Efficiency Gains by Modifying GMM Estimation in Linear Models under Heteroskedasticity

Jan F. Kiviet* and Qu Feng†

Version of 5 October, 2015‡
JEL-code: C01, C13, C26.

Keywords: *efficiency, generalized method of moments, instrument strength, heteroskedasticity, conditional moment assumptions*

Abstract

While coping with nonsphericality of the disturbances, standard GMM suffers from a blind spot for exploiting the most effective instruments when these are obtained directly from unconditional rather than conditional moment assumptions. For instance, standard GMM counteracts that exogenous regressors are used as their own optimal instruments. This is easily seen after transmuting GMM for linear models into IV in terms of transformed variables. It is demonstrated that modified GMM (MGMM), exploiting straight-forward modifications of the instruments, can achieve substantial efficiency gains and bias reductions, even under mild heteroskedasticity. A feasible MGMM implementation and its standard error estimates are examined and compared with standard GMM and IV for a range of typical models for cross-section data, both by simulation and by empirical illustration.

*Emeritus Professor of Econometrics, Amsterdam School of Economics, University of Amsterdam, Valckenierstraat 65, 1018 XE Amsterdam, The Netherlands (j.f.kiviet@uva.nl) and Visting Professor at the Division of Economics, School of Humanities and Social Sciences, Nanyang Technological University, 14 Nanyang Drive, Singapore 637332 (jfkiviet@ntu.edu.sg).

†Division of Economics, School of Humanities and Social Sciences, Nanyang Technological University, 14 Nanyang Drive, Singapore 637332 (qfeng@ntu.edu.sg).

‡We would like to thank Manuel Arellano, Paul Bekker, Joerg Breitung, Geert Dhaene, Guido Kuersteiner, Choon-Geol Moon, Liangjun Su, Tom Wansbeek and participants at various seminars and conferences for their comments and suggestions on earlier versions of this study.

1. Introduction

For over three decades GMM (generalized method of moments) excels as the generic orthogonality conditions based optimal technique for semiparametric estimation. It subsumes the majority of linear and nonlinear econometric estimators. The status of GMM seems undisputable: the standard GMM technique of Hansen (1982) deals optimally with both any nonsphericality of the disturbances and any overidentification, given the actual set of instrumental variables that is being used.

However, as discussed in Newey (1993) and Arellano (2003a,b), this does not necessarily imply that the conditional moment conditions as such are being exploited optimally. These allow to exploit as instruments also non-trivial transformations of instrumental variables. In line with Davidson and MacKinnon (2004), we will show that especially in models with heteroskedasticity a GLS-like model transformation leads to orthogonality conditions from which much stronger instruments may emerge. These will not only lead to reduced bias and variance, but also to a better correspondence between actual distribution in finite samples and its asymptotic approximation. This all supports a strategy for finding instrumental variables close to the original IV (instrumental variables) approach suggested by Sargan (1958, 1959).[1]

For the sake of simplicity we will focus here on both unfeasible and feasible standard and modified implementations of GMM just in linear cross-sectional models with heteroskedasticity. However, we conjecture that the results will have implications for more general models too. A similar analysis regarding serial correlation (the major worry in Sargan's approach) in models for dependent observations is left for future research. Our analysis of possibilities to improve on the efficiency of the standard implementation of GMM under heteroskedasticity leads to the conclusion that in practice one should aim to weigh observations first in order to get as close to homoskedasticity as possible. Not before, but after that, one should design a matrix of instruments according to the adopted orthogonality conditions in terms of the weighted variables. Similar conclusions have been drawn at various places in different contexts in the literature, see Bowden and Turkington (1984), White (1986), Baltagi and Li (1992), Wooldridge (2010 p.351, 2013 p.516) and Iglesias and Phillips (2012). Models with endogenous regressors and heteroskedasticity have also been considered in another line of research (for a recent contribution see the references in and the study by Hausman et al., 2012), but there

[1] 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).

the purpose is not to improve estimator efficiency, but to make the variance estimates of suboptimal estimators robust in the presence of heteroskedasticity.

To reveal the actual efficiency gains of the modified implementation of GMM requires some well-designed simulation investigations. For feasible implementations, a parametric setup, where use is made of a proper specification of the determining factors of the (conditional) heteroskedasticity, yields substantial efficiency gains in comparison with the standard GMM estimator. We find that under nonextreme circumstances root mean squared errors may reduce by about 50%. When this can be estimated consistently one can employ instruments which asymptotically attain the efficiency bound achieved by the unfeasible optimal instruments.

The structure of this study is as follows. In section 2 we show how GMM can be interpreted in terms of transformed IV. Optimality of instruments is also discussed. Section 3 considers a modified GMM estimator and a feasible implementation. It also discusses the estimation of the variance of coefficient estimators. Section 4 presents a Monte Carlo design for a typical family of simultaneous heteroskedastic cross-section models followed in Section 5 by simulation results which demonstrate the huge potential efficiency gains by modifying GMM. In Section 6 alternative implementations of feasible GMM are applied to empirical data to illustrate the practical consequences of the suggested modifications. Section 7 concludes.

2. GMM and optimality under heteroskedasticity in linear models

We consider the linear regression model $y_i = x_i'\beta + \varepsilon_i$, $i = 1, ..., n$, or in matrix form

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

where the $n \times K$ regressor matrix $X = (x_1 \ ... \ x_n)'$ is supposed to have full column rank. Some regressors may be endogenous, hence possibly $E(x_i\varepsilon_i) \neq 0$. Available are also L instrumental variables z_i based on moment assumptions

$$E(\varepsilon_i|z_i) = 0. \tag{2.2}$$

$Z = (z_1 \ ... \ z_n)'$ is an $n \times L$ matrix of rank $L \geq K$. Provided $Z'X$ has rank K the IV (or 2SLS) estimator is given by

$$\hat{\beta}_{IV} = (X'P_Z X)^{-1} X'P_Z y, \tag{2.3}$$

where $P_A = A(A'A)^{-1}A'$ for any full column rank matrix A.

If $\varepsilon = (\varepsilon_1 \ ... \ \varepsilon_n)' \sim (0, \Sigma)$, with a nonspherical $n \times n$ matrix Σ, then the preferred estimator of β is obtained by GMM, given by

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

where Ω can be any non-zero scalar multiple of Σ. It is convenient to define

$$\Omega = \Sigma/\sigma_\varepsilon^2 \text{ with } \sigma_\varepsilon^2 = tr(\Sigma)/n, \text{ so that } tr(\Omega) = n. \tag{2.5}$$

The GMM estimator is optimal in the sense that (under standard asymptotics) the variance of its limiting distribution

$$n^{1/2}(\hat{\beta}_{GMM} - \beta) \xrightarrow{d} N(0, \operatorname{plim} n[X'Z(Z'\Sigma Z)^{-1}Z'X]^{-1}) \tag{2.6}$$

is minimal in a matrix sense. GMM simplifies to IV when $\Sigma = \sigma_\varepsilon^2 I$ or $\Omega = I$.

In the special case that $L = K$ the matrix $Z'X$ is invertible and, irrespective of the value of Ω, we obtain $\hat{\beta}_{GMM} = (Z'X)^{-1}Z'y = \hat{\beta}_{IV}$. When the regressors are exogenous, substituting X for Z we find, also when $\Omega \neq I$,

$$\hat{\beta}_{GMM} = (X'X)^{-1}X'y = \hat{\beta}_{OLS}. \tag{2.7}$$

However, it is well know that in that case the optimal estimator is

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y. \tag{2.8}$$

Apparently, the optimality of GMM is not universal. We will clarify this disturbing result[2] below, but first we shall demonstrate that in linear models GMM is equivalent to applying IV to a properly transformed model.

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

$$y^* = X^*\beta + \varepsilon^*, \tag{2.9}$$

where $y^* = \Psi y$, $X^* = \Psi X$ and $\varepsilon^* = \Psi\varepsilon \sim (0, \sigma_\varepsilon^2 I)$. Define $Z^\dagger = (\Psi')^{-1}Z$. Because $\varepsilon^* \sim (0, \sigma_\varepsilon^2 I)$ we may estimate the transformed model (2.9) by IV, which yields

$$(X^{*\prime}P_{Z^\dagger}X^*)^{-1}X^{*\prime}P_{Z^\dagger}y^* = [X'Z(Z'\Omega Z)^{-1}Z'X]^{-1}X'Z(Z'\Omega Z)^{-1}Z'y = \hat{\beta}_{GMM}, \tag{2.10}$$

[2] Our clarification is less abstract than those presented by Davidson and MacKinnon (2004, p.358), Cameron and Trivedi (2005, p.747) and Wooldridge (2010, p.542).

where $P_{Z^\dagger} = (\Psi')^{-1}Z(Z'\Omega Z)^{-1}Z'\Psi^{-1}$. Hence, GMM corresponds algebraically to applying IV to a "GLS-like" transformed model, exploiting instruments $Z^\dagger = (\Psi')^{-1}Z$, instead of $Z^* = \Psi Z$. Since $E(Z^{\dagger\prime}\varepsilon^*) = E(Z'\varepsilon) = 0$, the instruments $Z^\dagger$ used for estimating the transformed model (2.9) exploit still exactly the original moment conditions.

