.227 |
| | | 0.80 | -0.123 | 0.198 | 0.233 | -0.112 | 0.209 | 0.237 | -0.112 | 0.203 | 0.232 | -0.067 | 0.593 | 0.597 |
| T=6 | 13 | 0.20 | -0.066 | 0.080 | 0.104 | -0.054 | 0.078 | 0.095 | -0.064 | 0.079 | 0.102 | -0.050 | 0.078 | 0.093 |
| | | 0.50 | -0.073 | 0.086 | 0.113 | -0.055 | 0.083 | 0.100 | -0.060 | 0.084 | 0.103 | -0.039 | 0.092 | 0.100 |
| | | 0.80 | -0.064 | 0.089 | 0.110 | -0.044 | 0.084 | 0.094 | -0.047 | 0.086 | 0.098 | -0.007 | 0.565 | 0.565 |
| T=9 | 19 | 0.20 | -0.054 | 0.059 | 0.080 | -0.042 | 0.056 | 0.070 | -0.051 | 0.057 | 0.077 | -0.037 | 0.050 | 0.063 |
| | | 0.50 | -0.058 | 0.063 | 0.085 | -0.042 | 0.059 | 0.072 | -0.047 | 0.060 | 0.076 | -0.031 | 0.052 | 0.061 |
| | | 0.80 | -0.048 | 0.063 | 0.079 | -0.031 | 0.057 | 0.065 | -0.034 | 0.059 | 0.068 | -0.012 | 0.057 | 0.058 |
| | L | $\beta$ | | | | | | | | | | | | |
| T=3 | 7 | 1.43 | 0.495 | 0.778 | 0.922 | 0.489 | 0.837 | 0.970 | 0.500 | 0.793 | 0.937 | 0.498 | 0.752 | 0.902 |
| | | 0.93 | 0.419 | 0.643 | 0.768 | 0.403 | 0.675 | 0.786 | 0.413 | 0.650 | 0.770 | 0.468 | 1.303 | 1.384 |
| | | 0.31 | 0.276 | 0.504 | 0.574 | 0.265 | 0.511 | 0.576 | 0.279 | 0.495 | 0.568 | 0.385 | 1.358 | 1.411 |
| T=6 | 13 | 1.43 | 0.284 | 0.321 | 0.429 | 0.247 | 0.318 | 0.403 | 0.276 | 0.315 | 0.419 | 0.219 | 0.297 | 0.369 |
| | | 0.93 | 0.244 | 0.278 | 0.370 | 0.202 | 0.270 | 0.337 | 0.216 | 0.268 | 0.344 | 0.160 | 0.318 | 0.356 |
| | | 0.31 | 0.151 | 0.216 | 0.264 | 0.125 | 0.203 | 0.238 | 0.135 | 0.203 | 0.243 | 0.128 | 1.055 | 1.063 |
| T=9 | 19 | 1.43 | 0.220 | 0.223 | 0.313 | 0.180 | 0.211 | 0.278 | 0.210 | 0.215 | 0.300 | 0.143 | 0.180 | 0.230 |
| | | 0.93 | 0.190 | 0.197 | 0.273 | 0.151 | 0.183 | 0.237 | 0.167 | 0.185 | 0.249 | 0.108 | 0.151 | 0.186 |
| | | 0.31 | 0.115 | 0.150 | 0.189 | 0.092 | 0.135 | 0.163 | 0.103 | 0.138 | 0.172 | 0.065 | 0.137 | 0.152 |

*$R = 10000$ simulation replications. Design parameter values: $N = 200$, $SNR = 3$, $DEN_y = 1.0$, $EVF_x = 0.0$, $\bar\rho_{x\varepsilon} = 0.3$, $\xi = 0.8$, $\kappa = 0.00$, $\sigma_\varepsilon = 1$, $q = 1$, $\phi = 1.0$. These yield the DGP parameter values: $\pi_\lambda = 0.00$, $\pi_\eta = 0.00$, $\sigma_v = 0.60$, $\sigma_\eta = 1.0*(1-\gamma)$, $\rho_{v\varepsilon} = 0.5$ (and $\bar\rho_{x\eta} = 0.00$, $\bar\rho_{x\lambda} = 0.00$).

**Table 14.** P5ft-EC*

Feasible t-test Arellano-Bond: actual significance level

| $\bar{\rho}_{x\varepsilon}=0.3$ | $L$ | $\gamma$ | AB1 | AB1aR | AB1cR | AB2a | AB2aW | AB2c | AB2cW | AB1 | AB1aR | AB1cR | AB2a | AB2aW | AB2c | AB2cW | MAB |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | $\theta=0$ | | | | | | | $\theta=1$ | | | |
| $T=3$ | 4 | 0.20 | 0.118 | 0.129 | 0.112 | 0.134 | 0.134 | 0.115 | 0.116 | 0.252 | 0.155 | 0.118 | 0.175 | 0.164 | 0.134 | 0.134 | 0.706 |
| | | 0.50 | 0.135 | 0.138 | 0.129 | 0.144 | 0.137 | 0.135 | 0.125 | 0.283 | 0.173 | 0.140 | 0.195 | 0.175 | 0.161 | 0.154 | 0.685 |
| | | 0.80 | 0.126 | 0.128 | 0.124 | 0.132 | 0.123 | 0.127 | 0.119 | 0.274 | 0.170 | 0.146 | 0.191 | 0.166 | 0.168 | 0.158 | 0.632 |
| $T=6$ | 10 | 0.20 | 0.154 | 0.161 | 0.137 | 0.181 | 0.148 | 0.145 | 0.140 | 0.420 | 0.250 | 0.196 | 0.316 | 0.203 | 0.225 | 0.211 | 0.400 |
| | | 0.50 | 0.142 | 0.140 | 0.135 | 0.162 | 0.128 | 0.133 | 0.127 | 0.404 | 0.231 | 0.204 | 0.289 | 0.177 | 0.203 | 0.187 | 0.343 |
| | | 0.80 | 0.104 | 0.101 | 0.099 | 0.119 | 0.091 | 0.098 | 0.095 | 0.340 | 0.173 | 0.157 | 0.220 | 0.127 | 0.154 | 0.147 | 0.248 |
| $T=9$ | 16 | 0.20 | 0.152 | 0.152 | 0.132 | 0.188 | 0.141 | 0.139 | 0.136 | 0.453 | 0.258 | 0.218 | 0.359 | 0.201 | 0.237 | 0.224 | 0.297 |
| | | 0.50 | 0.141 | 0.138 | 0.130 | 0.173 | 0.124 | 0.129 | 0.124 | 0.440 | 0.245 | 0.215 | 0.327 | 0.177 | 0.206 | 0.194 | 0.261 |
| | | 0.80 | 0.103 | 0.102 | 0.097 | 0.130 | 0.089 | 0.095 | 0.094 | 0.353 | 0.175 | 0.161 | 0.247 | 0.118 | 0.145 | 0.140 | 0.181 |

| | $L$ | $\beta$ | AB1 | AB1aR | AB1cR | AB2a | AB2aW | AB2c | AB2cW | AB1 | AB1aR | AB1cR | AB2a | AB2aW | AB2c | AB2cW | MAB |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $T=3$ | 4 | 1.43 | 0.115 | 0.127 | 0.112 | 0.130 | 0.131 | 0.115 | 0.112 | 0.253 | 0.154 | 0.114 | 0.175 | 0.167 | 0.132 | 0.128 | 0.788 |
| | | 0.93 | 0.130 | 0.141 | 0.129 | 0.142 | 0.141 | 0.130 | 0.122 | 0.274 | 0.173 | 0.135 | 0.196 | 0.183 | 0.155 | 0.147 | 0.758 |
| | | 0.31 | 0.110 | 0.118 | 0.109 | 0.121 | 0.118 | 0.110 | 0.107 | 0.259 | 0.154 | 0.130 | 0.176 | 0.159 | 0.148 | 0.139 | 0.701 |
| $T=6$ | 10 | 1.43 | 0.165 | 0.166 | 0.157 | 0.191 | 0.153 | 0.163 | 0.155 | 0.429 | 0.255 | 0.217 | 0.336 | 0.219 | 0.243 | 0.227 | 0.445 |
| | | 0.93 | 0.141 | 0.141 | 0.138 | 0.166 | 0.134 | 0.136 | 0.131 | 0.413 | 0.234 | 0.207 | 0.305 | 0.195 | 0.218 | 0.202 | 0.409 |
| | | 0.31 | 0.093 | 0.096 | 0.089 | 0.114 | 0.096 | 0.090 | 0.089 | 0.319 | 0.154 | 0.136 | 0.213 | 0.138 | 0.149 | 0.140 | 0.308 |
| $T=9$ | 16 | 1.43 | 0.156 | 0.158 | 0.148 | 0.193 | 0.144 | 0.151 | 0.145 | 0.463 | 0.270 | 0.238 | 0.375 | 0.215 | 0.262 | 0.248 | 0.324 |
| | | 0.93 | 0.142 | 0.141 | 0.139 | 0.179 | 0.132 | 0.137 | 0.132 | 0.443 | 0.244 | 0.226 | 0.350 | 0.193 | 0.226 | 0.214 | 0.302 |
| | | 0.31 | 0.096 | 0.098 | 0.096 | 0.126 | 0.097 | 0.094 | 0.093 | 0.348 | 0.161 | 0.147 | 0.249 | 0.129 | 0.149 | 0.145 | 0.226 |

Feasible t-test Blundell-Bond: actual significance level

| $\bar{\rho}_{x\varepsilon}=0.3$ | $L$ | $\gamma$ | BB1 | BB1aR | BB1cR | BB2a | BB2aW | BB2c | BB2cW | BB1 | BB1aR | BB1cR | BB2a | BB2aW | BB2c | BB2cW | MBB |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | $\theta=0$ | | | | | | | $\theta=1$ | | | |
| $T=3$ | 7 | 0.20 | 0.100 | 0.101 | 0.083 | 0.126 | 0.112 | 0.109 | 0.146 | 0.263 | 0.149 | 0.113 | 0.198 | 0.147 | 0.146 | 0.154 | 0.640 |
| | | 0.50 | 0.116 | 0.112 | 0.100 | 0.127 | 0.104 | 0.113 | 0.132 | 0.301 | 0.166 | 0.142 | 0.210 | 0.150 | 0.165 | 0.168 | 0.622 |
| | | 0.80 | 0.104 | 0.098 | 0.091 | 0.103 | 0.079 | 0.089 | 0.089 | 0.297 | 0.145 | 0.135 | 0.188 | 0.127 | 0.147 | 0.139 | 0.587 |
| $T=6$ | 13 | 0.20 | 0.111 | 0.115 | 0.086 | 0.130 | 0.098 | 0.099 | 0.108 | 0.359 | 0.195 | 0.140 | 0.247 | 0.141 | 0.159 | 0.160 | 0.265 |
| | | 0.50 | 0.113 | 0.111 | 0.098 | 0.116 | 0.087 | 0.092 | 0.101 | 0.367 | 0.193 | 0.163 | 0.236 | 0.129 | 0.149 | 0.154 | 0.240 |
| | | 0.80 | 0.094 | 0.092 | 0.083 | 0.096 | 0.073 | 0.074 | 0.075 | 0.332 | 0.153 | 0.130 | 0.185 | 0.098 | 0.109 | 0.110 | 0.205 |
| $T=9$ | 19 | 0.20 | 0.115 | 0.118 | 0.095 | 0.145 | 0.100 | 0.102 | 0.104 | 0.399 | 0.215 | 0.165 | 0.294 | 0.149 | 0.169 | 0.168 | 0.206 |
| | | 0.50 | 0.115 | 0.115 | 0.101 | 0.132 | 0.091 | 0.092 | 0.098 | 0.405 | 0.209 | 0.176 | 0.279 | 0.135 | 0.152 | 0.155 | 0.189 |
| | | 0.80 | 0.095 | 0.093 | 0.085 | 0.105 | 0.070 | 0.075 | 0.077 | 0.339 | 0.155 | 0.138 | 0.216 | 0.098 | 0.102 | 0.105 | 0.140 |

| | $L$ | $\beta$ | BB1 | BB1aR | BB1cR | BB2a | BB2aW | BB2c | BB2cW | BB1 | BB1aR | BB1cR | BB2a | BB2aW | BB2c | BB2cW | MBB |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $T=3$ | 7 | 1.43 | 0.097 | 0.095 | 0.086 | 0.122 | 0.101 | 0.110 | 0.198 | 0.265 | 0.151 | 0.130 | 0.205 | 0.152 | 0.162 | 0.216 | 0.709 |
| | | 0.93 | 0.111 | 0.118 | 0.109 | 0.126 | 0.102 | 0.115 | 0.172 | 0.304 | 0.180 | 0.164 | 0.223 | 0.160 | 0.185 | 0.221 | 0.709 |
| | | 0.31 | 0.101 | 0.111 | 0.098 | 0.111 | 0.089 | 0.092 | 0.131 | 0.288 | 0.158 | 0.151 | 0.195 | 0.140 | 0.163 | 0.197 | 0.744 |
| $T=6$ | 13 | 1.43 | 0.123 | 0.128 | 0.112 | 0.142 | 0.109 | 0.117 | 0.142 | 0.381 | 0.215 | 0.179 | 0.280 | 0.165 | 0.194 | 0.223 | 0.330 |
| | | 0.93 | 0.118 | 0.120 | 0.115 | 0.131 | 0.100 | 0.108 | 0.124 | 0.387 | 0.213 | 0.192 | 0.266 | 0.157 | 0.185 | 0.212 | 0.319 |
| | | 0.31 | 0.089 | 0.094 | 0.085 | 0.107 | 0.085 | 0.079 | 0.091 | 0.330 | 0.165 | 0.146 | 0.226 | 0.135 | 0.146 | 0.163 | 0.319 |
| $T=9$ | 19 | 1.43 | 0.127 | 0.127 | 0.117 | 0.158 | 0.113 | 0.121 | 0.126 | 0.423 | 0.239 | 0.203 | 0.323 | 0.173 | 0.206 | 0.213 | 0.252 |
| | | 0.93 | 0.124 | 0.126 | 0.121 | 0.146 | 0.107 | 0.112 | 0.121 | 0.425 | 0.228 | 0.211 | 0.318 | 0.166 | 0.196 | 0.213 | 0.246 |
| | | 0.31 | 0.095 | 0.097 | 0.094 | 0.127 | 0.094 | 0.086 | 0.093 | 0.355 | 0.167 | 0.153 | 0.262 | 0.135 | 0.150 | 0.165 | 0.226 |

*$R = 10000$ simulation replications. Design parameter values: $N = 200$, $SNR = 3$, $DEN_y = 1.0$, $EVF_x = 0.0$, $\bar{\rho}_{x\varepsilon} = 0.3$, $\xi = 0.8$, $\kappa = 0.00$, $\sigma_\varepsilon = 1$, $q = 1$, $\phi = 1.0$. These yield the DGP parameter values: $\pi_\lambda = 0.00$, $\pi_\eta = 0.00$, $\sigma_v = 0.60$, $\sigma_\eta = 1.0*(1-\gamma)$, $\rho_{v\varepsilon} = 0.5$ (and $\bar{\rho}_{x\eta} = 0.00$, $\bar{\rho}_{x\lambda} = 0.00$).

lowing. For $T = 3$ and $\gamma = 0.2$ the rejection probability of the $JAB$ and $JBB$ tests does not rise very much when $\bar{\rho}_{x\varepsilon}$ moves from 0 to 0.3, whereas for $T = 9$, $\bar{\rho}_{x\varepsilon} = 0.3$ this rejection probability is often larger than 0.7 when $\gamma \geq 0.5$ and often larger than 0.9 for $\gamma = 0.8$. Hence, only for particular $T$, $\gamma$ and $\theta$ parametrizations the probability to detect inconsistency seems reasonable, whereas the major consequence of inconsistency, which is serious estimator bias, is relatively invariant regarding $T$, $\gamma$ and $\theta$.

