Judging Contending Estimators by Simulation: Tournaments in Dynamic Panel Data Models

Jan F. Kiviet

Faculty of Economics and Econometrics, University of Amsterdam, and Tinbergen Institute.

Judging contending estimators by simulation: tournaments in dynamic panel data models

JAN F. KIVIET[*]

(Universiteit van Amsterdam & Tinbergen Institute)

8 December 2005

An early version of this paper was presented at the
Conference in Memory of Professor Michael Magdalinos (1949-2002)
held in Athens (Greece) on 20-21 November 2003

Abstract

An attempt is made to set rules for a fair and fruitful competition between alternative inference methods based on their performance in simulation experiments. This leads to a list of eight methodologic aspirations. Against their background we criticize aspects of many simulation studies that have been used in the past to compare competing estimators for dynamic panel data models. To illustrate particular pitfalls some further Monte Carlo results are produced, obtained from a simulation design inspired by an analysis of the (non-)invariance properties of estimators and occasionally by available higher-order asymptotic results. We focus on the very specific case of alternative implementations of one and two step generalized method of moments (GMM) estimators in homoskedastic stable zero-mean panel AR(1) models with random individual specific effects. We compare a few implementations, including GMM sytem estimators with alternative weight matrices, and illustrate that an impartial evaluation of the outcome of a Monte Carlo based contest requires evidence – both analytical and empirical – on the completeness, orthogonality and relevance of the simulation design.

1   Introduction

The joint occurrence in dynamic panel data models of individual specific effects and of lagged dependent variables complicates the statistical inference on the model parameters considerably. A great number of alternative techniques for the estimation of dynamic panel data models have been suggested over the last few decades, see inter alia Balestra and Nerlove (1966), Anderson and Hsiao (1981, 1982), Holtz-Eakin et al. (1988), Arellano and Bond (1991), Ahn and Schmidt (1995, 1997), Blundell and Bond (1998) and Hsiao et al. (2002). As a rule these techniques claim particular desirable asymptotic properties under specific circumstances. Various Monte Carlo studies have been undertaken in order to find out how well (variants of) these methods work in finite samples, see inter alia Nerlove (1967, 1971), Bhargava and Sargan (1983), Arellano and Bond (1991), Kiviet (1995), Blundell and Bond (1998), Judson and Owen (1999), Blundell et al. (2000), Harris and Mátyás (2000, 2004), Andrews and Lu (2001), Doornik et al. (2002), Hsiao et al. (2002), Alvarez and Arellano (2003), Bun and Kiviet (2003, 2005), Doran and Schmidt (2005). The main purpose of this study is to clarify that in most of these simulation studies the focus has been too narrow, at least regarding particular aspects, to enable fair and fully informative conclusions on the qualities of the alternative inference procedures examined. Progress in this line of research could have been more efficient, as we shall argue, if the designs of these Monte Carlo studies had been less restrictive and more transparent.

Usually the finite sample distribution of individual coefficient estimators and test statistics does not involve just the parameters affecting their asymptotic distribution. However, exact or almost exact finite sample results are usually hard to obtain by analytic derivation. Therefore, examination of the effects of nuisance parameters and initial conditions on inference in finite samples constitutes the main motivation for performing simulation experiments. As we shall illustrate, the design of such experiments requires a justification built on both analytical and empirical considerations regarding the choices made on the model specifications included in the simulation study, the particular parameter values chosen and any further conditions set and possibly varied. Although any Monte Carlo design will have limitations regarding its size and scope, we shall show that in addition there are particular qualitative aspects – say aspects of proper Monte Carlo methodology – which should always be respected in order to make a simulation exercise really worthwhile. Already for many decades a great number of studies in econometric theory are supplemented by simulation findings, but usually without much explicit reference to simulation methodology. Monte Carlo has become part of the standard toolkit, and seems to be very user-friendly, because researchers use it without reference to any user-manual. In fact, not much has been published in econometrics about Monte Carlo methodology since Hendry (1984). Indeed, this is still the state-of-the-art study. It provides the necessary background to Monte Carlo simulation, but just as far as it concerns the fundamentals of the techniques for assessing in an efficient and effective way by random experimentation on a programmable computer the finite sample characteristics of one particular estimator or test procedure. However, it does not address explicitly the various issues that are relevant when a further purpose of the Monte Carlo study is to make a comparison between different inference techniques. Here we will argue that particular improvements in the practice of designing Monte Carlo contests seems to be called for. We shall illustrate this in the context of the comparison by simulation of different method of moment estimators for dynamic panel data models.

We focus on a very simple example of the dynamic panel data model, viz. the stable first-order autoregressive panel relationship with an unknown intercept, random unobserved

individual effects, and i.i.d. disturbances. Hence, there are no further external regressors. Its full specification is

$$
\left.
\begin{array}{l}
y_{it} = \beta + \gamma y_{i,t-1} + \eta_i + \varepsilon_{it} \\[2mm]
y_{i0} = \alpha_0 + \alpha_1 \eta_i + \alpha_2 \varepsilon_{i0} \\[2mm]
\eta_i \sim \text{i.i.d.}(0, \sigma_\eta^2) \\[2mm]
\varepsilon_i = (\varepsilon_{i0}, \varepsilon_{i1}, ..., \varepsilon_{iT})' \sim \text{i.i.d.}(0, \sigma_\varepsilon^2 I_{T+1})
\end{array}
\right\}
\qquad i = 1, ..., N; \ t = 1, ..., T.
\qquad (1)
$$

The model has two random error components, viz. the individual specific effects η_i and the white-noise innovations ε_{it}. For any $i, j \in \{1, ..., N\}$ the vector ε_i and scalar effect η_j are independent. The start-up values y_{i0} are mutually independent, are determined by the two types of random error components and a non-random component too, and are independent of all ε_{jt} for $t > 0$. Note that we have seven unknown parameters: α_0, α_1, α_2, β, γ, σ_η^2 and σ_ε^2. These are all similar for the N cross-section units. We shall only examine here the stable case $|\gamma| < 1$, which will yield weakly-stationary y_{it} series (time-invariant mean and auto-covariances) under particular initial conditions only. Inference on the parameter of primary interest γ can only be based on the N observed time-series $\{y_{it}; \ t = 0, ..., T\}$, which are identically and independently distributed over the N cross-sections. Various alternative implementations of GMM estimators are available for the estimation of model (1). In illustrations below some of these will be used in our attempt to constitute some explicit methodological standards for designing Monte Carlo experiments when the purpose is to compare and to rate alternative competing inference techniques.

The structure of this paper is the following. In Section 2 we list a number of important general qualitative aspects for a simulation study, when the aim is to draw conclusions on the relative and absolute qualities of alternative inference methods under relevant circumstances. Before we can fully substantiate by illustrations the importance of the rules we set for such studies, we have to discuss some further details on panel AR(1) models and their estimation. Therefore, in Section 3, we first examine some consequences for data transformations occurring in GMM procedures which stem from particular specifications of the start-up values y_{i0} via the coefficients α_0, α_1 and α_2. Next, in Section 4, we introduce the range of estimators to be considered here. Special attention is paid to the weight matrices used in GMM implementations. In Section 5 we discuss aspects of a number of the earlier Monte Carlo studies referred to above and confront them with the rules discussed in Section 2. To illustrate some of their deficiencies we produce a few further Monte Carlo results from an alternative design. Finally, Section 6 concludes.

2 Rules for simulation contests

Inevitably any simulation study is limited in scope and detail. One simply cannot produce results for all parameter values deemed relevant and for all inference techniques that might be of interest for the type of model under study. On the contrary, if one does not want to put off consumers of the results of simulation studies, one has to restrict the number and size of resulting tables severely. Also one should try to condense the information as much as possible to enhance its palatability, for instance by using graphical methods or possibly by producing

so-called response surfaces[1], although these never became widely popular. Hence, it is simply unavoidable to put restrictions on the design of the experiments, such as the chosen density of the grid of discrete numerical values of both the model parameters and further design parameters, such as: the examined actual sample sizes, nominal significance levels of tests, the chosen number of included exogenous variables and the actual parametrizations of their generating schemes, etc. Apart from that, also the actual number of executed Monte Carlo replications for each separate experiment will be limited, implying that Monte Carlo results do not deliver the exact characterizations of estimators and test statistics, such as their moments and quantiles, but only estimates of these which have error margins. These aspects and various methods to economize on computer time, while reducing at the same time as much as possible the specificity and imprecision of the simulation results, are all addressed in Hendry (1984), when the focus is to examine one single specific estimator or test statistic for a particular class of data generating process.

However, when comparisons are to be made between competing inference techniques, then next to the primarily quantitative aspects just indicated there are particular more qualitative facets to the justification of the chosen design, to the range of techniques included and to the final presentation of the results of the executed experiments, which are of paramount importance too. Their neglect may in fact have much more serious consequences than the unavoidable imposed restrictions on the primarily quantitative aspects. Below, we list eight such more qualitative aspects. First we merely give a succinct characterization, and in the remainder of this section we give some further explanation. A more tangible clarification follows later when we discuss and criticize aspects of earlier simulation studies on panel AR(1) models and produce some further illustrative simulation results.

Methodologic aspirations for an adequate and impartial design of simulation studies that aim to rank various alternative inference techniques:

1. *Explicit exploitation of any invariance properties;*
2. *Exploration of the non-invariance properties, both analytically and experimentally;*
3. *No dimensional restrictions on the relevant nuisance parameter space;*
4. *"Orthogonal" reparametrization of the parameter space to enhance interpretability;*
5. *Well-argued choice of the examined design parameter values;*
6. *Any contending techniques should play both at home and away;*
7. *Inclusion of (non-)operational full information techniques;*
8. *Full documentation with respect to accuracy, interpretability and reproducibility.*

Most of the points mentioned above are strongly interrelated. Some of these aspirations do also refer to the simulation efficiency issues discussed in Hendry (1984); some are also mentioned in Davidson and MacKinnon (1993, Chapter 21).

The first six points focus on the construction and delineation of the set of designs to be examined in the simulation experiments. Point 1 is simply a matter of simulation efficiency, and is easily illustrated as follows. For most techniques for analyzing γ in model (1) the actual value of β is inconsequential under particular initial conditions. Of course, it is useless then to perform Monte Carlo experiments for various values of β, as long as we are only interested in the qualities of inference methods for γ. Moreover, the study should mention that the actual value chosen for β, for instance zero, does not limit the scope for conclusions from this design. As a complement to a successful extraction of all – if any – invariance properties,

---

[1]In designing reponse surfaces, see Hendry (1984), one is faced with inevitable specification problems that might corrupt the simulation findings; see Davidson and MacKinnon (1993) for further references.

the nuisance parameters, which inflict non-invariance properties, as mentioned in point 2, remain. Again, simulation efficiency is enhanced when theoretical evidence is exploited on the effects of particular parameters on the statistics of interest, as we shall illustrate later. As a rule, however, such analytical small sample evidence is scarce or incomplete, and the primary goal of the simulation study is exactly to disclose experimentally whether and how the various parameters do have effects in finite samples. This requires a deliberate strategy when choosing the actual parameter values in the experiments, in order to be able to reveal the essentials of the actual dependencies. This is reiterated by point 3. If restrictions are imposed on the dimensionality of the experimental design then it is impossible to reveal the independent effects from all parameters separately, and consequently what is supposed to be an impartial simulation contest between various estimators may actually be a handicap race for particular competitors.

