Exploiting strong instruments unduly neglected by standard GMM

Jan F. Kiviet[*†] and Qu Feng[*]

preliminary and confidential draft: 17 March 2012
JEL-code: C01, C13, C23, C26.

Keywords: *efficiency, generalized method of moments, instrument strength, non-spherical disturbances, orthogonality conditions*

Abstract

While coping with non-sphericity of the disturbances, GMM suffers from a blind spot for exploiting the strongest possible instruments. In this study it is demonstrated that its established optimality is achieved over a too inflexible interpretation of the adopted moment restrictions. In particular, GMM does not automatically respect the golden rule that exogenous regressors establish their own strongest possible instruments. For some typical cross-section and dynamic panel data models it is shown by simulation that under moderate heteroskedasticity straight-forward modifications of the exploited set of instruments, which respect just the very same moment conditions, can achieve very substantial reductions in both bias and variance in finite samples. That adapting GMM as proposed here has profound positive effects on the significance of inference in models with non-spherical disturbances is illustrated by estimating a micro employment equation for a panel of UK companies.

1. Introduction

Since three decades, GMM (generalized method of moments) excels as the generic orthogonality conditions based optimal technique for limited information semiparametric estimation. It subsumes the majority of linear and nonlinear econometric estimators. Not starting off from tight fully parametric distributional assumptions, consequently GMM is not asymptotically efficient, but in many situations this may well be compensated by its greater robustness. Applications of GMM are numerous, especially for the analysis of continuously varying dependent variables both in micro econometric studies of cross-sectional or panel data models and in the macro econometric analysis of

[*]Division of Economics, School of Humanities and Social Sciences, Nanyang Technological University, 14 Nanyang Drive, Singapore 637332; jfkiviet@ntu.edu.sg, qfeng@ntu.edu.sg.

[†]Department of Quantitative Economics, Amsterdam School of Economics, University of Amsterdam, Valckenierstraat 65, 1018 XE Amsterdam, The Netherlands; j.f.kiviet@uva.nl.

time-series, in particular when stationarity assumptions can be made. The status of GMM seems undisputable. This paper will not criticize GMM as such either, but it will demonstrate that the standard implementation of GMM is naive, because it interprets the adopted population moment conditions too narrowly when these are converted into sample moments. In short, the standard unfeasible GMM implementation originating in Hansen (1982) suggests that, after the orthogonality conditions have been formulated, the GMM technique will deal optimally with both any overidentification and the specified nonsphericity of the disturbances. We will show that this does not imply that the orthogonality conditions will be exploited optimally. When the model is first reparametrized such that disturbances are homoskedastic and serially uncorrelated then basically the same orthogonality conditions can as a rule lead to the employment of much stronger instruments, which will not only yield a smaller asymptotic variance, but also better correspondence between actual distribution in finite samples and its asymptotic approximation. This has implications too for feasible implementations of GMM. In fact, we will argue that a strategy close to the original IV (instrumental variables) approach suggested by Sargan (1958, 1959) seems generally better. For an overview which puts the IV and GMM approaches into historical perspective see Arellano (2002) and for a monograph on GMM see Hall (2005).

That weakness of instruments has serious consequences for inference has been thoroughly investigated since about two decades already, but is still prominent on the research agenda. See, for instance, Staiger and Stock (1997) and also Andrews and Stock (2007) and its references. This still growing literature, though, has not really been extended yet to cover also the case of GMM, but is largely confined to simple IV estimation of linear models with iid (independently and identically distributed) errors. Although not filling this gap, the major contribution of this study is the following. It shows that for models with nonspherical disturbances, but otherwise linear in the coefficients of jointly dependent and (weakly) exogenous regressors, all aspects of GMM can in principle be understood in terms of IV applied to a model with spherical disturbances in transformed variables. However, this correspondence is less straight-forward as in the similar but more basic case of GLS (generalized least-squares), which is equivalent to OLS (ordinary least-squares) applied to transformed variables, because the instruments are affected in a topsy-turvy manner. This latter aspect has in general devastating effects on the strength of the instruments, unless one is aware of the phenomenon and takes precautions, by implementing GMM less guilelessly.

GMM usually involves iterative estimation, even when the model is linear in its regression coefficients. This is because it requires the assessment of an optimal weighting matrix, which should be proportional to a consistent estimator of the variance of the limiting distribution of the employed orthogonality conditions. In practice this estimate has to be based on residuals which are consistent for the disturbances. Usually, the iteration is started off from IV estimation, although this is suboptimal when its implicit assumption of spherical disturbances is invalid. As we will demonstrate, the standard implementation of GMM and the alternatives that we will suggest, will in principle depart from this same starting position. However, in next iterations standard GMM will generally employ weaker instruments, whereas we will suggest several alternatives which should prevent this. These alternatives either replace these weaker instruments through transformation by possibly stronger ones, or simply extend the standard set of instruments by some of these possibly stronger ones, whereas all these instruments are still referring to exactly the same original set of population orthogonality conditions. Such an extension has the advantage of ensuring a smaller asymptotic variance, but at the same time carries the risk of less attractive finite sample properties, because more instruments may induce larger finite sample bias, see for instance Donald and Newey (2000). Hence, to find out what the major practical consequences are requires some well-designed simulation investigations. When performing these, initially we will not bother much about the iterations that are required in practice, but simply focus on unfeasible ideal implementations of GMM. In these we directly exploit the in practice unavailable optimal weighting matrix, because our first interest is in finding out what the actual potential differences are between the various implementations. Next we focus on feasible implementations of the potentially most optimal adaptations of standard GMM. In this paper we focus mainly on the effects on GMM of heteroskedasticity, and leave a full analysis regarding serial correlation (the major worry in Sargan's approach) for future research, although we give hints here already on how we would approach it.

Our analysis of the options on how to improve on standard GMM analysis, both in cross-sections and in (short) dynamic panel data models involving both endogenous and exogenous regressors together with (conditional) cross-sectional heteroskedasticity, leads to the conclusion that in practice one should aim to weight observations first to get as close to homoskedasticity as possible. Not before, but after that (as already done by many researchers on sensible intuitive grounds), one should design a matrix of instruments according to the adopted orthogonality conditions that matches with the weighted variables. In any next GMM iteration steps the required transformation of variables should have deliberate implications for the chosen transformation of the instruments. In dynamic panel data model analysis the removal of unobserved time-constant individual heterogeneity by taking first differences has the unpleasant side effect that any heteroskedasticity occurs jointly with moving average errors causing originally predetermined regressors to be no longer weakly exogenous. In this context GMM is usually applied in its standard form, see Holtz-Eakin *et al.* (1988) and Arellano and Bond (1991), mainly using lagged internal variables as instruments. Often it has been felt that these estimators lack precision. Especially after a first iteration, their root mean squared errors do not develop in the aspired direction. This is one of the explications that substantial efforts have been made to enhance their precision by taking further orthogonality conditions into account (often of rather doubtful and hard to substantiate validity[1]), see Ahn and Schmidt (1995) and Blundell and Bond (1998). We demonstrate here that from the second stage on, provided the heteroskedasticity is substantial, inadvertently the employed instruments are made weak, which explains their relatively poor performance, and concurrent poor asymptotic approximation by standard methods, see Windmeijer (2005). We show that a simple transformation of the instrument matrix yields a sharp reduction of root mean squared errors.

The structure of this study is as follows. In section 2 we show how GMM can be interpreted in terms of transformed IV, which immediately demonstrates that it leads in principle to using unnecessarily weak instruments. Section 3 lists various options for adapting GMM in order to remedy the problem. Section 4 presents simulation results

[1]When we speak of a valid instrument this refers to the validity of the corresponding orthogonality condition. The degree by which a (valid) instrument is effective (or not) with respect to achieving attractive estimator variance will always be addressed here by instrument strength (or weakness).

on standard and adapted GMM obtained from a design by which a typical family of simultaneous heteroskedastic cross-section models can be represented. Section 5 examines the options to modify standard GMM for dynamic panel data models, and Section 6 presents illustrative simulation results on various variants of panel GMM. In Section 7 some alternative implementations of feasible GMM are applied to a classic empirical dynamic panel data set to illustrate the practical consequences of our modifications, and Section 8 concludes.

2. The relationship between GMM and IV

GMM in its current standard form is basically just equivalent to applying IV to a properly transformed model upon exploiting instruments also subjected to a transformation related to –but different from– the model transformation. To demonstrate this, we consider the standard single linear regression model

$$y = X\beta + \varepsilon, \tag{2.1}$$

where the $n \times K$ full column rank regressor matrix may contain some endogenous regressors, hence in principle $E(X'\varepsilon) \neq 0$. An identifying set of moment conditions is available, in the form of $E(Z'\varepsilon) = 0$, where Z is an $n \times L$ full column rank matrix with $L \geq K$. Provided $Z'X$ has rank K the IV or Two-Stage Least-Squares estimator is given by

$$\hat{\beta}_{IV} = [X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}Z'y. \tag{2.2}$$

By defining for any full column rank matrix A the projection matrix $P_A = A(A'A)^{-1}A'$ the expression for $\hat{\beta}_{IV}$ can be condensed to $(X'P_ZX)^{-1}X'P_Zy$, and even further to the so-called second stage OLS regression $(\hat{X}'\hat{X})^{-1}\hat{X}'y$ by denoting the fitted first-stage regression results as $\hat{X} = P_ZX$. Under sufficient regularity the IV estimator is consistent and asymptotically normal.[2] Sufficient for asymptotic optimality is $\varepsilon_i \mid z_i \sim iid(0, \sigma_\varepsilon^2)$ for $i = 1, ..., n$, where $L \times 1$ vector z_i is the transpose of the i^{th} row of Z. Then one can derive

$$n^{1/2}(\hat{\beta}_{IV} - \beta) \xrightarrow{d} N(0, \sigma_\varepsilon^2 \operatorname{plim} n(X'P_ZX)^{-1}). \tag{2.3}$$

This result is no longer valid when $\varepsilon \sim (0, \sigma_\varepsilon^2\Omega)$ with $\Omega \neq I$ a symmetric positive definite matrix which, without loss of generality, may be scaled such that $tr(\Omega) = n$. Assuming Z and ε are such that we have $E(\varepsilon_i \mid z_i, z_{i-1}, ..., z_1) = 0$, $E(\varepsilon_i^2 \mid z_i, z_{i-1}, ..., z_1) = \sigma_\varepsilon^2\Omega_{ii}$ and, for $i > t$, $E(\varepsilon_i\varepsilon_t \mid z_i, z_{i-1}, ..., z_1) = \sigma_\varepsilon^2\Omega_{it}$ the currently preferred estimator of β is obtained by the Generalized Method of Moments (GMM), see Hansen (1982). It is given by

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

and is optimal in the sense that in aiming to bring the L sample moments $Z'(y - X\hat{\beta}_{GMM})$ jointly as closely as possible to zero, it weights them in a quadratic form in such a way that the variance of the limiting distribution of $\hat{\beta}_{GMM}$ is minimal in a matrix sense. One obtains

$$n^{1/2}(\hat{\beta}_{GMM} - \beta) \xrightarrow{d} N(0, \sigma_\varepsilon^2 \operatorname{plim} n[X'Z(Z'\Omega Z)^{-1}Z'X]^{-1}). \tag{2.5}$$

[2] All asymptotic results in this study are of a standard large sample nature, where $n \to \infty$, upon assuming stationarity of all regressors and instrumental variables.

When $\Omega = I$ (2.4) and (2.5) simplify and specialize to (2.2) and (2.3) respectively. Estimator (2.4) is unfeasible unless Ω is known, but asymptotically equivalent feasible implementations are available.

Whereas the GLS estimator can be obtained by applying OLS to a suitably transformed model, GMM corresponds algebraically to applying IV to a similarly transformed model, exploiting instruments which are transformed by a linear transformation which is actually the inverse of the transpose of the model transformation. This is seen as follows.

Let $\Omega^{-1} = \Psi'\Psi$, where Ψ has full rank but is generally non-unique. Premultiplication of (2.1) by Ψ yields the transformed model

$$y^* = X^*\beta + \varepsilon^*, \tag{2.6}$$

where $y^* = \Psi y$, $X^* = \Psi X$ and $\varepsilon^* = \Psi\varepsilon \sim (0, \sigma^2\Psi\Omega\Psi')$ with $\Psi\Omega\Psi' = \Psi(\Psi'\Psi)^{-1}\Psi' = I$. In order to uphold respecting the very same moment conditions, we should use different instruments now, which we shall denote as $Z^\dagger$. They are such that for any β and Ψ we have $Z^{\dagger\prime}(y^* - X^*\beta) = Z^{\dagger\prime}\Psi(y - X\beta) = Z'(y - X\beta)$, which implies $Z^{\dagger\prime}\Psi = Z'$ or

$$Z^\dagger = (\Psi')^{-1}Z. \tag{2.7}$$

Defining $\hat{X}^* = P_{Z^\dagger}X^* = (\Psi')^{-1}Z(Z'\Omega Z)^{-1}Z'X$ it can easily be verified that

$$\hat{\beta}^*_{IV} = (\hat{X}^{*\prime}\hat{X}^*)^{-1}\hat{X}^{*\prime}y^* = \hat{\beta}_{GMM} \tag{2.8}$$

indeed. Note that $Z^{\dagger\prime}\varepsilon^* = Z'\varepsilon$, so $E(Z'\varepsilon) = 0$ does imply $E(Z^{\dagger\prime}\varepsilon^*) = 0$.