Summarizing our results for effect stationary models we note the following. We established that finite sample inaccuracies of the asymptotic techniques seriously aggravate when either $\sigma_\eta/(1-\gamma) \gg \sigma_\varepsilon$ or under simultaneity. For both problems it helps to collapse instruments, and the first problem is mitigated and the second problem detected with higher probability by instrumenting according to W rather than X. Neglected simultaneity leads to seemingly accurate but seriously biased coefficient estimators, whereas asymptotically valid inference on simultaneous dynamic relationships if often not very accurate either. Even when the more efficient BB estimator is used with Windmeijer corrected standard errors, the bias in both $\gamma$ and $\beta$ is very substantial and test sizes are seriously distorted. Some further pilot simulations disclosed that $N$ should be much and

**Table 15.** P5fJ-EC*

Feasible Sargan-Hansen test: rejection probability

| $\bar{\rho}_{x\varepsilon}=0.3$ | df | | | $\gamma$ | $\theta=0$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | AB | BB | Inc | | $JAB_a^{(2,1)}$ | $JBB_a^{(2,1)}$ | $JES_a^{(2,1)}$ | $JAB_c^{(2,1)}$ | $JBB_c^{(2,1)}$ | $JES_c^{(2,1)}$ | $JMAB$ | $JMBB$ | $JESM$ |
| $T=3$ | 2 | 4 | 2 | 0.20 | 0.047 | 0.046 | 0.060 | 0.044 | 0.041 | 0.056 | 0.316 | 0.247 | 0.235 |
| | | | | 0.50 | 0.060 | 0.055 | 0.059 | 0.054 | 0.048 | 0.055 | 0.319 | 0.312 | 0.293 |
| | | | | 0.80 | 0.074 | 0.063 | 0.059 | 0.071 | 0.057 | 0.051 | 0.305 | 0.406 | 0.385 |
| $T=6$ | 8 | 10 | 2 | 0.20 | 0.069 | 0.065 | 0.058 | 0.057 | 0.054 | 0.053 | 0.065 | 0.295 | 0.368 |
| | | | | 0.50 | 0.071 | 0.064 | 0.058 | 0.061 | 0.054 | 0.048 | 0.065 | 0.308 | 0.383 |
| | | | | 0.80 | 0.065 | 0.061 | 0.054 | 0.053 | 0.053 | 0.051 | 0.057 | 0.354 | 0.432 |
| $T=9$ | 14 | 16 | 2 | 0.20 | 0.065 | 0.060 | 0.053 | 0.052 | 0.051 | 0.048 | 0.059 | 0.302 | 0.400 |
| | | | | 0.50 | 0.065 | 0.059 | 0.056 | 0.050 | 0.050 | 0.051 | 0.052 | 0.308 | 0.405 |
| | | | | 0.80 | 0.054 | 0.053 | 0.054 | 0.045 | 0.046 | 0.051 | 0.046 | 0.317 | 0.413 |
| $\bar{\rho}_{x\varepsilon}=0.3$ | df | | | $\gamma$ | $\theta=1$ | | | | | | | | |
| | AB | BB | Inc | | $JAB_a^{(2,1)}$ | $JBB_a^{(2,1)}$ | $JES_a^{(2,1)}$ | $JAB_c^{(2,1)}$ | $JBB_c^{(2,1)}$ | $JES_c^{(2,1)}$ | $JMAB$ | $JMBB$ | $JESM$ |
| $T=3$ | 2 | 4 | 2 | 0.20 | 0.036 | 0.042 | 0.068 | 0.032 | 0.045 | 0.072 | 0.339 | 0.454 | 0.422 |
| | | | | 0.50 | 0.044 | 0.050 | 0.074 | 0.038 | 0.055 | 0.077 | 0.344 | 0.505 | 0.475 |
| | | | | 0.80 | 0.067 | 0.057 | 0.069 | 0.061 | 0.064 | 0.074 | 0.345 | 0.566 | 0.530 |
| $T=6$ | 8 | 10 | 2 | 0.20 | 0.065 | 0.064 | 0.072 | 0.056 | 0.057 | 0.074 | 0.090 | 0.649 | 0.691 |
| | | | | 0.50 | 0.071 | 0.066 | 0.067 | 0.066 | 0.063 | 0.070 | 0.088 | 0.663 | 0.704 |
| | | | | 0.80 | 0.060 | 0.058 | 0.064 | 0.063 | 0.063 | 0.065 | 0.071 | 0.671 | 0.711 |
| $T=9$ | 14 | 16 | 2 | 0.20 | 0.063 | 0.058 | 0.063 | 0.053 | 0.058 | 0.072 | 0.065 | 0.701 | 0.748 |
| | | | | 0.50 | 0.063 | 0.056 | 0.066 | 0.060 | 0.062 | 0.072 | 0.060 | 0.707 | 0.750 |
| | | | | 0.80 | 0.053 | 0.048 | 0.061 | 0.053 | 0.056 | 0.066 | 0.051 | 0.710 | 0.753 |

*$R=10000$ simulation replications. Design parameter values: $N=200$, $SNR=3$, $DEN_y=1.0$, $EVF_x=0.0$, $\bar{\rho}_{x\varepsilon}=0.3$, $\xi=0.8$, $\kappa=0.00$, $\sigma_\varepsilon=1$, $q=1$, $\phi=1.0$. These yield the DGP parameter values: $\pi_\lambda=0.00$, $\pi_\eta=0.00$, $\sigma_v=0.60$, $\sigma_\eta=1.0*(1-\gamma)$, $\rho_{v\varepsilon}=0.5$ (and $\bar{\rho}_{x\eta}=0.00$, $\bar{\rho}_{x\lambda}=0.00$).

much larger than 200 in order to find much more reasonable asymptotic approximation errors.

## 5.2 Nonstationarity

Next we examine the effects of a value of $\phi$ different from unity while $s=-50$. We will just consider setting $\phi=0.5$ and perturbing $x_{i0}$ and $y_{i0}$ according to (4.25), so that their dependence on the effects is initially 50% away from stationarity so that BB estimation is inconsistent. That this occurred we will indicate in the parametrization code by changing P into $P^\phi$. Comparing the results for $P^\phi 0$-XA with those for P0-XA, where $\phi=1$ (effect stationarity), we note from Table 19 ($P^\phi 0$fc-XA) a rather moderate positive bias in the BB estimators for both $\gamma$ and $\beta$ when both $T$ and $\gamma$ are small. Despite the inconsistency of BB the bias is very mild for larger $T$ and for smaller $\gamma$. This can be explained, because convergence towards effect stationarity does occur when time proceeds, while this convergence is faster for smaller $\gamma$. In fact, we note that the RMSE of inconsistent BB1, BB2a and BB2c is often smaller than that for consistent AB1, AB2a and AB2c when $T$ and $\gamma$ are not both small. With respect to the AB estimators we find little to no difference compared to the results under stationarity. Table 20 ($P^\phi 0$ft-XA) shows that when $\gamma=0.8$ the BB2cW coefficient test on $\gamma$ yields very mild overrejection, while AB2aW and AB2cW seriously overreject. For smaller values of $\gamma$ it is the other way around. After collapsing (not tabulated here) similar but more moderate patterns are found, due to the mitigated bias which goes again with slightly increased standard errors. Hence, for this case we find that if $T$ is not too small and $\gamma$ not very large we should not worry too much when applying BB even if effect stationarity does not strictly hold for the initial observations. As it happens, we note from Table 21 ($P^\phi 0$fJ-XA) that the rejection probabilities of the JES tests are such that they are relatively low when BB inference is preferable to AB inference, and relatively high when either $T$ or $\gamma$ are low and $\phi=0.5$. This pattern is much more pronounced for the JES tests than for the JBB

**Table 16.** P5fJ-EA*

Feasible Sargan-Hansen test: rejection probability

| $\bar{\rho}_{x\varepsilon}=0.3$ | df | | | $\gamma$ | $\theta = 0$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | AB | BB | Inc | | $JAB_a^{(2,1)}$ | $JBB_a^{(2,1)}$ | $JES_a^{(2,1)}$ | $JAB_c^{(2,1)}$ | $JBB_c^{(2,1)}$ | $JES_c^{(2,1)}$ | $JMAB$ | $JMBB$ | $JESM$ |
| $T=3$ | 4 | 7 | 3 | 0.20 | 0.067 | 0.053 | 0.052 | 0.047 | 0.034 | 0.036 | 0.315 | 0.287 | 0.271 |
| | | | | 0.50 | 0.078 | 0.065 | 0.055 | 0.055 | 0.041 | 0.038 | 0.324 | 0.331 | 0.313 |
| | | | | 0.80 | 0.096 | 0.074 | 0.062 | 0.077 | 0.049 | 0.039 | 0.324 | 0.396 | 0.376 |
| $T=6$ | 28 | 37 | 9 | 0.20 | 0.082 | 0.074 | 0.058 | 0.047 | 0.041 | 0.037 | 0.061 | 0.389 | 0.436 |
| | | | | 0.50 | 0.089 | 0.081 | 0.058 | 0.057 | 0.043 | 0.035 | 0.059 | 0.396 | 0.444 |
| | | | | 0.80 | 0.088 | 0.076 | 0.057 | 0.061 | 0.041 | 0.028 | 0.053 | 0.418 | 0.467 |
| $T=9$ | 70 | 85 | 15 | 0.20 | 0.057 | 0.044 | 0.057 | 0.045 | 0.044 | 0.035 | 0.056 | 0.420 | 0.475 |
| | | | | 0.50 | 0.057 | 0.050 | 0.062 | 0.050 | 0.049 | 0.037 | 0.057 | 0.423 | 0.480 |
| | | | | 0.80 | 0.057 | 0.045 | 0.057 | 0.049 | 0.039 | 0.027 | 0.046 | 0.426 | 0.484 |

| $\bar{\rho}_{x\varepsilon}=0.3$ | df | | | $\gamma$ | $\theta = 1$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | AB | BB | Inc | | $JAB_a^{(2,1)}$ | $JBB_a^{(2,1)}$ | $JES_a^{(2,1)}$ | $JAB_c^{(2,1)}$ | $JBB_c^{(2,1)}$ | $JES_c^{(2,1)}$ | $JMAB$ | $JMBB$ | $JESM$ |
| $T=3$ | 4 | 7 | 3 | 0.20 | 0.052 | 0.047 | 0.059 | 0.039 | 0.040 | 0.053 | 0.342 | 0.513 | 0.483 |
| | | | | 0.50 | 0.065 | 0.057 | 0.064 | 0.043 | 0.048 | 0.063 | 0.343 | 0.555 | 0.523 |
| | | | | 0.80 | 0.083 | 0.065 | 0.068 | 0.058 | 0.055 | 0.069 | 0.342 | 0.597 | 0.573 |
| $T=6$ | 28 | 37 | 9 | 0.20 | 0.065 | 0.056 | 0.062 | 0.042 | 0.048 | 0.057 | 0.075 | 0.721 | 0.743 |
| | | | | 0.50 | 0.074 | 0.062 | 0.062 | 0.055 | 0.059 | 0.063 | 0.076 | 0.728 | 0.750 |
| | | | | 0.80 | 0.081 | 0.059 | 0.057 | 0.071 | 0.065 | 0.062 | 0.071 | 0.739 | 0.760 |
| $T=9$ | 70 | 85 | 15 | 0.20 | 0.030 | 0.018 | 0.049 | 0.046 | 0.054 | 0.060 | 0.054 | 0.767 | 0.787 |
| | | | | 0.50 | 0.032 | 0.021 | 0.048 | 0.060 | 0.070 | 0.073 | 0.057 | 0.771 | 0.789 |
| | | | | 0.80 | 0.032 | 0.016 | 0.043 | 0.070 | 0.078 | 0.067 | 0.048 | 0.775 | 0.794 |

*$R = 10000$ simulation replications. Design parameter values: $N = 200$, $SNR = 3$, $DEN_y = 1.0$, $EVF_x = 0.0$, $\bar{\rho}_{x\varepsilon} = 0.3$, $\xi = 0.8$, $\kappa = 0.00$, $\sigma_\varepsilon = 1$, $q = 1$, $\phi = 1.0$. These yield the DGP parameter values: $\pi_\lambda = 0.00$, $\pi_\eta = 0.00$, $\sigma_v = 0.60$, $\sigma_\eta = 1.0 * (1 - \gamma)$, $\rho_{v\varepsilon} = 0.5$ (and $\bar{\rho}_{x\eta} = 0.00$, $\bar{\rho}_{x\lambda} = 0.00$).

tests. However, it is also the case in P$^\phi$0 that collapsing mitigates this welcome quality of the JES tests to warn against unfavorable consequences of effect nonstationarity on BB inference.

From P$^\phi$1-XA, in which the individual effects are much more prominent, we find that $\phi = 0.5$ has curious effects on AB and BB results. For effect stationarity ($\phi = 1$) we already noted more bias for AB than under P0. For $\gamma$ large, this bias is even more serious when $\phi = 0.5$, despite the consistency of AB. For BB estimation the reduction of $\phi$ leads to much larger bias and much smaller stdv, with the effect that RMSE values for inconsistent BB are usually much worse than for AB, but are often slightly better, except for BB2c, when $\gamma = 0.8$. All BB coefficient tests for $\gamma$ have size close or equal to 1 under P$^\phi$1-XA and the AB tests for $\gamma = 0.8$ overreject very seriously as well. Under P$^\phi$1-XC the bias of AB is reasonable except for $\gamma = 0.8$. The bias of BB has decreased but is still enormous, although it still has preferable RMSE when $\gamma = 0.8$. Especially regarding tests on $\gamma$ BB fails. For both the a and c versions the JES test has high rejection probability to detect $\phi \neq 1$, except when $\gamma$ is large. The relatively low rejection probability of JES tests obtained after collapsing when $\gamma = 0.8$ and $\phi = 0.5$ again indicates that despite its inconsistency BB has similar or smaller RMSE than AB for that specific case.