Assume that $X = (X_1, X_2)$, where X_1 contains $K_1 > 0$ exogenous regressors, i.e., $E(\varepsilon \mid X_1) = 0$, and X_2 contains $K_2 = K - K_1 \geq 0$ possibly endogenous regressors. Because $E(\varepsilon \mid X_1) = 0$ implies $E(\varepsilon^* \mid X_1^*) = E(\Psi\varepsilon \mid \Psi X_1) = 0$ it is obvious that when estimating the transformed model (2.9) by IV we should preferably include X_1^* in the matrix of instruments. However, standard GMM would include X_1 in the matrix of instruments for the untransformed model, implying that $X_1^\dagger = (\Psi')^{-1}X_1$ is a component of the instrument matrix it employs to the transformed model. Note, though, that it is unlikely that in the first-stage regression of X^* on the instruments $Z^\dagger = (X_1^\dagger\ Z_2^\dagger)$ a perfect fit will be realized for X_1^*, whereas this occurs when including X_1^* in the instrument set for the transformed model.

Standard GMM does not automatically use regressors which are exogenous in the transformed model as instruments for estimating the transformed model. The optimality of standard GMM is achieved over the sample moment conditions expressed in terms of the chosen instruments Z. The above illustrates that efficiency gains seem possible by allowing to consider transformations of Z as well.

Newey (1993) and Arellano (2003a, Appendix B) address the optimality of instruments in a setup, where (y_i, x_i', z_i') are i.i.d. and the conditional moment assumption $E(y_i - x_i'\beta | z_i) = 0$ is made. Then the optimal instruments[3] are shown to be given by the $K \times 1$ vector

$$\bar{g}(z_i) = E(x_i|z_i)/E(\varepsilon_i^2|z_i). \tag{2.11}$$

The corresponding unfeasible optimal (UO) GMM estimator uses $\bar{g}(z_i)$ as instruments. It has asymptotic variance

$$V_{UO} = \{E(x_i|z_i)[E(\varepsilon_i^2|z_i)]^{-1}E(x_i'|z_i)\}^{-1} = E(\varepsilon_i^2|z_i)\{\bar{g}(z_i)\bar{g}(z_i)'\}^{-1}. \tag{2.12}$$

Newey (1993) shows that this variance V_{UO} is a lower bound for the asymptotic variance of all IV/GMM estimators for the present model, exploiting (2.2), and even in a stronger sense, achieving the semi-parametric efficiency bound (Chamberlain, 1987). When z_i contains some exogenous elements of x_i, so $x_i = (x_{1i}', x_{2i}')'$ and $z_i = (x_{1i}', z_{2i}')'$, then

$$E(x_i'|z_i) = (x_{1i}', E(x_{2i}'|z_i))\,,$$

[3] See also Cameron and Trivedi (2005, p.188).

which indicates that the optimal instrument matrix should have a component proportional to $\Omega^{-1}X_1$, which vindicates our earlier more informal derivations.

The optimal instruments for the linear model under conditions $E(\varepsilon_i|z_i) = 0$ are also discussed in Davidson and MacKinnon (2004, Chapter 9) for dependent data. They show that the matrix of optimal instruments is given by $\sigma_\varepsilon^{-2}\Omega^{-1}\bar{X}$, where $\bar{X} = (\bar{x}_1, \cdots, \bar{x}_n)'$ and $\bar{x}_i$ is defined as $E(x_i|\Xi_i)$, where Ξ_i contains the set of all deterministic functions of the elements of $(z_i, ..., z_1)$. It easily follows that an expression for the asymptotic variance of the GMM estimator using these optimal instruments is

$$\sigma_\varepsilon^2 \operatorname{plim} n(\bar{X}'\Omega^{-1}\bar{X})^{-1},$$

which corresponds to (2.12). Since $\bar{X}$ is not observable, Davidson and MacKinnon (2004, p.361) suggest to use instruments $\Omega^{-1}Z$. When the span of $\bar{X}$ is a subset of that of Z, the GMM estimator using instruments $\Omega^{-1}Z$ has the same asymptotic variance, implying that $\Omega^{-1}Z$ should be used as optimal instrument matrix. This estimator is only feasible when Ω is known.

3. Modified GMM and implementations

When Σ and Ω are diagonal, say $\Sigma = \operatorname{diag}(\sigma_1, ..., \sigma_n)$ with $\sigma_i = \sigma_\varepsilon^2\omega_i$ and $\sum_{i=1}^n \omega_i = n$ with all $\omega_i > 0$, then the moment conditions $E(z_i\varepsilon_i) = 0$ are equivalent with $E(z_i\varepsilon_i/\omega_i) = 0$. Hence, a possible modification of GMM is using instrument matrix $\Omega^{-1}Z$, rather than Z. This yields the modified GMM (MGMM) estimator

$$\hat{\beta}_{MGMM} = [X'\Omega^{-1}Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}X]^{-1}X'\Omega^{-1}Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}y, \tag{3.1}$$

or equivalently, $\hat{\beta}_{MGMM} = [X'\Sigma^{-1}Z(Z'\Sigma^{-1}Z)^{-1}Z'\Sigma^{-1}X]^{-1}X'\Sigma^{-1}Z(Z'\Sigma^{-1}Z)^{-1}Z'\Sigma^{-1}y$ with limiting distribution

$$n^{1/2}(\hat{\beta}_{MGMM} - \beta) \xrightarrow{d} N(0, \operatorname{plim} n[X'\Sigma^{-1}Z(Z'\Sigma^{-1}Z)^{-1}Z'\Sigma^{-1}X]^{-1}). \tag{3.2}$$

Since $\hat{\beta}_{MGMM} = (X^{*\prime}P_{Z^*}X^*)^{-1}X^{*\prime}P_{Z^*}y^*$, MGMM corresponds algebraically to applying IV to the transformed model, using instruments $Z^* = \Psi Z$. MGMM[4] would be more efficient than GMM when the difference between the matrices $\operatorname{plim} n^{-1}X^{*\prime}P_{Z^*}X^*$ and $\operatorname{plim} n^{-1}X^{*\prime}P_{Z^\dagger}X^*$ is positive (semi-)definite.

When Ω is unknown, we could – as in feasible GLS – parametrize the functional form of the skedastic function, for instance as $\sigma_i = h(z_i'\gamma)$. From consistent estimates

[4] Bowden and Turkington (1984, p.69) consider this estimator, but suppose it to be of limited value.

$\hat{\sigma}_i = h(z_i'\hat{\gamma})$ a consistent estimate $\hat{\Sigma}$ can be obtained.[5] This enables to obtain the feasible parametric MGMM estimator, which uses the instruments $\hat{\Sigma}^{-1}Z$ to the untransformed model, and is given by

$$\hat{\beta}_{FpMGMM} = [X'\hat{\Sigma}^{-1}Z(Z'\hat{\Sigma}^{-1}Z)^{-1}Z'\hat{\Sigma}^{-1}X]^{-1}X'\hat{\Sigma}^{-1}Z(Z'\hat{\Sigma}^{-1}Z)^{-1}Z'\hat{\Sigma}^{-1}y. \qquad (3.3)$$

Here $E(\varepsilon_i^2) = h(z_i'\gamma)$ establishes a moment condition that is nonlinear in β, which could be exploited directly when deriving a method of moments estimator. In our implementations, however, we use a simple 2-step procedure, in which $\hat{\gamma}$ is obtained by regressing $h^{-1}(\hat{\varepsilon}_i^2)$ on z_i, where $\hat{\varepsilon}_i = y_i - x_i'\hat{\beta}_{IV}$ is consistent for the disturbance ε_i.