This conversion of GMM into IV also simplifies the proof of the asymptotic normality of GMM considerably. When choosing Ψ lower-triangular then $(\Psi^{-1})'$ will be lower-triangular too, and it follows from $E(\varepsilon_i \mid z_i, ..., z_1) = 0$ that $E(\varepsilon_i^* \mid z_i^\dagger, ..., z_1^\dagger) = 0$ too, and we also have $E(\varepsilon_i^{*2} \mid z_i^\dagger, ..., z_1^\dagger) = \sigma_\varepsilon^2$ and (for $i > t$) $E(\varepsilon_i^*\varepsilon_t^* \mid z_i^\dagger, ..., z_1^\dagger) = 0$. Therefore, necessary requirements are fulfilled to employ the central limit theorem to $n^{-1/2}Z^{\dagger\prime}\varepsilon^*$, giving

$$n^{-1/2}Z'\varepsilon = n^{-1/2}Z^{\dagger\prime}\varepsilon^* = n^{-1/2}\sum\nolimits_{i=1}^n z_i^\dagger\varepsilon_i^* \xrightarrow{d} N(0, \sigma_\varepsilon^2 \operatorname{plim} n^{-1}Z^{\dagger\prime}Z^\dagger). \tag{2.9}$$

Here we used $E(z_i^\dagger\varepsilon_i^*) = E[E(z_i^\dagger\varepsilon_i^* \mid z_i^\dagger)] = E[z_i^\dagger E(\varepsilon_i^* \mid z_i^\dagger)] = E(z_i^\dagger \times 0) = 0$ and

$$\begin{aligned} Var(n^{-1/2}\sum\nolimits_{i=1}^n z_i^\dagger\varepsilon_i^*) &= n^{-1}E\left[\sum\nolimits_{i=1}^n\sum\nolimits_{t=1}^n (z_i^\dagger\varepsilon_i^*)(z_t^\dagger\varepsilon_t^*)'\right] = n^{-1}\sum\nolimits_{i=1}^n E(z_i^\dagger z_i^{\dagger\prime}\varepsilon_i^{*2}) \\ &= n^{-1}\sum\nolimits_{i=1}^n E[z_i^\dagger z_i^{\dagger\prime}E(\varepsilon_i^{*2} \mid z_i^\dagger, ..., z_1^\dagger)] = \sigma_\varepsilon^2 n^{-1}\sum\nolimits_{i=1}^n E(z_i^\dagger z_i^{\dagger\prime}), \end{aligned}$$

whereas invoking the law of large numbers yields

$$\operatorname{plim} n^{-1}Z^{\dagger\prime}Z^\dagger = \operatorname{plim} n^{-1}\sum\nolimits_{i=1}^n z_i^\dagger z_i^{\dagger\prime} = \lim n^{-1}\sum\nolimits_{i=1}^n E(z_i^\dagger z_i^{\dagger\prime}).$$

This provides the building blocks for finding

$$\begin{aligned} n^{1/2}(\hat{\beta}^*_{IV} - \beta) &= n[X^{*\prime}Z^\dagger(Z^{\dagger\prime}Z^\dagger)^{-1}Z^{\dagger\prime}X^*]^{-1}X^{*\prime}Z^\dagger(Z^{\dagger\prime}Z^\dagger)^{-1} \times n^{-1/2}Z^{\dagger\prime}\varepsilon^* \\ &\xrightarrow{d} N(0, \sigma_\varepsilon^2 \operatorname{plim} n[X^{*\prime}Z^\dagger(Z^{\dagger\prime}Z^\dagger)^{-1}Z^{\dagger\prime}X^*]^{-1}), \end{aligned} \tag{2.10}$$

which is equivalent to (2.5).

As is well-known, in the special case $L = K$ we have

$$\begin{aligned}\hat{\beta}_{GMM} &= \hat{\beta}^*_{IV} = (Z^{\dagger\prime}X^*)^{-1}Z^{\dagger\prime}y^* \\ &= [Z'(\Psi')^{-1}\Psi'X]^{-1}Z'(\Psi')^{-1}\Psi'y = (Z'X)^{-1}Z'y \\ &= \hat{\beta}_{IV}. \qquad (2.11)\end{aligned}$$

Then all exploited moment conditions can be satisfied in the sample, so that the weighting matrix has no effect, and therefore Ω cannot play a role in optimizing the estimator. However, it does of course still affect the variance of its limiting distribution, which is $\sigma^2_\varepsilon \operatorname{plim} n(Z'X)^{-1}Z'\Omega Z(X'Z)^{-1}$ in that case.

In the specific case $E(x^*_i\varepsilon^*_i) = 0$, we can use the X^* variables as instruments in model (2.6) and then the resulting method of moments estimator is

$$\begin{aligned}(X^{*\prime}X^*)^{-1}X^{*\prime}y^* &= (X'\Psi'\Psi X)^{-1}X'\Psi'\Psi y = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \\ &= \hat{\beta}_{GLS}. \qquad (2.12)\end{aligned}$$

This is obviously the optimal estimator under those circumstances. Nevertheless, when $E(x_i\varepsilon_i) = 0$ and $\Omega \neq I$, then mechanically following the GMM recipe (which boils down to IV because $L = K$) would involve using the instruments $(\Psi')^{-1}X$ to the transformed model (2.6) yielding the estimator

$$(X'\Psi^{-1}\Psi X)^{-1}X'\Psi^{-1}\Psi y = (X'X)^{-1}X'y = \hat{\beta}_{OLS}, \qquad (2.13)$$

which we know to be sub-optimal. This confusing result is mentioned in Davidson and MacKinnon (2006, p.358). They suggest to interpret GLS as employing instruments $\Omega^{-1}X$ to the untransformed model (2.1). Although it is indeed the case that then the IV and GMM estimators produce $\hat{\beta}_{GLS}$, they provide no rationale for choosing these instruments and thus preferring to embrace here suddenly the orthogonality conditions $E(X'\Omega^{-1}\varepsilon) = 0$ instead of $E(X'\varepsilon) = 0$. In the next section we will provide such a rationale by arguing that GMM is only optimal within the unnecessarily restricted context of the initial set of instruments used, whereas after the transformation of the model the very same orthogonality conditions allow to use alternative instruments, which in principle (but not necessarily) are stronger than those exploited by standard GMM.

3. Adapting GMM to enhance instrument strength

That in the very specific case $E(x_i\varepsilon_i) = 0$, where all regressors are predetermined, GMM delivers a sub-optimal estimator is due to the fact that the GMM principle is only geared towards dealing with any non-sphericity ($\Omega \neq I$) of the disturbances, and is not aiming at the same time to achieve optimal strength of the set of instruments to be exploited. This does not only have consequences in the less interesting case where there are no endogenous regressors and one would never repose on GMM but directly apply (feasible) GLS. To substantiate this, we will from now on assume that $X = (X_1\ X_2)$, where X_1 contains $K_1 \geq 0$ predetermined regressors, hence $E(X_1'\varepsilon) = 0$, and X_2 contains $K_2 \geq 0$ possibly endogenous regressors, so $E(X_2'\varepsilon)$ may differ from a zero vector. Moreover, $X_1 = (X_{10}\ X_{11})$, where X_{10} contains $K_{10} \geq 0$ exogenous regressors, so $E(\varepsilon \mid X_{10}) = 0$, whereas regarding the $K_{11} \geq 0$ regressors in X_{11} we just know that they are predetermined. Of course $K_1 = K_{10} + K_{11}$ and $K = K_1 + K_2 > 0$, so X_{10}, X_{11} and X_2 can be void, but not all three at the same time.

Because $E(\varepsilon \mid X_{10}) = 0$ implies $E(\varepsilon^* \mid X_{10}^*) = 0$ we should preferably use $X_{10}^* = \Psi X_{10}$ as instruments in the transformed model (2.6), whereas GMM would use $X_{10}^\dagger = (\Psi')^{-1} X_{10}$ if Z includes X_{10}, which would normally be the case. It is however highly unlikely that regressing in the first stage X_{10}^* on $Z^\dagger = (\Psi')^{-1} Z$ will yield a perfect fit, whereas this would occur (and then greatly benefits the asymptotic variance) if the valid instruments X_{10}^* would be used. Moreover, given the way in which one usually assembles instruments for the potentially endogenous regressors X_2, namely by collecting predetermined variables Z which yield a good fit $P_Z X_2$ for X_2, it seems quite likely that first stage regressions of $\Psi X_2 = X_2^*$ on $\Psi Z = Z^*$ yield as a rule a much better fit than those of ΨX_2 on $(\Psi')^{-1} Z = Z^\dagger$, as used by standard GMM.

In the case of pure heteroskedasticity Ω is diagonal and Ψ can be chosen diagonal. Then it is obvious that $E(Z'\varepsilon) = 0$ not only implies $E(Z^{\dagger\prime}\varepsilon^*) = 0$, but also $E(Z'\varepsilon^*) = 0$, as well as $E(Z^{*\prime}\varepsilon^*) = 0$. So, regarding the regressors X_1 it is obvious that in the transformed model the instruments X_1^* will be superior to $X_1^\dagger$. This special case clearly demonstrates the inflexibility of standard GMM. It translates the L orthogonality conditions $E(z_i \varepsilon_i) = 0$, $i = 1, ..., n$, directly into "use instrument matrix Z (and stick to it)", whereas the very same orthogonality conditions can also be expressed as $E(c_i z_i \varepsilon_i) = 0$, $\forall c_i \neq 0$ with c_i scalar. These would induce using the instrument matrix CZ, where $C = diag(c_1, ..., c_n)$ and matrix C can be chosen such that it enhances the instrument strength.

When Ω is non-diagonal, which may be relevant especially in a time-series or panel data context, it is usually cumbersome to find valid instruments, *i.e.* instruments z_i which are weakly exogenous with respect to the disturbances ε_i of the untransformed model (2.1). In such cases $E(Z'\varepsilon) = 0$ requires either strict exogeneity of all instruments, or a special form of predeterminedness in combination with a band matrix form of Ω, namely such that for some positive integer $q < n-1$ all elements Ω_{ij} for which $|i-j| > q$ have value zero, whereas z_i may depend on $\varepsilon_{i-q-1-l}$ but does not on ε_{i-q+l} for $l \geq 0$. This situation is obtained when the disturbances establish a (possibly heteroskedastic) q^{th} order stochastic process of moving average type, for which

$$\varepsilon = \Phi\varepsilon^*, \text{ where } \varepsilon_i^* \sim iid(0, \sigma_\varepsilon^2), \tag{3.1}$$

and Φ is an $n \times n$ lower-triangular band matrix such that $\phi_{ij} = 0$ for all $i > j + q$ (and for all $i < j$, of course), giving $E(\varepsilon\varepsilon') = \sigma_\varepsilon^2 \Phi\Phi'$, so $\Omega = \Phi\Phi'$. Since the above implies for $l \geq 0$ that $E(z_{i-l}\varepsilon_i) = 0$, we deduce that $E(z_{i-l}\varepsilon_i^*) = 0$ too. And, assuming that process (3.1) is invertible, then $\Psi = \Phi^{-1}$ is a lower-triangular matrix as well. Therefore row z_i^* of $Z^* = \Psi Z$ is a linear combination of $z_1, ..., z_i$, thus $E(z_{i-l}\varepsilon_i^*) = 0$ implies $E(z_{i-l}^*\varepsilon_i^*) = 0$. Therefore, we find that generally when $\varepsilon \sim (0, \sigma_\varepsilon^2 \Omega)$ with Ω positive definite

$$E(Z'\varepsilon) = 0 \text{ implies } E(Z^{\dagger\prime}\varepsilon^*) = 0, \;\; E(Z'\varepsilon^*) = 0 \text{ and } E(Z^{*\prime}\varepsilon^*) = 0. \tag{3.2}$$

This provides ample opportunities for implementing GMM differently, exploiting in the transformed model $y_i^* = x_i^{*\prime}\beta + \varepsilon_i^*$ the instruments z_i^* additional to, or replacing the probably weaker standard instruments $z_i^\dagger$, possibly also supplemented by z_i and lagged versions of these all.

In many such cases of dynamic possibly autoregressive longitudinal relationships, however, actual practice is that one does not start by seeking valid instruments for the yet untransformed model, but directly tries to parametrize a seriously dynamic model with serially uncorrelated disturbances. Next, when this seems possible, one may start to seek instruments which should be valid and preferably strong for the already transformed model and apply IV, thus avoiding GMM. Such a strategy, which comes close to what Sargan (1958,1959) suggested, runs again into the problems with standard GMM highlighted here, if in a next stage the problem of heteroskedasticity is addressed by reverting to standard GMM again.