Next we consider the simultaneous model again. In case P$^\phi$5-EA estimator AB is consistent and BB again inconsistent. Nevertheless, for all $\gamma$ and $T$ values examined in Table 22 (P$^\phi$5fc-EA), AB has a more severe bias than BB, whereas BB has smaller stdv values at the same time and thus has smaller RMSE for all $\gamma$ and $T$ values examined. The size control of coefficient tests is worse for AB, but for BB it is appalling too, where BB2aW, with estimated type I error probabilities ranging from 5% to 70%, is often preferable to BB2cW. The 2-step JAB tests behave reasonably (wrongly indicate inconsistency with a probability rather close to 5%), whereas the JBB tests reject with probabilities in the 3-38% range, and JES in the 3-69% range. By collapsing the RMSE of AB generally reduces when $T \geq 6$ and for BB especially when $\gamma = 0.8$. BB has again

**Table 17.** P5fJ-WA*

Feasible Sargan-Hansen test: rejection probability

| $\bar{\rho}_{x\varepsilon}=0.3$ | df | | | | $\theta = 0$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | AB | BB | Inc | $\gamma$ | $JAB_a^{(2,1)}$ | $JBB_a^{(2,1)}$ | $JES_a^{(2,1)}$ | $JAB_c^{(2,1)}$ | $JBB_c^{(2,1)}$ | $JES_c^{(2,1)}$ | $JMAB$ | $JMBB$ | $JESM$ |
| $T=3$ | 6 | 10 | 4 | 0.20 | 0.206 | 0.247 | 0.137 | 0.135 | 0.151 | 0.105 | 0.325 | 0.331 | 0.313 |
| | | | | 0.50 | 0.228 | 0.303 | 0.175 | 0.185 | 0.218 | 0.130 | 0.333 | 0.344 | 0.326 |
| | | | | 0.80 | 0.414 | 0.537 | 0.286 | 0.402 | 0.467 | 0.223 | 0.370 | 0.388 | 0.363 |
| $T=6$ | 33 | 43 | 10 | 0.20 | 0.555 | 0.743 | 0.355 | 0.512 | 0.594 | 0.232 | 0.291 | 0.538 | 0.534 |
| | | | | 0.50 | 0.578 | 0.821 | 0.478 | 0.614 | 0.758 | 0.372 | 0.335 | 0.543 | 0.530 |
| | | | | 0.80 | 0.879 | 0.983 | 0.592 | 0.963 | 0.994 | 0.637 | 0.724 | 0.622 | 0.554 |
| $T=9$ | 78 | 94 | 16 | 0.20 | 0.536 | 0.633 | 0.145 | 0.796 | 0.857 | 0.297 | 0.621 | 0.598 | 0.553 |
| | | | | 0.50 | 0.549 | 0.728 | 0.233 | 0.860 | 0.947 | 0.518 | 0.667 | 0.602 | 0.552 |
| | | | | 0.80 | 0.877 | 0.963 | 0.215 | 0.999 | 1.000 | 0.797 | 0.975 | 0.702 | 0.573 |

| $\bar{\rho}_{x\varepsilon}=0.3$ | df | | | | $\theta = 1$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | AB | BB | Inc | $\gamma$ | $JAB_a^{(2,1)}$ | $JBB_a^{(2,1)}$ | $JES_a^{(2,1)}$ | $JAB_c^{(2,1)}$ | $JBB_c^{(2,1)}$ | $JES_c^{(2,1)}$ | $JMAB$ | $JMBB$ | $JESM$ |
| $T=3$ | 6 | 10 | 4 | 0.20 | 0.110 | 0.130 | 0.099 | 0.077 | 0.111 | 0.112 | 0.339 | 0.555 | 0.531 |
| | | | | 0.50 | 0.120 | 0.155 | 0.115 | 0.095 | 0.142 | 0.133 | 0.359 | 0.572 | 0.543 |
| | | | | 0.80 | 0.182 | 0.252 | 0.178 | 0.171 | 0.278 | 0.235 | 0.391 | 0.622 | 0.588 |
| $T=6$ | 33 | 43 | 10 | 0.20 | 0.217 | 0.316 | 0.189 | 0.184 | 0.270 | 0.199 | 0.284 | 0.758 | 0.755 |
| | | | | 0.50 | 0.233 | 0.382 | 0.237 | 0.244 | 0.395 | 0.293 | 0.271 | 0.755 | 0.754 |
| | | | | 0.80 | 0.465 | 0.670 | 0.319 | 0.580 | 0.824 | 0.580 | 0.625 | 0.788 | 0.766 |
| $T=9$ | 78 | 94 | 16 | 0.20 | 0.141 | 0.138 | 0.093 | 0.303 | 0.416 | 0.255 | 0.597 | 0.793 | 0.780 |
| | | | | 0.50 | 0.138 | 0.171 | 0.125 | 0.382 | 0.589 | 0.411 | 0.569 | 0.792 | 0.780 |
| | | | | 0.80 | 0.349 | 0.401 | 0.122 | 0.835 | 0.968 | 0.731 | 0.923 | 0.816 | 0.784 |

*$R = 10000$ simulation replications. Design parameter values: $N = 200$, $SNR = 3$, $DEN_y = 1.0$, $EVF_x = 0.0$, $\bar{\rho}_{x\varepsilon} = 0.3$, $\xi = 0.8$, $\kappa = 0.00$, $\sigma_\varepsilon = 1$, $q = 1$, $\phi = 1.0$. These yield the DGP parameter values: $\pi_\lambda = 0.00$, $\pi_\eta = 0.00$, $\sigma_v = 0.60$, $\sigma_\eta = 1.0 * (1 - \gamma)$, $\rho_{v\varepsilon} = 0.5$ (and $\bar{\rho}_{x\eta} = 0.00$, $\bar{\rho}_{x\lambda} = 0.00$).

**Table 18.** P5fJ-XA*

Feasible Sargan-Hansen test: rejection probability

| $\bar{\rho}_{x\varepsilon}=0.3$ | df | | | | $\theta = 0$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | AB | BB | Inc | $\gamma$ | $JAB_a^{(2,1)}$ | $JBB_a^{(2,1)}$ | $JES_a^{(2,1)}$ | $JAB_c^{(2,1)}$ | $JBB_c^{(2,1)}$ | $JES_c^{(2,1)}$ | $JMAB$ | $JMBB$ | $JESM$ |
| $T=3$ | 9 | 13 | 4 | 0.20 | 0.199 | 0.248 | 0.150 | 0.156 | 0.188 | 0.123 | 0.290 | 0.353 | 0.347 |
| | | | | 0.50 | 0.199 | 0.266 | 0.161 | 0.176 | 0.214 | 0.124 | 0.297 | 0.354 | 0.346 |
| | | | | 0.80 | 0.380 | 0.479 | 0.233 | 0.399 | 0.444 | 0.182 | 0.339 | 0.388 | 0.373 |
| $T=6$ | 48 | 58 | 10 | 0.20 | 0.436 | 0.604 | 0.295 | 0.494 | 0.570 | 0.218 | 0.294 | 0.530 | 0.527 |
| | | | | 0.50 | 0.424 | 0.669 | 0.409 | 0.547 | 0.673 | 0.322 | 0.303 | 0.516 | 0.515 |
| | | | | 0.80 | 0.769 | 0.922 | 0.459 | 0.947 | 0.983 | 0.516 | 0.710 | 0.590 | 0.534 |
| $T=9$ | 114 | 130 | 16 | 0.20 | 0.154 | 0.110 | 0.063 | 0.754 | 0.811 | 0.261 | 0.588 | 0.575 | 0.542 |
| | | | | 0.50 | 0.144 | 0.135 | 0.104 | 0.783 | 0.885 | 0.440 | 0.590 | 0.569 | 0.538 |
| | | | | 0.80 | 0.388 | 0.368 | 0.069 | 0.998 | 1.000 | 0.660 | 0.968 | 0.648 | 0.552 |

| $\bar{\rho}_{x\varepsilon}=0.3$ | df | | | | $\theta = 1$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | AB | BB | Inc | $\gamma$ | $JAB_a^{(2,1)}$ | $JBB_a^{(2,1)}$ | $JES_a^{(2,1)}$ | $JAB_c^{(2,1)}$ | $JBB_c^{(2,1)}$ | $JES_c^{(2,1)}$ | $JMAB$ | $JMBB$ | $JESM$ |
| $T=3$ | 9 | 13 | 4 | 0.20 | 0.103 | 0.128 | 0.103 | 0.088 | 0.133 | 0.124 | 0.312 | 0.587 | 0.574 |
| | | | | 0.50 | 0.108 | 0.136 | 0.103 | 0.098 | 0.149 | 0.131 | 0.313 | 0.591 | 0.577 |
| | | | | 0.80 | 0.175 | 0.220 | 0.141 | 0.187 | 0.291 | 0.219 | 0.359 | 0.631 | 0.607 |
| $T=6$ | 48 | 58 | 10 | 0.20 | 0.148 | 0.215 | 0.169 | 0.183 | 0.250 | 0.164 | 0.278 | 0.763 | 0.763 |
| | | | | 0.50 | 0.145 | 0.248 | 0.218 | 0.223 | 0.336 | 0.243 | 0.246 | 0.757 | 0.760 |
| | | | | 0.80 | 0.324 | 0.488 | 0.251 | 0.576 | 0.771 | 0.507 | 0.624 | 0.783 | 0.765 |
| $T=9$ | 114 | 130 | 16 | 0.20 | 0.013 | 0.004 | 0.049 | 0.290 | 0.368 | 0.189 | 0.561 | 0.795 | 0.787 |
| | | | | 0.50 | 0.011 | 0.006 | 0.069 | 0.339 | 0.484 | 0.319 | 0.492 | 0.792 | 0.785 |
| | | | | 0.80 | 0.033 | 0.016 | 0.052 | 0.815 | 0.940 | 0.619 | 0.924 | 0.811 | 0.784 |

*$R = 10000$ simulation replications. Design parameter values: $N = 200$, $SNR = 3$, $DEN_y = 1.0$, $EVF_x = 0.0$, $\bar{\rho}_{x\varepsilon} = 0.3$, $\xi = 0.8$, $\kappa = 0.00$, $\sigma_\varepsilon = 1$, $q = 1$, $\phi = 1.0$. These yield the DGP parameter values: $\pi_\lambda = 0.00$, $\pi_\eta = 0.00$, $\sigma_v = 0.60$, $\sigma_\eta = 1.0 * (1 - \gamma)$, $\rho_{v\varepsilon} = 0.5$ (and $\bar{\rho}_{x\eta} = 0.00$, $\bar{\rho}_{x\lambda} = 0.00$).

smaller RMSE than AB. The rejection rates of the JBB and JES tests are substantially lower now, which seems bad because the invalid (first-differenced) instruments are less often detected, but this may nevertheless be appreciated because it induces to prefer less inaccurate BB inference to AB inference. After collapsing the size distortions of BB2aW and BB2cW are less extreme too, now ranging from 5-33%, but the RMSE values for BB may suffer due to collapsing, especially when $\gamma$ and $T$ are small. The RMSE values for BB under $P^\phi5$-WA and $P^\phi5$-XA are usually much worse than those for AB under $P^\phi5$-EA. Hence, although the invalid instruments for the level equation are not necessarily a curse when endogeneity of $x_{it}$ is respected, they should not be used when they are invalid for two reasons ($\phi \neq 1, \rho_{x\varepsilon} \neq 0$). That neither AB nor BB should

**Table 19. P⁰0fc-XA***

Feasible coefficient estimators for Arellano-Bond

| $\bar\rho_{\varepsilon\varepsilon}=0.0$ | L | γ | θ=0 AB1 Bias | Stdv | RMSE | AB2a Bias | Stdv | RMSE | AB2c Bias | Stdv | RMSE | θ=1 AB1 Bias | Stdv | RMSE | AB2a Bias | Stdv | RMSE | AB2c Bias | Stdv | RMSE | MAB Bias | Stdv | RMSE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| T=3 | 11 | 0.20 | -0.012 | 0.058 | 0.060 | -0.012 | 0.061 | 0.062 | -0.012 | 0.059 | 0.061 | -0.024 | 0.089 | 0.092 | -0.020 | 0.084 | 0.086 | -0.023 | 0.084 | 0.087 | -0.023 | 0.086 | 0.089 |
|  |  | 0.50 | -0.022 | 0.076 | 0.079 | -0.022 | 0.079 | 0.082 | -0.021 | 0.077 | 0.080 | -0.044 | 0.116 | 0.124 | -0.038 | 0.110 | 0.117 | -0.043 | 0.110 | 0.118 | -0.043 | 0.113 | 0.121 |
|  |  | 0.80 | -0.075 | 0.133 | 0.153 | -0.075 | 0.140 | 0.159 | -0.073 | 0.135 | 0.154 | -0.146 | 0.199 | 0.246 | -0.136 | 0.193 | 0.236 | -0.144 | 0.189 | 0.237 | -0.143 | 0.194 | 0.241 |
| T=6 | 50 | 0.20 | -0.009 | 0.029 | 0.030 | -0.009 | 0.032 | 0.034 | -0.009 | 0.029 | 0.031 | -0.020 | 0.046 | 0.050 | -0.016 | 0.041 | 0.044 | -0.017 | 0.041 | 0.045 | -0.015 | 0.037 | 0.040 |
|  |  | 0.50 | -0.017 | 0.034 | 0.038 | -0.017 | 0.038 | 0.042 | -0.017 | 0.035 | 0.038 | -0.035 | 0.053 | 0.064 | -0.030 | 0.049 | 0.057 | -0.031 | 0.048 | 0.057 | -0.027 | 0.044 | 0.052 |
|  |  | 0.80 | -0.054 | 0.052 | 0.075 | -0.055 | 0.059 | 0.080 | -0.053 | 0.053 | 0.075 | -0.105 | 0.080 | 0.132 | -0.093 | 0.076 | 0.120 | -0.098 | 0.072 | 0.122 | -0.089 | 0.070 | 0.113 |
| T=9 | 116 | 0.20 | -0.008 | 0.021 | 0.023 | -0.008 | 0.024 | 0.025 | -0.008 | 0.022 | 0.023 | -0.017 | 0.034 | 0.038 | -0.015 | 0.032 | 0.035 | -0.015 | 0.029 | 0.033 | -0.011 | 0.025 | 0.027 |
|  |  | 0.50 | -0.014 | 0.024 | 0.028 | -0.014 | 0.026 | 0.030 | -0.014 | 0.024 | 0.028 | -0.029 | 0.037 | 0.047 | -0.026 | 0.035 | 0.044 | -0.025 | 0.032 | 0.041 | -0.020 | 0.028 | 0.035 |
|  |  | 0.80 | -0.041 | 0.033 | 0.053 | -0.043 | 0.037 | 0.056 | -0.041 | 0.033 | 0.053 | -0.079 | 0.050 | 0.094 | -0.075 | 0.049 | 0.089 | -0.073 | 0.044 | 0.085 | -0.064 | 0.042 | 0.076 |

| $\bar\rho_{\varepsilon\varepsilon}=0.0$ | L | β | θ=0 AB1 Bias | Stdv | RMSE | AB2a Bias | Stdv | RMSE | AB2c Bias | Stdv | RMSE | θ=1 AB1 Bias | Stdv | RMSE | AB2a Bias | Stdv | RMSE | AB2c Bias | Stdv | RMSE | MAB Bias | Stdv | RMSE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| T=3 | 11 | 1.43 | 0.003 | 0.100 | 0.100 | 0.004 | 0.103 | 0.103 | 0.003 | 0.101 | 0.101 | 0.006 | 0.158 | 0.158 | 0.005 | 0.145 | 0.145 | 0.005 | 0.147 | 0.147 | 0.005 | 0.148 | 0.148 |
|  |  | 0.93 | 0.002 | 0.099 | 0.099 | 0.002 | 0.103 | 0.103 | 0.002 | 0.100 | 0.100 | 0.004 | 0.157 | 0.157 | 0.003 | 0.144 | 0.144 | 0.003 | 0.146 | 0.146 | 0.004 | 0.147 | 0.147 |
|  |  | 0.31 | -0.002 | 0.097 | 0.098 | -0.002 | 0.101 | 0.101 | -0.002 | 0.099 | 0.099 | -0.004 | 0.152 | 0.152 | -0.006 | 0.140 | 0.140 | -0.005 | 0.142 | 0.143 | -0.004 | 0.142 | 0.142 |
| T=6 | 50 | 1.43 | 0.006 | 0.055 | 0.055 | 0.006 | 0.060 | 0.061 | 0.006 | 0.055 | 0.055 | 0.013 | 0.087 | 0.088 | 0.010 | 0.077 | 0.077 | 0.011 | 0.078 | 0.079 | 0.009 | 0.067 | 0.068 |
|  |  | 0.93 | 0.007 | 0.053 | 0.053 | 0.007 | 0.059 | 0.059 | 0.007 | 0.053 | 0.054 | 0.014 | 0.085 | 0.086 | 0.011 | 0.074 | 0.075 | 0.012 | 0.076 | 0.077 | 0.011 | 0.065 | 0.066 |
|  |  | 0.31 | 0.004 | 0.051 | 0.051 | 0.004 | 0.057 | 0.057 | 0.004 | 0.052 | 0.052 | 0.007 | 0.082 | 0.082 | 0.005 | 0.072 | 0.072 | 0.006 | 0.073 | 0.074 | 0.006 | 0.063 | 0.063 |
| T=9 | 116 | 1.43 | 0.007 | 0.040 | 0.041 | 0.008 | 0.045 | 0.045 | 0.007 | 0.041 | 0.041 | 0.014 | 0.065 | 0.067 | 0.013 | 0.060 | 0.062 | 0.012 | 0.057 | 0.058 | 0.010 | 0.046 | 0.047 |
|  |  | 0.93 | 0.009 | 0.039 | 0.040 | 0.009 | 0.043 | 0.044 | 0.009 | 0.039 | 0.040 | 0.017 | 0.062 | 0.065 | 0.016 | 0.058 | 0.060 | 0.015 | 0.054 | 0.056 | 0.012 | 0.044 | 0.046 |
|  |  | 0.31 | 0.006 | 0.037 | 0.037 | 0.007 | 0.041 | 0.041 | 0.006 | 0.037 | 0.037 | 0.012 | 0.059 | 0.060 | 0.011 | 0.055 | 0.056 | 0.011 | 0.051 | 0.052 | 0.009 | 0.042 | 0.043 |