The standard GMM estimator (2.4) is unfeasible for unspecified heteroskedasticity, whereas its feasible nonparametric alternative is

$$\hat{\beta}_{FnpGMM} = [X'Z\hat{S}_{zz}^{-1}Z'X]^{-1}X'Z\hat{S}_{zz}^{-1}Z'y, \qquad (3.4)$$

where $\hat{S}_{zz} = n^{-1}\sum_{i=1}^{n}\hat{\varepsilon}_i^2 z_i z_i'$ is a consistent estimator for $Z'\Sigma Z/n$ using IV residuals $\hat{\varepsilon}_i$.

In the simulations to follow, we will focus on the relatively simple case of heteroskedasticity in cross-sections. First, we examine standard GMM and its modification MGMM for the unrealistic situation that Ω is supposed to be known. Exploiting the true value of Ω will disclose what the potential differences will be between the for practitioners more interesting feasible implementations. Next, we examine what the actual losses are in terms of RMSE (root mean squared error) when Ω is assessed in the implementations described above. We will also make comparisons with IV, and with the under simultaneity inconsistent estimators ordinary and weighted least squares (OLS and WLS).

In addition to comparing bias, actual standard deviation, and root mean squared error, we will examine also how well the various consistent estimators are able to assess their actual efficiency, by examining the bias in the estimates provided by the standard expressions for their standard errors. These calculations will be based on square roots of the diagonal elements of the following list of variance estimators:

$$\widehat{Var}(\hat{\beta}_{GMM}) = \hat{\sigma}_\varepsilon^2[X'Z(Z'\Omega Z)^{-1}Z'X]^{-1},\ \hat{\sigma}_\varepsilon^2 = (y - X\hat{\beta}_{GMM})'(y - X\hat{\beta}_{GMM})/n$$
$$\widehat{Var}(\hat{\beta}_{FnpGMM}) = n[X'Z\hat{S}_{zz}^{-1}Z'X]^{-1}$$
$$\widehat{Var}(\hat{\beta}_{MGMM}) = \hat{\sigma}_\varepsilon^2[X^{*\prime}P_{Z^*}X^*]^{-1},\ \hat{\sigma}_\varepsilon^2 = (y^* - X^*\hat{\beta}_{MGMM})'(y^* - X^*\hat{\beta}_{MGMM})/n$$
$$\widehat{Var}(\hat{\beta}_{FpMGMM}) = [X'\hat{\Sigma}^{-1}Z(Z'\hat{\Sigma}^{-1}Z)^{-1}Z'\hat{\Sigma}^{-1}X]^{-1},\ \hat{\sigma}_i = h(z_i'\hat{\gamma})$$
$$\widehat{Var}(\hat{\beta}_{IV}) = \hat{\sigma}_\varepsilon^2[X'P_ZX]^{-1},\ \hat{\sigma}_\varepsilon^2 = (y - X\hat{\beta}_{IV})'(y - X\hat{\beta}_{IV})/n$$
$$\widehat{Var}_{np}(\hat{\beta}_{IV}) = n[X'P_ZX]^{-1}X'Z(Z'Z)^{-1}\hat{S}_{zz}(Z'Z)^{-1}Z'X[X'P_ZX]^{-1}$$

[5] Newey (1993) discusses how the assessment of asymptotically optimal instruments $\Omega^{-1}Z$ should be approached in a nonparametric context.

$\widehat{Var}(\hat{\beta}_{IV})$ is improper in case $\Omega \neq I$, but a nonparametric robustification for this variance is provided by $\widehat{Var}_{np}(\hat{\beta}_{IV})$.

4. Simulation design for a heteroskedastic cross-section model

We shall design a data generating process (DGP) in which we can easily change the seriousness and characteristics of the heteroskedasticity, the degree of simultaneity, the strength of the instruments, the significance of individual regressors and the general fit 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 DGP 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 $j = 3,4,5;\ i = 1,...,n$. Of course, the two endogenous variables x_{i3} and y_i, and the pattern of the heteroskedasticity $(\omega_1,...,\omega_n)$ in the disturbances ε_i, where the ω_i are the diagonal elements of Ω, have to be designed such that these seem very realistic for typical cross-section applications.

The structural form equation will be generated as

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i \tag{4.1}$$

and the reduced form equation for x_{i3} by

$$x_{i3} = \pi_{31} + \pi_{32} x_{i2} + \pi_{33} z_{i3} + \pi_{34} z_{i4} + \pi_{35} z_{i5} + v_i, \tag{4.2}$$

where

$$v_i = \sigma_v \omega_i^{1/2} v_i^{\circ}, \text{ with } v_i^{\circ} \sim iidN(0,1), \tag{4.3}$$

$$\varepsilon_i = \sigma_\varepsilon \omega_i^{1/2} \left[\rho v_i^{\circ} + (1-\rho^2)^{1/2} \varepsilon_i^{\circ}\right], \text{ with } \varepsilon_i^{\circ} \sim iidN(0,1). \tag{4.4}$$

Hence, the reduced form disturbances $v_i \sim N(0, \sigma_v^2 \omega_i)$ and the structural equation disturbances $\varepsilon_i \sim N(0, \sigma_\varepsilon^2 \omega_i)$ are affected by the same heteroskedasticity pattern. Parameter $\rho \in (-1, +1)$ is the correlation coefficient of ε_i and v_i and expresses the degree of simultaneity.

Using $n^{-1}\sum_{i=1}^n E(v_i^2) = \sigma_v^2 n^{-1} \sum_{i=1}^n \omega_i = \sigma_v^2$, the joint strength of the three external instruments is determined by the scaled concentration parameter inspired scalar quantity

$$\mu^2 = \frac{n}{3} \frac{\pi_{33}^2 + \pi_{34}^2 + \pi_{35}^2}{\sigma_v^2}, \tag{4.5}$$

which implies $Var(x_{i3}) = \pi_{32}^2 + \sigma_v^2(3\mu^2/n + \omega_i)$. All the external instruments and exogenous regressor x_{i2} have normalized variance unity. Thus, treating endogenous regressor x_{i3} similarly requires $\pi_{32}^2 + \sigma_v^2(3\mu^2/n + 1) = 1$. Therefore we will choose

$$\sigma_v^2 = \frac{1 - \pi_{32}^2}{3\mu^2/n + 1}. \tag{4.6}$$

Note that $\pi_{32}^2 < 1$, since π_{32} is the correlation coefficient determining the multicollinearity between the structural form regressors x_{i2} and x_{i3}. To have all three external instruments equally weak or strong, we should take

$$\pi_{33} = \pi_{34} = \pi_{35} = |\mu\sigma_v| / \sqrt{n}. \tag{4.7}$$

The heteroskedasticity pattern will follow a so-called multiplicative form determined by the $iidN(0,1)$ series x_{i2} and z_{i3} (hence by one internal and one external instrument), but could be driven by another in practice unobserved independent variable $\eta_i^\circ \sim iidN(0,1)$ as well. A parameter $\phi \geq 0$ determines the seriousness of the heteroskedasticity, where $\phi = 0$ implies homoskedasticity. When $\phi > 0$ a parameter λ, with $0 \leq \lambda \leq 1$, determines the relative importance of the observed variables (x_{i2} and z_{i3}) and the unobserved variable (η_i°) regarding the heteroskedasticity. And, if $\lambda > 0$, a parameter κ, with $0 \leq \kappa \leq 1$, determines the relative importance of x_{i2} and z_{i3} regarding any heteroskedasticity. This is achieved by generating the variable

$$g_i = -\phi^2/2 + \phi\{\lambda^{1/2}[\kappa^{1/2}x_{i2} + (1-\kappa)^{1/2}z_{i3}] + (1-\lambda)^{1/2}\eta_i^\circ\} \sim iidN(-\phi^2/2, \phi^2), \tag{4.8}$$

and taking

$$\omega_i = \exp(g_i), \tag{4.9}$$

which follows a lognormal distribution with $E(\omega_i) = 1$ and $Var(\omega_i) = \exp(\phi^2) - 1$. Since about 99% of the drawings g_i will be in the interval $[-\phi^2/2 - 2.58\phi, -\phi^2/2 + 2.58\phi]$, also 99% of the drawings ω_i will fall in the interval

$$[\exp(-\phi^2/2 - 2.58\phi), \exp(-\phi^2/2 + 2.58\phi)]. \tag{4.10}$$

Table 4.1 presents the bounds of these intervals, both for ω_i and for $\omega_i^{1/2}$, for particular values of ϕ. From these we learn that $\phi \geq 1$ implies pretty serious heteroskedasticity, whereas we may qualify it mild when $\phi < 0.3$, say.