From the above, an obvious specific alternative to standard GMM emerges. Above we demonstrated that $E(Z'\varepsilon) = 0$ not just implies $E(Z^{\dagger\prime}\varepsilon^*) = 0$, but also $E(Z^{*\prime}\varepsilon^*) = 0$, whereas in the transformed model the instruments Z^* will in principle be stronger than $Z^\dagger$, which are employed by standard GMM. Using Z^* instead of $Z^\dagger$ we will label here kGMM (keen GMM), which gives the coefficient estimator

$$\begin{aligned}\hat{\beta}_{kGMM} &= (X^{*\prime}P_{Z^*}X^*)^{-1}X^{*\prime}P_{Z^*}y^* \\ &= [X'\Omega^{-1}Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}X]^{-1}X'\Omega^{-1}Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}y. \quad (3.3)\end{aligned}$$

This has limiting distribution

$$n^{1/2}(\hat{\beta}_{kGMM} - \beta) \xrightarrow{d} N(0, \sigma_\varepsilon^2 \operatorname{plim} n[X'\Omega^{-1}Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}X]^{-1}). \quad (3.4)$$

Its asymptotic variance matrix is not necessarily more attractive than that of standard GMM. We made that clear already for the case where all regressors are exogenous. When choosing $Z = X$, standard GMM has the OLS variance $\sigma_\varepsilon^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$, whereas kGMM yields the more efficient GLS variance $\sigma_\varepsilon^2(X'\Omega^{-1}X)^{-1}$. But, when choosing $Z = \Omega^{-1}X$, then standard GMM produces GLS, whereas the variance of kGMM is now $\sigma_\varepsilon^2(X'\Omega^{-2}X)^{-1}X'\Omega^{-3}X(X'\Omega^{-2}X)^{-1}$, which, invoking the Gauss-Markov theorem, must be less efficient because in this special case the estimator is linear in y and unbiased. So, although the one is not uniformly superior to the other, it seems very likely that when $\Omega \neq I$ and Z is selected such that it has better explanatory power for X than $\Omega^{-1}Z$ has, that in that case kGMM will actually exploit stronger instruments, which in finite samples should lead then to smaller bias and smaller dispersion.

Note that the transformations of the instrument matrix considered above do not involve postmultiplication of Z by a full rank square matrix, which would lead to linear transformation of the columns of Z (and have no effect on IV and GMM), but involves premultiplication, leading to linear transformation of the rows of Z, which does alter the space spanned by the columns of Z resulting in different IV and GMM estimators.

3.1. Various intermediate forms of adapted GMM

So, we claim that the efficiency of standard GMM can often be enhanced by adapting the set of instruments used in the transformed model, and not automatically employ $Z^\dagger$. Instead of just replacing $Z^\dagger$ by Z^*, such an adaptation could also be an extension of $Z^\dagger$ by a finite number of instruments. Any extension will always yield improved asymptotic efficiency compared to standard GMM, though not necessarily smaller mean squared errors in finite samples. When all or some of the instruments in $Z^\dagger$ are replaced by others, the effects should be examined case by case.

We can categorize the various options now as follows. Let in model (2.1), where $\varepsilon \sim (0, \sigma_\varepsilon^2 \Omega)$ and $X = (X_1\ X_2)$ as specified above, the $n \times L$ matrix Z of instruments be $Z = (Z_1\ Z_2)$, where $Z_1 = X_1$ contains the K_1 internal instruments, and Z_2 establishes the L_2 external instruments, whereas $L_2 \geq K_2$. GMM uses, when implemented as estimating the transformed model (2.6) by IV, the L instruments $Z^\dagger = (\Psi')^{-1}Z$. In the simulations to follow we will examine the various alternatives indicated in Table 3.1 (with between brackets the total number of instruments). They are labelled M1 through M7.

Table 3.1

Overview of alternative extended or modified sets of instruments, labelled M1 through M7 for IV estimation of the transformed model (2.6)

M1:	M2:	M3:	
$Z^\dagger$, Z^*, Z $[3L]$	$Z^\dagger$, Z^* $[2L]$	$Z^\dagger$, X_1^* $[L + K_1]$	
M4:	M5:	M6:	M7:
$Z_2^\dagger$, Z^*, Z $[2L + L_2]$	$Z_2^\dagger$, Z^* $[L + L_2]$	$Z_2^\dagger$, X_1^* $[L]$	Z^* $[L]$

The upper row lists the extensions, so they all include $Z^\dagger$. Some of these have been removed in the modifications given in the lower row. In M4 through M6 the instruments $Z_1^\dagger$ have been removed, and we do not expect that this will lead to much efficiency loss, because $Z_1^* = X_1^*$ is included in the set of instruments. M2 skips the instruments Z from M1 (as M5 does from M4); although valid, we do not expect Z to offer much strength additional to Z^* in the transformed model. In the next column with M3 and M6 Z_2^* has been removed from M2 and M5 respectively, which will possibly lead to efficiency loss, because we expect that in many cases the instruments Z_2^* will be stronger than $Z_2^\dagger$, especially when Z_2 has been selected because of its explanatory power regarding X_2. Removing the standard GMM instruments $Z^\dagger$ completely and just sticking to Z^* yields the most drastic modification M7 which we labelled kGMM already, see (3.3).

The alternative implementations of adapted GMM put forward in Table 3.1 are formulated as IV estimators for the transformed model. It is obvious that they correspond to alternative GMM implementations to be applied to the untransformed model when these exploit the sets of instruments of Table 3.1 after premultiplication by Ψ'. This leads to Table 3.2.

Table 3.2

Overview of alternative extended or modified sets of instruments, labelled M1 through M7 for GMM estimation of the untransformed model (2.1)

M1:	M2:	M3:	
Z, $\Omega^{-1}Z$, $\Psi'Z$ $[3L]$	Z, $\Omega^{-1}Z$ $[2L]$	Z, $\Omega^{-1}X_1$ $[L + K_1]$	
M4:	M5:	M6:	M7:
Z_2, $\Omega^{-1}Z$, $\Psi'Z$ $[2L + L_2]$	Z_2, $\Omega^{-1}Z$ $[L + L_2]$	Z_2, $\Omega^{-1}X_1$ $[L]$	$\Omega^{-1}Z$ $[L]$

In case $K_2 = 0$ (no endogenous regressors), thus X_2 is void (and Z_2 irrelevant since there is no need for external regressors) Table 3.2 makes clear what the rationale is for switching from the OLS orthogonality conditions $E(X'\varepsilon) = 0$ to $E[(\Omega^{-1}X)'\varepsilon] = 0$, which yield GLS. This paves the way to achieve genuine optimality, whereas exploiting just $E(X'\varepsilon) = 0$, as propounded by GMM, disregards the obvious way to enhance the efficiency. However, the representation as if we should use instruments $\Omega^{-1}X$ is not really enlightening, because it hides that there are two issues at stake, namely coping with non-sphericity and choosing optimal instruments. Intuition is better served by distinguishing between coping with non-sphericity through model transformation, giving the spherical disturbances $\Psi\varepsilon$, for which in the model with exogenous X the most adequate instruments are ΨX, instead of $(\Psi')^{-1}X$. So, the preferable moment conditions $E[(\Omega^{-1}X)'\varepsilon] = 0$ are easier interpreted through the (algebraically equivalent) expression $E[(\Psi X)'\Psi\varepsilon] = 0$.

3.2. Consequences for feasible GMM

In the foregoing we examined alternative GMM estimators upon assuming that Ω is known, which in reality is usually not the case. In practice some form of feasible GMM (FGMM) has to be employed. For standard GMM the most basic feasible procedure (2-step GMM), which is asymptotically equivalent to GMM, goes as follows. First (possibly wrongly) taking Ω to be I, and thus applying sub-optimal IV, residuals $\hat{\varepsilon}_i = y_i - x_i'\hat{\beta}_{IV}$ are obtained, which are consistent for the disturbances. Next, in the case of pure heteroskedasticity, FGMM estimates of the coefficients and their variance are obtained by replacing in the formulas $\sigma_\varepsilon^2 Z'\Omega Z$ by $\sum_{i=1}^n \hat{\varepsilon}_i^2 z_i z_i'$. Obviously 2-step kGMM estimates are obtained by replacing the expressions $Z'\Omega^{-1}Z$, $Z'\Omega^{-1}X$ and $Z'\Omega^{-1}y$ by $\sum_{i=1}^n \hat{\varepsilon}_i^{-2} z_i z_i'$, $\sum_{i=1}^n \hat{\varepsilon}_i^{-2} z_i x_i'$ and $\sum_{i=1}^n \hat{\varepsilon}_i^{-2} z_i y_i$ respectively. Only when $L = K$ the same effect is achieved if one simply uses the standard IV formula substituting $\hat{\varepsilon}_i^{-2} z_i$ for z_i. Cases where Ω is nondiagonal require a more subtle approach; we return to that when discussing dynamic panel data models.

When performing just two steps standard GMM and kGMM (or another adaptation) will already be different, and if standard GMM is used in the habitual way (not adapting the instruments in order to achieve kGMM results) then usually the modified estimator which uses stronger instruments will be more efficient. That means that especially after more steps, when residuals have been obtained by the more efficient adapted estimator, further efficiency gains may be achieved. In the simulations to follow, we will primarily examine nonfeasible GMM which exploits the true value of Ω in order to find out what the genuine differences are between standard GMM and the various suggested extensions and modifications. This will indicate what the potential differences will be between feasible standard and modified implementations of GMM. Only for situations where these differences are most promising, we will simulate feasible variants.

4. Simulation results for a heteroskedastic cross-section model

We shall design a DGP in which we can easily change the seriousness and characteristics of the heteroskedasticity, the strength of the instruments, the degree of simultaneity and the significance of the relationship. To assure that the first two moments of IV estimators exist we choose the degree of overidentification to be 2. In the relationship under study we allow for the presence of an intercept, another exogenous regressor and one possibly endogenous regressor, hence $K_1 = 2$, $K_2 = 1$ and $K = 3$. The two exogenous regressors, which are also used as instruments ($L_1 = K_1$), are $x_{i1} = 1$ and $x_{i2} \sim iidN(0,1)$. The three external instruments ($L_2 = 3$) are generated too as mutually independent $z_{ij} \sim iidN(0,1)$ for $j = 3,4,5$; $i = 1,...,n$. Until we mention otherwise, the experiments are such that the four random exogenous variables are generated only once, so they are kept fixed over the replications of the Monte Carlo simulation. In order to make the results less dependent on their arbitrariness we have rescaled these four vectors such that they do have mean zero and variance unity in the sample of n observations and are mutually orthogonal. To verify to what degree the results still depend on the arbitrary draws of the exogenous variables, we will compare results for a few different arbitrary draws. In order to generate the two endogenous variables and the pattern of the heteroskedasticity such that these are very realistic for typical cross-section applications we proceed as follows. We assume that any heteroskedasticity is of the so-called multiplicative nature and related to x_{i2} and z_{i3} only. A parameter $\phi \geq 0$ determines the seriousness of the heteroskedasticity (where $\phi = 0$ implies homoskedasticity) and a parameter κ, with $0 \leq \kappa \leq 1$, determines the relative importance of x_{i2} and z_{i3} regarding any heteroskedasticity. For the diagonal elements of Ω, indicated by $\omega_i = \Omega_{ii}$, we generate values as follows. Consider the variable

$$h_i^e(c) = \exp(h_i(c)), \text{ with } h_i(c) = c + \phi[\kappa^{1/2}x_{i2} + (1-\kappa)^{1/2}z_{i3}] \sim iidN(c, \phi^2). \tag{4.1}$$

This follows a lognormal distribution with

$$E[h_i^e(c)] = \exp[c + \phi^2/2], \ Var[h_i^e(c)] = [\exp(\phi^2) - 1]\exp(2c + \phi^2).$$

So, upon taking $c = -\phi^2/2$, the expectation of $h_i^e(-\phi^2/2)$ will be 1 and its variance $2\exp(\phi) - 1$. About 99% of the drawings from $h_i(-\phi^2/2)$ will be in the interval $[-\phi^2/2 - 2.58\phi, -\phi^2/2 + 2.58\phi]$. Thus, for the corresponding variable $h_i^e(-\phi^2/2)$, which will establish ω_i, 99% of the drawings will fall in the interval

$$[\exp(-\phi^2/2 - 2.58\phi), \exp(-\phi^2/2 + 2.58\phi)]. \tag{4.2}$$

For particular values of ϕ intervals are found as indicated in Table 4.1.