Feasible coefficient estimators for Blundell-Bond

| $\bar\rho_{\varepsilon\varepsilon}=0.0$ | L | γ | θ=0 BB1 Bias | Stdv | RMSE | BB2a Bias | Stdv | RMSE | BB2c Bias | Stdv | RMSE | θ=1 BB1 Bias | Stdv | RMSE | BB2a Bias | Stdv | RMSE | BB2c Bias | Stdv | RMSE | MBB Bias | Stdv | RMSE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| T=3 | 16 | 0.20 | 0.043 | 0.049 | 0.066 | 0.047 | 0.056 | 0.073 | 0.054 | 0.053 | 0.076 | 0.036 | 0.073 | 0.082 | 0.056 | 0.075 | 0.094 | 0.061 | 0.074 | 0.096 | 0.060 | 0.078 | 0.098 |
|  |  | 0.50 | 0.024 | 0.056 | 0.060 | 0.037 | 0.058 | 0.069 | 0.039 | 0.055 | 0.067 | 0.011 | 0.084 | 0.084 | 0.036 | 0.079 | 0.087 | 0.039 | 0.078 | 0.087 | 0.062 | 0.213 | 0.222 |
|  |  | 0.80 | -0.013 | 0.071 | 0.073 | 0.007 | 0.071 | 0.071 | 0.008 | 0.068 | 0.068 | -0.040 | 0.108 | 0.116 | -0.010 | 0.103 | 0.104 | 0.005 | 0.101 | 0.101 | 0.050 | 1.104 | 1.106 |
| T=6 | 61 | 0.20 | 0.016 | 0.026 | 0.031 | 0.017 | 0.029 | 0.034 | 0.026 | 0.028 | 0.039 | 0.009 | 0.041 | 0.042 | 0.016 | 0.038 | 0.042 | 0.036 | 0.040 | 0.054 | 0.028 | 0.037 | 0.047 |
|  |  | 0.50 | 0.007 | 0.029 | 0.029 | 0.017 | 0.032 | 0.036 | 0.028 | 0.030 | 0.041 | -0.006 | 0.045 | 0.045 | 0.009 | 0.042 | 0.043 | 0.032 | 0.041 | 0.052 | 0.041 | 0.042 | 0.059 |
|  |  | 0.80 | -0.020 | 0.036 | 0.042 | -0.001 | 0.037 | 0.037 | 0.010 | 0.033 | 0.034 | -0.049 | 0.055 | 0.074 | -0.028 | 0.051 | 0.058 | 0.007 | 0.045 | 0.045 | 0.042 | 0.400 | 0.402 |
| T=9 | 133 | 0.20 | 0.008 | 0.020 | 0.021 | 0.008 | 0.021 | 0.023 | 0.016 | 0.021 | 0.026 | 0.002 | 0.031 | 0.031 | 0.003 | 0.030 | 0.030 | 0.023 | 0.029 | 0.037 | 0.014 | 0.025 | 0.029 |
|  |  | 0.50 | 0.001 | 0.021 | 0.021 | 0.004 | 0.023 | 0.023 | 0.019 | 0.022 | 0.029 | -0.011 | 0.033 | 0.034 | -0.007 | 0.031 | 0.032 | 0.025 | 0.029 | 0.038 | 0.023 | 0.028 | 0.036 |
|  |  | 0.80 | -0.022 | 0.026 | 0.034 | -0.013 | 0.027 | 0.030 | 0.009 | 0.023 | 0.025 | -0.048 | 0.039 | 0.062 | -0.042 | 0.038 | 0.056 | 0.007 | 0.030 | 0.031 | 0.021 | 0.045 | 0.050 |

| $\bar\rho_{\varepsilon\varepsilon}=0.0$ | L | β | θ=0 BB1 Bias | Stdv | RMSE | BB2a Bias | Stdv | RMSE | BB2c Bias | Stdv | RMSE | θ=1 BB1 Bias | Stdv | RMSE | BB2a Bias | Stdv | RMSE | BB2c Bias | Stdv | RMSE | MBB Bias | Stdv | RMSE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| T=3 | 16 | 1.43 | 0.036 | 0.098 | 0.105 | 0.053 | 0.104 | 0.117 | 0.044 | 0.099 | 0.108 | 0.038 | 0.153 | 0.158 | 0.064 | 0.149 | 0.162 | 0.058 | 0.146 | 0.158 | 0.043 | 0.144 | 0.150 |
|  |  | 0.93 | 0.025 | 0.095 | 0.099 | 0.041 | 0.100 | 0.108 | 0.037 | 0.095 | 0.102 | 0.029 | 0.149 | 0.152 | 0.047 | 0.139 | 0.147 | 0.048 | 0.138 | 0.146 | 0.050 | 0.213 | 0.219 |
|  |  | 0.31 | 0.008 | 0.095 | 0.095 | 0.012 | 0.099 | 0.100 | 0.012 | 0.095 | 0.095 | 0.013 | 0.147 | 0.147 | 0.016 | 0.136 | 0.137 | 0.019 | 0.136 | 0.138 | 0.029 | 1.741 | 1.741 |
| T=6 | 61 | 1.43 | 0.004 | 0.054 | 0.054 | 0.007 | 0.060 | 0.061 | 0.003 | 0.055 | 0.055 | 0.008 | 0.086 | 0.086 | 0.010 | 0.078 | 0.079 | 0.004 | 0.078 | 0.078 | -0.001 | 0.067 | 0.067 |
|  |  | 0.93 | 0.007 | 0.052 | 0.052 | 0.010 | 0.057 | 0.058 | 0.008 | 0.052 | 0.053 | 0.014 | 0.083 | 0.084 | 0.015 | 0.074 | 0.076 | 0.011 | 0.074 | 0.074 | 0.003 | 0.064 | 0.065 |
|  |  | 0.31 | 0.006 | 0.050 | 0.050 | 0.006 | 0.054 | 0.054 | 0.006 | 0.048 | 0.049 | 0.010 | 0.079 | 0.080 | 0.009 | 0.070 | 0.071 | 0.009 | 0.068 | 0.069 | 0.008 | 0.207 | 0.207 |
| T=9 | 133 | 1.43 | -0.001 | 0.040 | 0.040 | 0.001 | 0.043 | 0.043 | -0.004 | 0.040 | 0.041 | 0.005 | 0.064 | 0.064 | 0.005 | 0.061 | 0.061 | -0.006 | 0.056 | 0.057 | -0.004 | 0.047 | 0.047 |
|  |  | 0.93 | 0.005 | 0.038 | 0.038 | 0.005 | 0.042 | 0.042 | -0.000 | 0.038 | 0.038 | 0.012 | 0.061 | 0.062 | 0.011 | 0.058 | 0.059 | -0.000 | 0.053 | 0.053 | -0.003 | 0.045 | 0.045 |
|  |  | 0.31 | 0.006 | 0.036 | 0.036 | 0.006 | 0.039 | 0.039 | 0.003 | 0.035 | 0.035 | 0.011 | 0.057 | 0.059 | 0.011 | 0.054 | 0.055 | 0.005 | 0.048 | 0.048 | 0.003 | 0.045 | 0.045 |

*R = 10000 simulation replications. Design parameter values: $N = 200$, $SNR = 3$, $DEN_y = 1.0$, $EVF_x = 0.0$, $\bar\rho_{x\varepsilon} = 0.0$, $\xi = 0.8$, $\kappa = 0.00$, $\sigma_\varepsilon = 1$, $q = 1$, $\phi = 0.5$.
These yield the DGP parameter values: $\pi_\lambda = 0.00$, $\pi_\eta = 0.00$, $\sigma_v = 0.60$, $\sigma_\eta = 1.0*(1-\gamma)$, $\rho_{v\varepsilon} = 0.0$ (and $\bar\rho_{x\lambda} = 0.00$).

46

**Table 20.** P$^\phi$0ft-XA*

Feasible t-test Arellano-Bond: actual significance level

| $\bar\rho_{x\varepsilon}=0.0$ | | | $\theta=0$ | | | | | | | $\theta=1$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $L$ | $\gamma$ | AB1 | AB1aR | AB1cR | AB2a | AB2aW | AB2c | AB2cW | AB1 | AB1aR | AB1cR | AB2a | AB2aW | AB2c | AB2cW | MAB |
| $T=3$ | 11 | 0.20 | 0.063 | 0.066 | 0.059 | 0.085 | 0.065 | 0.066 | 0.061 | 0.225 | 0.088 | 0.073 | 0.140 | 0.075 | 0.086 | 0.069 | 0.570 |
| | | 0.50 | 0.070 | 0.073 | 0.066 | 0.092 | 0.068 | 0.073 | 0.068 | 0.250 | 0.103 | 0.088 | 0.159 | 0.087 | 0.106 | 0.086 | 0.600 |
| | | 0.80 | 0.117 | 0.116 | 0.112 | 0.139 | 0.103 | 0.118 | 0.109 | 0.353 | 0.191 | 0.170 | 0.254 | 0.147 | 0.196 | 0.171 | 0.699 |
| $T=6$ | 50 | 0.20 | 0.062 | 0.064 | 0.054 | 0.164 | 0.061 | 0.060 | 0.056 | 0.250 | 0.092 | 0.067 | 0.361 | 0.079 | 0.077 | 0.065 | 0.206 |
| | | 0.50 | 0.076 | 0.078 | 0.070 | 0.182 | 0.072 | 0.077 | 0.072 | 0.307 | 0.125 | 0.101 | 0.411 | 0.099 | 0.114 | 0.096 | 0.251 |
| | | 0.80 | 0.188 | 0.182 | 0.176 | 0.319 | 0.142 | 0.182 | 0.172 | 0.549 | 0.322 | 0.282 | 0.624 | 0.236 | 0.312 | 0.277 | 0.479 |
| $T=9$ | 116 | 0.20 | 0.071 | 0.073 | 0.063 | 0.324 | 0.067 | 0.069 | 0.065 | 0.274 | 0.101 | 0.074 | 0.693 | 0.093 | 0.088 | 0.077 | 0.154 |
| | | 0.50 | 0.099 | 0.097 | 0.091 | 0.361 | 0.084 | 0.097 | 0.092 | 0.349 | 0.148 | 0.117 | 0.734 | 0.138 | 0.134 | 0.116 | 0.202 |
| | | 0.80 | 0.255 | 0.243 | 0.241 | 0.552 | 0.196 | 0.247 | 0.234 | 0.655 | 0.404 | 0.363 | 0.893 | 0.376 | 0.412 | 0.379 | 0.478 |
| | $L$ | $\beta$ | AB1 | AB1aR | AB1cR | AB2a | AB2aW | AB2c | AB2cW | AB1 | AB1aR | AB1cR | AB2a | AB2aW | AB2c | AB2cW | MAB |
| $T=3$ | 11 | 1.43 | 0.052 | 0.052 | 0.052 | 0.071 | 0.053 | 0.056 | 0.052 | 0.215 | 0.070 | 0.064 | 0.117 | 0.063 | 0.071 | 0.056 | 0.583 |
| | | 0.93 | 0.050 | 0.052 | 0.050 | 0.073 | 0.053 | 0.056 | 0.051 | 0.216 | 0.072 | 0.064 | 0.118 | 0.063 | 0.071 | 0.054 | 0.584 |
| | | 0.31 | 0.050 | 0.050 | 0.050 | 0.072 | 0.052 | 0.055 | 0.050 | 0.215 | 0.069 | 0.063 | 0.117 | 0.061 | 0.072 | 0.056 | 0.580 |
| $T=6$ | 50 | 1.43 | 0.051 | 0.054 | 0.050 | 0.146 | 0.055 | 0.056 | 0.052 | 0.221 | 0.068 | 0.059 | 0.315 | 0.060 | 0.068 | 0.058 | 0.211 |
| | | 0.93 | 0.053 | 0.055 | 0.053 | 0.148 | 0.055 | 0.058 | 0.053 | 0.227 | 0.070 | 0.061 | 0.314 | 0.064 | 0.069 | 0.059 | 0.215 |
| | | 0.31 | 0.054 | 0.054 | 0.054 | 0.143 | 0.052 | 0.058 | 0.054 | 0.232 | 0.066 | 0.064 | 0.315 | 0.057 | 0.069 | 0.058 | 0.227 |
| $T=9$ | 116 | 1.43 | 0.050 | 0.053 | 0.048 | 0.294 | 0.056 | 0.054 | 0.051 | 0.232 | 0.067 | 0.056 | 0.654 | 0.068 | 0.065 | 0.056 | 0.137 |
| | | 0.93 | 0.053 | 0.055 | 0.052 | 0.299 | 0.059 | 0.058 | 0.054 | 0.241 | 0.073 | 0.065 | 0.657 | 0.071 | 0.070 | 0.060 | 0.148 |
| | | 0.31 | 0.058 | 0.056 | 0.059 | 0.299 | 0.057 | 0.062 | 0.057 | 0.242 | 0.068 | 0.069 | 0.657 | 0.066 | 0.072 | 0.061 | 0.154 |