Table 4.1	Heteroskedasticity for different values of ϕ			
ϕ	bounds of 99% intervals			
	$\omega_i^{1/2}$		ω_i	
0.2	0.76	1.28	0.59	1.64
0.4	0.57	1.61	0.33	2.59
0.6	0.42	1.98	0.18	3.93
0.8	0.30	2.39	0.09	5.72
1.0	0.21	2.83	0.05	8.00
1.2	0.15	3.28	0.02	10.76
1.4	0.10	3.73	0.01	13.90
1.6	0.07	4.15	0.00	17.25

The design defined above, with the deliberate choices (4.6) and (4.7), has yet, apart from the sample size n, 11 free parameters, namely: β_1, β_2, β_3 and σ_ε; ρ, π_{31}, π_{32} and μ^2; and ϕ, λ and κ. For all the estimation techniques to be examined their estimation errors are invariant with respect to β. For instance, $\hat{\beta}_{OLS} - \beta = (X'X)^{-1}X'\varepsilon$ and $\hat{\beta}_{IV} - \beta = (X'P_ZX)^{-1}X'P_Z\varepsilon$, etc. Therefore, the bias, variance and mean squared errors of the coefficient estimates will for all techniques be invariant with respect to β too. So, from that point of view we may choose any values for the structural coefficients. However, since the estimation errors do depend on ε and hence on σ_ε, by imposing the habitual normalization $\sigma_\varepsilon = 1$, the magnitude of the estimation errors will be affected and thus the findings for the bias, variance and mean squared errors too. This is of less concern, though, when we focus on the relative magnitude of the estimation errors for the different techniques to be examined. Moreover, we can choose values for β_2 and β_3 such that, in combination with $\sigma_\varepsilon = 1$, values (averaged over all replications) for t and F tests or for R^2 are found which are not uncommon in cross-section regression analysis. A loosely defined population coefficient of determination for simultaneous heteroskedastic model (4.1) is given by

$$R_p^2 = 1 - \frac{\sum_{i=1}^n Var(\varepsilon_i)}{\sum_{i=1}^n [Var(\beta_2 x_{i2} + \beta_3 x_{i3}) + Var(\varepsilon_i)]} = 1 - \frac{1}{\beta_2^2 + 2\beta_2\beta_3\pi_{32} + \beta_3^2 + 1}. \quad (4.11)$$

Taking equal values for the two slope coefficients then yields

$$\beta_2^2 = \beta_3^2 = R_p^2/[2(1 - R_p^2)(1 + \pi_{32})].$$

For $0.1 \leq R_p^2 \leq 0.2$ and $0 \leq \pi_{32} \leq 0.8$ this implies positive solutions in the range $0.18 \leq \beta_2 = \beta_3 \leq 0.35$. What we will do is simply take $\beta_2 = \beta_3 = 0.25$ and calculate over all the replications of the Monte Carlo the average of particular statistics which enable to monitor the practical relevance of the cases examined. In these cases we take particular combinations from the grid given in Table 4.2.

Table 4.2	Grid of design parameter values in the simulation		
$n = \{50, 200\}$			
$\beta_1 = 0$	$\beta_2 = 0.25$	$\beta_3 = 0.25$	$\sigma_\varepsilon = 1$
$\rho = \{0.1, 0.5\}$	$\pi_{31} = 0$	$\pi_{32} = \{0, 0.8\}$	$\mu^2 = \{2, 10, 50\}$
$\phi = \{0.5, 1\}$	$\lambda = \{0, 0.2, 0.5, 1\}$	$\kappa = \{0.2, 0.5, 0.8\}$	

Choosing $\beta_1 = 0$ and $\pi_{31} = 0$, but incorporating the intercept always in the model and in the matrix of instruments, will not affect the findings regarding inference on β_2 and β_3. We limit the analysis to just 12 different cases as defined in Table 4.3.

Table 4.3 Examined cases of design parameter value combinations, and results on σ_v, R^2 and $F_{3,n-L}$ ($\beta_1 = 0$, $\beta_2 = \beta_3 = 0.25$, $\sigma_\varepsilon = 1$, $\pi_{31} = 0$)

Case	n	ρ	π_{32}	μ^2	ϕ	λ	κ	σ_v	R^2	$F_{3,n-L}$
A	200	0.5	0	50	1	1	0.5	0.76	0.35 (0.07)	52.56 (13.36)
B	200	0.5	0	10	1	1	0.5	0.93	0.43 (0.08)	11.47 (4.61)
C	200	0.5	0	2	1	1	0.5	0.99	0.45 (0.08)	3.24 (2.10)
D	200	0.5	0	10	0.5	1	0.5	0.93	0.43 (0.06)	11.16 (4.21)
E	200	0.1	0	10	1	1	0.5	0.93	0.17 (0.07)	11.47 (4.61)
F	200	0.5	0.8	10	1	1	0.5	0.56	0.43 (0.07)	11.47 (4.61)
G	200	0.5	0	10	1	1	0.2	0.93	0.43 (0.08)	11.58 (4.77)
H	200	0.5	0	10	1	1	0.8	0.93	0.43 (0.08)	11.35 (4.44)
I	50	0.5	0	10	1	1	0.5	0.79	0.39 (0.14)	11.92 (6.01)
J	200	0.5	0	10	1	0.5	0.5	0.93	0.43 (0.07)	11.27 (4.35)
K	200	0.5	0	10	1	0.2	0.5	0.93	0.43 (0.06)	11.16 (4.21)
L	200	0.5	0	10	1	0	0.5	0.93	0.43 (0.05)	11.07 (4.11)

Table 4.3 also presents the value of σ_v that results according to (4.6). The final two columns contain the average (with standard deviation between parentheses) over all replications of the standard OLS R^2 statistic (thus neglecting the simultaneity and any heteroskedasticity) in the structural model (4.1) and of the $F_{3,n-L}$ test statistic on the joint significance of the external instruments in the reduced form equation (4.2) when estimated by OLS (thus again neglecting any heteroskedasticity). Both measures have their drawbacks, but they are only used here to give a rough impression of major characteristics of the DGPs, namely their fit and the strength of the external instruments. Note that there is a reasonable correspondence between the values of μ^2 and the average $F_{3,n-L}$ statistic. Due to the inconsistency of OLS results in the structural model, only in case E (where the simultaneity is very mild) the average R^2 statistic is in the range of the aimed at R_p^2 value. In the next section further evidence will be discussed regarding the empirical relevance of the chosen designs. The simulation estimates have been obtained

from 10,000 Monte Carlo replications. In each replication new independent realizations have been drawn for ε_i°, x_{i2}, z_{i3}, z_{i4}, z_{i5}, v_i° and η_i°, so the Monte Carlo averages estimate unconditional moments.[6]

5. Simulation results

In Table 5.1 we collect results for the cases A, B and C. They all concern the larger sample size, have serious heteroskedasticity, substantial simultaneity, and no multicollinearity between x_{i2} and x_{i3}. These three cases just differ in the strength of the three external instruments, as can be seen from Table 4.3.[7]

Table 5.1 Simulation results on coefficient estimates for cases A, B and C

		β_2				β_3			
Case		bias	st.dv	rmse	rrmse	bias	st.dv	rmse	rrmse
A:	GMM	0.002	0.081	0.081	1.000	0.005	0.115	0.115	1.000
	FnpGMM	0.002	0.081	0.081	0.998	0.005	0.114	0.114	0.999
	MGMM	0.001	0.045	0.045	0.551	0.002	0.068	0.068	0.590
	FpMGMM	0.001	0.047	0.047	0.572	0.002	0.070	0.070	0.610
	IV	0.001	0.086	0.086	1.051	0.005	0.118	0.118	1.027
	OLS	0.001	0.078	0.078	0.958	0.375	0.091	0.386	3.366
	WLS	0.003	0.046	0.046	0.565	0.225	0.058	0.232	2.028
B:	GMM	0.002	0.082	0.082	1.000	0.020	0.211	0.212	1.000
	FnpGMM	0.002	0.082	0.082	0.999	0.020	0.211	0.212	1.001
	MGMM	0.001	0.045	0.045	0.547	0.007	0.123	0.123	0.582
	FpMGMM	0.001	0.047	0.047	0.568	0.008	0.127	0.127	0.602
	IV	-0.000	0.086	0.086	1.048	0.022	0.215	0.216	1.022
	OLS	0.001	0.075	0.075	0.912	0.465	0.095	0.474	2.243
	WLS	0.003	0.044	0.044	0.532	0.386	0.062	0.390	1.846
C:	GMM	0.005	0.089	0.089	1.000	0.117	0.528	0.541	1.000
	FnpGMM	0.004	0.091	0.091	1.018	0.118	0.553	0.566	1.046
	MGMM	0.001	0.046	0.046	0.512	0.037	0.268	0.271	0.500
	FpMGMM	0.001	0.047	0.047	0.529	0.040	0.276	0.279	0.516
	IV	0.001	0.091	0.091	1.028	0.124	0.528	0.542	1.002
	OLS	0.001	0.074	0.074	0.829	0.492	0.097	0.502	0.928
	WLS	0.002	0.040	0.040	0.454	0.471	0.062	0.475	0.878

[6] Advantages and disadvantages of conditioning or not on exogenous variables in simulation experiments are discussed in Kiviet (2012).