Table 4.1

Heteroskedasticity for different values of ϕ

ϕ	bounds of 99% intervals			
	$\omega_i^{1/2}$		ω_i	
.2	.76	1.28	.59	1.64
.4	.57	1.61	.33	2.59
.6	.42	1.98	.18	3.93
.8	.30	2.39	.09	5.72
1.0	.21	2.83	.05	8.00
1.2	.15	3.28	.02	10.76
1.4	.10	3.73	.01	13.90
1.6	.07	4.15	.00	17.25

To ensure that we have $\sum_{i=1}^{n}\omega_i = n$ we rescale the random series $h_i^e(-\phi^2/2)$ by dividing by its sample mean. Like all exogenous variables the series ω_i is kept fixed over the replications. From Table 4.1 we learn that $\phi \geq 1$ implies pretty serious heteroskedasticity, whereas we may qualify it mild when $\phi < .3$, say.

The reduced form equation for x_{i3} is

$$x_{i3} = \pi_{31} + \pi_{32}x_{i2} + \pi_{33}z_{i3} + \pi_{34}z_{i4} + \pi_{35}z_{i5} + v_{i3}, \tag{4.3}$$

where we take

$$v_{i3} = \omega_i^{1/2}v_{i3}^{\circ}, \text{ with } v_{i3}^{\circ} \sim iidN(0,1).$$

Then the joint strength of the three external instruments will be determined by the concentration parameter inspired scalar quantity

$$\mu^2 = n(\pi_{33}^2 + \pi_{34}^2 + \pi_{35}^2)/3. \tag{4.4}$$

Choosing all three instruments equally weak or strong, we should take

$$\pi_{33} = \pi_{34} = \pi_{35} = n^{-1/2}\left|\mu\right|.$$

This implies $Var(x_{i3}) = \pi_{32}^2 + 3\mu^2/n + \omega_i$.

The structural form equation will be generated as

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i, \tag{4.5}$$

where

$$\varepsilon_i = \sigma_\varepsilon \omega_i^{1/2}\left[\rho v_{i3}^{\circ} + (1-\rho^2)^{1/2}\varepsilon_i^{\circ}\right], \text{ with } \varepsilon_i^{\circ} \sim iidN(0,1).$$

So, $\rho \in (-1,+1)$ is the correlation coefficient of ε_i and v_{i3}, which in this context expresses the simultaneity. Because the 4 random exogenous instruments will be kept fixed over the replications of the Monte Carlo, the correlation between x_{i3} and ε_i will also be ρ.

Without loss of generality we may choose $\pi_{31} = 0$ and $\beta_1 = 0$ and hence, given ϕ, μ^2 and ρ, the remaining free parameters are π_{32}, β_2, β_3 and σ_ε. Note that π_{32} determines the multicollinearity between the structural form regressors x_{i2} and x_{i3}. Without loss of generality we may fix one of the remaining three parameters of the structural form equation. We will take $\sigma_\varepsilon = 1$. Now the choice of β_2 and β_3 will determine the signal-noise ratio of the structural form equation. We will experiment a bit with choosing β_2 and β_3, starting by giving them both value .5, and check whether the fit in the resulting DGPs seems realistic for cross-sections. By that we mean standard errors of the estimated coefficients which are a fraction of their corresponding true coefficient value in the range 0.1-2, say. Since $\sigma_\varepsilon^2 = 1$ we will assess what we will call the realized signal-noise ratio by

$$SNR = \frac{1}{n-1}\sum_{i=1}^{n}[\beta_2(x_{i2} - \bar{x}_2) + \beta_3(x_{i3} - \bar{x}_3)]^2. \tag{4.6}$$

From this we can obtain $FIT = SNR/(1+SNR)$. We will check whether this has an average over the simulation replications in the range of about 0.1-0.6, which seems a reasonable range for cross-sections. In summary, we will examine various cases from the grid

$$\begin{array}{ll} n = \{50, 200\} & \rho = \{.1, .5\} \\ \phi = \{.5, 1\} & \pi_{32} = \{0, .4, .8\} \\ \kappa = \{.2, .5, .8\} & \beta_2 = \{.3, .5\} \\ \mu^2 = \{2, 10, 50\} & \beta_3 = \{.3, .5\} \end{array}$$

In Table 4.2 we collect a first set of results for the larger sample size, serious heteroskedasticity (the minimum and maximum values of $\omega_i^{1/2}$ were .21 and 4.13 respectively), strong instruments, substantial simultaneity, no multicollinearity between x_{i2} and x_{i3} and values of the coefficients such that a moderate fit results. The heading mentions the average (and the standard deviation between parentheses) of the FIT defined below (4.6) and of the $F_{3,n-L}$ test statistic on the joint significance of the external instruments in the reduced form equation for x_{i3} when estimated by OLS (thus neglecting the heteroskedasticity). Both measures have their drawbacks, but they are only used here to get a rough impression of major characteristics of the DGP. FGMM refers to the 2-step feasible version of standard GMM.

Table 4.2

Simulation results for $n = 200$, $\phi = 1$, $\kappa = .5$, $\mu^2 = 50$, $\rho = .5$, $\pi_{32} = 0$, $\beta_2 = \beta_3 = 0.5$; giving simulation average for FIT of .41 (.03) and for $F_{3,n-L}$ of 52.5 (13.2)

	β_2				β_3			
	Bias	Std.Err	RMSE	Rel.Eff	Bias	Std.Err	RMSE	Rel.Eff
GMM	.002	.090	.090	1.000	.003	.094	.094	1.000
AGMM-M1	-.000	.047	.047	.524	.014	.052	.054	.574
AGMM-M2	-.000	.047	.047	.526	.007	.053	.053	.566
AGMM-M3	-.000	.047	.047	.526	.007	.071	.072	.764
AGMM-M4	-.000	.047	.047	.525	.011	.052	.053	.570
AGMM-M5	.000	.047	.047	.526	.005	.053	.053	.566
AGMM-M6	.000	.047	.047	.528	.003	.094	.094	1.00
AGMM-M7	.000	.047	.047	.527	.000	.053	.053	.569
FGMM	.001	.089	.089	.991	.003	.093	.093	.993
IV	-.001	.092	.092	1.023	-.000	.099	.098	1.050
OLS	.001	.080	.080	.897	.282	.073	.292	3.111
WLS	-.005	.044	.044	.492	.175	.039	.179	1.912

We find that all techniques examined produce little bias for this case, except the two inconsistent estimators OLS and WLS which show substantial bias for the coefficient of the endogenous regressor. For β_2 all adapted GMM techniques have much lower standard error than classic GMM and therefore produce stunning improvements in RMSE over standard GMM. This is also the case for β_3 except for M6, which proves to be equivalent to the standard GMM estimator.[3] Also M3, which like M6 lacks Z_2^* in its instrument set, performs less impressive with respect to β_3. The columns Rel.Eff contain the RMSE divided by the RMSE of GMM. What the effects are of taking π_{32} nonzero can be seen from Table 4.3.

[3]We have tried to prove this formally, but were not successful yet.

Table 4.3

Simulation results for $n = 200$, $\phi = 1$, $\kappa = .5$, $\mu^2 = 50$, $\rho = 0.5$, $\pi_{32} = .8$, $\beta_2 = \beta_3 = 0.5$; giving simulation average for FIT of .57 (.02) and for $F_{3,n-L}$ of 52.4 (13.2)

	β_2				β_3			
	Bias	Std.Err	RMSE	Rel.Eff	Bias	Std.Err	RMSE	Rel.Eff
GMM	-.001	.095	.095	1.000	.003	.094	.094	1.000
AGMM-M1	-.011	.063	.064	.678	.014	.052	.054	.574
AGMM-M2	-.006	.064	.064	.674	.007	.053	.053	.566
AGMM-M3	-.006	.075	.075	.790	.007	.071	.072	.764
AGMM-M4	-.009	.064	.064	.676	.011	.052	.053	.570
AGMM-M5	-.004	.064	.064	.675	.005	.053	.053	.566
AGMM-M6	-.002	.090	.090	.950	.003	.094	.094	1.000
AGMM-M7	-.000	.064	.064	.677	.000	.053	.053	.569
FGMM	-.001	.095	.095	1.003	.003	.093	.093	.993
IV	-.001	.095	.095	1.004	-.000	.099	.098	1.050
OLS	-.225	.092	.243	2.560	.282	.073	.292	3.111
WLS	-.145	.055	.155	1.628	.175	.039	.179	1.912

The results regarding β_3 are found to be invariant with respect to multicollinearity (which we can prove analytically) but not those regarding β_2. For the latter less impressive results are obtained through adapting the instrument set, especially so for M3 and M6. In the results to follow we will no longer monitor these less successful variants closely. To focus on reasonably realistic cases we will set π_{32} at the moderate value of .4 in all the calculations to follow. Because the value of FIT is rather high for cross-sections we will also reduce the values of β_2 and β_3 to .3 from now on and at the same time we will decrease the value of μ^2 to the transitional value of 10 in order to see what happens when the instruments are almost weak.

Table 4.4

Simulation results for $n = 200$, $\phi = 1$, $\kappa = .5$, $\mu^2 = 10$, $\rho = 0.5$, $\pi_{32} = .4$, $\beta_2 = \beta_3 = 0.3$; giving simulation average for FIT of .23 (.02) and for $F_{3,n-L}$ of 11.3 (4.8)

	β_2				β_3			
	Bias	Std.Err	RMSE	Rel.Eff	Bias	Std.Err	RMSE	Rel.Eff
GMM	-.005	.100	.100	1.000	.019	.211	.212	1.000
AGMM-M1	-.027	.064	.069	.690	.065	.107	.126	.592
AGMM-M2	-.015	.066	.067	.672	.037	.113	.119	.560
AGMM-M3	-.015	.079	.080	.799	.035	.155	.159	.750
AGMM-M4	-.022	.065	.068	.681	.054	.110	.122	.576
AGMM-M5	-.010	.067	.067	.672	.025	.116	.118	.557
AGMM-M6	-.008	.098	.099	.984	.019	.211	.212	1.000
AGMM-M7	-.002	.068	.068	.680	.005	.120	.120	.566
FGMM	-.005	.101	.101	1.004	.018	.211	.211	.995
IV	-.006	.101	.101	1.012	.006	.223	.223	1.049
OLS	-.172	.084	.191	1.910	.433	.095	.443	2.089
WLS	-.150	.048	.158	1.573	.365	.055	.369	1.738

From Table 4.4 we note that weaker instruments lead to slightly more bias, especially for the variants which use many instruments. This is all in agreement with established understanding of the behavior in finite samples of IV and GMM estimators. Especially the standard error of the estimators for β_3 has increased, but it is remarkable how robust the values of the relative efficiencies are, indicating that by adapting GMM without making any extra orthogonality conditions, but just exploiting those already made in a better way, RMSE reductions of about 40% are up for grabs. The inconsistent estimators OLS and especially WLS have an (almost always) smaller standard error than all GMM implementations, but their huge bias makes them unfit. Note that FGMM barely beats IV. Hence, making an attempt to assess and exploit the heteroskedasticity is not necessarily worthwhile, although we should add that in practice such an assessment would be required anyhow for the actual consistent estimation of the variance of IV. Next we will examine a case with really weak external instruments.

Table 4.5

Simulation results for $n = 200$, $\phi = 1$, $\kappa = .5$, $\mu^2 = 2$, $\rho = 0.5$, $\pi_{32} = .4$, $\beta_2 = \beta_3 = 0.3$; giving simulation average for FIT of .22 (.02) and for $F_{3,n-L}$ of 3.13 (2.2)

	β_2				β_3			
	Bias	Std.Err	RMSE	Rel.Eff	Bias	Std.Err	RMSE	Rel.Eff
GMM	-.045	.185	.190	1.000	.135	.507	.525	1.000
AGMM-M1	-.090	.088	.126	.661	.223	.186	.290	.553
AGMM-M2	-.062	.098	.116	.610	.153	.215	.264	.502
AGMM-M3	-.064	.132	.147	.772	.159	.306	.344	.656
AGMM-M4	-.080	.091	.122	.638	.198	.196	.279	.531
AGMM-M5	-.047	.105	.115	.603	.115	.233	.260	.495
AGMM-M6	-.055	.209	.216	1.136	.135	.507	.525	1.000
AGMM-M7	-.012	.125	.126	.659	.031	.288	.289	.551
FGMM	-.045	.187	.192	1.011	.129	.508	.524	.997
IV	-.045	.193	.198	1.040	.107	.536	.547	1.042
OLS	-.193	.086	.211	1.110	.485	.103	.495	.944
WLS	-.189	.048	.195	1.023	.465	.061	.469	.893

Especially the estimates of the coefficient of the endogenous regressor are seriously biased now, but less so for M7 (kGMM). This variant benefits here from using relatively few instruments. Standard GMM uses as many, but because these are weaker they do not only yield more bias, but also much larger standard errors. Regarding β_3 the standard error of GMM is even larger than the true coefficient value, and its bias is about half the coefficient value. Using many more instruments, M1 has much lower standard errors, but at the expense of very serious bias. Although M7 suffers from the weakness of the instruments too it might still yield reasonably useful inference, whereas that does not seem possible here for standard GMM. M6 gives worse estimates of β_2 than standard GMM. For the cases considered hence far M5 and M7 produce the best results, although no technique is uniformly superior. Below we will get back to the $\mu^2 = 10$ case and examine the effects of ϕ, ρ, κ and n. Therefore Table 4.6 should be compared especially with Table 4.4 because they only differ regarding the value of ϕ.