Feasible t-test Blundell-Bond: actual significance level

| $\bar\rho_{x\varepsilon}=0.0$ | | | $\theta=0$ | | | | | | | $\theta=1$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $L$ | $\gamma$ | BB1 | BB1aR | BB1cR | BB2a | BB2aW | BB2c | BB2cW | BB1 | BB1aR | BB1cR | BB2a | BB2aW | BB2c | BB2cW | MBB |
| $T=3$ | 16 | 0.20 | 0.166 | 0.170 | 0.144 | 0.258 | 0.163 | 0.236 | 0.207 | 0.257 | 0.118 | 0.087 | 0.322 | 0.165 | 0.213 | 0.182 | 0.659 |
| | | 0.50 | 0.084 | 0.089 | 0.070 | 0.191 | 0.131 | 0.140 | 0.131 | 0.219 | 0.083 | 0.061 | 0.245 | 0.120 | 0.135 | 0.120 | 0.635 |
| | | 0.80 | 0.058 | 0.059 | 0.040 | 0.103 | 0.061 | 0.057 | 0.045 | 0.240 | 0.090 | 0.060 | 0.196 | 0.071 | 0.085 | 0.064 | 0.615 |
| $T=6$ | 61 | 0.20 | 0.092 | 0.102 | 0.079 | 0.255 | 0.091 | 0.176 | 0.153 | 0.219 | 0.080 | 0.049 | 0.445 | 0.090 | 0.205 | 0.171 | 0.317 |
| | | 0.50 | 0.053 | 0.060 | 0.046 | 0.252 | 0.092 | 0.196 | 0.166 | 0.212 | 0.068 | 0.045 | 0.435 | 0.074 | 0.184 | 0.157 | 0.419 |
| | | 0.80 | 0.083 | 0.087 | 0.071 | 0.199 | 0.055 | 0.080 | 0.061 | 0.382 | 0.176 | 0.127 | 0.499 | 0.088 | 0.085 | 0.060 | 0.399 |
| $T=9$ | 133 | 0.20 | 0.068 | 0.076 | 0.057 | 0.403 | 0.066 | 0.127 | 0.116 | 0.210 | 0.067 | 0.044 | 0.730 | 0.064 | 0.164 | 0.143 | 0.186 |
| | | 0.50 | 0.049 | 0.055 | 0.042 | 0.380 | 0.057 | 0.174 | 0.155 | 0.233 | 0.075 | 0.054 | 0.729 | 0.067 | 0.182 | 0.154 | 0.283 |
| | | 0.80 | 0.133 | 0.136 | 0.118 | 0.450 | 0.075 | 0.094 | 0.069 | 0.511 | 0.267 | 0.209 | 0.866 | 0.217 | 0.085 | 0.059 | 0.287 |
| | $L$ | $\beta$ | BB1 | BB1aR | BB1cR | BB2a | BB2aW | BB2c | BB2cW | BB1 | BB1aR | BB1cR | BB2a | BB2aW | BB2c | BB2cW | MBB |
| $T=3$ | 16 | 1.43 | 0.065 | 0.069 | 0.066 | 0.133 | 0.085 | 0.081 | 0.070 | 0.230 | 0.083 | 0.073 | 0.214 | 0.106 | 0.106 | 0.098 | 0.600 |
| | | 0.93 | 0.058 | 0.061 | 0.058 | 0.117 | 0.080 | 0.075 | 0.067 | 0.226 | 0.077 | 0.067 | 0.190 | 0.089 | 0.097 | 0.088 | 0.605 |
| | | 0.31 | 0.049 | 0.052 | 0.050 | 0.089 | 0.059 | 0.061 | 0.059 | 0.220 | 0.073 | 0.062 | 0.163 | 0.069 | 0.083 | 0.077 | 0.632 |
| $T=6$ | 61 | 1.43 | 0.050 | 0.054 | 0.049 | 0.177 | 0.051 | 0.057 | 0.056 | 0.219 | 0.070 | 0.054 | 0.387 | 0.065 | 0.070 | 0.073 | 0.211 |
| | | 0.93 | 0.052 | 0.055 | 0.051 | 0.179 | 0.056 | 0.056 | 0.056 | 0.227 | 0.070 | 0.058 | 0.395 | 0.070 | 0.072 | 0.074 | 0.215 |
| | | 0.31 | 0.054 | 0.054 | 0.053 | 0.172 | 0.055 | 0.056 | 0.059 | 0.230 | 0.069 | 0.060 | 0.395 | 0.066 | 0.069 | 0.069 | 0.239 |
| $T=9$ | 133 | 1.43 | 0.047 | 0.051 | 0.046 | 0.342 | 0.048 | 0.054 | 0.058 | 0.221 | 0.065 | 0.052 | 0.719 | 0.063 | 0.066 | 0.074 | 0.137 |
| | | 0.93 | 0.050 | 0.052 | 0.049 | 0.347 | 0.052 | 0.052 | 0.057 | 0.229 | 0.069 | 0.059 | 0.721 | 0.067 | 0.066 | 0.074 | 0.136 |
| | | 0.31 | 0.055 | 0.055 | 0.055 | 0.357 | 0.057 | 0.057 | 0.060 | 0.236 | 0.070 | 0.065 | 0.734 | 0.067 | 0.067 | 0.069 | 0.146 |

be used in P5 under W and X will be indicated with highest probability under WC, and then this probability is larger than 0.8 for $JBB_a^{(2,1)}$ only when $T$ is high and for $JAB_a^{(2,1)}$ only when both $T$ and $\gamma$ are high.

Summarizing our findings regarding effect nonstationarity, we have established that although $\phi \neq 1$ renders BB estimators inconsistent, especially when $T$ is not small and $\gamma$ not large BB inference nevertheless beats consistent AB, provided possible endogeneity of $x_{it}$ is respected. The JES test seems to have the remarkable property to be able to guide towards the technique with smallest RMSE instead of the technique exploiting the valid instruments.

# 6 Empirical results

The above findings will be employed now in a re-analysis of the data and some of the techniques studied in Ziliak (1997). The main purpose of that article was to expose the downward bias in GMM as the number of moment conditions expands. This is done by

**Table 21.** $P^{\phi}0fJ\text{-}XA^*$

Feasible Sargan-Hansen test: rejection probability

| $\bar{\rho}_{x\varepsilon}=0.0$ | df | | | $\gamma$ | $\theta=0$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | AB | BB | Inc | | $JAB_a^{(2,1)}$ | $JBB_a^{(2,1)}$ | $JES_a^{(2,1)}$ | $JAB_c^{(2,1)}$ | $JBB_c^{(2,1)}$ | $JES_c^{(2,1)}$ | JMAB | JMBB | JESM |
| $T=3$ | 9 | 13 | 4 | 0.20 | 0.046 | 0.317 | 0.548 | 0.044 | 0.281 | 0.503 | 0.248 | 0.374 | 0.371 |
| | | | | 0.50 | 0.050 | 0.116 | 0.185 | 0.045 | 0.091 | 0.140 | 0.249 | 0.349 | 0.347 |
| | | | | 0.80 | 0.061 | 0.062 | 0.057 | 0.055 | 0.043 | 0.033 | 0.257 | 0.389 | 0.383 |
| $T=6$ | 48 | 58 | 10 | 0.20 | 0.035 | 0.160 | 0.474 | 0.027 | 0.204 | 0.643 | 0.032 | 0.429 | 0.485 |
| | | | | 0.50 | 0.036 | 0.082 | 0.222 | 0.028 | 0.072 | 0.216 | 0.032 | 0.409 | 0.463 |
| | | | | 0.80 | 0.044 | 0.045 | 0.069 | 0.030 | 0.026 | 0.020 | 0.039 | 0.411 | 0.459 |
| $T=9$ | 114 | 130 | 16 | 0.20 | 0.007 | 0.006 | 0.179 | 0.021 | 0.129 | 0.634 | 0.022 | 0.440 | 0.502 |
| | | | | 0.50 | 0.006 | 0.004 | 0.116 | 0.022 | 0.062 | 0.264 | 0.021 | 0.428 | 0.486 |
| | | | | 0.80 | 0.009 | 0.003 | 0.053 | 0.025 | 0.023 | 0.021 | 0.025 | 0.418 | 0.474 |

| $\bar{\rho}_{x\varepsilon}=0.0$ | df | | | $\gamma$ | $\theta=1$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | AB | BB | Inc | | $JAB_a^{(2,1)}$ | $JBB_a^{(2,1)}$ | $JES_a^{(2,1)}$ | $JAB_c^{(2,1)}$ | $JBB_c^{(2,1)}$ | $JES_c^{(2,1)}$ | JMAB | JMBB | JESM |
| $T=3$ | 9 | 13 | 4 | 0.20 | 0.036 | 0.289 | 0.505 | 0.037 | 0.240 | 0.412 | 0.294 | 0.622 | 0.605 |
| | | | | 0.50 | 0.039 | 0.101 | 0.169 | 0.041 | 0.101 | 0.146 | 0.299 | 0.611 | 0.601 |
| | | | | 0.80 | 0.056 | 0.048 | 0.060 | 0.060 | 0.057 | 0.063 | 0.296 | 0.614 | 0.605 |
| $T=6$ | 48 | 58 | 10 | 0.20 | 0.017 | 0.224 | 0.660 | 0.020 | 0.192 | 0.596 | 0.037 | 0.742 | 0.769 |
| | | | | 0.50 | 0.019 | 0.082 | 0.327 | 0.022 | 0.071 | 0.201 | 0.035 | 0.739 | 0.764 |
| | | | | 0.80 | 0.023 | 0.024 | 0.074 | 0.032 | 0.035 | 0.046 | 0.043 | 0.744 | 0.766 |
| $T=9$ | 114 | 130 | 16 | 0.20 | 0.001 | 0.004 | 0.350 | 0.016 | 0.134 | 0.642 | 0.025 | 0.771 | 0.793 |
| | | | | 0.50 | 0.001 | 0.002 | 0.219 | 0.016 | 0.056 | 0.264 | 0.023 | 0.769 | 0.791 |
| | | | | 0.80 | 0.000 | 0.001 | 0.068 | 0.024 | 0.025 | 0.039 | 0.028 | 0.772 | 0.793 |

$^*R = 10000$ simulation replications. Design parameter values: $N = 200$, $SNR = 3$, $DEN_y = 1.0$, $EVF_x = 0.0$, $\bar{\rho}_{x\varepsilon} = 0.0$, $\xi = 0.8$, $\kappa = 0.00$, $\sigma_\varepsilon = 1$, $q = 1$, $\phi = 0.5$. These yield the DGP parameter values: $\pi_\lambda = 0.00$, $\pi_\eta = 0.00$, $\sigma_v = 0.60$, $\sigma_\eta = 1.0*(1-\gamma)$, $\rho_{v\varepsilon} = 0.0$ (and $\bar{\rho}_{x\eta} = 0.00$, $\bar{\rho}_{x\lambda} = 0.00$).

estimating a static life-cycle labor-supply model for a ten year balanced panel of males, and comparing for various implementations of 2SLS and GMM the coefficient estimates and their estimated standard errors when exploiting expanding sets of instruments. We find this approach rather naive for various reasons: ($a$) the difference between empirical coefficient estimates will at best provide a very poor proxy to any underlying difference in bias; ($b$) standard asymptotic variance estimates of IV estimators are known to be very poor representations of true estimator uncertainty;[11] ($c$) the whole analysis is based on just one sample and possibly the model is seriously misspecified.[12] The latter issue also undermines conclusions drawn in Ziliak (1997) on overrejection by the $J$ test, because it is of course unknown in which if any of his empirical models the null hypothesis is true. To avoid such criticism we designed the controlled experiments in the two foregoing sections on the effects of different sets of instruments on various relevant inference techniques. And now we will examine how these simulation results can be exploited to underpin actual inference from the data set used by Ziliak.

This data set originates from waves XII-XXI and the years 1979-1988 of the PSID. The subjects are $N = 532$ continuously married working men aged 22-51 in 1979. Ziliak (1997) employs the static model[13]

$$\ln h_{it} = \beta \ln w_{it} + z_{it}'\gamma + \eta_i + \varepsilon_{it}, \qquad (6.1)$$

where $h_{it}$ is the observed annual hours of work, $w_{it}$ the hourly real wage rate, $z_{it}$ a vector of four characteristics (kids, disabled, age, age-squared), $\eta_i$ an individual effect and $\varepsilon_{it}$ the idiosyncratic error term. It is assumed that $\ln w_{it}$ may be an endogenous regressor and that all variables included in $z_{it}$ are predetermined. The parameter of interest is $\beta$ and its GMM estimates range from approximately 0.07 to 0.52, depending on the

---

[11]See findings in Kiviet (2013) and in many of its references.

[12]Baltagi et al. (2005) study a similar life-cycle labor-supply model for physicians in Norway. They consider a dynamic model, and this rejects the static specification used by Ziliak (1997).

[13]This static model is also used extensively for illustrative purposes in Cameron and Trivedi (2005).

# Table 22. P⁰5fc-EA*

**Feasible coefficient estimators for Arellano-Bond**

| $\bar\rho_{\varepsilon\varepsilon}=0.3$ | | | θ=0 AB1 | | | θ=0 AB2a | | | θ=0 AB2c | | | θ=1 AB1 | | | θ=1 AB2a | | | θ=1 AB2c | | | MAB | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | L | $\gamma$ | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE |
| T=3 | 6 | 0.20 | -0.124 | 0.158 | 0.201 | -0.123 | 0.160 | 0.202 | -0.135 | 0.158 | 0.208 | -0.169 | 0.193 | 0.256 | -0.167 | 0.191 | 0.254 | -0.182 | 0.184 | 0.259 | -0.177 | 0.191 | 0.260 |
| | | 0.50 | -0.171 | 0.199 | 0.263 | -0.171 | 0.204 | 0.266 | -0.175 | 0.209 | 0.273 | -0.241 | 0.251 | 0.349 | -0.243 | 0.251 | 0.350 | -0.253 | 0.247 | 0.354 | -0.250 | 0.247 | 0.352 |
| | | 0.80 | -0.179 | 0.235 | 0.296 | -0.182 | 0.242 | 0.303 | -0.173 | 0.244 | 0.299 | -0.292 | 0.335 | 0.444 | -0.311 | 0.332 | 0.455 | -0.312 | 0.327 | 0.452 | -0.298 | 0.328 | 0.443 |
| T=6 | 30 | 0.20 | -0.096 | 0.053 | 0.110 | -0.099 | 0.058 | 0.114 | -0.106 | 0.055 | 0.119 | -0.144 | 0.070 | 0.160 | -0.135 | 0.069 | 0.151 | -0.149 | 0.065 | 0.163 | -0.145 | 0.063 | 0.158 |
| | | 0.50 | -0.122 | 0.065 | 0.138 | -0.128 | 0.072 | 0.147 | -0.124 | 0.067 | 0.141 | -0.187 | 0.087 | 0.206 | -0.178 | 0.087 | 0.198 | -0.183 | 0.082 | 0.200 | -0.197 | 0.083 | 0.214 |
| | | 0.80 | -0.116 | 0.074 | 0.137 | -0.121 | 0.083 | 0.147 | -0.111 | 0.075 | 0.134 | -0.203 | 0.109 | 0.230 | -0.191 | 0.110 | 0.221 | -0.196 | 0.102 | 0.221 | -0.192 | 0.105 | 0.219 |
| T=9 | 72 | 0.20 | -0.094 | 0.035 | 0.100 | -0.096 | 0.040 | 0.104 | -0.101 | 0.036 | 0.107 | -0.142 | 0.048 | 0.150 | -0.134 | 0.047 | 0.142 | -0.144 | 0.043 | 0.150 | -0.125 | 0.039 | 0.131 |
| | | 0.50 | -0.110 | 0.041 | 0.118 | -0.116 | 0.047 | 0.125 | -0.112 | 0.043 | 0.120 | -0.169 | 0.057 | 0.179 | -0.163 | 0.056 | 0.172 | -0.163 | 0.052 | 0.171 | -0.159 | 0.049 | 0.167 |
| | | 0.80 | -0.095 | 0.045 | 0.105 | -0.101 | 0.052 | 0.113 | -0.092 | 0.046 | 0.103 | -0.165 | 0.068 | 0.179 | -0.157 | 0.067 | 0.171 | -0.157 | 0.062 | 0.169 | -0.150 | 0.062 | 0.162 |

| | L | $\beta$ | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| T=3 | 6 | 1.43 | 0.500 | 0.656 | 0.825 | 0.501 | 0.668 | 0.835 | 0.531 | 0.661 | 0.848 | 0.677 | 0.816 | 1.060 | 0.685 | 0.804 | 1.056 | 0.719 | 0.778 | 1.060 | 0.682 | 0.788 | 1.043 |
| | | 0.93 | 0.477 | 0.582 | 0.753 | 0.477 | 0.591 | 0.759 | 0.484 | 0.607 | 0.777 | 0.660 | 0.741 | 0.992 | 0.668 | 0.728 | 0.988 | 0.690 | 0.723 | 0.999 | 0.663 | 0.717 | 0.977 |
| | | 0.31 | 0.345 | 0.495 | 0.604 | 0.344 | 0.502 | 0.608 | 0.332 | 0.509 | 0.608 | 0.534 | 0.695 | 0.877 | 0.554 | 0.669 | 0.868 | 0.568 | 0.672 | 0.880 | 0.533 | 0.675 | 0.860 |
| T=6 | 30 | 1.43 | 0.350 | 0.196 | 0.401 | 0.366 | 0.215 | 0.425 | 0.372 | 0.201 | 0.423 | 0.515 | 0.257 | 0.576 | 0.495 | 0.255 | 0.557 | 0.529 | 0.240 | 0.581 | 0.484 | 0.227 | 0.535 |
| | | 0.93 | 0.339 | 0.181 | 0.384 | 0.356 | 0.200 | 0.409 | 0.336 | 0.187 | 0.385 | 0.508 | 0.240 | 0.562 | 0.489 | 0.239 | 0.544 | 0.498 | 0.228 | 0.547 | 0.489 | 0.216 | 0.534 |
| | | 0.31 | 0.219 | 0.148 | 0.265 | 0.228 | 0.162 | 0.280 | 0.207 | 0.151 | 0.256 | 0.368 | 0.214 | 0.425 | 0.347 | 0.208 | 0.404 | 0.357 | 0.201 | 0.410 | 0.312 | 0.181 | 0.361 |
| T=9 | 72 | 1.43 | 0.326 | 0.118 | 0.347 | 0.340 | 0.135 | 0.366 | 0.341 | 0.121 | 0.362 | 0.485 | 0.166 | 0.512 | 0.462 | 0.161 | 0.490 | 0.489 | 0.148 | 0.511 | 0.409 | 0.132 | 0.429 |
| | | 0.93 | 0.313 | 0.111 | 0.333 | 0.330 | 0.127 | 0.353 | 0.312 | 0.115 | 0.332 | 0.472 | 0.157 | 0.498 | 0.454 | 0.153 | 0.479 | 0.456 | 0.142 | 0.478 | 0.413 | 0.128 | 0.433 |
| | | 0.31 | 0.193 | 0.089 | 0.213 | 0.202 | 0.100 | 0.226 | 0.184 | 0.090 | 0.205 | 0.327 | 0.135 | 0.354 | 0.307 | 0.128 | 0.332 | 0.312 | 0.122 | 0.335 | 0.258 | 0.104 | 0.279 |