Although the parameters as they appear in the model specification may provide a straight-forward base for the nuisance parameter space, it can be very useful to find a transformation for this base, such that it becomes both easier to select empirically relevant values for the nuisance parameters and to interpret their effects. As we shall illustrate below this base should consist of autonomous parameters and thus be orthogonal in a particular sense, as is stated in point 4. The related point 5 re-emphasizes that the actual choice of and variation in the design parameter values should be thoroughly justified, both in the light of the before mentioned goal to reveal yet unknown (in)dependencies, but also with respect to empirical relevance. Since it is simply practically impossible to examine a grid of parameter values which reasonably well represents the full usually unbounded parameter space, it seems much more important to make sure that the chosen grid covers that part of the parameter space which seems empirically most relevant. Points 4 and 5 can be illustrated as follows for model (1). Below, we shall argue more extensively that it is very beneficial for the simulation study if one does not select particular fixed values for α_1, but for another parameter ϕ, and select for instance $\phi \in \{0, 0.5, 1\}$ and $\gamma \in \{0, 0.4, 0.8\}$ to determine $\alpha_1 = \phi(1 - \gamma)$. Now ϕ is a base parameter "orthogonal" to γ expressing and procuring the degree in which the start-up value y_{i0} has attained the equilibrium level of the individual effect component in y_{it}, irrespective of the value of γ.

Points 6 and 7 have to do with the techniques to be included in a simulation contest in relation to further particulars of the simulation design. Both points are easily clarified by illustration in the context of model (1) too. Comparing the behavior of various 1-step GMM implementations for this model differing in, for instance, the weight matrix used and in the nature and number of moment conditions exploited, it seems worthwhile to generate data under various situations covering all particular situations that in turn renders each of the examined weight matrices optimal or not, and the exploited moment conditions either valid or not. Point 6 highlights that a fair competition requires that the range of simulation designs examined should be such that all contending techniques can demonstrate their qualities under the conditions for which they were specifically designed, but have to expose as well their possible failures under the conditions that may better suit the other techniques examined. Point 7 approaches the same issue from opposite direction. Whereas point 6 says that, given the techniques included, the various designs should be such that all techniques have to perform under each others most favorable circumstances, point 7 states, that given the designs that are to be covered because of their practical relevance, one should also include techniques that exploit a considerable amount of the information incorporated in the simulated data generating processes, even if such techniques appear non-operational from a practical point of view. Their inclusion is useful nevertheless, because it generates information on the costs of

being deprived of particular information. For instance, it seems useful in the context of 1-step GMM estimation to include techniques that exploit the optimal weight matrix, even if this includes parameters which are unknown in practice, simply because that yields a yardstick against which the performance of the operational techniques can be judged, whereas it offers insights also into what at best can be achieved by an operational 2-step GMM estimator.

The final point 8 refers to obvious quality criteria such as reporting all relevant information regarding the reproducibility of the simulation study, and with respect to the accuracy of the resulting Monte Carlo estimates. It also stresses the importance that the presentation of Monte Carlo studies should allow a full and proper interpretation of their findings, mentioning clearly its unavoidable limitations. To further substantiate and illustrate the above, some further analysis of the panel AR(1) model is required first.

3   Initial conditions

According to the scheme $y_{it} = \beta + \gamma y_{i,t-1} + \eta_i + \varepsilon_{it}$ all y_{it} contain, in addition to $\gamma y_{i,t-1}$, three types of components, viz. a deterministic component β and two random error components, the individual specific η_i and the idiosyncratic ε_{it} respectively. Therefore it seems reasonable and fully general that we assumed the y_{i0} to be generated by three such types of components as well. By defining ε_{i0} in (1) all conceivable options for a specification of the start-up values are represented by $y_{i0} = \alpha_0 + \alpha_1 \eta_i + \alpha_2 \varepsilon_{i0}$, where $\alpha = (\alpha_0, \alpha_1, \alpha_2)'$ are three extra unobservable parameters. By repeated substitution we find that model (1) implies for $t > 0$

$$
\begin{aligned}
y_{it} &= \beta \sum_{s=0}^{t-1} \gamma^s + \alpha_0 \gamma^t + \left(\sum_{s=0}^{t-1} \gamma^s + \alpha_1 \gamma^t \right) \eta_i + \sum_{s=0}^{t-1} \gamma^s \varepsilon_{i,t-s} + \gamma^t \alpha_2 \varepsilon_{i0} \\
&= \left[\frac{\beta}{1-\gamma} + \gamma^t \left(\alpha_0 - \frac{\beta}{1-\gamma} \right) \right] + \left[\frac{1}{1-\gamma} + \gamma^t \left(\alpha_1 - \frac{1}{1-\gamma} \right) \right] \eta_i \\
&\quad + \left(\sum_{s=0}^{t-1} \gamma^s \varepsilon_{i,t-s} + \gamma^t \alpha_2 \varepsilon_{i0} \right),
\end{aligned}
\tag{2}
$$

exposing for each y_{it} its three constituent components, viz. a non-random part, an individual effects part, and a component determined by the idiosyncratic disturbances $(\varepsilon_{i0}, ..., \varepsilon_{it})$.

From the first component of the final expression of (2) we find that y_{it} will have constant mean through time only if

$$
\alpha_0 = \frac{\beta}{1-\gamma}.
\tag{3}
$$

The second component shows that the impact of η_i on y_{it} will be constant through time only if

$$
\alpha_1 = \frac{1}{1-\gamma}.
\tag{4}
$$

When (4) is not fulfilled this will have far reaching consequences, because of the following. An important issue in the estimation of dynamic panel data models with individual effects is whether it is possible to remove these individual effects from the model, or from the regressors, or from any variable that may be used as an instrument. A transformation often employed in this context is first-differencing. Although it may seem, at first sight, that taking first differences in (1), which results (for $t > 1$) in

$$
\Delta y_{it} = \gamma \Delta y_{i,t-1} + \Delta \varepsilon_{it},
\tag{5}
$$

has completely removed the individual effects, this is only really the case under (4). This is seen as follows. From (2) we find (for $t > 1$)

$$
\begin{aligned}
\Delta y_{it} \;=\; & \gamma^{t-1}\left[\beta - (1-\gamma)\alpha_0\right] + \gamma^{t-1}\left[1 - (1-\gamma)\alpha_1\right]\eta_i \\
& + \varepsilon_{it} - (1-\gamma)\left(\sum_{s=1}^{t-1}\gamma^{s-1}\varepsilon_{i,t-s} + \gamma^{t-1}\alpha_2\varepsilon_{i0}\right).
\end{aligned}
\tag{6}
$$

Hence, estimators involving Δy_{it} will in general not be invariant with respect to the elements of α, nor to β, γ, σ_ε^2 and σ_η^2. However, from (6) we find that under (3) and (4) Δy_{it} is invariant with respect to α_0, α_1, β and σ_η^2. For other transformations that are used in this context, such as taking differences from individual means (taken over time) or orthogonal forward deviations (see Arellano and Bover, 1995), one easily finds similar invariance properties, provided (3) and (4) hold.

From (2) we can also derive that (for $t \geq 0$)

$$
\mathrm{Var}(y_{it}) = \left[\frac{1}{1-\gamma} + \gamma^t\left(\alpha_1 - \frac{1}{1-\gamma}\right)\right]^2\sigma_\eta^2 + \left[\frac{1}{1-\gamma^2} + \gamma^{2t}\left(\alpha_2^2 - \frac{1}{1-\gamma^2}\right)\right]\sigma_\varepsilon^2.
\tag{7}
$$

Therefore, variance constancy of y_{it} through time requires both (4) and

$$
\alpha_2 = \pm\sqrt{\frac{1}{1-\gamma^2}},
\tag{8}
$$

and then yields

$$
\mathrm{Var}(y_{it}) = \frac{\sigma_\eta^2}{(1-\gamma)^2} + \frac{\sigma_\varepsilon^2}{1-\gamma^2}, \qquad t = 0, ..., T.
\tag{9}
$$

When the two conditions (4) and (8) hold jointly then we obtain for the auto-covariance (for $0 \leq s \leq t$)

$$
\mathrm{Cov}(y_{it}, y_{i,t-s}) = \frac{\sigma_\eta^2}{(1-\gamma)^2} + \frac{\gamma^s\sigma_\varepsilon^2}{1-\gamma^2}.
\tag{10}
$$

Like (9) these are not determined by t either, thus the three conditions on α jointly imply weak-stationarity of y_{it}.

From the above we conclude that the distribution of the untransformed y_{it} (for $t > 0$) is always determined by β, γ, σ_η^2 and σ_ε^2, irrespective of the properties of y_{i0} (the values of the α parameters), whereas for particular α values, viz. under the three above special conditions, Δy_{it} is determined by γ and σ_ε^2 only. Because in panel data often T is rather small and asymptotics concerns $N \to \infty$, the effects of the initial conditions are not asymptotically diminishing, and hence they are of major importance.

In addition we want to remark that the three special conditions on α, though mathematically convenient, are not necessarily very realistic cases. There does not seem much reason to assume that in actual empirical panel data observations the accumulated impact of the disturbances, of the random individual effects and of the deterministic components all three happen to be constant in magnitude (or variance) through time. Therefore, in principle, in a Monte Carlo study of model (1) one should vary α_0, α_1, α_2, β, γ, σ_η^2 and σ_ε^2, which after scaling with respect to for instance σ_ε^2 implies dimensionality six of the parameter space. However, in most earlier studies all α parameters have been set at their stationarity values. Although weak stationarity implies invariance of inference on γ with respect to β for some

techniques only, β is often set at zero nevertheless. Moreover, $\sigma_\eta^2/\sigma_\varepsilon^2$ is usually fixed at unity[2] and only γ is varied, so that the dimensionality of the nuisance parameter space is restricted from six to just one.

Below, when we address the three stationarity assumptions regarding the initial values that may hold in addition to the *stability condition* $|\gamma| < 1$, we will indicate (3) as *deterministic stationarity*, which implies that $\mathsf{E}(y_{it})$ is constant; (4) will be called *accumulated effect stationarity*, which implies that $\mathsf{E}(y_{it} \mid \eta_i) - \mathsf{E}(y_{it})$ is constant; and (8) will be called *accumulated noise stationarity*, which implies that $\mathrm{Var}(y_{it} \mid \eta_i)$ is constant for all i and t.