Table 4.6

Simulation results for $n = 200$, $\phi = .5$, $\kappa = .5$, $\mu^2 = 10$, $\rho = 0.5$, $\pi_{32} = .4$, $\beta_2 = \beta_3 = 0.3$; giving simulation average for FIT of .23 (.01) and for $F_{3,n-L}$ of 11.1 (4.2)

	β_2				β_3			
	Bias	Std.Err	RMSE	Rel.Eff	Bias	Std.Err	RMSE	Rel.Eff
GMM	-.005	.103	.103	1.000	.016	.195	.195	1.000
AGMM-M1	-.047	.083	.096	.932	.117	.139	.182	.931
AGMM-M2	-.028	.087	.092	.894	.071	.152	.167	.855
AGMM-M3	-.017	.094	.096	.930	.044	.173	.179	.915
AGMM-M4	-.040	.085	.094	.913	.099	.144	.175	.895
AGMM-M5	-.020	.089	.092	.892	.050	.158	.166	.847
AGMM-M6	-.006	.101	.101	.984	.016	.195	.195	1.000
AGMM-M7	-.004	.094	.094	.913	.011	.170	.171	.873
FGMM	-.005	.104	.104	1.010	.015	.197	.197	1.008
IV	-.004	.103	.103	1.005	.010	.197	.197	1.007
OLS	-.176	.070	.186	1.812	.434	.066	.439	2.244
WLS	-.172	.061	.182	1.770	.421	.058	.425	2.176

In Table 4.6 $\phi = .5$ and this gave $\omega_i^{1/2}$ values with a minimum and a maximum of .49 and 2.17 respectively. When the heteroskedasticity is more moderate the added strength of the instruments used by the adapted GMM techniques is much more moderate too, so that especially the variants that add many instruments suffer from bias. Again M5 and M7 come out favorably. In what follows we put $\phi = 1$ again and examine the effects of the remaining design parameters.

Table 4.7

Simulation results for $n = 200$, $\phi = 1$, $\kappa = .2$, $\mu^2 = 10$, $\rho = 0.5$, $\pi_{32} = .4$, $\beta_2 = \beta_3 = 0.3$; giving simulation average for FIT of .22 (.02) and for $F_{3,n-L}$ of 11.3 (4.9)

	β_2				β_3			
	Bias	Std.Err	RMSE	Rel.Eff	Bias	Std.Err	RMSE	Rel.Eff
GMM	-.005	.097	.097	1.000	.019	.209	.210	1.000
AGMM-M1	-.028	.063	.069	.708	.068	.110	.130	.618
AGMM-M2	-.016	.064	.066	.686	.039	.116	.123	.584
AGMM-M3	-.016	.079	.081	.833	.039	.162	.166	.793
AGMM-M4	-.023	.063	.067	.698	.057	.113	.126	.602
AGMM-M5	-.011	.065	.066	.685	.026	.119	.122	.580
AGMM-M6	-.008	.096	.096	.993	.019	.209	.210	1.000
AGMM-M7	-.002	.067	.067	.694	.005	.127	.124	.589
FGMM	-.005	.097	.097	1.006	.018	.209	.210	1.000
IV	-.006	.098	.098	1.016	.008	.222	.222	1.058
OLS	-.172	.074	.188	1.939	.433	.090	.442	2.108
WLS	-.152	.045	.158	1.638	.370	.055	.374	1.784

Table 4.8

Simulation results for $n = 200$, $\phi = 1$, $\kappa = .8$, $\mu^2 = 10$, $\rho = 0.5$, $\pi_{32} = .4$, $\beta_2 = \beta_3 = 0.3$; giving simulation average for FIT of .23 (.02) and for $F_{3,n-L}$ of 11.3 (4.7)								
	β_2				β_3			
	Bias	Std.Err	RMSE	Rel.Eff	Bias	Std.Err	RMSE	Rel.Eff
GMM	-.004	.111	.111	1.000	.018	.206	.207	1.000
AGMM-M1	-.026	.065	.070	.631	.062	.104	.121	.586
AGMM-M2	-.014	.067	.068	.618	.035	.110	.116	.559
AGMM-M3	-.013	.079	.080	.723	.032	.150	.154	.743
AGMM-M4	-.021	.066	.069	.624	.052	.107	.118	.572
AGMM-M5	-.010	.068	.068	.618	.024	.113	.115	.557
AGMM-M6	-.008	.098	.099	.890	.018	.206	.207	1.000
AGMM-M7	-.002	.069	.069	.624	.004	.117	.117	.565
FGMM	-.005	.110	.110	.996	.017	.205	.205	.993
IV	-.005	.113	.113	1.019	.006	.213	.213	1.032
OLS	-.172	.094	.196	1.768	.433	.096	.443	2.144
WLS	-.149	.050	.157	1.422	.360	.054	.364	1.758

Tables 4.7 and 4.8 show that the effect of κ is very moderate. Giving it again the value of .5 we change the simultaneity in Table 4.9.

Table 4.9

Simulation results for $n = 200$, $\phi = 1$, $\kappa = .5$, $\mu^2 = 10$, $\rho = 0.1$, $\pi_{32} = .4$, $\beta_2 = \beta_3 = 0.3$; giving simulation average for FIT of .23 (.02) and for $F_{3,n-L}$ of 11.3 (4.8)								
	β_2				β_3			
	Bias	Std.Err	RMSE	Rel.Eff	Bias	Std.Err	RMSE	Rel.Eff
GMM	.001	.102	.102	1.000	.001	.213	.213	1.000
AGMM-M1	-.005	.066	.067	.654	.011	.117	.112	.528
AGMM-M2	-.002	.067	.067	.662	.006	.115	.115	.543
AGMM-M3	-.002	.081	.081	.792	.006	.158	.158	.742
AGMM-M4	-.004	.067	.067	.656	.009	.113	.113	.533
AGMM-M5	-.001	.068	.068	.666	.004	.117	.117	.550
AGMM-M6	-.001	.100	.100	.982	.001	.213	.213	1.000
AGMM-M7	.000	.069	.069	.674	-.001	.120	.120	.564
FGMM	.000	.102	.102	1.002	.001	.217	.212	.996
IV	.000	.102	.102	1.004	-.001	.226	.223	1.047
OLS	-.033	.096	.101	.995	.086	.109	.138	.651
WLS	-.030	.054	.061	.603	.072	.061	.095	.446

Little changes for the GMM procedures when there is hardly simultaneity, although the ordering of the adapted procedures is different now, with M1 outperforming all the others. Compared with other adapted GMM procedures, M6 is less attractive for estimating β_2. The inconsistent procedures show much less bias now, so that regarding RMSE both OLS and WLS beat standard GMM with a wide margin, and WLS even surpasses all the adapted GMM procedures. This, and also the poor results for IV, shows that transforming the instruments under heteroskedasticity is essential for gaining efficiency, but strict orthogonality with respect to the transformed disturbances of the employed instruments should not necessarily be a matter of principle. Just using the strongest possible (though mildly invalid) instruments, as WLS does, can be preferable to using a great number of weaker though valid instruments, as all the GMM procedures do. In Table 4.10 we look into the case of Table 4.4 again, but now for a much smaller sample size.

Table 4.10

Simulation results for $n = 50$, $\phi = 1$, $\kappa = .5$, $\mu^2 = 10$, $\rho = 0.5$, $\pi_{32} = .4$, $\beta_2 = \beta_3 = 0.3$; giving simulation average for FIT of .22 (.04) and for $F_{3,n-L}$ of 12.2 (5.9)

	β_2				β_3			
	Bias	Std.Err	RMSE	Rel.Eff	Bias	Std.Err	RMSE	Rel.Eff
GMM	-.004	.180	.180	1.000	.024	.202	.203	1.000
AGMM-M1	-.034	.124	.129	.717	.081	.113	.139	.684
AGMM-M2	-.020	.127	.129	.716	.049	.119	.129	.635
AGMM-M3	-.017	.136	.137	.761	.041	.154	.159	.784
AGMM-M4	-.029	.125	.128	.715	.069	.115	.134	.661
AGMM-M5	-.014	.128	.129	.720	.035	.122	.127	.626
AGMM-M6	-.010	.149	.149	.829	.024	.202	.203	1.00
AGMM-M7	-.004	.131	.131	.729	.012	.128	.128	.633
FGMM	-.005	.183	.183	1.019	.024	.201	.202	.996
IV	-.014	.188	.188	1.048	.016	.217	.217	1.071
OLS	-.120	.171	.209	1.164	.303	.141	.334	1.646
WLS	-.094	.114	.148	0.823	.220	.088	.237	1.166

Table 4.10 shows that also in a much smaller sample the alternatively transformed instruments are much better than those used by standard GMM. The improvement over GMM is less impressive here as for the $n = 200$ case. However, these results are hard to compare, because the higher-order sample moments of the exogenous regressors will have been different, and also the range of values obtained for $\omega_i^{1/2}$ is not the same (here between .34 and 2.84).

To examine the effect of conditioning on exogenous variables we have reproduced the case of Table 4.4 for various different arbitrary realizations of the exogenous variables, and therefore for the ω_i series. Although that does make the figures different, occasionally even by margins of around 20%, we found no evidence of a systematic disturbing influence on our general findings from the specific series that we used for all the foregoing Tables. Finally, we shall examine the case of Table 4.4 in an alternative simulation design, where all exogenous random variables have been redrawn every replication, without any standardizing or orthogonalizing.

Table 4.11 Unconditional simulation

Simulation results for $n = 200$, $\phi = 1$, $\kappa = .5$, $\mu^2 = 10$, $\rho = 0.5$, $\pi_{32} = .4$, $\beta_2 = \beta_3 = 0.3$; giving simulation average for FIT of .22 (.02) and for $F_{3,n-L}$ of 11.3 (4.5)

	β_2				β_3			
	Bias	Std.Err	RMSE	Rel.Eff	Bias	Std.Err	RMSE	Rel.Eff
GMM	-.006	.106	.106	1.000	.019	.198	.199	1.000
AGMM-M1	-.026	.061	.066	.627	.066	.105	.123	.620
AGMM-M2	-.015	.063	.065	.613	.039	.110	.117	.585
AGMM-M3	-.013	.076	.077	.728	.033	.151	.155	.777
AGMM-M4	-.022	.062	.066	.620	.055	.107	.120	.603
AGMM-M5	-.010	.064	.065	.612	.027	.112	.115	.578
AGMM-M6	-.007	.092	.092	.872	.019	.199	.200	1.003
AGMM-M7	-.003	.065	.065	.617	.008	.116	.116	.583
FGMM	-.006	.106	.106	1.001	.018	.198	.199	1.000
IV	-.009	.107	.108	1.017	.019	.204	.204	1.026
OLS	-.174	.081	.191	1.809	.434	.088	.443	2.222
WLS	-.142	.049	.150	1.419	.360	.057	.364	1.829

From Table 4.11 we see that the unconditional results are not very much different from the results obtained for a specific set of exogenous variables. Nevertheless, they support again the general conclusion that especially the variants M5 and M7, but also M2, yield substantial "free-lunch" (because no additional orthogonality assumptions are being made) reductions in standard deviation and possibly in bias too. The remarkable feature of M7 (keen GMM) is that it achieves its substantial improvements over standard GMM by using exactly the same number of instruments. Variants M2 and M5 both use all the instruments exploited in kGMM, but also maintain instruments from standard GMM, which proved to be beneficial in various of the tables just presented. Such adaptations will be labelled kfGMM below, because they are both keen and faithful (regarding classic GMM). Especially kGMM and kfGMM variants will be examined further now in more complex models.

5. Adapting GMM for dynamic panel data models

Here we consider a case where heteroskedasticity may occur jointly with simultaneity, moving-average errors and weakly-exogenous regressors and instruments. Consider the balanced linear first-order dynamic panel data model

$$y_{it} = \eta_i + \tau_t + \gamma y_{i,t-1} + \beta' x_{it} + \varepsilon_{it}, \text{ with } \varepsilon_{it} = \sigma_\varepsilon \omega_i^{1/2} \varepsilon_{it}^\circ, \tag{5.1}$$

where $i = 1, ..., N$ indexes the individual subjects in the sample and $t = 1, ..., T$ the time periods. In the right-hand side of (5.1) the η_i are unobserved individual-specific time-constant effects, τ_t are unobserved time-specific individual-constant effects, x_{it} is a $k \times 1$ vector of observed explanatory variables (not containing a constant), which individually may be either exogenous, predetermined or endogenous, depending on their correlation structure with the unobserved idiosyncratic disturbances ε_{it}. Because $\varepsilon_{it}^\circ \sim iid(0, 1)$ the disturbances ε_{it} are independent, both between time periods and between individuals.