**Feasible coefficient estimators for Blundell-Bond**

| $\bar\rho_{\varepsilon\varepsilon}=0.3$ | | | θ=0 BB1 | | | θ=0 BB2a | | | θ=0 BB2c | | | θ=1 BB1 | | | θ=1 BB2a | | | θ=1 BB2c | | | MBB | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | L | $\gamma$ | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE |
| T=3 | 10 | 0.20 | 0.075 | 0.107 | 0.131 | 0.099 | 0.127 | 0.161 | 0.107 | 0.127 | 0.166 | 0.030 | 0.134 | 0.137 | 0.067 | 0.152 | 0.166 | 0.064 | 0.153 | 0.166 | 0.109 | 0.189 | 0.218 |
| | | 0.50 | 0.007 | 0.111 | 0.111 | 0.032 | 0.123 | 0.127 | 0.038 | 0.122 | 0.128 | -0.048 | 0.147 | 0.155 | -0.014 | 0.157 | 0.158 | -0.013 | 0.160 | 0.160 | 0.036 | 0.160 | 0.160 |
| | | 0.80 | -0.068 | 0.113 | 0.132 | -0.047 | 0.123 | 0.132 | -0.040 | 0.120 | 0.126 | -0.124 | 0.160 | 0.202 | -0.103 | 0.172 | 0.201 | -0.092 | 0.169 | 0.192 | -0.029 | 0.900 | 0.900 |
| T=6 | 40 | 0.20 | -0.001 | 0.047 | 0.047 | 0.002 | 0.054 | 0.055 | 0.023 | 0.060 | 0.064 | -0.048 | 0.062 | 0.079 | -0.024 | 0.065 | 0.069 | -0.004 | 0.071 | 0.071 | 0.004 | 0.075 | 0.075 |
| | | 0.50 | -0.026 | 0.048 | 0.054 | -0.001 | 0.057 | 0.057 | 0.019 | 0.055 | 0.058 | -0.080 | 0.065 | 0.103 | -0.040 | 0.068 | 0.079 | -0.015 | 0.068 | 0.069 | 0.017 | 0.077 | 0.079 |
| | | 0.80 | -0.052 | 0.046 | 0.069 | -0.027 | 0.049 | 0.056 | -0.014 | 0.044 | 0.047 | -0.102 | 0.069 | 0.122 | -0.069 | 0.064 | 0.094 | -0.041 | 0.060 | 0.072 | 0.015 | 1.687 | 1.687 |
| T=9 | 88 | 0.20 | -0.035 | 0.032 | 0.047 | -0.033 | 0.035 | 0.048 | -0.024 | 0.038 | 0.045 | -0.083 | 0.045 | 0.095 | -0.072 | 0.044 | 0.085 | -0.050 | 0.047 | 0.069 | -0.053 | 0.044 | 0.069 |
| | | 0.50 | -0.050 | 0.033 | 0.060 | -0.037 | 0.037 | 0.052 | -0.011 | 0.038 | 0.040 | -0.104 | 0.047 | 0.114 | -0.086 | 0.047 | 0.098 | -0.042 | 0.047 | 0.063 | -0.034 | 0.050 | 0.060 |
| | | 0.80 | -0.055 | 0.032 | 0.063 | -0.038 | 0.034 | 0.051 | -0.015 | 0.030 | 0.033 | -0.104 | 0.047 | 0.114 | -0.086 | 0.045 | 0.097 | -0.038 | 0.039 | 0.055 | -0.018 | 0.044 | 0.047 |

| | L | $\beta$ | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE | Bias | Stdv | RMSE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| T=3 | 10 | 1.43 | -0.017 | 0.473 | 0.473 | -0.066 | 0.544 | 0.548 | -0.105 | 0.526 | 0.536 | 0.169 | 0.582 | 0.606 | 0.113 | 0.660 | 0.670 | 0.118 | 0.636 | 0.647 | 0.110 | 0.628 | 0.638 |
| | | 0.93 | 0.159 | 0.376 | 0.408 | 0.126 | 0.409 | 0.428 | 0.096 | 0.402 | 0.413 | 0.327 | 0.492 | 0.591 | 0.295 | 0.523 | 0.601 | 0.289 | 0.520 | 0.594 | 0.327 | 0.604 | 0.687 |
| | | 0.31 | 0.208 | 0.287 | 0.355 | 0.186 | 0.301 | 0.354 | 0.162 | 0.295 | 0.337 | 0.342 | 0.404 | 0.530 | 0.332 | 0.420 | 0.536 | 0.316 | 0.409 | 0.517 | 0.377 | 2.290 | 2.321 |
| T=6 | 40 | 1.43 | 0.120 | 0.171 | 0.208 | 0.116 | 0.195 | 0.227 | 0.056 | 0.200 | 0.208 | 0.291 | 0.216 | 0.362 | 0.238 | 0.227 | 0.329 | 0.192 | 0.232 | 0.301 | 0.152 | 0.229 | 0.275 |
| | | 0.93 | 0.165 | 0.144 | 0.219 | 0.126 | 0.164 | 0.207 | 0.076 | 0.156 | 0.173 | 0.326 | 0.193 | 0.378 | 0.256 | 0.198 | 0.323 | 0.205 | 0.190 | 0.280 | 0.149 | 0.195 | 0.246 |
| | | 0.31 | 0.155 | 0.111 | 0.191 | 0.126 | 0.114 | 0.170 | 0.100 | 0.105 | 0.145 | 0.276 | 0.159 | 0.319 | 0.228 | 0.149 | 0.273 | 0.194 | 0.139 | 0.239 | 0.158 | 1.937 | 1.943 |
| T=9 | 88 | 1.43 | 0.186 | 0.107 | 0.215 | 0.184 | 0.118 | 0.219 | 0.154 | 0.123 | 0.197 | 0.351 | 0.149 | 0.381 | 0.324 | 0.145 | 0.355 | 0.274 | 0.146 | 0.310 | 0.246 | 0.135 | 0.281 |
| | | 0.93 | 0.201 | 0.095 | 0.222 | 0.180 | 0.107 | 0.209 | 0.122 | 0.105 | 0.161 | 0.358 | 0.136 | 0.383 | 0.323 | 0.134 | 0.350 | 0.240 | 0.129 | 0.272 | 0.195 | 0.127 | 0.233 |
| | | 0.31 | 0.154 | 0.074 | 0.171 | 0.136 | 0.079 | 0.157 | 0.099 | 0.070 | 0.122 | 0.272 | 0.112 | 0.294 | 0.245 | 0.105 | 0.266 | 0.183 | 0.092 | 0.205 | 0.147 | 0.087 | 0.171 |

*$R = 10000$ simulation replications. Design parameter values: $N = 200$, $SNR = 3$, $DEN_y = 1.0$, $EVF_x = 0.0$, $\bar\rho_{x\varepsilon} = 0.3$, $\xi = 0.8$, $\kappa = 0.00$, $\sigma_\varepsilon = 1$, $q = 1$, $\phi = 0.5$. These yield the DGP parameter values: $\pi_\lambda = 0.00$, $\pi_\eta = 0.00$, $\sigma_\upsilon = 0.60$, $\sigma_\eta = 1.0 * (1-\gamma)$, $\rho_{\upsilon\varepsilon} = 0.5$ (and $\bar\rho_{x\eta} = 0.00$, $\bar\rho_{x\lambda} = 0.00$).

49

number of instruments employed.

After some experimentation we inferred that lagged reactions play a significant role in this relationship and that in fact a general second-order linear dynamic specification is required in order to pass the diagnostic tests which are provided by default in the Stata package xtabond2, see Roodman (2009a). This model, also allowing for time-effects, is given by

$$
\begin{aligned}
\ln h_{it} &= \sum_{l=1}^{2} \gamma_l \ln h_{i,t-l} + \sum_{l=0}^{2} (\beta_l^w \ln w_{i,t-l} + \beta_l^k kids_{i,t-l} + \beta_l^d disab_{i,t-l}) \\
&\quad + \beta^a age_{i,t} + \beta^{aa} age_{i,t}^2 + \tau_t + \eta_i + \varepsilon_{it}.
\end{aligned}
\tag{6.2}
$$

Note that lags of age and its square would lead to multicollinearity problems. The inclusion of second-order lags yields $T = 7$ in estimating the first-differenced model.

Table 23 presents estimation and test results for model (6.2) obtained by employing various of the techniques examined in the earlier sections. All results have been obtained by Stata/SE11.2 and xtabond2, supplemented with code for calculating $\hat{\sigma}_\eta^2$ and $\hat{\sigma}_\varepsilon^2$. In column (1) 2-step Arellano-Bond GMM estimators are presented (omitting the time-effect results) in which all instruments are being used that follow from assuming that all regressors are predetermined, which gives 170 instruments. For the AR and $J$ tests given in the bottom lines the $p$-values are presented. Hence, in column (1) the (first-differenced) residuals do exhibit 1st order serial correlation (as they should) but no significant 2nd order problems emerge. The presented $J$ test is $JAB_a^{(2,1)}$, which according to the simulations may have a tendency to underreject, and to have moderate rejection probability when a regressor has incorrectly not been categorized as endogenous, especially under heteroskedasticity. Therefore its high $p$-value is not necessarily reassuring regarding validity of the instruments. When abstaining from 7 instruments such that wage is treated as endogenous, while still using as weighting matrix a submatrix of the weights under predeterminedness, the $p$-value of $JAB_a^{(2,1)}$ is still 0.577 and the resulting incremental $J$ test for wage being predetermined is 7.13, which has $p$-value 0.416. Hence, from the pure asymptotic point of view these coefficient estimates and their Windmeijer-corrected standard errors could be taken serious, because yet no alarming signs of misspecifcation have been detected. All coefficient estimates with a $t$-ratio above 2 are indicated with a double asterix, and a single asterix when between 1 and 2; estimated standard errors are given between parentheses. Half of the time-dummies have a $t$-ratio around 1 and the others much smaller. Different though asymptotically equivalent tests for their joint significance (not presented here) yield $\chi^2(7)$ statistics between 8 and 9. So, although not significantly different from zero, simply removing them from the model seems a bit rash.

Next, we will examine the effects of the Windmeijer correction in column (2) and of treating wage not as predetermined, but as endogenous, in column (3). For the squared (not first-differenced) residuals of results (1) and (3) we also ran informal auxiliary regressions (with and without individual effects) on all the predetermined regressors of model (6.2). Due to several highly significant regressors these make obviously clear that $\varepsilon_{it}$ is heteroskedastic and not of the simple cross-sectional type. Hence, 1-step estimates which follow in columns (4) through (7) should be judged on the basis of robust standard errors only.

Column (2) presents crude asymptotic standard errors, which are as expected much smaller for all coefficients. From the simulations we know that especially under het-

**Table 23.** Empirical findings for (restricted versions of) model (6.2) by AB estimation

| | (1) AB2aW WA | (2) AB2a WA | (3) AB2aW EA | (4) AB1aR WA | (5) AB1aR EA | (6) AB1aR WC | (7) AB1aR EC | (8) AB2aW EC |
|---|---|---|---|---|---|---|---|---|
| $\gamma_1$ | 0.178** | 0.178** | 0.199** | 0.177** | 0.206** | 0.237** | 0.238** | 0.242** |
| | (0.057) | (0.012) | (0.061) | (0.064) | (0.069) | (0.096) | (0.101) | (0.078) |
| $\gamma_2$ | 0.050* | 0.050** | 0.080** | 0.033* | 0.068* | 0.062** | 0.055* | 0.056* |
| | (0.032) | (0.009) | (0.027) | (0.031) | (0.028) | (0.026) | (0.028) | (0.030) |
| $\beta_0^w$ | 0.403** | 0.403** | 0.462** | 0.566** | 0.607** | 0.254* | 0.131 | 0.203 |
| | (0.157) | (0.041) | (0.153) | (0.174) | (0.172) | (0.179) | (0.253) | (0.284) |
| $\beta_1^w$ | 0.121* | 0.121** | -0.053 | 0.154* | -0.012 | 0.009 | 0.045 | 0.124* |
| | (0.062) | (0.024) | (0.100) | (0.073) | (0.116) | (0.106) | (0.115) | (0.110) |
| $\beta_2^w$ | 0.001 | 0.001 | -0.063* | -0.003 | -0.077* | -0.058 | -0.049 | 0.002 |
| | (0.035) | (0.016) | (0.050) | (0.044) | (0.062) | (0.068) | (0.068) | (0.071) |
| $\beta_0^k$ | -0.037* | -0.037** | -0.033* | -0.051** | -0.049** | -0.044* | -0.040 | -0.024* |
| | (0.020) | (0.011) | (0.017) | (0.024) | (0.023) | (0.023) | (0.024) | (0.023) |
| $\beta_1^k$ | -0.002 | -0.002 | 0.000 | 0.007 | 0.010 | -0.004 | -0.006 | 0.004 |
| | (0.009) | (0.007) | (0.009) | (0.011) | (0.011) | (0.012) | (0.013) | (0.011) |
| $\beta_2^k$ | 0.009 | 0.009* | 0.008 | 0.008 | 0.006 | 0.010 | 0.010 | 0.005 |
| | (0.009) | (0.006) | (0.010) | (0.012) | (0.012) | (0.021) | (0.014) | (0.011) |
| $\beta_0^d$ | 0.027 | 0.027* | 0.027 | -0.005 | 0.000 | 0.049 | 0.047 | -0.039 |
| | (0.043) | (0.021) | (0.044) | (0.047) | (0.045) | (0.077) | (0.077) | (0.061) |
| $\beta_1^d$ | 0.027 | 0.027* | 0.027 | 0.011 | 0.016 | 0.059 | 0.061 | -0.036 |
| | (0.042) | (0.017) | (0.042) | (0.050) | (0.051) | (0.074) | (0.078) | (0.050) |
| $\beta_2^d$ | 0.038* | 0.038** | 0.046 | 0.042* | 0.054* | 0.087** | 0.083** | 0.046* |
| | (0.025) | (0.011) | (0.027) | (0.030) | (0.033) | (0.031) | (0.029) | (0.024) |
| $\beta^a$ | -0.001 | -0.001 | 0.003 | -0.007 | -0.003 | 0.006 | 0.008 | -0.016 |
| | (0.019) | (0.012) | (0.019) | (0.021) | (0.020) | (0.021) | (0.021) | (0.018) |
| $\beta^{aa}$ | -0.000 | -0.000 | -0.000 | -0.000 | -0.000 | 0.000 | -0.000 | 0.000 |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| | | | | | | | | |
| $K$ | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 |
| $L$ | 170 | 170 | 163 | 170 | 163 | 44 | 43 | 43 |
| $AR(1)$ | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 |
| $AR(2)$ | 0.549 | 0.479 | 0.515 | 0.255 | 0.167 | 0.239 | 0.218 | 0.617 |
| $J$ | 0.573 | 0.573 | 0.656 | 0.573 | 0.656 | 0.295 | 0.221 | 0.221 |
| $\hat{\sigma}_\eta^2$ | 0.048 | 0.048 | 0.027 | 0.079 | 0.048 | 0.017 | 0.015 | 0.026 |
| $\hat{\sigma}_\varepsilon^2$ | 0.054 | 0.054 | 0.056 | 0.056 | 0.058 | 0.055 | 0.055 | 0.056 |
| $TM^w$ | 0.680 | 0.680 | 0.480 | 0.908 | 0.713 | 0.292 | 0.173 | 0.452 |

eroskedasticity we should distrust the uncorrected estimated standard errors of column (2). Column (3) shows that although some estimated standard errors substantially increase also many remain almost the same or even decrease when we treat log wage as endogenous. Especially the findings in column (3) on the coefficients $\gamma_1$ and $\gamma_2$ strongly reject the static specifications maintained in Ziliak (1997) and in Cameron and Trivedi (2005) for the same data. On the other hand the estimated values are such that the simulation results on large $\gamma$ values do not seem relevant here. Testing the endogeneity by the difference between the two obtained $JAB_a^{(2,1)}$ statistics yields 10.52 (with $p$-value 0.161) leaving us still in doubt whether we should simply accept predeterminedness of wage, or err on the side of caution and treat it as endogenous. The rather indeterminate test results on endogeneity are in agreement with the finding that the coefficient estimates in columns (1) and (3) are only slightly different. They imply an immediate wage elasticity of hours worked ($\hat{\beta}_0^w$) of 0.40 and 0.46 respectively, and an estimated long-run elasticity $TM^w = (\hat{\beta}_0^w + \hat{\beta}_1^w + \hat{\beta}_2^w)/(1 - \hat{\gamma}_1 - \hat{\gamma}_2)$ of 0.68 and (assuming endogeneity

of wage) 0.48. A static model self-evidently would produce equivalent point-estimates for the immediate and the long-run wage elasticity of hours worked. The estimates in columns (1) and (3) suggest a value of $DEN_y^\eta$ of about 1.22 or 0.96 respectively, which we interpret such that the relatively unfavorable simulation results for case P1 where it is 4 do not apply here.