4 Various GMM implementations

4.1 Generic framework

The GMM (system) estimators we will examine all fit into the following simple generic setup, which allows further regressors. After appropriate manipulation (transformation and stacking) of the panel data observations one has

$$y_i^{**} = X_i\theta + \upsilon_i, \ i = 1, ..., N, \tag{11}$$

where y_i^{**} is a $T^{**} \times 1$ vector that may contain $\Delta y_i^* = (\Delta y_{i2}, ..., \Delta y_{iT})'$ or $y_i^* = (y_{i2}, ..., y_{iT})'$ or both stacked, hence T^{**} is either $T^* = T - 1$ or $2T^*$, and υ_i contains the corresponding vector of error components. The $T^{**} \times K^{**}$ matrix X_i and $K^{**} \times 1$ coefficient vector θ follow straightforwardly from the choice regarding y_i^{**}. In case of our panel AR(1) model θ equals $(\beta, \gamma)'$ if y_i^{**} contains y_i^* (i.e. $K^{**} = 2$) and if it just contains Δy_i^* then $\theta = \gamma$ (i.e. $K^{**} = 1$), with $X_i = \Delta y_{i,-1}^*$.

The unknown vector θ is estimated by employing $L \geq K^{**}$ moment conditions that hold for $i = 1, ..., N$, viz.

$$\mathsf{E}[Z_i'(y_i^{**} - X_i\theta)] = 0, \tag{12}$$

where Z_i, which will be substantiated in the next subsection, is $T^{**} \times L$. Exploiting the assumption that the individuals are i.i.d. the GMM estimator using the $L \times L$ semi-positive definite weight matrix W is found by minimizing the quadratic form

$$\left(\sum_{i=1}^N Z_i'(y_i^{**} - X_i\theta) \right)' W \left(\sum_{i=1}^N Z_i'(y_i^{**} - X_i\theta) \right), \tag{13}$$

which yields

$$\hat{\theta}_W = (X'ZWZ'X)^{-1}X'ZWZ'y^{**}, \tag{14}$$

where $y^{**} = (y_1^{**\prime}, ..., y_N^{**\prime})'$, $X = (X_1', ..., X_N')'$ and $Z = (Z_1', ..., Z_N')'$. We only consider cases where convenient regularity conditions hold, including the existence of

$$\plim_{N \to \infty} \frac{1}{N} Z'X \qquad \text{and} \qquad \plim_{N \to \infty} \frac{1}{N} Z'Z,$$

whereas $\mathrm{rank}(N^{-1}Z'X)$ and $\mathrm{rank}(N^{-1}Z'Z)$ are K^{**} and L with probability 1, respectively, implying that the estimator $\hat{\theta}_W$ exists. According to (12) the instruments are valid, thus $\plim_{N \to \infty} N^{-1} \sum_{i=1}^N Z_i'\upsilon_i = 0$, so that $\hat{\theta}_W$ is consistent.

---

[2]Notable exceptions are Nerlove (1967, 1971), initiating this line of research. Nerlove uses in his design the so-called "intra-class correlation", the transformed parameter $\rho \equiv \sigma_\eta^2/(\sigma_\varepsilon^2 + \sigma_\eta^2)$, which he varies over its entire domain $0 \leq \rho < 1$.

The asymptotically efficient GMM estimator[3] in the class of estimators exploiting instruments Z is obtained if W is chosen such that, after appropriate scaling, it has probability limit proportional to the inverse of the covariance of the limiting distribution of $N^{-1/2} \sum_{i=1}^{N} Z_i' v_i$. This implies

$$W^{opt} \propto \left(\plim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} Z_i' v_i v_i' Z_i \right)^{-1}, \tag{15}$$

and we shall denote the asymptotically efficient GMM estimator as $\hat{\theta}_{W^{opt}}$.

For the special case $v_i \sim$ i.i.d.$(0, \sigma_v^2 I_{T^{**}})$ one would obtain $W_{iid}^{opt} \propto (Z'Z)^{-1}$, which yields the familiar 2SLS or GIV result

$$\hat{\theta}_{GIV} = [X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}Z'y^{**}, \tag{16}$$

which in case $K^{**} = L$ simplifies to the simple instrumental variable estimator

$$\hat{\theta}_{IV} = (Z'X)^{-1}Z'y^{**}. \tag{17}$$

However, $v_i \sim$ i.i.d.$(0, \sigma_v^2 I_{T^{**}})$ does usually not hold in panel data, either due to the occurrence of the two error components, or due to the effects of the transformation applied to remove the individual specific component. Also (but we shall not consider these cases here) there may be further complications, such as cross-sectional heteroskedasticity, cross-sectional dependence or serial correlation.

If, under particular assumptions on the distribution of v_i, the matrix W^{opt} is not directly available then one may use some arbitrary initial weight matrix W that produces a consistent (though inefficient) 1-step GMM estimator $\hat{\theta}_W$, and then exploit the 1-step residuals $\hat{v}_i = y_i^{**} - X_i \hat{\theta}_W$ to construct the empirical weight matrix

$$\hat{W}^{opt} \propto \left(\frac{1}{N} \sum_{i=1}^{N} Z_i' \hat{v}_i \hat{v}_i' Z_i \right)^{-1}, \tag{18}$$

which yields the 2-step GMM estimator

$$\hat{\theta}_{\hat{W}^{opt}} = (X'Z\hat{W}^{opt}Z'X)^{-1}X'Z\hat{W}^{opt}Z'y^{**}. \tag{19}$$

This 2-step GMM estimator $\hat{\theta}_{\hat{W}^{opt}}$ is asymptotically equivalent to $\hat{\theta}_{W^{opt}}$, and hence it is efficient in the class of estimators exploiting instruments Z.

Note that, provided the moment conditions are valid, $\hat{\theta}_{GIV}$ is consistent thus could be employed as a 1-step GMM estimator. When $K^{**} = L$ using a weight matrix is redundant, because the criterion function (13) will be zero for any W, because all moment conditions can be imposed on the sample observations.

4.2   Instruments for panel AR(1) models

We now consider a range of GMM estimators that have been suggested for linear dynamic panel data models, but we specialize them to our very specific model (1). It is obvious that estimating model (1) by OLS will yield inconsistent estimators, because it follows from (2) that

$$\mathsf{E}[y_{i,t-1}(\varepsilon_{it} + \eta_i)] = \sigma_\eta^2 \left[\frac{1}{1-\gamma} + \gamma^t \left(\alpha_1 - \frac{1}{1-\gamma} \right) \right], \tag{20}$$

---

[3]see Hansen (1982) and, in the context of dynamic panel data, Arellano (2003) and Baltagi (2005).

and this differs from zero for any α_1 and $|\gamma| < 1$, unless $\sigma_\eta^2 = 0$. Also the least-squares estimator for γ obtained after removing the individual effects from the model by taking deviations from the mean per individual over the time-series observations, known as the least-squares dummy variable estimator (LSDV) or within groups estimator, is inconsistent unless $T \to \infty$.

To obtain valid moment conditions many techniques actually estimate the transformed model

$$\Delta y_{it} = \gamma \Delta y_{i,t-1} + \Delta \varepsilon_{it}, \qquad i = 1, ..., N; \ t = 2, ..., T. \tag{21}$$

Applying least-squares would again yield an inconsistent estimator (also for $T \to \infty$) because

$$\mathsf{E}(\Delta y_{i,t-1} \Delta \varepsilon_{it}) = -\mathsf{E}(y_{i,t-1} \varepsilon_{i,t-1}) = -\sigma_\varepsilon^2 \neq 0,$$

but, given $\varepsilon_{it} \sim$ i.i.d.$(0, \sigma_\varepsilon^2)$, it is obvious that in (21) the $(T-2)(T-1)/2$ moment conditions

$$\mathsf{E}(\Delta y_{ir} \Delta \varepsilon_{it}) = 0, \qquad r = 1, ..., t-2; \ t = 3, ..., T, \tag{22}$$

can be exploited. In fact, the even more extensive set of $(T-1)T/2$ conditions

$$\mathsf{E}(y_{ir} \Delta \varepsilon_{it}) = 0, \qquad r = 0, ..., t-2; \ t = 2, ..., T \tag{23}$$

hold. Note that all conditions of set (22), where the instruments are lagged first differences, are implied by set (23), where the instruments are lagged levels. So, we cannot exploit these two sets jointly, because they would lead to a Z matrix that does not have full column rank. Set (23) implies for the ith partition $\Delta \varepsilon_i^* = (\Delta \varepsilon_{i2}, ..., \Delta \varepsilon_{iT})'$ of the stacked disturbance vector of model (21) the corresponding block Z_i of the $N(T-1) \times (T-1)T/2$ matrix of instruments Z that is given (and here denoted) by

$$Z_i^{AB} = \begin{bmatrix} y_{i,0} & 0, 0 & \cdots & 0' & 0' \\ 0 & y_{i,0}, y_{i,1} & & 0' & 0' \\ \vdots & & \ddots & & \vdots \\ 0 & 0, 0 & & y_{i,0}, ..., y_{i,T-3} & 0' \\ 0 & 0, 0 & \cdots & 0' & y_{i,0}, ..., y_{i,T-2} \end{bmatrix}. \tag{24}$$

For the instrument matrix Z^{AB} to have full column rank it is required that $N \geq T/2$ and $T \geq 2$. Exploiting the instrument set Z_i^{AB} was suggested by Arellano and Bond (1991).

It has been found, however, that using Z^{AB} may lead to bias problems, especially when N is moderate and T not very small. Bun and Kiviet (2005) found that the leading term of this bias has magnitude $O(N^{-1}T^0)$. They suggest using fewer instruments in order to mitigate the bias problems, and prove that GMM using a subset of $2T - 3$ instruments, viz.

$$Z_i^{BK} = \begin{bmatrix} y_{i,0} & 0, 0 & \cdots & 0, 0 & 0, 0 \\ 0 & y_{i,0}, y_{i,1} & & 0, 0 & 0, 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0, 0 & & y_{i,T-4}, y_{i,T-3} & 0, 0 \\ 0 & 0, 0 & \cdots & 0, 0 & y_{i,T-3}, y_{i,T-2} \end{bmatrix}, \tag{25}$$

reduces the order of the bias to $O(N^{-1}T^{-1})$. Note that $Z_i^{BK} = Z_i^{AB} C_{AB}^{BK}$, where C^{BK} is a $(T-1)T/2 \times (2T-3)$ matrix that annihilates instruments from Z_i^{AB}.

Anderson and Hsiao (1981, 1982) originated removing the individual effects from the model by first-differencing and then applying IV to (21). They suggested using either a

lagged first-difference as instrument, i.e. exploiting (22) just for $r = t - 2$, or a lagged level variable as instrument, i.e. exploiting (23) just for $r = t - 2$, giving either

$$Z_i^{AHd} = \begin{bmatrix} 0 \\ \Delta y_{i,1} \\ \vdots \\ \Delta y_{i,T-2} \end{bmatrix} \text{ or } Z_i^{AHl} = \begin{bmatrix} y_{i,0} \\ y_{i,1} \\ \vdots \\ y_{i,T-2} \end{bmatrix} \tag{26}$$

as instruments. These entail a much more drastic reduction of the valid instruments available than Z_i^{BK}. Note that $Z_i^{AHl} = Z_i^{BK} C_{BK}^{AHl}$ and $Z_i^{AHd} = Z_i^{BK} C_{BK}^{AHd}$, where both C_{BK}^{AHl} and C_{BK}^{AHd} are simple transformation matrices of just one column.