[7] Regarding the empirical relevance of the models examined in cases A, B and C it is useful to examine the quantities $\beta_2/sd(\hat{\beta}_2)$ and $\beta_3/sd(\hat{\beta}_3)$ for the different techniques. This ratio has correspondences with the inverse of the coefficient of variation and with a stylized t-ratio. For case A it is found to be in the range (2.12, 5.43), for case B in the range (1.19, 5.68) and for C in (0.45, 6.25). Note that in case C the estimators based on external weak instruments have for the estimate of β_3 such a large standard deviation in comparison to the true coefficient value of 0.25 that they do not enable to produce very useful inference.

Because the regressors x_2 and x_3 are uncorrelated it can easily be derived that even the inconsistent coefficient vector estimators (both the least-squares variants) yield a consistent estimator for element β_2, the coefficient of the exogenous regressor x_2. So, it should not surprise that all estimators produce almost unbiased results for β_2, also when instruments are weak. Regarding β_3 it are only the consistent estimators that have moderate bias, provided the instruments are not weak. For the weak instrument case C also the consistent estimators of β_3 show substantial bias, in particular IV and those based on standard GMM. The most remarkable result, however, is that the MGMM and FpMGMM estimators are substantially more efficient than standard GMM. In these three cases they reduce the RMSE by about 40 or 50%, irrespective of the strength of the instruments, as can be seen easily from the rrmse (relative RMSE) columns which presents the RMSE divided by that of unfeasible standard GMM.

Further note that the nonparametric feasible implementation of standard GMM is very close to the unfeasible estimator. Also the parametric feasible version of MGMM works well. It is also noteworthy that we find that standard GMM is in fact not all that much better than IV, illustrating that the efficiency gain due to taking Ω into account is largely offset by the fact that the instruments used by the standard implementation of GMM are weaker than those used by IV. This weakening is prevented by MGMM through weighing the instruments by the same weights as used for the variables in the transformed model, inducing remarkable reductions both in bias and standard deviation. Another interesting finding from Table 5.1 is that on the basis of their RMSE the inconsistent estimators often outperform consistent GMM and occasionally even MGMM (just for β_2). This is in line with results reported on OLS and IV in Kiviet (2013): weakness of instruments is often more detrimental to estimator accuracy than invalidity of instruments.

In all the further cases to be examined, we shall keep $\mu^2 = 10$, so the instruments are not very weak, but certainly not strong. Thus, from now on, case B should be considered the reference case. Case D in Table 5.2 is similar to case B, apart from the seriousness of the heteroskedasticity. We see that $\phi = 0.5$ leads to similar though more moderate relative differences, with efficiency gains by MGMM still around 15%. All remaining cases have $\phi = 1$ again. Table 5.2 also contains cases E and F. In case E the simultaneity is mild. This is seen to have no effects on the relative performance of MGMM, but now the inconsistent estimators have minor bias for both coefficients and therefore they have better RMSE. Case F differs from B just regarding the occurrence of substantial multicollinearity between x_{i2} and x_{i3}, which reduces the value of σ_v. Note that the rrmse results on β_3 are invariant regarding π_{23}, but not those for β_2. The least-squares based estimators for β_2 are now inconsistent too. All standard deviations are much higher and the bias in estimates for β_3 infects the estimates of β_2 with bias. As before, MGMM and FpMGMM perform best.

Table 5.2 Simulation results on coefficient estimates for cases D, E and F

		β_2				β_3			
Case		bias	st.dv	rmse	rrmse	bias	st.dv	rmse	rrmse
D:	GMM	0.002	0.075	0.075	1.000	0.019	0.204	0.205	1.000
	FnpGMM	0.001	0.075	0.075	1.007	0.019	0.206	0.207	1.009
	MGMM	0.001	0.063	0.063	0.844	0.014	0.176	0.177	0.863
	FpMGMM	0.001	0.065	0.065	0.872	0.015	0.182	0.182	0.888
	IV	0.001	0.075	0.075	1.006	0.019	0.205	0.206	1.003
	OLS	0.001	0.066	0.066	0.883	0.466	0.070	0.471	2.297
	WLS	0.001	0.057	0.057	0.757	0.450	0.063	0.455	2.217
E:	GMM	0.001	0.083	0.083	1.000	0.003	0.211	0.211	1.000
	FnpGMM	0.001	0.083	0.083	1.000	0.002	0.211	0.211	1.000
	MGMM	0.001	0.045	0.045	0.547	0.002	0.124	0.124	0.589
	FpMGMM	0.001	0.047	0.047	0.568	0.002	0.128	0.128	0.607
	IV	0.000	0.087	0.087	1.051	0.003	0.216	0.216	1.023
	OLS	0.001	0.085	0.085	1.021	0.093	0.108	0.143	0.677
	WLS	0.001	0.045	0.045	0.541	0.078	0.065	0.101	0.479
F:	GMM	-0.024	0.280	0.281	1.000	0.034	0.351	0.353	1.000
	FnpGMM	-0.024	0.280	0.281	0.999	0.033	0.351	0.353	1.001
	MGMM	-0.009	0.169	0.170	0.603	0.012	0.205	0.205	0.582
	FpMGMM	0.001	0.047	0.047	0.568	0.002	0.128	0.128	0.607
	IV	-0.029	0.284	0.285	1.014	0.036	0.359	0.360	1.022
	OLS	-0.619	0.145	0.636	2.259	0.775	0.159	0.791	2.243
	WLS	-0.511	0.092	0.520	1.847	0.643	0.103	0.651	1.846

Cases G and H in Table 5.3 differ from B only in κ, so in whether either x_{i2} or z_{i3} is the major source of the heteroskedasticity. The effects of κ are found to be moderate and both cases again show spectacular efficiency gains by (feasible parametric) MGMM. The only difference between cases I and B is the smaller sample size. This clearly mitigates the gains by MGMM over standard GMM, but they are still around 30%.

Table 5.3		Simulation results on coefficient estimates for cases G, H and I							
		β_2				β_3			
Case		bias	st.dv	rmse	rrmse	bias	st.dv	rmse	rrmse
G:	GMM	0.002	0.075	0.075	1.000	0.021	0.215	0.216	1.000
	FnpGMM	0.002	0.075	0.075	0.998	0.021	0.215	0.216	1.001
	MGMM	0.001	0.044	0.044	0.594	0.007	0.123	0.123	0.571
	FpMGMM	0.001	0.046	0.046	0.613	0.008	0.127	0.127	0.588
	IV	0.000	0.077	0.077	1.035	0.024	0.223	0.224	1.039
	OLS	0.001	0.068	0.068	0.904	0.465	0.096	0.475	2.199
	WLS	0.002	0.043	0.043	0.572	0.386	0.063	0.391	1.810
H:	GMM	0.002	0.091	0.091	1.000	0.019	0.204	0.205	1.000
	FnpGMM	0.002	0.091	0.091	0.998	0.019	0.204	0.205	1.001
	MGMM	0.001	0.045	0.045	0.500	0.007	0.123	0.123	0.601
	FpMGMM	0.001	0.047	0.047	0.522	0.008	0.127	0.127	0.621
	IV	0.000	0.094	0.094	1.036	0.020	0.206	0.207	1.010
	OLS	0.001	0.081	0.081	0.897	0.465	0.095	0.474	2.315
	WLS	0.002	0.044	0.044	0.489	0.385	0.061	0.390	1.903
I:	GMM	-0.002	0.165	0.165	1.000	0.027	0.247	0.248	1.000
	FnpGMM	-0.003	0.168	0.168	1.014	0.027	0.250	0.251	1.010
	MGMM	-0.002	0.101	0.101	0.610	0.010	0.162	0.162	0.652
	FpMGMM	-0.003	0.116	0.116	0.701	0.013	0.180	0.180	0.725
	IV	-0.008	0.177	0.177	1.072	0.029	0.258	0.260	1.044
	OLS	-0.004	0.157	0.158	0.953	0.390	0.176	0.428	1.721
	WLS	0.004	0.099	0.099	0.597	0.262	0.122	0.289	1.163