Next we estimate the same model again, but now by robust 1-step Arellano-Bond, assuming all variables to be predetermined in column (4) and wage to be endogenous in column (5). Most AB1aR standard errors are a bit larger than those for the corresponding AB2aW. This is in agreement with the differences between true standard deviations that we noted within simulations for P0 and P5. Although these simulations did not produce a serious difference between the coefficient bias of AB2 and AB1 the results in column (4) yield a much larger estimated $TM^w$ of 0.908. The test on overidentification provided by xtabond2[14] (after estimating by AB1aR) is again $JAB_a^{(2,1)}$ and not $JAB_a^{(1,1)}$.

Next we re-estimate the models of columns (4) and (5) by AB1aR, though employing substantially fewer instruments by collapsing. The results in columns (6) and (7) show some remarkable changes, both regarding coefficient estimates (yielding substantially lower $TM^w$) and regarding standard errors, which are often larger (as expected) but sometimes much lower. According to the simulation results both (6) and (7) will produce seriously biased coefficient estimates in the presence of a genuinely endogenous regressor, and both virtually unbiased estimators if all regressors are actually predetermined. The two versions of the incremental test (with 1 degree of freedom) on endogeneity give here $p$-values of 0.100 and 0.078 respectively. The collapsed AB2aW results in column (8) are less extreme and more in line with those of column (3). An unpleasant finding in columns (7) and (8) is that they lead to rejection by the incremental $J$ test of validity as instruments of the time dummies. Since these are exogenous by nature this raises doubts on the adequacy of the model specification. Such doubts may also be fed by the fact that all columns, except (8), suggest that a disability leads to increased labor supply, although this might perhaps be realistic in the US.

We also employed BB2aW estimation, for E and W, and for A and C. When using all instruments we found $p$-values for the incremental test on effect stationarity of 0.061 and 0.045. However, after collapsing these are 0.314 and 0.252. This is a pattern that we noted in our simulations and gives reasons to believe that we should reject effect stationarity and take this test outcome as a warning that BB inference may be here less accurate than AB inference. The BB coefficient estimates imply an insignificant but positive immediate wage elasticity, whereas the long-run elasticity for all four is close to zero or even negative. This could well be due to the huge coefficient bias that we noted in our experiments with $\phi = 0.5$, especially those with an endogenous regressor.

Hence, supposing that for these data wage is endogenous indeed, that the effects are nonstationary and heteroskedasticity is present, our cases $P^\phi 5$-EA and $P^\phi 5$-EC with moderate $\gamma$, $T = 6$ and $\theta = 1$ seem most relevant. For these we learned that collapsing has some advantages, but otherwise there is little to choose between AB1a and AB2a, because both yield substantial negative bias (about -50%) for $\gamma$ and huge positive bias (about +30%) for $\beta$ with actual test size of corrected $t$-tests close to 25%, although in

---

[14]From our simulations we found that $JAB^{(1,0)}$, which is the test addressed as Sargan by xtabond2, has a rejection probability close to 1 under heteroskedasticity. So, it is useless after AB2 estimation and even misleading due to the comment "Not robust, but not weakened by many instruments".

this data set $N$ is actually larger than 200. Nevertheless it seems a truism to conclude that there is great urge for developing more refined inference procedures for structural dynamic panel data models.

# 7  Major findings

In social science the quantitative analysis of many highly relevant problems requires structural dynamic panel data methods. These allow the observed data to have at best a quasi-experimental nature, whereas the causal structure and the dynamic interactions in the presence of unobserved heterogeneity have yet to be unraveled. When the cross-section dimension of the sample is not very small, employing GMM techniques seems most appropriate in such circumstances. This is also practical since corresponding software packages are widely available. However, not too much is known yet about the actual accuracy in practical situations on the abundance of different not always asymptotically equivalent implementations of estimators and test procedures. This study aims to demarcate the areas in the parameter space where the asymptotic approximations to the properties of the relevant inference techniques in this context have either shown to be reliable beacons or are actually often misguiding marsh fires.

In this context we provide a rather rigorous treatment of many major variants of GMM implementations as well as for the inference techniques on testing the validity of particular orthogonality assumptions and restrictions on individual coefficient values. Special attention is given to the consequences of the joint presence in the model of time-constant and individual-constant unobserved effects, covariates that may be strictly exogenous, predetermined or endogenous, and disturbances that may show particular forms of heteroskedasticity. Also the implications regarding initial conditions for separate regressors with respect to individual effect stationarity are analyzed in great detail, and various popular options that aim to mitigate bias by reducing the number of exploited internal instruments are elucidated. In addition, as alternatives to those used in current standard software, less robust weighting matrices and additional variants of Sargan-Hansen test implementations are considered, as well as the effects of particular modifications of the instruments under heteroskedasticity.

Next, a simulation study is designed in which all the above variants and details are being parametrized and categorized, which leads to a data generating process involving 10 parameters, for which, under 6 different settings regarding sample size and initial conditions, 60 different grid points are examined. For each setting and various of the grid points 13 different choices regarding the set of instruments have been used to examine 12 different implementations of GMM coefficient estimates, giving rise to 24 different implementations of $t$-tests and 27 different implementations of Sargan-Hansen tests. From all this only a pragmatically selected subset of results is actually presented in this paper.

The major conclusion from the simulations is that, even when the cross-section sample size is several hundreds, the quality of this type of inference depends heavily on a great number of aspects of which many are usually beyond the control of the investigator, such as: magnitude of the time-dimension sample size, speed of dynamic adjustment, presence of any endogenous regressors, type and severity of heteroskedasticity, relative prominence of the individual effects and (non)stationarity of the effect impact of any

of the explanatory variables. The quality of inference also depends seriously on choices made by the investigator, such as: type and severity of any reductions applied regarding the set of instruments, choice between (robust) 1-step or (corrected) 2-step estimation, employing a modified GMM estimator, the chosen degree of robustness of the adopted weighting matrix, the employed variant of coefficient tests and of (incremental) Sargan-Hansen tests in deciding on the endogeneity of regressors, the validity of instruments and on the (dynamic) specification of the relationship in general.

Our findings regarding the alternative approaches of modifying instruments and exploiting different weighting matrices are as follows for the examined case of cross-sectional heteroskedasticity. Although the unfeasible form of modification does yield very substantial reductions in both bias and variance, for the straight-forward feasible implementation examined here the potential efficiency gains do not materialize. The robust weighting matrix, which also allows for possible time-series heteroskedasticity, performs often as well as (and sometimes even better than) a specially designed less robust version, although the latter occasionally demonstrates some benefits for incremental Sargan-Hansen tests.

Furthermore we can report to practitioners: ($a$) when the effect-noise-ratio is large, the performance of all GMM inference deteriorates; ($b$) the same occurs in the presence of a genuine (or a supervacaneously treated as) endogenous regressor; ($c$) in many settings the coefficient restrictions tests show serious size problems which usually can be mitigated by a Windmeijer correction, although for $\gamma$ large or under simultaneity serious overrejection remains unless $N$ is very much larger than 200; ($d$) the limited effectiveness of the Windmeijer correction is due to the fact that the positive or negative bias in coefficient estimates is often more serious than the negative bias in the variance estimate; ($e$) limiting to some degree the number of instruments usually reduces bias and therefore improves size properties of coefficient tests, though at the potential cost of power loss because efficiency usually suffers; ($f$) for the case of an autoregressive strictly exogenous regressor we noted that it is better to not just instrument it by itself, but also by some of its lags because this improves inference, especially regarding the lagged dependent variable coefficient; ($g$) to mitigate size problems of the overall Sargan-Hansen overidentification tests the set of instruments should be reduced, possibly by collapsing, and then, after 2-step estimation, none of the examined alternative variants did systematically beat the one using the standard robust weighting matrix based on 1-step residuals; ($h$) collapsing also reduces size problems of the incremental Sargan-Hansen effect stationarity test; ($i$) except under simultaneity, the GMM estimator which exploits instruments which are invalid under effect nonstationarity may nevertheless perform better than the estimator abstaining from these instruments; ($j$) the rejection probability of the incremental Sargan-Hansen test for effect stationarity is such that it tends to direct the researcher towards applying the most accurate estimator, even if this is inconsistent.

When re-analyzing a popular empirical data set in the light of the above simulation findings we note in particular that actual dynamic feedbacks may be much more subtle than those that can be captured by just including a lagged dependent variable regressor, which at present seems the most common approach to model dynamics in panels. In theory the omission of further lagged regressor variables should result in rejections by Sargan-Hansen test statistics, but their power suffers when many valid and some invalid orthogonality conditions are tested jointly, instead of by deliberately chosen sequences of incremental tests or direct variable addition tests. Hopefully tests for serial correlation,

which we intentionally left out of this already overloaded study, provide an extra help to practitioners in guiding them towards well-specified models.

# References

Ahn, S.C., Schmidt, P., 1995. Efficient estimation of models for dynamic panel data. *Journal of Econometrics* 58, 3-29.

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

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

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

Baltagi, B.H., Bratberg, E., Holmås, T.H., 2005. A panel data study of physicians' labor supply: the case of Norway, *Health Economics* 14, 1035-1045.

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

Blundell, R., Bond, S.R., Windmeijer, F., 2001. Estimation in dynamic panel data models: Improving on the performance of the standard GMM estimator. *Advances in Economtrics* 15(1), 53-91.

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

Bowsher, C.G., 2002. On testing overidentifying restrictions in dynamic panel data models. *Economics Letters* 77, 211–220.

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

Bun, M.J.G., Carree, M.A., 2005b. Correction "Bias-corrected estimation in dynamic panel data models". *Journal of Business and Economic Statistics* 23, 363.

Bun, M.J.G., Carree, M.A., 2006. Bias-corrected estimation in dynamic panel data models with heteroscedasticity. *Economics Letters* 92, 220-227.

Bun, M.J.G., Kiviet, J.F., 2002. On the diminishing retruns of higher-order terms in asymptotic expansions of bias. *Economics Letters* 79, 145-152.

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

Bun, M.J.G., Sarafidis, V., 2013. Dynamic panel data models. UvA-Econometrics discussion paper 2013/01.

Cameron, A.C., Trivedi, P.K., 2005. *Microeconometrics: Methods and Applications.* Cambridge University Press.

Dheane, G., Jochmans, K., 2012. An adjusted profile likelihood for non-stationary panel data models with fixed effects. Mimeo.

Everaert, G., 2013. Orthogonal to backward mean transformation for dynamic panel data models. *The Econometrics Journal* 16, 179-221.

Flannery, M.J., Watson Hankins, K., 2013. Estimating dynamic panel models in corporate finance. *Journal of Corporate Finance* 19, 1-19.

Gouriéroux, C., Phillips, P.C.B., Yu, J., 2010. Indirect inference for dynamic panel models. *Journal of Econometrics* 157, 68-77.

Han, C., Phillips, P.C.B., 2013. First difference maximum likelihood and dynamic panel estimation. *Journal of Econometrics* 175, 33-45.

Hahn, J., Hausman, J., Kuersteiner, G., 2007. Long difference instrumental variables estimation for dynamic panel models with fixed effects. *Journal of Econometrics* 140, 574-617.

Hahn, J., Kuersteiner, G., 2002. Asymptotically unbiased inference for a dynamic panel model with fixed effects when both $n$ and $T$ are large. *Econometrica* 70, 1639-1657.

Harris, M., Kostenko, W., Mátthyás, L., Timol, I., 2009. The robustness of estimators for dynamic panel data to misspecification. *The Singapore Economic Review* 54, 399-426.

Hayakawa, K., 2009. On the effects of mean-stationarity in dynamic panel data models. *Journal of Econometrics* 153, 133-135.

Hayakawa, K., 2010. The effects of dynamic feedbacks on LS and MM estimator accuracy in panel data models: Some additional results. *Journal of Econometrics* 159, 202-208.

Hayashi, Fumio, 2000. *Econometrics*. Princeton, Princeton University Press.

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

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

Juodis, A., 2013. A note on bias-corrected estimation in dynamic panel data models. *Economics Letters* 118, 435-438.

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

Kiviet, J.F., 2007. Judging contending estimators by simulation: Tournaments in dynamic panel data models. Chapter 11 (pp.282-318) in Phillips, G.D.A., Tzavalis, E. (Eds.). *The Refinement of Econometric Estimation and Test Procedures; Finite Sample and Asymptotic Analysis*. Cambridge University Press.

Kiviet, J.F., 2012. Monte Carlo Simulation for Econometricians. *Foundations and Trends in Econometrics* 5 (1-2), 1-181.

Kiviet, J.F., 2013. Identification and inference in a simultaneous equation under alternative information sets and sampling schemes. *The Econometrics Journal* 16, S24-S59.

Kiviet, J.F., Feng, Q., 2014. Efficiency gains by modifying GMM estimation in linear models under heteroskedasticity. Mimeo.

Kripfganz, S., Schwarz, C., 2013. Estimation of linear dynamic panel data models with time-invariant regressors. Mimeo.

Kruiniger, H., 2008. Maximum likelihood estimation and and inference methods for the covariance stationary panel AR(1)/unit root model. *Journal of Econometrics* 144, 447-464.

Okui, R., 2009. The optimal choice of moments in dynamic panel data models. *Journal of Econometrics* 151, 1-16.

Roodman, D., 2009a. How to do xtabond2: An introduction to difference and system GMM in Stata. *The Stata Jounal* 9, 86-136.

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

Windmeijer, F., 2005. A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics* 126, 25-51.

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

# Appendices

# A   Corrected variance estimation for 2-step GMM

Windmeijer (2005) provides a correction to the standard expression for the estimated variance of the 2-step GMM estimator in general nonlinear models and next specializes his results for models with linear moment conditions and finally for linear (panel data) models. Here we apply his approach directly to the standard linear model of section 2.1 where $\hat{\beta}^{(2)}$ is based on weighting matrix $Z'\hat{\Omega}^{(1)}Z$, where $\hat{\Omega}^{(1)}$ depends on $\hat{u}^{(1)}$ and thus on $\hat{\beta}^{(1)}$.