Blundell and Bond (1998), on the other hand, attempted to mitigate the bias problems associated with Z^{AB} by extending the set of instruments. Upon checking the validity of lags of differenced variables as instruments in the untransformed model (1) in levels, one finds using (6)

$$\mathsf{E}[\Delta y_{ir}(\eta_i + \varepsilon_{it})] = \gamma^{r-1} \left[1 - (1-\gamma)\alpha_1\right] \sigma_\eta^2, \qquad r = 1, ..., t-1; \ t = 2, ..., T. \tag{27}$$

Hence, under the condition of stationary accumulated effects (4) these expectations are zero, and then they imply the $(T-1)T/2$ moment conditions

$$\mathsf{E}[\Delta y_{ir}(\eta_i + \varepsilon_{it})] = 0, \qquad r = 1, ..., t-1; \ t = 2, ..., T. \tag{28}$$

This set can be transformed linearly into two sub-sets of $T-1$ and $(T-2)(T-1)/2$ conditions respectively, viz.

$$\left. \begin{aligned} \mathsf{E}[\Delta y_{i,t-1}(\eta_i + \varepsilon_{it})] &= 0, \qquad t = 2, ..., T, \\ \mathsf{E}[\Delta y_{ir} \Delta \varepsilon_{it}] &= 0, \qquad\quad r = 1, ..., t-2; \ t = 3, ..., T. \end{aligned} \right\} \tag{29}$$

Note that the second sub-set conforms to (22) and hence is found to be implied by (23) already. Nevertheless, we see that assuming stationary accumulated effects generates $T - 1$ moment conditions in addition to (22). Arellano and Bover (1995) and Blundell and Bond (1998)[4] exploit these in a system comprising both equations

$$\left. \begin{aligned} \Delta y_{it} &= \gamma \Delta y_{i,t-1} + \Delta \varepsilon_{it} \\ y_{it} &= \beta + \gamma y_{i,t-1} + u_{it} \end{aligned} \right\} \tag{30}$$

for $t = 2, ..., T$, with $u_{it} = \eta_i + \varepsilon_{it}$. Here the ith block has the disturbances

$$v_i = (\Delta \varepsilon_{i2}, ..., \Delta \varepsilon_{iT}, u_{i2}, ..., u_{iT})', \tag{31}$$

and the instrument matrix has i^{th} block

$$Z_i^{BB} = \begin{bmatrix} Z_i^{AB} & & & O & \\ & \Delta y_{i1} & 0 & \cdots & 0 \\ & 0 & \Delta y_{i2} & & 0 \\ O & \vdots & & \ddots & \vdots \\ & 0 & 0 & \cdots & \Delta y_{i,T-1} \end{bmatrix}, \tag{32}$$

---

[4]They speak about mean-stationarity instead of accumulated effect stationarity. We prefer the latter, because (4) does not lead to constant (conditional) mean in a model where (3) does not hold.

whereas $Z^{BB} = (Z_1^{BB'}, ..., Z_N^{BB'})'$ is $2N(T-1) \times (T+2)(T-1)/2$. Applying GMM to the system (30) exploiting the instruments (32) will be labelled below as GMMs. For Z^{BB} to have full column rank $T \geq 2$ and $N \geq (T+2)/4$ are required.

Due to the i.i.d. assumption regarding ε_i further (non-linear) moment conditions are valid in the dynamic panel data model, see Ahn and Schmidt (1995, 1997), but below we will only consider the instrument matrices (24), (25), (26) and (32).

4.3 Weight matrices

To establish the optimal weight matrix W^{opt} of (15) for the generic model (11) we have to find an expression that has probability limit equivalent to $\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \mathsf{E}(Z_i' v_i v_i' Z_i)$, which is a symmetric matrix. Note that when Z_i^{AB} is used the individual elements of the matrix $Z_i' v_i v_i' Z_i$ are all of the form $y_{ir} \Delta \varepsilon_{it} \Delta \varepsilon_{is} y_{ip}$, with $t, s = 2, ..., T$ and $r = 0, ..., t-2$; $p = 0, ..., s-2$. When Z_i^{BK} is used only the cases $r = \max(0, t-3), t-2$ and $p = \max(0, s-3), s-2$ occur. Because of the symmetry we shall focus on the expression $y_{ir} \Delta \varepsilon_{it} \Delta \varepsilon_{is} y_{ip}$ for the cases $t \geq s$ only. Using E_q to denote expectation conditional on information available at time period q, we obtain (under both time-series and cross-section homoskedasticity and independence of the disturbances) for $t = s$

$$\mathsf{E}_{s-2}(y_{ir} \Delta \varepsilon_{is} \Delta \varepsilon_{is} y_{ip}) = y_{ir} y_{ip} \mathsf{E}_{s-2}(\Delta \varepsilon_{is})^2 = 2\sigma_\varepsilon^2 y_{ir} y_{ip},$$

and for $t = s + 1$ we find

$$\mathsf{E}_{s-1}(y_{ir} \Delta \varepsilon_{i,s+1} \Delta \varepsilon_{is} y_{ip}) = y_{ir} y_{ip} \mathsf{E}_{s-1}[(\varepsilon_{i,s+1} - \varepsilon_{is})(\varepsilon_{is} - \varepsilon_{i,s-1})] = -\sigma_\varepsilon^2 y_{ir} y_{ip},$$

whereas $\mathsf{E}(y_{ir} \Delta \varepsilon_{it} \Delta \varepsilon_{is} y_{ip}) = 0$ for $t > s + 1$. Employing both the law of large numbers and the law of iterated expectations we have

$$\operatorname*{plim}_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \mathsf{E}_q(y_{ir} \Delta \varepsilon_{it} \Delta \varepsilon_{is} y_{ip}) = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \mathsf{E} \mathsf{E}_q(y_{ir} \Delta \varepsilon_{it} \Delta \varepsilon_{is} y_{ip})$$

$$= \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \mathsf{E}(y_{ir} \Delta \varepsilon_{it} \Delta \varepsilon_{is} y_{ip})$$

for any q. Using this with the obtained (conditional) expectations we establish

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \mathsf{E}(y_{ir} \Delta \varepsilon_{it} \Delta \varepsilon_{is} y_{ip}) = \sigma_\varepsilon^2 h_{ts} \operatorname*{plim}_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} y_{ir} y_{ip},$$

where h_{ts} is the typical element of the $(T-1) \times (T-1)$ matrix

$$H = \begin{bmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & -1 & 2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -1 \\ 0 & \dots & 0 & -1 & 2 \end{bmatrix}. \tag{33}$$

From the above we find that for GMM estimation of model (21), when the DGP (data generating process) is given by (1) and the instruments Z_i^k (for $k = AB, BK$) are being

employed, the optimal weight matrix W_k^{opt} is given by

$$W_k^{opt} \propto [\sum_{i=1}^{N} Z_i^{k\prime} H Z_i^k]^{-1} = [Z^{k\prime}(I_N \otimes H)Z^k]^{-1}. \tag{34}$$

Using it in 1-step GMM we will indicate as GMM_1^{AB} (or GMM_1^{BK}), whereas GMM_2^{AB} employs the residuals of GMM_1^{AB} in 2-step GMM and similarly for GMM_2^{BK}. Self-evidently, the weight matrix for the Anderson-Hsiao implementations is of no concern, because the number of instruments equals the number of regressors.

The derivation of the optimal weight matrix for the GMMs estimator is much more involved. Although its matrix $Z_i^{\prime} v_i v_i^{\prime} Z_i$ still contains the elements just examined, it now also has elements of the hybrid type $y_{ir}\Delta\varepsilon_{it}(\varepsilon_{is} + \eta_i)\Delta y_{i,s-1}$, for $t, s = 2, ..., T$ and $r = 0, ..., t-2$, and of the type $\Delta y_{i,t-1}(\varepsilon_{it} + \eta_i)(\varepsilon_{is} + \eta_i)\Delta y_{i,s-1}$, where $t, s = 2, ..., T$. To date the GMMs optimal weight matrix has only been derived for the specific no individual effects case $\sigma_\eta^2 = 0$ by Windmeijer (2000), who finds

$$W_{BB,\sigma_\eta^2=0}^{opt} \propto [Z^{BB\prime}(I_N \otimes D_{\sigma_\eta^2=0}^{opt})Z^{BB}]^{-1} \tag{35}$$

with

$$D_{\sigma_\eta^2=0}^{opt} = \begin{pmatrix} H & C \\ C' & I_{T-1} \end{pmatrix}, \tag{36}$$

where C is the $(T-1) \times (T-1)$ matrix

$$C = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ -1 & 1 & 0 & \ddots & \vdots \\ 0 & -1 & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & -1 & 1 \end{bmatrix}. \tag{37}$$

We re-establish this result with respect to the matrix C by observing that for $s > t$ and for $t > s + 1$ we have $\mathsf{E}(y_{ir}\Delta\varepsilon_{it}\varepsilon_{is}\Delta y_{i,s-1}) = 0$, whereas for $t = s$ we find

$$\mathsf{E}_{s-1}(y_{ir}\Delta\varepsilon_{is}\varepsilon_{is}\Delta y_{i,s-1}) = y_{ir}\Delta y_{i,s-1}\mathsf{E}_{s-1}(\varepsilon_{is}^2 - \varepsilon_{is}\varepsilon_{i,s-1}) = \sigma_\varepsilon^2 y_{ir}\Delta y_{i,s-1},$$

and for $t = s + 1$ we obtain

$$\mathsf{E}_{s-1}(y_{ir}\Delta\varepsilon_{i,s+1}\varepsilon_{is}\Delta y_{i,s-1}) = y_{ir}\Delta y_{i,s-1}\mathsf{E}_{s-1}(\varepsilon_{i,s+1}\varepsilon_{is} - \varepsilon_{is}^2) = -\sigma_\varepsilon^2 y_{ir}\Delta y_{i,s-1}.$$

The identity matrix in the South-East block of (36) follows from $\mathsf{E}(\Delta y_{i,t-1}\varepsilon_{it}\varepsilon_{is}\Delta y_{i,s-1}) = 0$ for $t \neq s$, whereas for $t = s$ we obtain $\mathsf{E}_{s-1}(\Delta y_{i,s-1}\varepsilon_{is}^2\Delta y_{i,s-1}) = \sigma_\varepsilon^2(\Delta y_{i,s-1})^2$.

When allowing for individual effects, we find with respect to the elements $\Delta y_{i,t-1}(\varepsilon_{it} + \eta_i)(\varepsilon_{is} + \eta_i)\Delta y_{i,s-1}$ for $t = s$ that

$$\mathsf{E}_{s-1}(\Delta y_{i,s-1}(\varepsilon_{is} + \eta_i)^2\Delta y_{i,s-1}) = (\Delta y_{i,s-1})^2(\sigma_\varepsilon^2 + \eta_i^2)$$

and for $t > s$

$$\begin{aligned} \mathsf{E}_{t-1}[\Delta y_{i,t-1}(\varepsilon_{it} + \eta_i)(\varepsilon_{is} + \eta_i)\Delta y_{i,s-1}] &= \Delta y_{i,t-1}\Delta y_{i,s-1}(\varepsilon_{is} + \eta_i)[\mathsf{E}_{t-1}(\varepsilon_{it}) + \eta_i] \\ &= \Delta y_{i,t-1}\Delta y_{i,s-1}(\varepsilon_{is}\eta_i + \eta_i^2). \end{aligned}$$

Since $\text{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \eta_i^2 = \sigma_\eta^2$ and $\text{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \varepsilon_{is}\eta_i = 0$ it follows that in case $\sigma_\eta^2 \neq 0$ we should replace the I_{T-1} matrix in the South-East block of (36) by $I_{T-1} + \frac{\sigma_\eta^2}{\sigma_\varepsilon^2} \iota_{T-1}\iota'_{T-1}$.