In Table 5.4 we examine cases that differ from B only in that $0 \leq \lambda < 1$, hence the heteroskedasticity does also depend now on a factor that one cannot capture in a feasible parametric technique. In case J $\lambda = 0.5$, in K it is 0.2 and $\lambda = 0$ in case L. In the latter case, where the heteroskedasticity is not related to any of the instruments, we note that the bias and variance results for GMM and IV are almost similar. This is due to

$$\begin{aligned} \operatorname{plim} n^{-1} Z'\Omega Z &= \operatorname{plim} n^{-1} \sum\nolimits_{i=1}^{n} \omega_i z_i z_i' \qquad (5.1) \\ &= \operatorname{plim} n^{-1} \sum\nolimits_{i=1}^{n} z_i z_i' + \operatorname{plim} n^{-1} \sum\nolimits_{i=1}^{n} (\omega_i - 1) z_i z_i' = \operatorname{plim} n^{-1} Z'Z, \end{aligned}$$

because here $\omega_i = \exp(-\phi^2/2 + \phi\eta_i^\circ)$ with $E(\omega_i \mid z_i) = 1$, thus the law of large numbers implies

$$\operatorname{plim} n^{-1} \sum\nolimits_{i=1}^{n} (\omega_i - 1) z_i z_i' = \lim n^{-1} \sum\nolimits_{i=1}^{n} E[E(\omega_i - 1 \mid z_i) z_i z_i'] = O.$$

A similar result yields the asymptotic equivalence of the standard and the heteroskedasticity consistent OLS variance estimators in models with just exogenous regressors which are unrelated with the disturbance variance. However, even for $\lambda = 0$, MGMM beats GMM and IV. That MGMM does not converge to IV in this case follows from

$$\begin{aligned}
\operatorname{plim} n^{-1}X'\Omega^{-1}Z &= \operatorname{plim} n^{-1}\sum\nolimits_{i=1}^{n} \omega_i^{-1}x_i z_i' \\
&= \operatorname{plim} n^{-1}\sum\nolimits_{i=1}^{n} x_i z_i' + \operatorname{plim} n^{-1}\sum\nolimits_{i=1}^{n}(\omega_i^{-1}-1)x_i z_i' \\
&= \operatorname{plim} n^{-1}X'Z + \operatorname{plim} n^{-1}\sum\nolimits_{i=1}^{n} E[E(\omega_i^{-1}-1 \mid z_i)x_i z_i'] \\
&\neq \operatorname{plim} n^{-1}X'Z, \qquad (5.2)
\end{aligned}$$

because $E(\omega_i^{-1}) \neq 1/E(\omega_i) = 1$. The results on the feasible MGMM estimators does reveal, however, that we have not managed yet to materialize this remarkable theoretical and experimental superiority of unfeasible MGMM for case L in a feasible implementation.

Table 5.4 Simulation results on coefficient estimates for cases J, K and L

Case		β_2 bias	st.dv	rmse	rrmse	β_3 bias	st.dv	rmse	rrmse
J:	GMM	0.002	0.078	0.078	1.000	0.020	0.207	0.207	1.000
	FnpGMM	0.002	0.078	0.078	1.005	0.019	0.208	0.209	1.007
	MGMM	0.001	0.055	0.055	0.708	0.011	0.153	0.153	0.740
	FpMGMM	0.001	0.058	0.058	0.750	0.012	0.160	0.161	0.776
	IV	0.001	0.079	0.079	1.017	0.020	0.208	0.209	1.008
	OLS	0.001	0.069	0.069	0.889	0.465	0.079	0.472	2.276
	WLS	0.002	0.051	0.051	0.652	0.428	0.062	0.433	2.085
K:	GMM	0.002	0.074	0.074	1.000	0.019	0.203	0.204	1.000
	FnpGMM	0.001	0.075	0.075	1.007	0.019	0.205	0.206	1.010
	MGMM	0.001	0.063	0.063	0.843	0.014	0.176	0.176	0.863
	FpMGMM	0.001	0.067	0.067	0.902	0.016	0.186	0.187	0.914
	IV	0.001	0.074	0.074	1.004	0.019	0.204	0.205	1.003
	OLS	0.001	0.065	0.065	0.880	0.466	0.071	0.471	2.309
	WLS	0.001	0.056	0.056	0.757	0.449	0.063	0.454	2.223
L:	GMM	0.001	0.071	0.071	1.000	0.018	0.200	0.201	1.000
	FnpGMM	0.001	0.072	0.072	1.007	0.018	0.203	0.204	1.012
	MGMM	0.001	0.068	0.068	0.957	0.017	0.193	0.194	0.962
	FpMGMM	0.001	0.074	0.074	1.035	0.018	0.207	0.207	1.030
	IV	0.001	0.071	0.071	1.000	0.018	0.201	0.201	1.001
	OLS	0.001	0.063	0.063	0.879	0.466	0.065	0.471	2.339
	WLS	0.001	0.060	0.060	0.847	0.461	0.063	0.466	2.314

Next we examine the estimated standard errors. In Tables 5.5 through 5.7 for both β_2 and β_3 first the Monte Carlo estimate of the true standard deviation is repeated and next the average over the replications of the square root of the variance estimators given at the end of section 3 (indicated as st.er) are presented, followed by their ratio, which directly indicates the degree of over or under assessment of the true standard deviation.

Table 5.5	Simulation results on standard errors, cases A, B, C, D						
		β_2			β_3		
Case		st.dv	st.er	ratio	st.dv	st.er	ratio
A:	GMM	0.081	0.084	1.038	0.115	0.117	1.019
	FnpGMM	0.081	0.078	0.959	0.114	0.109	0.951
	MGMM	0.045	0.044	0.993	0.068	0.067	0.996
	FpMGMM	0.047	0.024	0.522	0.070	0.037	0.526
	IV	0.086	0.071	0.828	0.118	0.108	0.915
	VnpIV	0.086	0.084	0.976	0.118	0.114	0.966
B:	GMM	0.082	0.100	1.218	0.211	0.246	1.170
	FnpGMM	0.082	0.079	0.969	0.211	0.199	0.945
	MGMM	0.045	0.045	0.997	0.123	0.123	0.998
	FpMGMM	0.047	0.025	0.526	0.127	0.067	0.529
	IV	0.086	0.072	0.843	0.215	0.197	0.917
	VnpIV	0.086	0.085	0.988	0.215	0.208	0.966
C:	GMM	0.089	0.131	1.475	0.528	0.628	1.189
	FnpGMM	0.091	0.089	0.986	0.553	0.468	0.846
	MGMM	0.046	0.046	1.017	0.268	0.266	0.993
	FpMGMM	0.047	0.026	0.546	0.276	0.147	0.534
	IV	0.091	0.083	0.912	0.528	0.464	0.879
	VnpIV	0.091	0.095	1.037	0.528	0.488	0.925
D:	GMM	0.075	0.092	1.224	0.204	0.240	1.176
	FnpGMM	0.075	0.074	0.983	0.206	0.197	0.956
	MGMM	0.063	0.064	1.011	0.176	0.175	0.994
	FpMGMM	0.065	0.034	0.519	0.182	0.093	0.512
	IV	0.075	0.072	0.961	0.205	0.198	0.967
	VnpIV	0.075	0.076	1.003	0.205	0.200	0.978

In Table 5.5 we find for cases A through C that standard unfeasible GMM (substituting Ω, but estimating σ_ε^2) is reasonable for strong instruments, but is more and more too pessimistic when instruments get weaker. This problem is primarily due to the estimation of σ_ε^2.[8] The nonparametric implementation of standard GMM seemingly works well, irrespective of the strength of the instruments. Unfeasible MGMM is fine,

[8] In additional simulations not presented here we found similar results when using estimator $\hat{\sigma}_\varepsilon^{*2} = (y^* - X^*\hat{\beta}_{GMM})'(y^* - X^*\hat{\beta}_{GMM})/n$, whereas using the true value gave reasonable results for $\mu^2 \geq 10$.

but its feasible parametric version is much too optimistic. The standard IV implementation, which neglects the heteroskedasticity completely, is found to be optimistic. The variant which aims to repair this by employing a nonparametrically robustified variance (VnpIV) works remarkably well.