However, there is cross-section heteroskedasticity if not all ω_i (with $\sum_{i=1}^n \omega_i = n$) are equal to unity. We will examine standard and adapted GMM estimators which are consistent for N large and T finite, and focus on structural relationships which are dynamically stable, requiring $|\gamma| < 1$. So, (non)stationarity of y_{it} should find its origin in the time-series properties of the series x_{it}.

Stacking the observations for individual i, we may also write

$$y_i = \eta_i \iota + \tau + \gamma y_{i,-1} + X_i \beta + \sigma_\varepsilon \omega_i^{1/2} \varepsilon_i^\circ, \tag{5.2}$$

where ι, τ, $y_{i,-1}$ and ε_i° are all $T \times 1$ vectors, ι just containing unit elements, $\tau = (\tau_1, ..., \tau_T)'$ and $y_{i,-1}$ and ε_i° stacking all $y_{i,t-1}$ and ε_{it}° for $t = 1, ..., T$, whereas X_i stacks the T rows x_{it}'. To remove the unknown $N \to \infty$ incidental parameters η_i from the model it should be transformed. We consider taking first differences and define the matrices

$$D = \begin{bmatrix} -1 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & -1 & 1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & -1 & 1 \end{bmatrix}, H = \begin{bmatrix} 2 & -1 & 0 & \cdots & \cdots & 0 \\ -1 & 2 & -1 & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ & & \ddots & \ddots & \ddots & -1 \\ 0 & \cdots & \cdots & 0 & -1 & 2 \end{bmatrix}, \tag{5.3}$$

where D is $(T-1) \times T$ and $H = DD'$ is $(T-1) \times (T-1)$. Obviously D performs the first difference operation and $D\iota_T = 0$. Also defining $\tilde{y}_i = Dy$, $\tilde{\tau} = D\tau$, $\tilde{y}_{i,-1} = Dy_{i,-1}$, $\tilde{X}_i = DX_i$ and $\tilde{\varepsilon}_i^\circ = D\varepsilon_i^\circ$, which all have $T-1$ rows, we have

$$\tilde{y}_i = W_i \delta + \sigma_\varepsilon \omega_i^{1/2} \tilde{\varepsilon}_i^\circ, \tag{5.4}$$

where $W_i = (I_{T-1}\ \tilde{y}_{i,-1}\ \tilde{X}_i)$ and coefficient vector $\delta = (\tilde{\tau}'\ \gamma\ \beta')'$ has $K = T + k$ elements. Note that not all individual T elements of τ can be identified now, but only (linear combinations of) the $T-1$ successive differences between them. Since $E(\tilde{\varepsilon}_i^\circ \tilde{\varepsilon}_i^{\circ\prime}) = H$ this transformation leads to disturbances which for each individual's remaining $T-1$ observations have MA(1) structure with a known first-order serial correlation coefficient of $-.5$.

If Z_i is the $(T-1) \times L$ matrix containing the observations regarding individual i on the L instrumental variables that will be exploited to the differenced model (5.4), then the optimal but unfeasible standard GMM estimator of δ can be written as

$$\begin{aligned} \hat{\delta}_{GMM} &= \left[\left(\sum\nolimits_{i=1}^N W_i' Z_i \right) \left(\sum\nolimits_{i=1}^N \omega_i Z_i' H Z_i \right)^{-1} \left(\sum\nolimits_{i=1}^N Z_i' W_i \right) \right]^{-1} \times \\ & \qquad \left(\sum\nolimits_{i=1}^N W_i' Z_i \right) \left(\sum\nolimits_{i=1}^N \omega_i Z_i' H Z_i \right)^{-1} \left(\sum\nolimits_{i=1}^N Z_i' \tilde{y}_i \right), \end{aligned} \tag{5.5}$$

where matrix Z_i is constructed as follows.

Partition Z_i of the full instrument matrix Z should of course include a partition I_{T-1}, which represents the orthogonality conditions $E(\tilde{\varepsilon}_i) = 0$. For an endogenous regressor $x_{it}^{(j)}$ $(j = 1, ..., k)$ we have $E(\tilde{x}_{it}^{(j)} \tilde{\varepsilon}_{it}) \neq 0$, so its current value cannot be used as an instrument in model (5.4). Similarly for a predetermined regressor, where $E(x_{jit} \varepsilon_{i,t-1}) \neq$

0 implies $E(\tilde{x}_{jit}\tilde{\varepsilon}_{i,t}) \neq 0$, and for the lagged-dependent variable, where $E(y_{i,t-1}\varepsilon_{i,t-1}) \neq 0$ implies $E(\tilde{y}_{i,t-1}\tilde{\varepsilon}_{i,t}) \neq 0$. However, by taking higher-order lags of endogenous and of predetermined regressors one may obtain valid internal instruments. The established GMM procedures for dynamic panel data models exploit that $E(y_{i,t-1-l}\Delta\varepsilon_{it}) = 0$ for $l = 1, ..., t-1$ and $t = 2, ..., T$. This allows to incorporate in Z_i the partitions

$$J_{T-1}y_{i0}, J_{T-2}y_{i1}, ..., J_1 y_{i,T-2}, \text{ where } J_t = (O_t \; I_t)', \; t = 1, ..., T-1. \tag{5.6}$$

Matrix J_t is of order $(T-1) \times t$ and such that $J_{T-1} = I_{T-1}$ whereas for $t < T-1$ matrix O_t is of order $(T-1-t) \times t$ with all its elements equal to zero. The full set (5.6) contributes no less than $T(T-1)/2$ columns to Z_i. Similar possibilities arise for a regressor $x_{it}^{(j)}$ $(j = 1, ..., k)$ for which we distinguish the three categories: either $E(x_{it}^{(j)}\varepsilon_{is}) = 0$ for $s > t$ (possibly endogenous), or $s \geq t$ (at least predetermined), or $\forall s, t$ (exogenous). Then one may incorporate in Z_i for estimating model (5.4) by GMM the partitions

$$J_{T-2}x_{i1}^{(j)}, ..., J_1 x_{i,T-2}^{(j)}, \text{ when possibly endogenous} \tag{5.7}$$
$$J_{T-1}x_{i1}^{(j)}, ..., J_1 x_{i,T-1}^{(j)}, \text{ when predetermined} \tag{5.8}$$
$$I_{T-1}x_{i1}^{(j)}, ..., I_{T-1} x_{iT}^{(j)}, \text{ when exogenous.} \tag{5.9}$$

For each element of regressor vector x_{it} these yield additional columns to Z_i, namely $(T-1)(T-2)/2$, or $T(T-1)/2$ or $T(T+1)$ columns respectively. Of course, not all these columns have to be included in Z_i. Often on just includes the initial columns of the separate partitions collected in (5.6) through (5.9) or restricts those of the latter even to just one linear combination of its $T(T+1)$ columns, namely $\tilde{x}_i^{(j)}$.

Let now Ψ_i be the non-unique matrix such that $(\omega_i H)^{-1} = \Psi_i'\Psi_i$, then we can construct the transformed variables $W_i^* = \Psi_i W_i$, $\tilde{y}_i^* = \Psi_i \tilde{y}_i$ and $Z_i^\dagger = (\Psi_i')^{-1}Z_i$. Given our derivations in Section 2, it is now straight-forward that GMM estimator (5.5) can also be obtained as the IV estimator

$$\begin{aligned} \hat{\delta}_{GMM} &= \hat{\delta}_{IV}^* = \left[\left(\sum\nolimits_{i=1}^N W_i^{*\prime} Z_i^\dagger \right) \left(\sum\nolimits_{i=1}^N Z_i^{\dagger\prime} Z_i^\dagger \right)^{-1} \left(\sum\nolimits_{i=1}^N Z_i^{\dagger\prime} W_i^* \right) \right]^{-1} \times \\ & \qquad \left(\sum\nolimits_{i=1}^N W_i^{*\prime} Z_i^\dagger \right) \left(\sum\nolimits_{i=1}^N Z_i^{\dagger\prime} Z_i^\dagger \right)^{-1} \left(\sum\nolimits_{i=1}^N Z_i^{\dagger\prime} \tilde{y}_i^* \right) \\ &= (W^{*\prime} P_{Z^\dagger} W^*)^{-1} W^{*\prime} P_{Z^\dagger} \tilde{y}^*, \end{aligned} \tag{5.10}$$

where W^* stacks all W_i^*, and similarly for $Z^\dagger$ and $\tilde{y}^*$. From this IV perspective it is obvious that unnecessary efficiency losses will be incurred if columns of W_i^* that would establish valid instruments for the transformed model

$$\tilde{y}_i^* = W_i^* \delta + \tilde{\varepsilon}_i^*, \tag{5.11}$$

where $\tilde{\varepsilon}_i^* = \Psi_i \tilde{\varepsilon}_i$ with $E(\tilde{\varepsilon}_i^* \tilde{\varepsilon}_i^{*\prime}) = \sigma_\varepsilon^2 I_{T-1}$, are not in the space spanned by the columns of $Z_i^\dagger$.

Let us examine what a lower-triangular transformation Ψ_i may look like. A transformed model (5.11) can be obtained by using a backward orthogonal deviation transformation to the model in levels (5.2), see Arellano and Bover (1995) for the related forward orthogonal deviations transformation. Consider

$$
\begin{aligned}
&\omega_i^{-1/2} S^{1/2} D^* \\
= \;& \omega_i^{-1/2} \begin{bmatrix} \frac{1}{2} & 0 & 0 & \cdots & 0 \\ 0 & \frac{2}{3} & 0 & \ddots & \vdots \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & \frac{T-1}{T} \end{bmatrix}^{1/2} \begin{bmatrix} -1 & 1 & 0 & \cdots & \cdots & 0 \\ -\frac{1}{2} & -\frac{1}{2} & 1 & \cdots & \cdots & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ -\frac{1}{T-2} & \cdots & \cdots & -\frac{1}{T-2} & 1 & 0 \\ -\frac{1}{T-1} & \cdots & \cdots & -\frac{1}{T-1} & -\frac{1}{T-1} & 1 \end{bmatrix} \\
= \;& \omega_i^{-1/2} \begin{bmatrix} \frac{1}{2} & 0 & 0 & \cdots & 0 \\ 0 & \frac{2}{3} & 0 & \ddots & \vdots \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & \frac{T-1}{T} \end{bmatrix}^{1/2} \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ \frac{1}{2} & 1 & \ddots & & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ \frac{1}{T-2} & \frac{2}{T-2} & \ddots & 1 & 0 \\ \frac{1}{T-1} & \frac{2}{T-1} & \cdots & \frac{T-2}{T-1} & 1 \end{bmatrix} D \\
= \;& \omega_i^{-1/2} S^{1/2} F D = \Psi_i D, \qquad (5.12)
\end{aligned}
$$

where we defined the $(T-1)\times T$ matrix D^* and the $(T-1)\times(T-1)$ matrices S, F and Ψ_i implicitly. Note that $D^* y_i$ takes y_{it} for $t = 2, ..., T$ in devation from the (backward) sample average over the indexes $s = 1, ..., t-1$. That $D^* = FD$ and

$$
Var(\Psi_i \tilde{\varepsilon}_i) = Var(\omega_i^{-1/2} S^{1/2} F D \varepsilon_i) = Var(\varepsilon_i^*) = \sigma_\varepsilon^2 I
$$

can easily be verified. This result implies that we can employ lower-triangular transformation matrix Ψ_i to the model in first differences (5.4) in order to get rid of its nonsphericity. Due to its lower-triangular nature it is obvious from (3.2) that using this very same transformation to the instruments Z_i, which are valid instruments for estimating the first-differenced model by GMM, will yield valid instruments $Z_i^* = \Psi_i Z_i$ that allow to estimate model (5.11) consistently by IV.