Without having analyzed the effects of $\sigma_\eta^2 \neq 0$ on the non-diagonal block, from the above we already note that, when the ratio $\sigma_\eta^2/\sigma_\varepsilon^2$ is unknown (which it usually is), the optimal weight matrix is infeasible. Therefore, various feasible but (even under i.i.d. disturbances) non-optimal weight matrices have been suggested. Blundell, Bond and Windmeijer (2000, footnote 11), and Doornik et al. (2002, p.9) in the computer program DPD, use in 1-step GMMs the operational weight matrix $W_{BB}^{DPD} \propto [Z^{BB'}(I_N \otimes D^{DPD})Z^{BB}]^{-1}$, with

$$D^{DPD} = \begin{pmatrix} H & O \\ O & I_{T-1} \end{pmatrix}. \tag{38}$$

The motivation for the chosen block diagonality of D^{DPD}, which does not lead to an interesting reduction of computational requirements, is unclear. There is no specific parametrization for which these weights are optimal. Blundell and Bond (1998) did use (see page 130, 7 lines from bottom) in their first step of 2-step GMMs

$$D^{GIV} = \begin{pmatrix} I_{T-1} & O \\ O & I_{T-1} \end{pmatrix} = I_{2T-2}, \tag{39}$$

which yields the simple GIV estimator. This is certainly not optimal, but it is easy and suits well perhaps as a first step (to be indicated below as GMMs_1^{GIV}) in a 2-step procedure (GMMs_2^{GIV}). Blundell and Bond (1998) mention that, in most of the cases they examined, GMMs_1^{GIV} gave similar results as GMMs_2^{GIV}, suggesting that the weight matrix to be used in combination with Z^{BB} under homoskedasticity seems of minor concern. However, in our less restrained simulation experiments below, we will find that different weight matrices can lead to huge differences in the performance of GMMs_1. Although we find too that GMMs_1^{GIV} and GMMs_2^{GIV} give very similar results, we also establish that both perform poorly in finite sample in comparison to other operational weight matrices.

A more promising operational alternative to D^{DPD} and D^{GIV} could be the following. Instead of capitalizing on the assumption $\sigma_\eta^2 = 0$, as in $D_{\sigma_\eta^2=0}^{opt}$, one could employ a weight matrix which presupposes a particular (not necessarily zero) value of the ratio $\sigma_\eta^2/\sigma_\varepsilon^2$, i.e.

$$D_{\sigma_\eta^2/\sigma_\varepsilon^2}^{subopt} = \begin{pmatrix} H & C \\ C' & I_{T-1} + \frac{\sigma_\eta^2}{\sigma_\varepsilon^2}\iota_{T-1}\iota'_{T-1} \end{pmatrix}. \tag{40}$$

We make explicit that the resulting weight matrix

$$W_{BB}^{subopt} \propto [Z^{BB'}(I_N \otimes D_{\sigma_\eta^2/\sigma_\varepsilon^2}^{subopt})Z^{BB}]^{-1} \tag{41}$$

is not optimal (unless $\sigma_\eta^2 = 0$) even when a correct non-zero value of $\sigma_\eta^2/\sigma_\varepsilon^2$ is substituted, because the non-diagonal blocks have been obtained under the assumption of no individual effects.

Above an already rather wide collection of competing 1-step and 2-step GMM estimators have been presented. They are consistent (provided the employed moment conditions are valid), and some may claim optimality properties under particular circumstances asymptotically for $N \to \infty$. Below we will make an initial attempt to rank the performance of these estimators by employing Monte Carlo simulation for the situation where both T and N are small or moderate.

5   A limited Monte Carlo contest

In the experiments below we have severely restricted ourselves regarding the generality of the Monte Carlo design, simply for practical reasons. The main purpose here is to illustrate particular pitfalls in estimator comparisons by Monte Carlo. This study is certainly not meant to produce the 'final' in a Monte Carlo tournament on dynamic panel data estimators. At this stage we only considered the fully homoskedastic case, where $\varepsilon_i \sim$ i.i.d.$(0, \sigma_\varepsilon^2 I_{T+1})$, whereas 2-step GMM is especially meant to cope with cross-section heteroskedasticity. According to aspiration 7, as set out before, we should have included estimators that exploit the homoskedasticity restriction, since it implies further moment conditions and corresponding asymptotically more efficient GMM implementations. However, if we had included them, we should also have performed experiments involving heteroskedasticity in order to satisfy aspiration 6 and make all estimators play both at home and away. Also, in all the present experiments ε_{it} and η_i are Gaussian. This situation too could have been exploited by including maximum-likelihood estimators in the set of competing techniques[5], but we didn't at this stage, so we do not fully respect aspirations 6 and 7. Not only did we not examine dynamic panel data models with any further exogenous regressors, but we also imposed $\beta = 0$ so that deterministic stationarity is in fact not an issue here, and all GMM estimators have $K^{**} = 1$. We also restricted ourselves to models with accumulated noise stationarity, i.e. (8) holds and $\alpha_2 = \pm(1 - \gamma^2)^{-1/2}$.

Thus, apart from illustrating some of the individual aspirartions for an impartial Monte Carlo contest, the main object of study will be limited here to examine whether under accumulated noise stationarity, homoskedasticity, cross-section independence and normality of both error components there is much impact on the bias and RMSE (root mean squared error) in zero-mean panel AR(1) models of:

(i) the effects of various different weight matrices in GMMs;

(ii) the effects of skipping from Z^{BB} the instruments related to accumulated effect stationarity, i.e. using Z^{AB};

(iii) the effects of skipping valid moment conditions from Z^{AB} and using either Z^{BK} or even Z^{AHl} or Z^{AHd}.

We will investigate just a few particular small values for N and T, only positive stable values of γ, a few different values of $\sigma_\eta^2/\sigma_\varepsilon^2$ and of α_1. Note that GMMs will be inconsistent for $\alpha_1 \neq (1 - \gamma)^{-1}$.

5.1   An orthogonal Monte Carlo design

Due to the above mentioned restrictions, we loose the parameters β, α_0 and α_2, and retain γ, α_1, σ_η^2, and σ_ε^2 only. Note that our data series y_{it} and Δy_{it} are now such that

$$\mathsf{E}(y_{it}) = 0 \qquad\qquad \mathsf{E}(\Delta y_{it}) = 0$$
$$\mathsf{E}(y_{it} \mid \eta_i) = \left[\tfrac{1}{1-\gamma} + \gamma^t\left(\alpha_1 - \tfrac{1}{1-\gamma}\right)\right]\eta_i \qquad \mathsf{E}(\Delta y_{it} \mid \eta_i) = \gamma^{t-1}[1 - (1-\gamma)\alpha_1]\eta_i$$
$$\mathrm{Var}(y_{it}) = \left[\tfrac{1}{1-\gamma} + \gamma^t\left(\alpha_1 - \tfrac{1}{1-\gamma}\right)\right]^2 \sigma_\eta^2 + \tfrac{\sigma_\varepsilon^2}{1-\gamma^2} \quad \mathrm{Var}(\Delta y_{it}) = \gamma^{2t-2}[1 - (1-\gamma)\alpha_1]^2 + \tfrac{2\sigma_\varepsilon^2}{1-\gamma^2}$$
$$\mathrm{Var}(y_{it} \mid \eta_i) = \tfrac{\sigma_\varepsilon^2}{1-\gamma^2} \qquad\qquad \mathrm{Var}(\Delta y_{it} \mid \eta_i) = \tfrac{2\sigma_\varepsilon^2}{1-\gamma^2}$$

Since all estimators to be examined include level and/or differenced variables in a highly non-linear way, we do not expect that more general invariance properties can be exploited in

---

[5]Hsiao et al. (2002) examines ML and GMM estimators under both normal and nonnormal errors.

the Monte Carlo than the one with respect to the scale parameter σ_ε. So, next to items such as N, T and the adopted normality of both random components, the only design parameters are γ, α_1 and $\sigma_\eta^2/\sigma_\varepsilon^2$. However, we shall not use these as a base for the Monte Carlo grid, but γ, ϕ and ψ, where

$$\phi \equiv \alpha_1(1-\gamma), \tag{42}$$

$$\psi^2 \equiv \frac{1+\gamma}{1-\gamma}\frac{\sigma_\eta^2}{\sigma_\varepsilon^2}. \tag{43}$$

Note that for $\phi = 1$ we have accumulated effect stationarity[6], for $\phi = 0$ the individual effect component in y_{i0} is zero, and for $\phi = 0.5$ the effect component in y_{i0} is 50% of the stationary magnitude $1/(1-\gamma)$. Hence, by fixing ϕ instead of α_1 we can control the effect component in y_{i0} in proportion to the stationary magnitude, irrespective of the chosen value for γ, whereas when we fix α_1 this characteristic would vary with γ. Therefore we may call the parameters γ and ϕ autonomous and orthogonal: γ is just about speed of adjustment and ϕ exclusively about the initial disequilibrium condition with respect to the individual effects, whereas α_1 is an ambiguous parameter because the consequences for y_{it} of its magnitude can only be understood when either γ or ϕ are known too.

The parameter ψ expresses the following basic characteristic of the observations from the DGP independently from γ and ϕ. We find that when $\phi = 1$ or otherwise for $t \to \infty$ when accumulated effect stationarity has been attained

$$\begin{aligned}
\text{Var}(y_{it}) &= \frac{\sigma_\eta^2}{(1-\gamma)^2} + \frac{\sigma_\varepsilon^2}{1-\gamma^2} = \left(\frac{1+\gamma}{1-\gamma}\frac{\sigma_\eta^2}{\sigma_\varepsilon^2} + 1\right)\frac{\sigma_\varepsilon^2}{1-\gamma^2} \\
&= (\psi^2 + 1)\frac{\sigma_\varepsilon^2}{1-\gamma^2}.
\end{aligned} \tag{44}$$

Hence, ψ^2 expresses under accumulated effect stationarity the magnitude of the variance component in $\text{Var}(y_{it})$ stemming from the individual effects in terms of the magnitude of the variance of $\text{Var}(y_{it})$ originating from the other error component, the accumulated noise. If we would fix $\sigma_\eta^2/\sigma_\varepsilon^2$ the characteristic ψ^2 would vary with γ. By fixing ψ we can control the relative size of the variance of the accumulated two error components in a stationary y_{it}, irrespective of the value of γ, by choosing

$$\sigma_\eta^2 = \frac{1-\gamma}{1+\gamma}\psi^2\sigma_\varepsilon^2. \tag{45}$$

Although $\sigma_\eta^2/\sigma_\varepsilon^2$ directly characterizes the relative magnitude of the incremental error components in the right-hand side of $y_{it} = \beta + \gamma y_{i,t-1} + \eta_i + \varepsilon_{it}$, it does not characterize the data, unless γ is known too. Therefore we find that ψ is an autonomous parameter (like γ and ϕ), whereas $\sigma_\eta/\sigma_\varepsilon$ and α_1 are not. The reparametrization makes it easier to chose and to cover the relevant values of the parameters and it will enhance the interpretability of the Monte Carlo results too, because we can change γ now while keeping the two other basic characteristics expressed by ϕ and ψ fixed, whereas otherwise changing γ would imply changing the relative prominence and closeness to stationarity of the accumulated individual effects too.[7]

---

[6]This parameter is also used in Andrews and Lu (2001, p.149).