Table 5.5 also contains case D (mild heteroskedasticity). Also when heteroskedasticity is mild there is yet no feasible variant for the superior estimator MGMM for which its standard errors are on average reasonably accurate for its true standard deviation.

Table 5.6 Simulation results on standard errors, cases E, F, G, H

		β_2			β_3		
Case		st.dv	st.er	ratio	st.dv	st.er	ratio
E:	GMM	0.083	0.083	1.004	0.211	0.205	0.974
	FnpGMM	0.083	0.080	0.971	0.211	0.200	0.950
	MGMM	0.045	0.045	0.993	0.124	0.123	0.990
	FpMGMM	0.047	0.025	0.528	0.128	0.068	0.531
	IV	0.087	0.074	0.844	0.216	0.198	0.920
	VnpIV	0.086	0.086	0.988	0.216	0.209	0.970
F:	GMM	0.280	0.328	1.169	0.351	0.410	1.170
	FnpGMM	0.280	0.266	0.949	0.351	0.332	0.945
	MGMM	0.169	0.169	0.998	0.205	0.204	0.998
	FpMGMM	0.175	0.093	0.528	0.212	0.112	0.529
	IV	0.284	0.272	0.960	0.359	0.329	0.917
	VnpIV	0.284	0.275	0.969	0.359	0.346	0.966
G:	GMM	0.075	0.098	1.309	0.215	0.270	1.255
	FnpGMM	0.075	0.073	0.978	0.215	0.203	0.944
	MGMM	0.044	0.044	0.999	0.123	0.123	0.999
	FpMGMM	0.046	0.024	0.530	0.127	0.067	0.530
	IV	0.077	0.072	0.932	0.223	0.197	0.885
	VnpIV	0.077	0.077	0.996	0.223	0.215	0.964
H:	GMM	0.091	0.103	1.136	0.204	0.224	1.096
	FnpGMM	0.091	0.087	0.959	0.204	0.194	0.951
	MGMM	0.045	0.045	0.994	0.123	0.122	0.996
	FpMGMM	0.047	0.025	0.523	0.127	0.067	0.529
	IV	0.094	0.073	0.774	0.206	0.197	0.957
	VnpIV	0.094	0.092	0.981	0.206	0.200	0.972

In Table 5.6 case E learns that also under mild simultaneity the parametric feasible FpMGMM estimates its variance poorly. Cases F, G and H support the wider validity of the earlier conclusions.

In Table 5.7 case I illustrates that all techniques tend more towards too optimistic standard error estimates when the sample size is smaller. From cases J through L we learn that the qualities of standard error estimates depend very little on the value of λ, except for IV. Note that for $\lambda = 0$ (instruments and heteroskedasticity unrelated) the asymptotic equivalence of GMM and IV already shows up at $n = 200$ for the coefficient estimates, but less so for their variance estimators.

Table 5.7 Simulation results on standard errors, cases I, J, K, L

		β_2			β_3		
Case		st.dv	st.er	ratio	st.dv	st.er	ratio
I:	GMM	0.165	0.157	0.953	0.247	0.231	0.935
	FnpGMM	0.168	0.142	0.846	0.250	0.214	0.858
	MGMM	0.101	0.096	0.950	0.162	0.152	0.940
	FpMGMM	0.116	0.055	0.475	0.180	0.087	0.487
	IV	0.177	0.143	0.810	0.258	0.230	0.893
	VnpIV	0.177	0.157	0.886	0.258	0.232	0.901
J:	GMM	0.078	0.096	1.238	0.207	0.246	1.190
	FnpGMM	0.078	0.076	0.976	0.208	0.198	0.952
	MGMM	0.055	0.055	1.007	0.153	0.152	0.995
	FpMGMM	0.058	0.030	0.515	0.160	0.082	0.513
	IV	0.079	0.072	0.915	0.208	0.198	0.950
	VnpIV	0.079	0.079	0.998	0.208	0.203	0.974
K:	GMM	0.074	0.092	1.242	0.203	0.243	1.196
	FnpGMM	0.075	0.073	0.983	0.205	0.196	0.957
	MGMM	0.063	0.063	1.013	0.176	0.174	0.991
	FpMGMM	0.067	0.034	0.510	0.186	0.094	0.504
	IV	0.074	0.072	0.971	0.204	0.198	0.972
	VnpIV	0.074	0.075	1.004	0.204	0.199	0.979
L:	GMM	0.071	0.085	1.193	0.200	0.231	1.152
	FnpGMM	0.072	0.071	0.989	0.203	0.195	0.960
	MGMM	0.068	0.069	1.016	0.193	0.190	0.987
	FpMGMM	0.074	0.037	0.506	0.207	0.102	0.496
	IV	0.071	0.072	1.014	0.201	0.198	0.988
	VnpIV	0.071	0.072	1.007	0.201	0.197	0.981

The major findings from these simulations are that if Ω were known MGMM would be much more attractive than GMM, uniformly over all designs examined, because it has smaller bias, much smaller true standard deviation and also its standard errors establish much more accurate estimates of its actual standard deviation. Moreover, it is found to be less vulnerable to weakness of the instruments chosen for the original model specification. However, although a feasible parametric implementation of MGMM is often almost as efficient, its standard asymptotic variance estimate is very seriously biased and underestimates its actual dispersion. The robust GMM implementation provides reasonably accurate standard errors, provided the sample is not too small and the instruments not too weak. However, simply sticking to IV estimation and using nonparametric heteroskedasticity robust standard errors (VnpIV) is almost equally effective as employing robust GMM, because the latter suffers from weakened instruments due to weighing the observations. In order to fully exploit in practice the impressive efficiency gains achieved by feasible parametric MGMM it is yet required to develop a more accurate assessment of its actual efficiency, possibly by bootstrapping. We leave this topic for future research.

6. Empirical illustrations

To illustrate our theoretical findings in practice, we first set out to extend the 2SLS and GMM comparison for an actual cross-section data set as presented in Wooldridge (2001) with operational MGMM findings. However, the wage equation analyzed in that study does not seem to be inflicted with much heteroskedasticity, so in such a situation IV, possibly using heteroskedasticity consistent variance estimates, is self-evidently the preferable technique. Heteroskedasticity seems evident in a data set stemming from Sander (1992) on the effect of women's schooling on fertility, also addressed in Wooldridge (2010, Problem 6.8, dataset fertil1).

The sample comprises 1129 US women. We regressed number of kids on education, age, age-squared a racial dummy and dummies for regions (east, west, north-central) and types of agglomeration (town, small-city, farm, other rural) and an intercept and year dummies. Education could be endogenous and is instrumented by years of education of the father and of the mother. Hence, the degree of overidentification is just one.[9] Some mothers may have yet relatively few years of education because they already care for children. This would explain the positive difference between the OLS and IV coefficient estimates. That positive difference may also be the result of omitted control variables which have a positive effect on fertility and are negatively correlated with years

[9] The F-test on the exclusion of the two external instruments in the reduced form equation for education is 155.8 so they are certainly not weak. However, the endogeneity problem does not seem severe. The DWH statistic has p-value 0.48, so many researchers would happily accept exogeneity of all regressors. However, the power of such a test is not always impressive, see Kiviet and Pleus (2015), so imposing exogeneity could be rash, especially because there are good theoretical reasons to assume endogeneity of years of schooling.

of education (or vice versa, such as, for instance, being aware of birth control methods).

Classic tests for heteroskedasticity after OLS estimation are all highly significant. Employing the instruments in the various GMM techniques examined here does not provide any evidence of instrument invalidity by the Sargan-Hansen J test, as can be seen from Table 6.1, which does not mention the results for the included demographic and year-dummy controls.

In Table 6.1 we ran a regression of the log of the squared IV residuals[10] on all the instruments, their squares and their cross-products, leaving out redundant contributions such as the squares of dummies and the cross-products of two annual dummies, of two regional dummies and of two agglomeration dummies (giving 149 regressors in addition to the constant).[11] We obtained a series of positive $\hat{\sigma}_i$ values by taking the exponential function of the fitted values of this (unrestricted) auxiliary regression. Next, by dividing these by their sample average we obtained a series $\hat{\omega}_i$ with sample average unity. This series has empirical quantiles $q^{0.005}(\hat{\omega}_i) = 0.094$ and $q^{0.995}(\hat{\omega}_i) = 7.970$, which matches with a ϕ in the range 0.8-1.0 according to Table 4.1.