Above we demonstrated that Z_i is composed of many partitions which are of the form $J_t w_{is}^{(j)}$ $(j = 1, ..., K)$. This implies that Z_i^* is composed of partitions $\Psi_i J_t w_{is}^{(j)} = w_{is}^{(j)} \Psi_i J_t = w_{is}^{(j)} \Psi_{i(t)}$, where $\Psi_{i(t)}$ contains the t final columns of Ψ_i. Because Ψ_i is lower-triangular $\Psi_{i(t)}$ has full column rank, and spans the same subspace as J_t. This implies that $\Psi_i J_t w_{is}^{(j)}$ spans the same subspace as $\omega_i^{-1/2} J_t w_{is}^{(j)}$. So, instead of taking $\Psi_i Z_i$ for Z_i^*, one may simply take $\omega_i^{-1/2} Z_i$. From this we deduce that the kGMM estimator can be obtained by applying GMM to the heteroskedasticity corrected first-differences model

$$
\omega_i^{-1/2} \tilde{y}_i = \omega_i^{-1/2} W_i \delta + \omega_i^{-1/2} \tilde{\varepsilon}_i,
$$

where $Var(\omega_i^{-1/2} \tilde{\varepsilon}_i) = \sigma_\varepsilon^2 H$, upon exploiting the instruments $\omega_i^{-1/2} Z_i$. This gives

$$
\begin{aligned}
\hat{\delta}_{kGMM} = \; & \left[\left(\sum_{i=1}^N \omega_i^{-1} W_i' Z_i \right) \left(\sum_{i=1}^N \omega_i^{-1} Z_i' H Z_i \right)^{-1} \left(\sum_{i=1}^N \omega_i^{-1} Z_i' W_i \right) \right]^{-1} \times \\
& \left(\sum_{i=1}^N \omega_i^{-1} W_i' Z_i \right) \left(\sum_{i=1}^N \omega_i^{-1} Z_i' H Z_i \right)^{-1} \left(\sum_{i=1}^N \omega_i^{-1} Z_i' \tilde{y}_i \right). \qquad (5.13)
\end{aligned}
$$

After first estimating δ consistently by the 1-step estimator

$$\hat{\delta}_{GMM1} = \left[\left(\sum_{i=1}^{N} W_i'Z_i\right)\left(\sum_{i=1}^{N} Z_i'HZ_i\right)^{-1}\left(\sum_{i=1}^{N} Z_i'W_i\right)\right]^{-1} \times \left(\sum_{i=1}^{N} W_i'Z_i\right)\left(\sum_{i=1}^{N} Z_i'HZ_i\right)^{-1}\left(\sum_{i=1}^{N} Z_i'\tilde{y}_i\right), \tag{5.14}$$

consistent residuals $\widehat{\tilde{\varepsilon}}_i^{(1)} = \tilde{y}_i - W_i\hat{\delta}_{GMM1}$ can be obtained, which yield

$$\hat{\sigma}_{\varepsilon,i}^{2(1)} = \widehat{\tilde{\varepsilon}}_i^{(1)\prime}\widehat{\tilde{\varepsilon}}_i^{(1)}/(2T-2). \tag{5.15}$$

From these the feasible and asymptotically optimal two-step estimators

$$\hat{\delta}_{GMM2} = \left[\left(\sum_{i=1}^{N} W_i'Z_i\right)\left(\sum_{i=1}^{N} \hat{\sigma}_{\varepsilon,i}^{2(1)} Z_i'HZ_i\right)^{-1}\left(\sum_{i=1}^{N} Z_i'W_i\right)\right]^{-1} \times \left(\sum_{i=1}^{N} W_i'Z_i\right)\left(\sum_{i=1}^{N} \hat{\sigma}_{\varepsilon,i}^{2(1)} Z_i'HZ_i\right)^{-1}\left(\sum_{i=1}^{N} Z_i'\tilde{y}_i\right) \tag{5.16}$$

and

$$\hat{\delta}_{kGMM2} = \left[\left(\sum_{i=1}^{N} W_i'Z_i/\hat{\sigma}_{\varepsilon,i}^{2(1)}\right)\left(\sum_{i=1}^{N} Z_i'HZ_i/\hat{\sigma}_{\varepsilon,i}^{2(1)}\right)^{-1}\left(\sum_{i=1}^{N} Z_i'W_i/\hat{\sigma}_{\varepsilon,i}^{2(1)}\right)\right]^{-1} \times \left(\sum_{i=1}^{N} W_i'Z_i/\hat{\sigma}_{\varepsilon,i}^{2(1)}\right)\left(\sum_{i=1}^{N} Z_i'HZ_i/\hat{\sigma}_{\varepsilon,i}^{2(1)}\right)^{-1}\left(\sum_{i=1}^{N} Z_i'\tilde{y}_i/\hat{\sigma}_{\varepsilon,i}^{2(1)}\right) \tag{5.17}$$

easily follow.

6. Simulation results for a dynamic panel data model

The model just discussed and various possible GMM implementations will be examined now for the case where $k = 1$, so

$$y_{it} = \eta_i + \tau_t + \gamma y_{i,t-1} + \beta x_{it} + \varepsilon_{it}, \text{ with } \varepsilon_{it} = \sigma_\varepsilon \omega_i^{1/2} \varepsilon_{it}^{\circ}, \tag{6.1}$$

and β has just one element. The individual and time effects are drawn as

$$\begin{aligned} \eta_i &= \sigma_\eta \eta_i^{\circ}, \ \eta_i^{\circ} \sim iidN(0,1), \\ \tau_t &= \sigma_\tau \tau_t^{\circ}, \ \tau_t^{\circ} \sim iidN(0,1). \end{aligned} \tag{6.2}$$

We will consider both the cases where these effects are fully random and hence redrawn each replication, and the case where they are kept fixed. The latter representing for instance the situation where the sample represents a panel of countries over a particular time span, and the former being more relevant for a micro panel where the sample represents a population with many more subjects than N. When the effects are kept constant over the replications the series η_i° and τ_t° are drawn only once, but standardized such that their sample average is zero and their sample variance unity.

The heteroskedasticity ω_i is designed in the same vein as in the pure cross-section model, but now such that it may depend on the individual specific effect η_i° of the relationship and on another independent individual specific effect $\lambda_i^\circ \sim iidN(0,1)$, which both will also affect the regressor x_{it}. Generating first the series

$$h_i(-\phi^2/2) = -\phi^2/2 + \phi[\kappa^{1/2}\eta_i^\circ + (1-\kappa)^{1/2}\lambda_i^\circ] \sim iidN(-\phi^2/2, \phi^2), \tag{6.3}$$

we can next obtain $\omega_i = \exp(h_i(-\phi^2/2))$, which has again lognormal distribution with expectation 1 and variance $2\exp(\phi) - 1$. When the η_i° are generated only once, we do the same for the λ_i°, upon standardizing and orthogonalizing them with respect to η_i°. As before, in the end we rescale the ω_i series such that it sums to n.

Regressor x_{it} is stable first-order autoregressive and is generated as

$$x_{it} = \xi x_{i,t-1} + \pi_\eta \eta_i^\circ + \pi_\lambda \lambda_i^\circ + \pi_\tau \tau_t^\circ + \sigma_v(1-\xi^2)^{1/2}\left[\rho\varepsilon_{it}^\circ + (1-\rho^2)^{1/2}v_{it}^\circ\right], \tag{6.4}$$

with $|\xi| < 1$ and $v_{it}^\circ \sim iidN(0,1)$. The latter drawings are all mutually independent from all ε_{it}° and also from the effects η_i°, λ_i° and τ_t°, which occur in the generating process for x_{it} with coefficients π_η, π_λ and π_τ respectively. Note that ρ, which should obey $|\rho| < 1$, again determines the simultaneity. Parameter σ_v serves to control the variance of x_{it}. Both variants of the autoregressive process for x_{it} require a start-up value. For that we choose

$$x_{is} = 0, \text{ with integer } s \leq 0. \tag{6.5}$$

The role of the start-up time point s is the following. The required initial values $y_{i,0}$ will be obtained by what is generally called "preheating", which means that, although we will only use the sample data for $t = 1, ..., T$ when estimating the model, the generation of the DGP will start off from a point in time $s \leq 0$, by fixing both (6.5) and $y_{i,s} = 0$, which are their unconditional expectations. When $|s|$ is chosen rather large, then for $t \geq 0$, $|\xi| < 1$ and $|\gamma| < 1$ the series for x_{it} and for y_{it} will have approached their stationary track and then an implementation of the so-called system estimator, see Blundell and Bond (1998), is generally seen as one of the most appropriate GMM estimators for this model. Otherwise, and when $|s|$ is rather small, moment conditions associated with the stationarity of the initial conditions y_{i0} and x_{i0} will not be satisfied and an implementation of the Arellano and Bond (1991) GMM estimator is generally seen as the one we should prefer, see also Holtz-Eakin *et al.* (1988). First, we will focus on various possible implementations of the Arellano-Bond classic GMM procedure, as indicated in Table 6.1.

Table 6.1

Overview of examined standard GMM implementations, labelled GMMa through GMMf, for model (6.1) and the number of instruments L in their exploited Z_i

	$\tilde{\tau}$	γ	β	L [for $T = 6$]
GMMa	I_{T-1}	(5.6)	$\tilde{x}_i$	
	$[T-1]$	$[T(T-1)/2]$	[1]	$(T^2+T)/2$ [=21]
GMMb	I_{T-1}	leading columns (5.6)	$\tilde{x}_i$	
	$[T-1]$	$[2T-3]$	[1]	$3T-3$ [=15]
GMMc	I_{T-1}	(5.6)	(5.8)	
	$[T-1]$	$[T(T-1)/2]$	$[T(T-1)/2]$	T^2-1 [=35]
GMMd	I_{T-1}	leading columns (5.6)	leading columns (5.8)	
	$[T-1]$	$[2T-3]$	$[2T-3]$	$5T-7$ [=33]
GMMe	I_{T-1}	(5.6)	(5.7)	
	$[T-1]$	$[T(T-1)/2]$	$[(T-1)(T-2)/2]$	T^2-T [=30]
GMMf	I_{T-1}	leading columns (5.6)	leading columns (5.7)	
	$[T-1]$	$[2T-3]$	$[2T-5]$	$5T-9$ [=21]

Hence, we distinguish six GMM variants and for all we will also examine an adapted variant which uses exactly the same number of instruments, though keenly transformed. These will be indicated as kGMMa through kGMMf. Note again that these do not exploit any extra information in the form of further orthogonality conditions or true values of parameters than exploited by standard GMM. At this stage we do not yet examine feasible versions of GMM and kGMM thus use for both the true heteroskedasticity parameters ω_1 through ω_N. Table 6.1 indicates that all variants include the time dummies I_{T-1} in Z_i because these are exogenous. Implementations a, c and e use the full set of lagged levels of the dependent variable as instruments, whereas b, d and f curtail this large number by taking only the initial two columns of $J_{T+1-t}y_{i,t-2}$ for $t = 2, ..., T$. Regarding instrumenting the regressor x_{it} we examine 5 alternatives. In a and b we use $\tilde{x}_i$ as its own instrument, apparently assuming that it is exogenous. In GMMc and GMMd instruments are used that would be valid if the regressor were predetermined. In GMMc all of these are used and in GMMd these have been curtailed Note that GMMa through GMMd, and also their adapted versions, will be inconsistent when $\rho \neq 0$. In GMMe and GMMf x_{it} is treated as endogenous. GMMe uses all available internal instruments and GMMf is again a curtailed version. In their kGMM counterparts all the regressors and instruments have been multiplied by $\omega_i^{-1/2}$. Note that our DGP does not entail any options for exploiting external instrumental variables (what is in close agreement with a line of conduct often followed in the practice of dynamic panel data analysis).

The Monte Carlo design has many parameters some of which we will keep constant most of the time. The first results to be presented below focus on the following chosen numerical combinations[4]:

$$\left.\begin{array}{l} N=200;\ T=6;\ s=-5; \\ \gamma\in\{0.3,\ 0.8\};\ \beta=1-\gamma; \\ \sigma_\varepsilon=1;\ \sigma_\eta=\sigma_\varepsilon(1-\gamma);\ \sigma_\tau=\sigma_\varepsilon(1-\gamma); \\ \xi=0.8;\ \sigma_v=3;\ \pi_\eta=\pi_\lambda=\pi_\tau=\sigma_v(1-\xi); \\ \rho=\{0,0.6\};\ \phi=\{0.5,1\};\ \kappa=0.5. \end{array}\right\} \tag{6.6}$$

So, x_{it} is stationary, though not strictly covariance stationary yet, and relatively smooth through time. First we focus on results based on just one arbitrary draw for the individual and time effects. For $\phi=1$ these gave rise to ω_i values varying between .06 and 9.9.

Table 6.2

Simulation results on standard GMM (kGMM) for $N=200$, $T=6$, $\rho=0$, $\phi=1$

	$\gamma=0.8$			$\beta=0.2$		
	Bias	Std.Err	RMSE	Bias	Std.Err	RMSE
a	-.208 (-.022)	.200 (.060)	.289 (.064)	-.011 (.001)	.024 (.011)	.026 (.011)
b	-.204 (-.015)	.248 (.062)	.321 (.064)	-.011 (.001)	.025 (.011)	.027 (.011)
c	-.072 (-.024)	.072 (.041)	.102 (.048)	-.084 (-.019)	.107 (.046)	.136 (.050)
d	-.083 (-.024)	.096 (.052)	.127 (.058)	-.121 (-.024)	.159 (.062)	.199 (.067)
e	-.085 (-.033)	.084 (.052)	.119 (.062)	-.088 (-.022)	.124 (.052)	.152 (.056)
f	-.099 (-.035)	.117 (.072)	.153 (.079)	-.139 (-.029)	.194 (.075)	.238 (.081)

Focussing on standard GMM first, we see from Table 6.2 (where $\rho=0$ so all estimators are consistent) that there is often substantial bias. Estimating β benefits when the single instrument $\tilde{x}_i$ is used at the expense of the precision of the γ estimate. Reducing the number of instruments is found to be beneficial neither for the bias nor for the standard errors. The same patterns show up for the kGMM results, but at a much more attractive level. It shows moderate bias and much smaller standard deviations, yielding RMSE values which would allow rather precise inferences on the true parameter values, whereas this seems impossible at the chosen parametrization by standard GMM. Obviously its much stronger instruments reduce both the bias and the standard deviation. To obtain more attractive results for standard GMM would require to increase the value of σ_v, by which the signal to noise ratio would improve, or by increasing N.