[7]Note that ψ differs from the design parameter $\mu = (1-\gamma)^{-1}\sigma_\eta/\sigma_\varepsilon$ introduced in the Monte Carlo in Kiviet (1995, p.65) and also practiced in Harris and Mátyás (2004). Like ψ this μ serves to break away

Most published Monte Carlo studies on (generalizations of) the above panel AR(1) model assume that all stationarity conditions hold. And in addition most of them consider only one value for $\sigma_\eta^2/\sigma_\varepsilon^2$ (often one), which is a further restriction on the dimensionality of the parameter space[8]. This restriction makes it quite likely that important properties of the estimators will be overlooked, and one cannot distinguish between the effects of increasing γ and decreasing ψ^2. Even if the value of $\sigma_\eta^2/\sigma_\varepsilon^2$ is varied, relationship (43) implies that from such simulations one cannot easily disentangle the effects due to a high (low) γ and those due to a high (low) ψ value[9].

What would be reasonable values for ψ? In Blundell and Bond (1998, model A) it varies from 1 when $\gamma = 0$ to 19 when $\gamma = 0.9$. In Blundell and Bond (1998, model B)[10] it is 1 again for $\gamma = 0$ but only 0.19 for $\gamma = 0.9$. Hence, neither in model A nor B ψ is held constant when γ is changed, which hampers the interpretation of the Monte Carlo results. It would be interesting to know the actual estimates of σ_η^2 and σ_ε^2 in the application of the AR(1) model in Bond (2002) or in similar empirical exercises. In Blundell, Bond, Windmeijer (2000), for instance p.16, it is discussed what happens "at high values of γ and high values of $\sigma_\eta^2/\sigma_\varepsilon^2$". Note that this implies doubly high values of ψ^2 in their model A. Thus, we should know whether it is really likely to occur that $\psi = 5\sqrt{1.8/0.2} = 15$ or even higher. Below, we will restrict ourselves to values $0 \le \psi \le 5$ and $0.1 \le \gamma \le 0.9$, implying $0 \le \sigma_\eta/\sigma_\varepsilon \le 4.52$.

Choosing in a Monte Carlo exclusively equivalent values for σ_η^2 and σ_ε^2 is deceiving from a theoretical point of view as well. Recently, Bun and Kiviet (2005) derived a first order asymptotic approximation to the bias of the GMMs estimator when use is made of the simple weight matrix involving $D^{GIV} = I$ in first-order autoregressive panel data models with another weakly exogenous regressor. When we specialize that result for the case of a pure AR(1) panel model we obtain that the leading term of the bias, which is $O(N^{-1})$, is in fact proportional to

$$\frac{1}{N}\frac{1}{1-\gamma}\left(1 - \frac{\sigma_\eta^2}{\sigma_\varepsilon^2}\right). \tag{46}$$

Hence, in all Monte Carlo studies restricted to the case $\sigma_\eta^2 = \sigma_\varepsilon^2$ this leading term is zero, which may explain the relatively small bias that has been established for the GMMs estimator

---

from the habitual designs where $\sigma_\eta/\sigma_\varepsilon$ is constant. Its rationalization bears on the arguments given in Nerlove (1967, footnote 2) and is also in line with the parameter orthogonalization achieved by Lancaster (2002, p.655). However, regarding designing a DGP we find that ψ has a more solid underpinning and interpretation than μ has. The latter compares the accumulated magnitude of the individual effect with that of the current idiosyncratic disturbance, whereas ψ^2 (already used in Bun and Kiviet, 2005) captures the accumulated variance of both error components and thus seems closer to a basic characteristic of the generated data orthogonal to γ.

[8]This offence of aspiration 3 occurs in, for instance, Bhargava and Sargan (1983), Arellano and Bond (1991), Blundell and Bond (1998), Blundell et al. (2000), Harris and Mátyás (2000), Andrews and Lu (2001), Bond (2002), Bowsher (2002), Doornik et al. (2002), Bond and Windmeijer (2005) and Windmeijer (2005). Little or no information on the actual magnitude of the individual effects in their experiments is given by Judson and Owen (1999), Hansen (2001) and Hsiao et al. (2002), which is at odds with aspiration 8.

[9]For instance, Alvarez and Arellano (2003, p.1133) examine $\gamma \in \{0.2, 0.5, 0.8\}$ and $\sigma_\eta^2/\sigma_\varepsilon^2 \in \{0.2, 1.0\}$. They infer that $\sigma_\eta^2/\sigma_\varepsilon^2$ has little effect, but since their design does not cover cases where ψ is large (say 5) and γ small, this limited grid cannot reveal the full effect of ψ and γ. Doran and Schmidt (2005) use a wider grid including $\sigma_\eta^2/\sigma_\varepsilon^2 = 4$, so at moderate $\gamma = 0.5$ they cover $\psi = 3.5$. They report, as do Alonso-Borrego and Arellano (1999), that the effects of increasing γ and $\sigma_\eta^2/\sigma_\varepsilon^2$ are qualitatively similar. Note that these findings underscore the non-orthogonality of these two parameters.

[10]Stephen Bond informed me that there is a typo in Blundell and Bond (1998, p.129-130) regarding the specification of simulation model B. Their published simulation results for model B have actually been obtained by (their notation) $y_{it} = \alpha y_{i,t-1} + (1-\alpha)\eta_i + v_{it}$, with start-up value $y_{it} = \eta_i + u_{i1}$, so that accumulated effect stationarity is obtained indeed (which would not when using the initial condition mentioned on p.130).

in such restricted simulation studies.

5.2   Some new Monte Carlo findings

Using the design suggested above I ran simulations (1000 replications for each parametrization, using the same η_i and ε_{it} realizations for different parameter values and different techniques) for various GMM implementations for $N = 100$ and $T = 3(+1)10$, over $\gamma = 0.1(+0.2)0.9$ and choosing $\phi \in \{1, 0.5\}$ and $\psi \in \{0 \text{ or } 0.5, 1, 5\}$. The chosen N may seem small for some applications, but it is large for other. Moreover, finding serious problems and quality differences for modest N values, usually quite well indicates where similar problems may still be looming for larger though finite N. Of course, for $N \to \infty$ there is no bias and the RMSE will be zero for any consistent GMM estimator.

Below two diagrams are presented above each other for all the particular parametrizations and implementations of GMM examined: the upper diagrams present relative bias, i.e. the Monte Carlo estimate of $\mathsf{E}(\hat{\gamma} - \gamma)/\gamma$; the lower diagrams depict relative efficiency, which is $\mathrm{RMSE}(\hat{\gamma})/\gamma$. Both are given in %. Where the relative bias is larger than, say, 25% in absolute value there is a serious bias problem, and where relative RMSE is larger than, say, 50% there seems a serious problem regarding efficient and useful inference. Naturally, extreme values of these relative measures will always occur for values of γ very close to zero. However, for a particular estimator (at the sample sizes examined) to show satisfying behavior it seems reasonable to require that its relative RMSE does not exceed 50% for, say, $\gamma > 0.2$, although this criterion should not be taken too strictly.

5.2.1   GMMs under accumulated effect stationarity ($\phi = 1$)

We first examine the system estimator GMMs for the situation where $\phi = 1$, hence the extra instruments for the equation in levels are valid. Given that observations are available on y_{it} for $t = 0, ..., T$, GMMs exploits $(T+2)(T-1)/2$ instruments, viz. $T(T-1)/2$ instruments as in the Arellano-Bond implementation plus $T-1$ non-redundant "level instruments". In Figure 1 we first investigate the simple weight matrix $D^{GIV} = I_{2T-2}$ in GMMs_1^{GIV} for $\psi = 0$, 1 and 5 (from left to right) respectively. Similar results for GMMs_2^{GIV}, which exploits the empirical weight matrix (18) based on GMMs_1^{GIV} residuals, are given in Figure 2. Both these estimators are examined in Blundell and Bond (1998) too. Note that in their design A $\psi = \sqrt{1.9/0.1} = 4.36$ when $\gamma = 0.9$ and $\psi = \sqrt{1.1/0.9} = 1.1$ when $\gamma = 0.1$, but they did not examine large ψ in combination with small or moderate γ (nor vice versa). From our figures we see that the effect on bias and RMSE of γ and T is almost similar, both for $\psi = 0$ and $\psi = 1$ and for one and two step GMMs^{GIV}, the latter being only slightly better. However, at $\psi = 5$ the sign of the bias has changed (which may explain the very small bias at $\gamma = 0.5$ and $\psi = \sqrt{1.5/0.5} = 1.73$ in Table 5 of Blundell and Bond, 1998) and the bias is dramatically high for moderate and small values of γ. At $\psi = 5$ the bias and RMSE results are satisfactory for large values of γ only (which are the only large ψ cases examined in earlier Monte Carlo's). In Table 5 of Blundell and Bond (1998), for $\gamma = 0.5$ and $\psi = 1.73$, the 1-step and 2-step estimators show both their consistency (by a small bias), and the 2-step estimator shows a just slightly lower variance. Examining the nuisance parameter space over an extra dimension, we find only minor differences between one and two step GMMs^{GIV} too; the 2-step procedure seems to lead to a very moderate improvement of efficiency over the whole parameter space. However, this correspondence should not be interpreted as though the weight matrix has only minor effects, as we will see from further experiments.

In Figures 3 and 4 we examine what the quality is of the operational weights used in DPD. We do not find much difference between the D^{DPD} and D^{GIV} results in 1-step GMMs, although for $\psi = 0$ we find that the DPD weights work slightly better than the simple GIV weights. However, for $\psi = 5$ and γ not very large, using D^{DPD} in the weight matrix works even worse than D^{GIV}.