Table 6.1	Empirical findings on fertility				
	OLS	IV	IV(Vnp)	FnpGMM	FpMGMM
educ	-0.128	-0.153	-0.153	-0.153	-0.168
	(0.018)	(0.039)	(0.041)	(0.041)	(0.033)
age	0.532	0.524	0.524	0.523	0.421
	(0.138)	(0.139)	(0.141)	(0.141)	(0.127)
agesq	-0.006	-0.006	-0.006	-0.006	-0.005
	(.002)	(0.002)	(0.002)	(0.002)	(0.001)
black	1.076	1.073	1.073	1.072	0.881
	(0.174)	(0.174)	(0.201)	(0.201)	(0.177)
J (p-val.)	-	0.88	0.88	0.88	0.90
(standard errors between parentheses)					

[10]The log of the squared IV residuals have skewness 0.26 and kurtosis 3.98. Hence, although they are significantly nonnormal, their distribution is not completely out of line with those of g_i in (4.8). Their sample mean is -0.47 and sample variance 0.66. Interpreting these as reflecting the expectation and variance of g_i they would imply ϕ to be 0.69 or 0.81 respectively.

[11]This regression yields an R^2 of only 0.128; this we associate with a value of λ in (4.8) as low as about 0.13. Although some individual coefficients in this auxiliary regression have substantial t-ratio's the overall F-test has p-value 0.61. So, one could conclude that heteroskedasticity determined by the instrumental variables is insignificant (although in a more parsimonious specification significant heteroskedasticity would emerge).

Although we should always realize that a difference in empirical standard errors does not necessarily represent a similar difference in true standard deviations, the most remarkable finding from Table 6.1 is undoubtedly that the standard errors of MGMM are in agreement with the in the simulations established superiority of FpMGMM over FnpGMM (when $\lambda > 0$).[12]

We also employed the various techniques to data analyzed in Wu et al. (2015). We re-analyzed one of its two structural equations for land and house prices using data for 2011 on the 35 major Chinese cities. Table 6.2 presents the effect of house price (hp) on land price, but does not mention results on further control variables, such as lagged budget deficit, construction costs, agricultural GDP and available land. External instruments used are disposable income, total population, sex ratio and expenditure on education, giving a degree of overidentification of 4. The DWH statistic has p-value 0.043, the 2SLS results yield a J-statistic with a p-value of 0.40, but in the reduced form equation the external instruments produce an F-value of only 6.87. Hence, it seems that house prices are endogenous and the employed instruments are valid though weak, although such a small sample of course hardly allows firm inferences of this nature. The auxiliary regression of the log of the squared 2SLS residuals on all 10 instruments yields an R^2 of 0.21 (we left out squares and cross-products because that would slurp all remaining degrees of freedom). From its fitted values we obtained a series for $\hat{\omega}_i$ as before. Again we note the attractive standard errors of FpMGMM.

Table 6.2 Empirical findings on land prices

	OLS	IV	IV(Vnp)	FnpGMM	FpMGMM
hp	1.179	1.706	1.706	1.321	1.638
	(0.309)	(0.426)	(0.461)	(0.476)	(0.317)
J (p-val.)	-	0.40	0.40	0.31	0.72

(standard errors between parentheses)

7. Conclusions

We reveal an inherent unfavorable and yet generally unperceived feature of GMM as it is currently usually implemented. Extracting from the assumed orthogonality conditions

[12] Given our findings regarding ϕ, λ and the strength of the instruments these empirical results can probably best be interpreted against the background of our simulation results for cases A and K, although the sample size in this application is much larger. In the simulation the cases A and K suggest similar performance regarding rmse by IV and the robust GMM technique. The similarity in the standard error results for FnpGMM and the nonparametrically robustified standard errors of IV, which are slightly larger than the (incorrect) standard IV standard errors, is in agreement with our simulation findings.

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 relatively simple precautions enable to neutralize this weakening process of the instruments. This also allows to improve on IV and equivalent GMM in just identified models with heteroskedasticity.

By simulation it is shown that empirically relevant forms of heteroskedasticity undermine the quality of standard GMM estimates and a modified implementation of GMM yields estimates that show both less bias and smaller standard deviations. 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 shown to be not exceptional when heteroskedasticity is prominent. We also examine the accuracy of empirical standard errors for the underlying true standard deviations.

In this paper we only examined GMM estimators for cross-sectional models that are linear in the regressor coefficients. However, we conjecture that these results have implications for general nonlinear models too[13], and also for the analysis of time-series data and especially for dynamic panel data models where GMM is used frequently. Next to our simulation findings, we also examined for empirical data sets what the practical consequences are. Interpreting these, one should keep in mind, that the synthetic simulation experiments produce accurate assessments of true bias and true standard deviations, whereas for the empirical findings any bias cannot be assessed, because the true parameter values are unknown, and the obtained estimated standard errors may be misleading for the underlying unknown true standard deviations. Nevertheless, we do find substantially smaller estimated standard errors and therefore we recommend the use of parametric MGMM, the modified form of feasible GMM as developed here.

References

Arellano, M., 2002. Sargan's instrumental variables estimation and the generalized

[13] To examine this, the findings in Lin and Chou (2015), who focus on nonlinear GMM based on instruments which are orthogonal to heteroskedastic error terms, should be extended to cases where the instruments are first scaled with respect to an assessment of the heteroskedasticity on the basis of 1-step GMM results.

method of moments. Journal of Business and Economic Statistics 20, 450-459.

Arellano, M., 2003a. *Panel Data Econometrics.* Oxford University Press: Advanced Texts in Econometrics, Oxford.

Arellano, M., 2003b. Modelling Optimal Instrumental Variables for Dynamic Panel Data Models". CEMFI Working Paper no. 0310.

Baltagi, B.H., Li, Q., 1992. A note on the estimation of simultaneous equations with error components. *Econometric Theory* 8, 113-119.

Bowden, J.B., Turkington, D.A., 1984. *Instrumental Variables.* Econometric Society Monograph in Quantitative Economics. Cambridge University Press.

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

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions, *Journal of Econometrics* 34, 305-334.

Davidson, R., MacKinnon, J.G., 2004. *Econometric Theory and Methods.* Oxford University Press.

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.

Hausman, J.A., Newey, W.K., Woutersen, T., Chao, J.C., Swanson, N.R., 2012. Instrumental variable estimation with heteroskedasticity and many instruments. *Quantitative Economics* 3, 211-255.

Iglesias, E.M., Phillips, G.D.A., 2012. Improved instrumental variables estimation of simultaneous equations under conditionally heteroskedastic disturbances. *Journal of Applied Econometrics* 27, 474-499.

Kiviet, J.F., 2012. Monte Carlo Simulation for Econometricians. *Foundations and Trends in Econometrics* 5:1-2, 1-181. Now Publishers, Boston - Delft.

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

Kiviet, J.F., Pleus, M., 2015. The performance of tests on endogeneity of subsets of explanatory variables scanned by simulation. *Mimeo.*

Lin, E.S., Chou, Ta-Sheng, 2015. Finite-sample refinement of GMM approach to nonlinear models under heteroskedasticity of unknown form. To appear in *Econometric Reviews.*

Newey, K.W., 1993. Efficient estimation of models with conditional moment restrictions. Chapter 16 in: *Handbook of Statistics, Volume 11, Econometrics* (G.S. Maddala, C.R. Rao and H.D. Vinod, eds.), 419-454, Elsevier Science Publishers.

Sander, W., 1992. The effect of women's schooling on fertility. *Economics Letters* 40, 229-233.

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.

Wansbeek, T.W., 2004. Correcting for heteroskedasticity of unspeified form (Exercise 04.1.2). *Econometric Theory* 20, 224.

White, H., 1986. Instrumental variables analogs of generalized least squares estimators. In: *Advances in Statistical Analysis and Statistical Computing Vol.1* (R.S. Mariano, ed.), 173-227. JAI Press.

Wooldridge, J.M., 2001. Applications of Genralized Method of Moments estimation. *Journal of Economic Perspectives* 15, 87-100.

Wooldridge, J.M., 2010. *Econometric Analysis of Cross Section and Panel Data*, second edition. The MIT Press, Cambridge USA.

Wooldridge, J.M., 2013. *Introductory Econometrics, A Modern Approach,* fifth edition. South-Western, Cengage Learning.

Wu, G.L., Feng, Q., Li, P., 2015. Does local governments' budget deficit push up housing prices in China? *China Economic Review* 35, 183-196.