Table 6.3

Simulation results on standard GMM (kGMM) for $N=200$, $T=6$, $\rho=0.6$, $\phi=1$

	$\gamma=0.8$			$\beta=0.2$		
	Bias	Std.Err	RMSE	Bias	Std.Err	RMSE
a	-.131 (.073)	.260 (.066)	.292 (.098)	.292 (.109)	.028 (.014)	.293 (.110)
b	-.055 (.094)	.407 (.076)	.410 (.121)	.302 (.106)	.040 (.015)	.304 (.107)
c	-.339 (-.252)	.064 (.055)	.345 (.258)	.123 (.025)	.140 (.079)	.186 (.083)
d	-.370 (-.286)	.081 (.068)	.379 (.294)	.026 (-.043)	.202 (.101)	.203 (.109)
e	-.038 (-.006)	.069 (.052)	.079 (.052)	.088 (.054)	.095 (.047)	.130 (.072)
f	-.013 (.012)	.088 (.072)	.089 (.073)	.083 (.057)	.138 (.069)	.161 (.089)

[4] In designing the Monte Carlo experiments in Sections 4 and 6 the chosen parametrizations and the established links between particular parameter values have been chosen deliberately in order to aim at orthogonalization and optimization as set out in Kiviet (2012).

The only design parameter with a different value in Table 6.3 is ρ. Only variants e and f are consistent now. The results show that simultaneity should better not be neglected. Variant e outperforms f and kGMM again leads to less bias and smaller standard deviations. Nevertheless, upon comparing the kGMM results of a and e regarding the standard errors it is striking to note that using relatively few instruments that fit the regressors well lead to relatively high precision even if some of the instruments are actually invalid.

Table 6.4

Simulation results on standard GMM (kGMM) for $N = 200$, $T = 6$, $\rho = 0.0$, $\phi = 1$

	$\gamma = 0.3$			$\beta = 0.7$		
	Bias	Std.Err	RMSE	Bias	Std.Err	RMSE
a	-.064 (-.026)	.123 (.077)	.139 (.082)	-.004 (.008)	.025 (.027)	.025 (.028)
b	-.079 (-.030)	.164 (.101)	.183 (.105)	-.006 (.009)	.026 (.034)	.027 (.036)
c	-.008 (-.003)	.029 (.017)	.030 (.017)	-.009 (-.002)	.070 (.030)	.071 (.030)
d	-.009 (-.003)	.034 (.020)	.035 (.020)	-.019 (-.003)	.097 (.037)	.099 (.037)
e	-.043 (-.016)	.070 (.043)	.082 (.046)	-.001 (.004)	.077 (.036)	.077 (.036)
f	-.041 (-.021)	.086 (.057)	.095 (.061)	-.021 (.000)	.108 (.042)	.110 (.042)

Table 6.4 looks again at $\rho = 0$ but now the value of γ is more moderate. Here we note that for implementations a and b GMM actually works rather well for β (but poorly for γ) and then kGMM does not outperform it. Nevertheless, on the whole, kGMMc performs best here. Of course, if we had adapted GMM such that we would simply add the transformed instruments to the standard ones the asymptotic variance could not be worse than that of standard GMM. Therefore, we will now examine such an implementation, that we already labelled kfGMM.

Table 6.5

Simulation results on standard GMM (kfGMM) for $N = 200$, $T = 6$, $\rho = 0.0$, $\phi = 1$

	$\gamma = 0.3$			$\beta = 0.7$		
	Bias	Std.Err	RMSE	Bias	Std.Err	RMSE
a	-.064 (-.009)	.123 (.028)	.139 (.030)	-.004 (.003)	.025 (.014)	.025 (.014)
b	-.079 (-.008)	.164 (.031)	.183 (.032)	-.006 (.002)	.026 (.014)	.027 (.014)
c	-.008 (-.005)	.029 (.015)	.030 (.015)	-.009 (-.001)	.070 (.020)	.071 (.020)
d	-.009 (-.004)	.034 (.016)	.035 (.016)	-.019 (-.001)	.097 (.022)	.099 (.022)
e	-.043 (-.017)	.070 (.030)	.082 (.034)	-.001 (.010)	.077 (.030)	.077 (.032)
f	-.041 (-.014)	.086 (.034)	.095 (.037)	-.021 (.009)	.108 (.034)	.110 (.035)

Note that doubling the number of instruments, although not always beneficial for the bias, improves the quality of the adaptation of GMM tremendously. Note that in Table 6.5 the RMSE values are reduced by at least 50% or occasionally by as much as 80%. Apparently one should not bother so much about the number of instruments. That standard GMM has much larger bias and standard deviation is not due to the large number of instruments, but primarily to their relative weakness. This weakness with respect to those of adapted GMM depends on the seriousness of the heteroskedasticity. Therefore, we will now look into a case where this is more moderate.

Table 6.6

Simulation results on standard GMM (kfGMM) for $N = 200$, $T = 6$, $\rho = 0$, $\phi = .5$						
	$\gamma = 0.8$			$\beta = 0.2$		
	Bias	Std.Err	RMSE	Bias	Std.Err	RMSE
a	-.162 (-.030)	.175 (.045)	.239 (.054)	-.009 (.001)	.024 (.015)	.026 (.015)
b	-.144 (-.022)	.207 (.046)	.252 (.051)	-.008 (.001)	.025 (.015)	.026 (.015)
c	-.059 (-.033)	.065 (.033)	.088 (.046)	-.071 (-.008)	.099 (.029)	.122 (.030)
d	-.069 (-.023)	.086 (.034)	.110 (.041)	-.101 (-.006)	.143 (.031)	.176 (.031)
e	-.067 (-.037)	.074 (.037)	.099 (.052)	-.071 (.001)	.113 (.034)	.133 (.034)
f	-.077 (-.025)	.102 (.038)	.128 (.046)	-.107 (.001)	.170 (.037)	.201 (.037)

Table 6.6 should be compared with Table 6.2. The only two differences are that now ϕ is .5, giving rise to ω_i values ranging from .28 to 3.6, and we examine kfGMM instead of kGMM. The quality of standard GMM has improved because the smaller ϕ has been less detrimental for the strength of its instruments. However, despite the milder heteroskedasticity, the doubling of the number of instruments in kfGMM makes that it outperforms standard GMM, again leading to RMSE values that are at most about half but sometimes just 20% of those of standard GMM. Corresponding t-values could be two to five times higher.

7. Empirical illustration

The proposed adapted GMM estimators kGMM and kfGMM will be applied now to the classic empirical example used in Arrelano and Bond (1991). It concerns an unbalanced panel of 140 UK companies over the years 1976 -1984 used to estimate a dynamic employment equation specified as

$$n_{it} = \alpha_1 n_{i,t-1} + \alpha_2 n_{i,t-2} + \beta' x_{it} + \tau_t + \eta_i + \varepsilon_{it}. \tag{7.1}$$

Here n_{it} is the log of employment in firm i at the end of year t and the vector x_{it} includes w_{it}, the log of real product wage, k_{it}, the log of gross capital stock and ys_{it}, the log of industry output, and their lags. Here we only present results for the case where the regressors x_{it} are assumed to be strictly exogenous. The model contains both individual and time effects. In the second and third columns of Table 7.1 we replicate the results in Arrelano and Bond (1991, Table 4, column a2).

Table 7.1

Various GMM estimates for the employment equation (7.1)

	2-step GMM-AB		2-step kGMM		2-step kfGMM	
regressor	coef.	std.err	coef.	std.err.	coef.	std.err.
$n_{i,t-1}$	.629	.090	.557	.110	.610	.036
$n_{i,t-2}$	-.065	.027	-.051	.028	-.087	.014
w_{it}	-.526	.054	-.332	.045	-.549	.024
$w_{i,t-1}$	.311	.094	.215	.055	.344	.035
k_{it}	.278	.045	.230	.025	.336	.015
$k_{i,t-1}$	.014	.053	.047	.033	-.010	.020
$k_{i,t-2}$	-.040	.026	-.042	.021	-.030	.013
ys_{it}	.592	.116	.636	.069	.565	.052
$ys_{i,t-1}$	-.566	.140	-.286	.097	-.610	.066
$ys_{i,t-2}$	.101	.113	-.029	.066	.050	.058

The 1-step Arrelano and Bond (GMM-AB) estimates were used to estimate the cross-sectional variances of the idiosyncratic errors ε_{it}. After normalization, the ω_i values ranged from .007 to 7.92, which corresponds about to $1 < \phi < 1.4$ (compare Table 4.1), showing that there is serious cross-sectional heteroskedasticity in the disturbances. With this estimated ω_i, we constructed new instrumental variables by multiplying them by ω_i^{-1}. As we explained before, each individuals exogenous instruments should also be premultiplied by H^{-1}, but we did not because of the complications stemming from the unbalancedness of the panel. Using the original Z_i yields the 2-step GMM-AB. Using the same algorithm with the adpated instruments yields 2-step kGMM. And using both the standard AB instruments and the transformed ones produces 2-step kfGMM. Especially the latter yield much smaller estimated standard errors, sometimes less than 50% of the standard results. These calculated standard errors are robust for time-series heteroskedasticity, as in the Arellano and Bond (1991) paper. Of course, these standard error estimates may be rather inaccurate, see Windmeijer (2005). Taking them serious nevertheless, we see that the kfGMM results provide enhanced evidence in favour of omitting regressor $ys_{i,t-2}$ from the specification and against omitting $k_{i,t-2}$.

8. Conclusions

We reveal an inherent unfavourable and yet generally unperceived feature of GMM as it is currently usually implemented. Extracting from the assumed orthogonality conditions instrumental variables such that they are reasonably effective (strong) for the regressors in the habitual sense, as understood for IV estimation, implies that these very same instruments will be much weaker in the context of GMM. This is, because implicitly GMM estimates a transformed model, in order to get rid of any non-sphericity of the disturbances, but at the same time this transformation affects the instruments in such a way that they will actually be much weaker than the researcher realizes. It is shown, however, that various relatively simple precautions enable to neutralize this weakening process of the instruments. Moreover, options are uncovered for profitable extensions of the set of instruments still in connection to just the originally adopted orthogonality conditions. By simulation it is shown that empirically relevant forms of heteroskedasticity undermine the quality of standard GMM estimates and that some of the suggested adapted forms of GMM yield estimates that show both less bias and smaller standard errors. Reductions of the root mean squared errors of the coefficient estimates of the alleged optimal standard GMM technique by a factor 2 or more are in fact not exceptional.

In this paper we only examined GMM estimators for models linear in the coefficients of the regressors; however, the results have implications for general nonlinear models too that deserve further investigation. We also show that in models with endogenous regressors and serial correlation of moving average form, the set of valid instruments can easily be extended by substantially stronger instruments, by seeking instruments not for the original model but for the model after a transformation that removes the serial correlation. Practical consequences of this have not been examined here yet.

For one classic empirical data set it has been examined what the practical consequences are. From that it seems that the conclusions drawn from our Monte Carlo designs are realistic, although one should keep in mind that the synthetic experiments produce accurate assessments of true bias and true standard deviation, whereas for the empirical result the bias cannot be assessed because the true parameter values are unknown, and the obtained estimated standard errors may be very misleading regarding the underlying true standard deviations. Nevertheless, we do find substantially smaller estimated standard errors and therefore we unreservedly recommend to use GMM in future as exposed here in a more keen way than presently is the custom.

References

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

Andrews, D.W.K., Stock, J.H., 2007. Inference with Weak Instruments, Chapter 6 in: Blundell, R., Newey, W.K., Persson, T. (eds.), *Advances in Economics and Econometrics, Theory and Applications*, 9th Congress of the Econometric Society, Vol. 3. Cambridge, UK: Cambridge University Press.

Arellano, M., 2002. Sargan's instrumental variables estimation and the generalized method of moments. Journal of Business and Economic Statistics 20, 450-459.

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.

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

Davidson, R., MacKinnon, J.G., 2006. *Econometric Theory and Methods.* New York, Oxford University Press.

Donald, S.G., Newey, W., 2001. Choosing the number of instruments. *Econometrica* 69, 1161-1191.

Hall, A.R., 2005. *Generalized Method of Moments.* Oxford University Press.

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

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

Kiviet, J.F., 2012. *Monte Carlo Simulation for Econometricians.* Forthcoming in: Foundations and Trends in Econometrics, Now Publishers, New York.

Sargan, J.D., 1958. The estimation of economic relationships using instrumental variables. *Econometrica* 26, 393-415.

Sargan, J.D., 1959. The estimation of relationships with autocorrelated residuals by the use of instrumental variables. *Journal of the Royal Statistical Society, Series B* 21, 91-105.

Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. *Econometrica* 65, 557-586.

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