Next we examine the effects on bias and RMSE of GMMs when using the non-operational weights W_{BB}^{subopt} given in (41), in which we substituted the true value of $\sigma_\eta^2/\sigma_\varepsilon^2$ in 1-step estimation, giving what we denote as $\text{GMMs}_1^{\sigma_\eta^2/\sigma_\varepsilon^2}$ (see Figure 5), and employing its residuals in 2-step estimation to give $\text{GMMs}_2^{\sigma_\eta^2/\sigma_\varepsilon^2}$ (see Figure 6). Note that for $\psi = 0$ (no individual effects) these are the optimal weights. We see that for $\sigma_\eta^2/\sigma_\varepsilon^2 = 0$ estimator $\text{GMMs}_1^{\sigma_\eta^2/\sigma_\varepsilon^2}$ has much smaller bias than both GMMs_1^{GIV} and GMMs_1^{DPD}, and also its RMSE is much smaller, especially for larger values of T (even more than four times smaller at $T = 10$ and $\gamma = 0.9$). Self-evidently at $\psi = 0$ $\text{GMMs}_2^{\sigma_\eta^2/\sigma_\varepsilon^2}$ does not outperform the already optimal $\text{GMMs}_1^{\sigma_\eta^2/\sigma_\varepsilon^2}$, which is also notably better than the asymptotically equivalent GMMs_2^{DPD} and GMMs_2^{GIV}, which have larger (double and triple respectively) RMSE at $\psi = 0$, $\gamma = 0.9$ and $T = 10$, $N = 100$. For $\psi > 0$ $\text{GMMs}_1^{\sigma_\eta^2/\sigma_\varepsilon^2}$ is sub-optimal, but we see that it has very moderate bias and behaves well, whereas both GIV and DPD weights lead to very poor performance when ψ is large and γ moderate or small. For this non-operational 1-step estimator we find too that (under i.i.d. disturbances) 2-step estimation does not yield worthwhile improvements.

Looking into the options for making the attractive properties of weight matrix W_{BB}^{subopt} operational, we examined using it while taking in 1-step estimation for the unknown $\sigma_\eta^2/\sigma_\varepsilon^2$ a value of 10, irrespective of the true value, and indicate it as GMMs_1^{10} in Figure 7. Note that in our simulations the actual value of $\sigma_\eta^2/\sigma_\varepsilon^2$ is zero when $\psi = 0$, and when $\psi = 1$ or 5, $\sigma_\eta^2/\sigma_\varepsilon^2$ decreases from 0.81 to 0.05 and from 20.5 to 1.32 respectively, when γ moves from 0.1 to 0.9. We see that the easy and operational GMMs_1^{10} procedure yields results almost as good as the non-operational $\text{GMMs}_1^{\sigma_\eta^2/\sigma_\varepsilon^2}$ when ψ is large. This is not the case when $\psi \leq 1$, but then it still has smaller RMSE than GMMs_1^{GIV} and GMMs_1^{DPD}, except for T very small and γ very large. Employing the GMMs_1^{10} residuals to obtain GMMs_2^{10} (see Figure 8) the performance of the estimator has improved slightly.

Hence, with respect to the system estimator we conclude that the quality of the weight matrix used in 1-step GMMs is of much more importance for its resulting bias and efficiency than was recognized previously. Also the widespread reputation of GMMs as yielding only very moderate bias needs correction: when GIV or DPD weights are used in the system estimator then a moderate bias is obtained only when γ is large.

5.2.2 GMM not exploiting all valid instruments, while $\phi = 1$

It seems interesting now to examine similar results for GMM in this context while omitting the $T - 1$ differenced instruments in the level equation. Still having $\alpha_1 = 1/(1 - \gamma)$, we find in Figure 9 for the Arellano-Bond implementation that it works best for intermediate values of γ. The bias is negative now for all ψ values and is not large for intermediate γ values. However, it is substantial for large and for small positive γ, especially for large ψ and small T. For large values of γ all GMMs implementations achieve a smaller RMSE. However, we also find that even in the case of accumulated effect stationarity, adding the level equation with differenced instruments and form a system, while using the poor weight matrices based on D^{GIV} or D^{DPD}, is counterproductive when γ is small and ψ large. Irrespective of the weights used in the first step, 2-step GMMs is asymptotically efficient. However in finite sample we

find, that when poor weights are used in 1-step GMMs, the Arellano-Bond implementation is often better.

From Figure 10 we note that omitting more valid instruments by using the matrix Z^{BK} yields slightly less bias and higher efficiency than for Z^{AB} only when $\psi \leq 1$, and the other way around when $\psi = 5$. The latter is surprising, because in Bun and Kiviet (2005) we found that when Z^{BK} is used instead of Z^{AB} then the leading term of the bias is of smaller order in T by a factor T. Earlier we found that it yields a smaller bias indeed in a model with further regressors and $T = 10$ and $N = 20$. Apparently, this does not occur in the pure AR(1) model for $T \leq 10$ at $N = 100$.

Reducing the number of instruments to just one and employing the Anderson-Hsiao implementation with the lagged level instrument Z_i^{AHl} of (26) leads to curious results presented in Figure 11. Here we notice a phenomenon that has already undermined so many earlier simulation findings. Because the number of instruments equals the number of regressors this estimator has no finite moments. Therefore, the Monte Carlo sample average of the 1000 estimates (and also the RMSE) do not converge to a constant for increasing number of Monte Carlo replications, but to a random variable. The ostensible small bias for $\psi \leq 1$ (apart from the outlying value at $T = 3$ and $\gamma = 0.9$) would change in a jumble as for $\psi = 5$ when we would use a much larger number of replications. The density at zero for the denominator in the expression of the estimator is apparently larger for smaller T, larger γ and larger ψ, but at $\psi \leq 1$ the probability of huge outliers in just 1000 trials is still moderate. The results for Z_i^{AHd} (not included in the figures) proved to be even more vulnerable in this respect. To represent the behavior of these estimators appropriately the median and interquartile range could and should have been used[11]. Note that the GMM^{AB} and GMM^{BK} estimators have an equal number of instruments and regressors for $T = 2$ only, so by examining $T \geq 3$ we avoided similar confusing results in the Figures 9 and 10. At $T = 3$ they exploit 3 instruments for one regressor and hence their first two moments exist.

5.2.3   Results for non-stationary accumulated effects ($\phi = 0.5$)

Next we examine various of the estimators when applied to data series where accumulated effect stationarity does not hold. Of course, this is of no interest when $\psi = 0$ and no effects are present. Therefore we examined $\psi = 0.5$ instead. Figure 12 shows that the finite sample properties of GMM$_1^{AB}$ are much worse than those of the $\phi = 1$ case when $\psi = 1$. When $\psi = 0.5$ or 5, however, GMM$_1^{AB}$ works much better under $\phi = 0.5$ than it does under $\phi = 1$. Similar results (not presented) were found for GMM$_1^{BK}$.

Finally we examine what happens when GMMs is applied when $\phi \neq 1$. Of course, for any weight matrix both 1-step and 2-step implementations are inconsistent now. In Figure 13 we see that the inconsistency of GMMs is evident only when ψ is not small and γ not very large. In fact, when $\psi = 1$ and γ large GMMs$_2^{DPD}$ behaves better than the consistent estimator GMM^{AB}. For larger ψ, however, GMM^{AB} is much more efficient. Figure 14 shows that GMMs$_1^{\sigma_\eta^2/\sigma_\varepsilon^2}$ seems less vulnerable when ψ is large. These results make clear that it is of great importance to test for the validity of accumulated effect stationarity, before its orthogonality conditions are imposed in estimation.

---

[11]Studies in which Monte Carlo estimates on moments of Anderson-Hsiao estimators have been presented which (due to the limited number of simulation replications) seem meaningful but are in fact all will o' the wisps include: Arellano and Bond (1991), Arellano and Bover (1995), Kiviet (1995), Judson and Owen (1999), Doornik et al. (2002) and Hsiao et al. (2002).

6 Concluding remarks

We performed Monte Carlo experiments in the context of a very specific simple dynamic panel data model and examined and compared the results of a few implementations of 1 and 2-step GMM, which differ in the number of moment conditions exploited and in the weight matrix employed. This Monte Carlo study as such was not designed to enable a serious competition between the full range of alternative techniques available for dynamic panel data models. It has only been used here to illustrate the importance of the eight distinct methodological aspirations listed in Section 2 for the design and conduct of a fair tournament. In particular, we illustrated that in Monte Carlo comparisons one should always examine what the restrictions are that have been put on the parameter space of the design, and whether these seem reasonable. We gave attention to the dimensionality of the parameter space of the Monte Carlo design, but also to the actual parameter values chosen in the experiments. Moreover, we demonstrated that it is very useful to create a base for the parameter space which is orthogonal with respect to separately interpretable characteristics of the generated data processes. Otherwise, the marginal effects of numerical changes in the design parameters are hard to disentangle. We also showed that it can be useful to examine non-operational techniques, which exploit information that is usually not available in practice, but self-evidently is in simulation experiments.

With respect to the qualities of various GMM estimators in zero-mean stable panel AR(1) models we found the following. Most existing studies have been misleading, because they did not include parametrizations where both the individual effects are prominent (ψ high) and the lagged dependent variable coefficient (γ) moderate. We demonstrate that these are cases where the quality of GMMs (the system estimator) is extremely dependent on the weight matrix used. Both the GIV and DPD weight matrices work poorly here, and performing a second iteration step is of little or no help. It is shown that aiming to get closer to the non-operational optimal weight matrix yields substantial improvements. This finding is in line with results obtained recently in Doran and Schmidt (2005). Earlier simulation studies paid little or no attention to data series where the initial conditions deviate from accumulated effect stationarity. Then GMMs is inconsistent and GMM_1^{AB} is the asymptotically efficient estimator. Although we re-establish that in finite sample this estimator may show substantial bias, we also find that GMM_1^{AB} may work surprisingly well. We note that the bias is affected non-monotonically by the long-run magnitude of the effects ψ when the effect in the initial observation is moderate ($\phi = 0.5$).

We should re-emphasize that our present experimental findings just pertain to the highly specific simple panel AR(1) model without any further explanatory variables. The simplicity of this DGP should in principle enable to obtain analytic evidence on the actual dependence on the model parameters of the quality of alternative instrument and weight matrices and resulting estimator efficiency. Blundell and Bond (1998) obtained such evidence for the simple case $T = 2$ (in our notation) where the weight matrix is not a complicating factor. Probably, due to the simplicity of the AR(1) DGP, a relatively good performance is achieved by instrument matrices incorporating very few lags, because it seems likely that higher-order lagged variables will establish weak instruments here. To rate estimators for empirically more relevant but much more complex DGP's (i.e. including higher-order lags, further weakly exogenous regressors, cross-sectional heteroskedasticity, non-normality, non-stationary initial values) the only practicable option seems to run appropriately designed simulation experiments, for which in our opinion the design in Bun and Kiviet (2005) establishes a reasonable starting point to meet the eight aspirations mentioned in Section 2. However, in conclu-

sion we want to remark that it seems highly unlikely that it will ever happen that single winners can be celebrated in fully fledged simulation contests between alternative inference techniques, because of the following truism: Techniques that build on many assumptions, though efficient when these assumptions do hold, usually are not robust to situations where they are false, whereas more robust techniques will not beat full information techniques on their home ground.

References

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

Ahn, S.C., Schmidt, P., 1997. Efficient estimation of dynamic panel data models: Alternative assumptions and simplified estimation. *Journal of Econometrics* 76, 309-321.