The nonlinear dependence of $\hat{\beta}^{(2)}$ on $\hat{\beta}^{(1)}$ can be made explicit by a linear approximation obtained by employing the well-known 'delta-method' to the vector function $f(\beta) = \{X'Z[Z'\Omega(\beta)Z]^{-1}Z'X\}^{-1}X'Z[Z'\Omega(\beta)Z]^{-1}Z'u$. Note that $\hat{\beta}^{(2)} = \beta_0 + f(\hat{\beta}^{(1)})$. Expanding the second term around $\beta_0$ yields

$$\hat{\beta}^{(2)} - \beta_0 \approx f(\beta_0) + \left.\frac{\partial f(\beta)}{\partial \beta'}\right|_{\beta=\beta_0} (\hat{\beta}^{(1)} - \beta_0), \tag{A.1}$$

where under sufficient regularity the omitted terms will be of small order. For $k = 1, ..., K$ we find

$$\begin{aligned}\frac{\partial f(\beta)}{\partial \beta_k} &= \frac{\partial \{X'Z[Z'\Omega(\beta)Z]^{-1}Z'X\}^{-1}}{\partial \beta_k} X'Z[Z'\Omega(\beta)Z]^{-1}Z'u \\ &+ \{X'Z[Z'\Omega(\beta)Z]^{-1}Z'X\}^{-1} \frac{\partial X'Z[Z'\Omega(\beta)Z]^{-1}Z'u}{\partial \beta_k},\end{aligned}$$

where

$$\begin{aligned}\frac{\partial \{X'Z[Z'\Omega(\beta)Z]^{-1}Z'X\}^{-1}}{\partial \beta_k} &= -\{X'Z[Z'\Omega(\beta)Z]^{-1}Z'X\}^{-1}X'Z\frac{\partial [Z'\Omega(\beta)Z]^{-1}}{\partial \beta_k}Z'X \\ &\times \{X'Z[Z'\Omega(\beta)Z]^{-1}Z'X\}^{-1},\end{aligned}$$

with

$$\frac{\partial [Z'\Omega(\beta)Z]^{-1}}{\partial \beta_k} = -[Z'\Omega(\beta)Z]^{-1}Z'\frac{\partial \Omega(\beta)}{\partial \beta_k}Z[Z'\Omega(\beta)Z]^{-1},$$

and

$$\frac{\partial \{X'Z[Z'\Omega(\beta)Z]^{-1}Z'u\}}{\partial \beta_k} = X'Z\frac{\partial [Z'\Omega(\beta)Z]^{-1}}{\partial \beta_k}Z'u.$$

In the latter we omit an extra term in $\partial u/\partial \beta_k = \partial(y - X\beta)/\partial \beta_k$ simply because we just want to extract the dependence of $\hat{\beta}^{(2)}$ on the operational weighting matrix.

Using the short-hand notation $A(\beta) = Z[Z'\Omega(\beta)Z]^{-1}Z'$ and $\Omega_k(\beta) = \partial\Omega(\beta)/\partial \beta_k$ we can establish from the above that

$$\partial f(\beta)/\partial \beta_k = -[X'A(\beta)X]^{-1}X'A(\beta)\Omega_k(\beta)\{I - A(\beta)X[X'A(\beta)X]^{-1}X'\}A(\beta)u.$$

This is the $k$-th column of the matrix $F(\beta) = \partial f(\beta)/\partial \beta'$ in (A.1). The latter can now be expressed as

$$\hat{\beta}^{(2)} - \beta_0 \approx \{[X'A(\beta_0)X]^{-1}X'A(\beta_0) + F(\beta_0)(X'P_ZX)^{-1}X'P_Z\}u. \qquad (A.2)$$

Because $F(\beta_0) = O_p(n^{-1/2})$ the second term is of smaller order.

This approximation to the estimation errors of $\hat{\beta}^{(2)}$ can be used to obtain a finite sample corrected variance estimate of $\hat{\beta}^{(2)}$. This is relatively easy if one conditions on some value for $F(\beta_0)$, say $\hat{F}$. Windmeijer chooses for the $k$-th column of $\hat{F}$ the vector

$$-[X'A(\hat{\beta}^{(1)})X]^{-1}X'A(\hat{\beta}^{(1)})\Omega_k(\hat{\beta}^{(1)})\{I - A(\hat{\beta}^{(1)})X[X'A(\hat{\beta}^{(1)})X]^{-1}X'\}A(\hat{\beta}^{(1)})\hat{u}^{(2)}.$$

Taking $\hat{u}^{(2)}$ instead of the asymptotically equivalent $\hat{u}^{(1)}$ leads to substantial simplification, because $X'A(\hat{\beta}^{(1)})\hat{u}^{(2)} = X'Z[Z'\Omega(\beta^{(1)})Z]^{-1}Z'(y - X\hat{\beta}^{(2)}) = 0$, giving

$$\hat{F} = (\hat{F}_{.1}, ..., \hat{F}_{.K}), \text{ with } \hat{F}_{.k} = -[X'A(\hat{\beta}^{(1)})X]^{-1}X'A(\hat{\beta}^{(1)})\Omega_k(\hat{\beta}^{(1)})A(\hat{\beta}^{(1)})\hat{u}^{(2)}. \qquad (A.3)$$

Note that when $L = K$ we have $\hat{F} = O$, because $Z'\hat{u}^{(1)} = Z'\hat{u}^{(2)} = 0$.

This all then yields for $L > K$ the corrected variance estimator

$$\widehat{Varc}(\hat{\beta}^{(2)}) = \widehat{Var}(\hat{\beta}^{(2)}) + \hat{F}\widehat{Var}(\hat{\beta}^{(2)}) + \widehat{Var}(\hat{\beta}^{(2)})\hat{F} + \hat{F}\widehat{Var}(\hat{\beta}^{(1)})\hat{F}', \qquad (A.4)$$

where

$$\widehat{Var}(\hat{\beta}^{(1)}) = \hat{\sigma}_u^2(X'P_ZX)^{-1}X'P_Z\hat{\Omega}P_ZX(X'P_ZX)^{-1}.$$

Note that in case $\Omega(\beta) = diag(u_1^2, ..., u_n^2)$ one has $\Omega_k(\hat{\beta}^{(1)}) = -2\hat{\beta}_k^{(1)}diag(\hat{u}_1^{(1)}x_{1k}, ..., \hat{u}_n^{(1)}x_{nk})$.

# B  Partialling out and GMM

The IV/2SLS result on partialling out directly generalizes for the MGMM estimator, provided this uses all the (transformed) predetermined regressors as instruments. In standard GMM the equivalence of predetermined regressors and a block of the instruments gets lost. Using the notation of (2.10) and considering the partitioned model leading to (2.8), we easily find its counterpart

$$\hat{\beta}_{1,GMM} = (\hat{X}_1^{\dagger*\prime}M_{\hat{X}_2^{\dagger*}}\hat{X}_1^{\dagger*})^{-1}\hat{X}_1^{\dagger*\prime}M_{\hat{X}_2^{\dagger*}}y^*, \qquad (B.1)$$

where

$$\hat{X}^{\dagger*} = (\hat{X}_1^{\dagger*} \ \hat{X}_2^{\dagger*}) = P_{Z^\dagger}(X_1^* \ X_2^*) = P_{(\Psi')^{-1}Z}(\Psi X_1 \ \Psi X_2).$$

In the special case of system (3.30) with instruments (3.32) we have $X_2 = Z_2 = (0', \iota_T')'$ and $Z_1'Z_2 = 0$, whereas under cross-sectional heteroskedasticity, due to $D\iota_T = 0$, the optimal weighting matrix is block-diagonal, hence $Z_1'\Omega Z_2 = 0$. Therefore $Z_1^{\dagger\prime}Z_2^\dagger = 0$

too, giving $P_{Z^\dagger} = P_{Z_1^\dagger} + P_{Z_2^\dagger}$. Now we find $\hat{X}_1^{\dagger*} = P_{Z^\dagger}X_1^* = (P_{Z_1^\dagger} + P_{Z_2^\dagger})X_1^*$ and $\hat{X}_2^{\dagger*} = (P_{Z_1^\dagger} + P_{Z_2^\dagger})\Psi Z_2 = P_{Z_2^\dagger}\Psi Z_2 = (\Psi')^{-1}Z_2(Z_2'\Omega Z_2)^{-1}Z_2'Z_2 = cZ_2^\dagger$, with $c$ some scalar, because $Z_2$ has just one column. Therefore, $M_{\hat{X}_2^{\dagger*}}\hat{X}_1^{\dagger*} = M_{Z_2^\dagger}(P_{Z_1^\dagger} + P_{Z_2^\dagger})X_1^* = P_{Z_1^\dagger}X_1^*$. Thus, in this particular case (when using an appropriate weighting matrix), we find

$$\hat{\beta}_{1,GMM} = (\hat{X}_1^{*\prime}M_{\hat{X}_2^*}\hat{X}_1^*)^{-1}\hat{X}_1^{*\prime}M_{\hat{X}_2^*}y^* = (\hat{X}_1^{*\prime}P_{Z_1^\dagger}X_1^*)^{-1}\hat{X}_1^{*\prime}P_{Z_1^\dagger}y.$$

Due to the block of zeros in $Z_1$ this is just the GMM estimator of the model in first differences.

# C    Extracting redundant moment conditions

Through linear transformation[15] we demonstrate that the sets of moment condions for the equation in levels and for the equation in first-differences have a non empty intersection. First we consider the moment conditions associated with the strictly exogenous regressors. For the equation in first differences these are $E(x_i^T\Delta\varepsilon_{it}) = E[\Delta\varepsilon_{it}(x_{i1}'...x_{iT}')'] = 0$, for $t = 2,...,T$. They can also be represented[16] by the combination $E[\Delta\varepsilon_{it}(\Delta x_{i2}'...\Delta x_{iT}')'] = 0$ and $E(x_{it}\Delta\varepsilon_{it}) = 0$. However, by a similar transformation[17] (here of the disturbances instead of the instruments), the conditions for the equation in levels $E[\Delta x_{ith}^\blacklozenge(\eta_i\iota_T + \varepsilon_i)] = 0$, where $h = 1,...,K_x^\blacklozenge$ (and again $t = 2,...,T$), can be represented by $E(\Delta x_{ith}^\blacklozenge\tilde{\varepsilon}_i) = 0$ and $E[\Delta x_{ith}^\blacklozenge(\eta_i + \varepsilon_{it})] = 0$. So, just the $K_x^\blacklozenge(T-1)$ orthogonality conditions $E[\Delta x_{it}^\blacklozenge(\eta_i + \varepsilon_{it})] = 0$ for $t = 2,...,T$ are additional due to effect stationarity of $K_x^\blacklozenge$ of the strictly exogenous regressors.

Similarly, the orthogonality conditions $E(w_i^{t-1}\Delta\varepsilon_{it}) = 0$, or $E(w_{is}\Delta\varepsilon_{it}) = 0$ for $s = 1,...,t-1$ with $t = 2,...,T$, can be represented by $E(w_{i,t-1}\Delta\varepsilon_{it}) = 0$ for $t = 2,...,T$ and $E(\Delta w_{is}\Delta\varepsilon_{it}) = 0$ for $t = 3,...,T$ and $s = 2,...,t-1$. On the other hand, the conditions $E[\Delta w_{it}^\blacklozenge(\eta_i + \varepsilon_{i,t+l})] = 0$ for $t > 1$ and $l > 0$ are actually $E[\Delta w_{is}^\blacklozenge(\eta_i + \varepsilon_{it})] = 0$ for $t = 2,...,T$ and $s = 2,...,t$, whereas these can be represented by $E(\Delta w_{is}^\blacklozenge\Delta\varepsilon_{it}) = 0$ for $t = 3,...,T$ and $s = 2,...,t-1$ and $E[\Delta w_{it}^\blacklozenge(\eta_i + \varepsilon_{it})] = 0$ for $t = 2,...,T$. Thus, only the $K_w^\blacklozenge(T-1)$ conditions $E[\Delta w_{it}^\blacklozenge(\eta_i + \varepsilon_{it})] = 0$ for $t = 2,...,T$ are additional.

Using the same logic, the orthogonality conditions $E(v_i^{t-2}\Delta\varepsilon_{it}) = 0$ for $t = 3,...,T$, which are actually $E(v_{is}\Delta\varepsilon_{it}) = 0$ for $t = 3,...,T$ and $s = 1,...,t-2$, can also be represented by $E(v_{i,t-2}\Delta\varepsilon_{it}) = 0$ for $t = 3,...,T$ and $E(\Delta v_{is}\Delta\varepsilon_{it}) = 0$ for $t = 4,...,T$ and $s = 2,...,t-2$. However, the conditions $E[\Delta v_{it}^\blacklozenge(\eta_i + \varepsilon_{i,t+1+l})] = 0$ for $t > 1$ and $l > 0$ are in fact $E[\Delta v_{is}^\blacklozenge(\eta_i + \varepsilon_{it})] = 0$ for $t = 3,...,T$ and $s = 2,...,t-1$, which can also be represented as $E(\Delta v_{is}^\blacklozenge\Delta\varepsilon_{it}) = 0$ for $t = 4,...,T$ and $s = 2,...,t-2$ and $E[\Delta v_{i,t-2}^\blacklozenge(\eta_i + \varepsilon_{it})] = 0$ for $t = 3,...,T$. Thus, we find that only the $K_v^\blacklozenge(T-2)$ conditions $E[\Delta v_{i,t-2}^\blacklozenge(\eta_i + \varepsilon_{it})] = 0$ for $t = 3,...,T$ are additional.

---

[15]In this Appendix we repeatedly use the result that the $p$ conditions $E(aCb) = 0$, where $a$ is a random scalar, $b$ a $p \times 1$ random vector and $C$ a deterministic nonsingular $p \times p$ matrix, are equivalent with the $p$ conditions $E(ab) = 0$, because $E(ab) = 0 \Leftrightarrow CE(ab) = 0 \Leftrightarrow E(aCb) = 0$.

[16]Here $C = (D' \, e_{T,t})' \otimes I_{K_x}$.

[17]Now $C = (D' \, e_{T,t})'$.

# D   Derivations for (4.17)

The results for $V_\eta$ and $V_\lambda$ are obvious. Those for $V_\zeta(i)$ and $V_\varepsilon(i)$ are obtained as follows. We use the standard result (reference ??)  that the variance of a general stationary ARMA(2,1) process

$$z_t = \frac{\psi(1 - \phi L)}{(1 - \gamma L)(1 - \xi L)} u_t, \tag{D.1}$$

where $u_t \sim IID(0, 1)$, is given by

$$Var(z_t) = \psi^2 \frac{(1 + \gamma\xi)(1 + \phi^2) - 2\phi(\gamma + \xi)}{(1 - \gamma\xi)(1 - \gamma^2)(1 - \xi^2)}. \tag{D.2}$$

Because we can rewrite (using $\sigma_\varepsilon = 1$)

$$[\beta\rho_{v\varepsilon}\sigma_v + (1 - \xi L)]\omega_i^{1/2} = (1 + \beta\rho_{v\varepsilon}\sigma_v)\omega_i^{1/2}\left(1 - \frac{\xi}{1 + \beta\rho_{v\varepsilon}\sigma_v}L\right),$$

the result for $V_\varepsilon(i)$ follows upon substituting $\psi = (1 + \beta\rho_{v\varepsilon}\sigma_v)\omega_i^{1/2}$ and $\phi = \xi/(1 + \beta\rho_{v\varepsilon}\sigma_v)$. For $V_\zeta(i)$ simply take $\phi = 0$ and $\psi = \beta\sigma_v(1 - \rho_{v\varepsilon}^2)^{1/2}\omega^{1/2}$.