Alonso-Borrego, C., Arellano, M., 1999. Symmetrically normalized instrumental-variable estimation using panel data. *Journal of Business & Economic Statistics* 17, 36-49.

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

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

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

Andrews, D.W.K., Lu, B., 2001. Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models. *Journal of Econometrics* 101, 123-164.

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

Arellano, M., Bond, S., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *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, 29-52.

Balestra, P., Nerlove, M., 1966. Pooling cross section and time series data in the estimation of a dynamic model: The demand for natural gas. Econometrica 34, 585-612.

Baltagi, B.H., 2005. *Econometric Analysis of Panel Data*, third edition. John Wiley & Sons.

Bhargava, A., Sargan, J.D., 1983. Estimating dynamic random effects models from panel data covering short time periods. *Econometrica* 51, 1635-1659.

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

Blundell, R., Bond, S., Windmeijer, F., 2000. Estimation in dynamic panel data models: improving on the performance of the standard GMM estimators. In: Baltagi, B.H. (Eds.), *Nonstationary Panels, Panel Cointegration, and Dynamic Panels*. Advances in Econometrics 15, Amsterdam: JAI Press, Elsevier Science, 53-91.

Bond, S., 2002. Dynamic panel data model: A guide to micro data methods and practice. *Portugese Economic Journal*, 1, 141-162.

Bond, S., Windmeijer, F., 2002. Projection estimators for autoregressive panel data models. *The Econometrics Journal* 5, 457-479.

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

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

Bun, M.J.G, Kiviet, J.F., 2003. On the diminishing returns of higher-order terms in asymptotic expansions of bias. *Economics Letters*, 79, 145-152.

Bun, M.J.G, Kiviet, J.F., 2005. The effects of dynamic feedbacks on LS and MM estimator accuracy in panel data models. Forthcoming in *Journal of Econometrics.*

Davidson, R., MacKinnon, J.G., 1993. *Estimation and Inference in Econometrics.* New York, Oxford University Press.

Doornik, J.A., Arellano, M., Bond, S., 2002. Panel data estimation using DPD for Ox. mimeo, University of Oxford.

Doran, H.E., Schmidt, P., 2005. GMM estimators with improved finite sample properties using principal components of the weighting matrix, with an application to the dynamic panel data model. Forthcoming in *Journal of Econometrics.*

Hansen, G., 2001. A bias-corrected least squares estimator of dynamic panel models. *Allgemeines Statistisches Archiv* 85, 127-140.

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

Harris, M.N., Mátyás, L., 2000. Performance of the operational Wansbeek-Bekker estimator for dynamic panel data models. *Applied Economics Letters* 7, 149-153.

Harris, M.N., Mátyás, L., 2004. A comparative analysis of different IV and GMM estimators of dynamic panel data models. *International Statistical Review* 72, 397-408.

Hendry, D.F., 1984. Monte Carlo experimentation in econometrics. Chapter 16 in: Griliches, Z., Intriligator, M.D. (eds.). *Handbook of Econometrics, Vol. II.* Amsterdam, Elsevier.

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

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

Judson, R.A., Owen, A.L., 1999. Estimating dynamic panel data models: a guide for macroeconomists. *Economics Letters* 65, 9-15.

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

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

Nerlove,M., 1967. Experimental evidence on the estimation of dynamic economic relations from a time-series of cross sections. *Economic Studies Quarterly* 18, 42-74.

Nerlove, M., 1971. Further evidence on the estimation of dynamic economic relations from a time-series of cross sections. *Econometrica* 39, 359-382.

Nickell, S., 1981. Biases in dynamic models with fixed effects. *Econometrica* 49, 1417-1426.

Windmeijer, F., 2000. Efficiency comparisons for a system GMM estimator in dynamic panel data models. In: R.D.H. Heijmans, D.S.G. Pollock and A. Satorra (eds.), Innovations in Multivariate Statistical Analysis, A Festschrift for Heinz Neudecker. Advanced Studies in Theoretical and Applied Econometrics, vol. 36, Dordrecht: Kluwer Academic Publishers (IFS working paper W98/1).

Windmeijer, F., 2005. A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics* 126, 25-51.
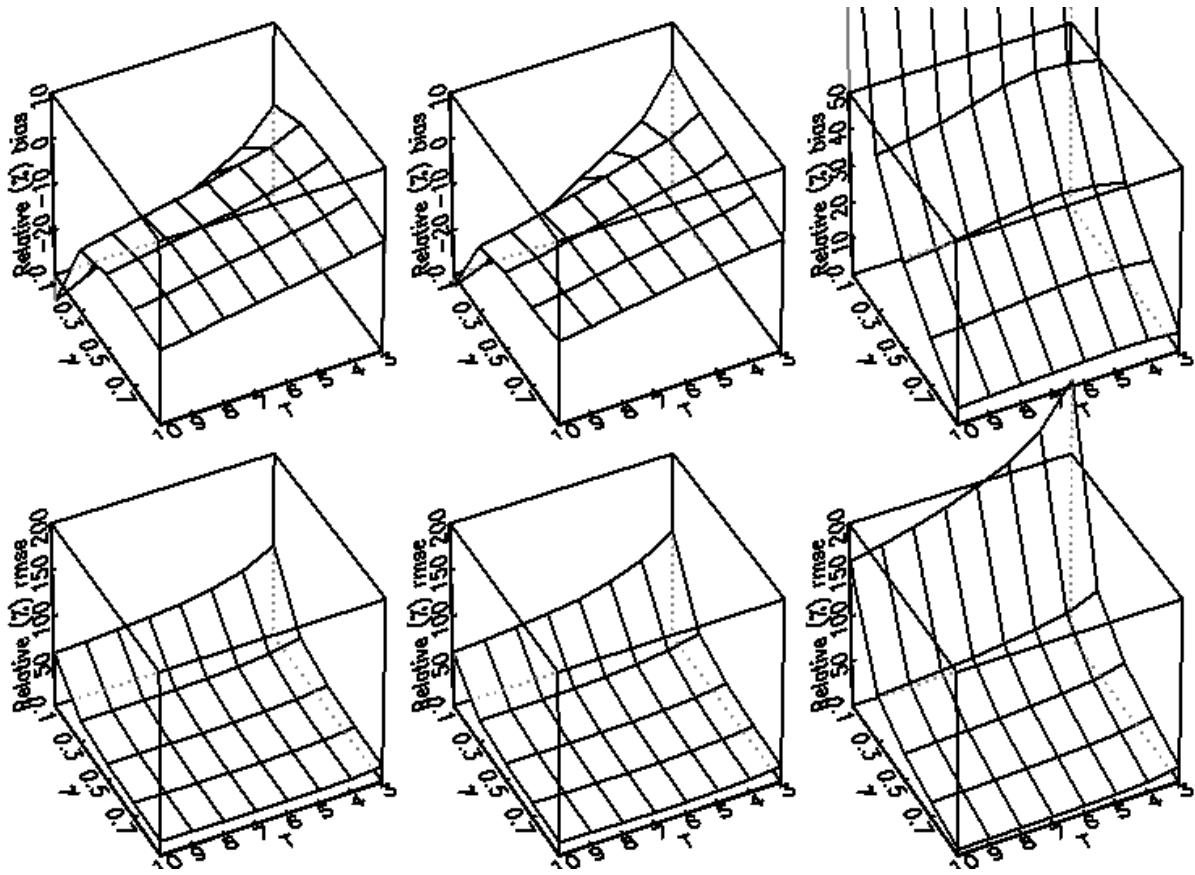
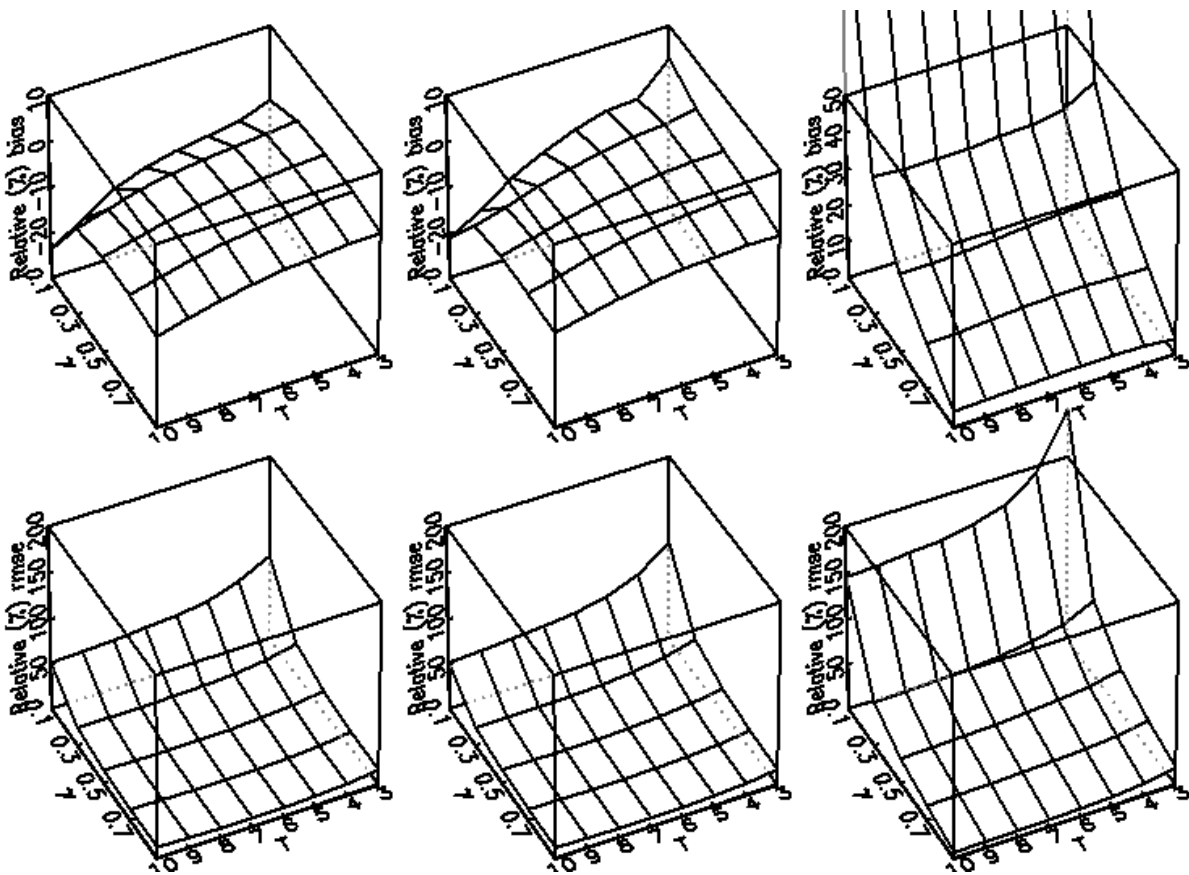Figure 1: GMMs_1^{GIV} at $N = 100$; $\phi = 1$; $\psi = 0, 1, 5$.

Figure 2: GMMs_2^{GIV} at $N = 100$; $\phi = 1$; $\psi = 0, 1, 5$.

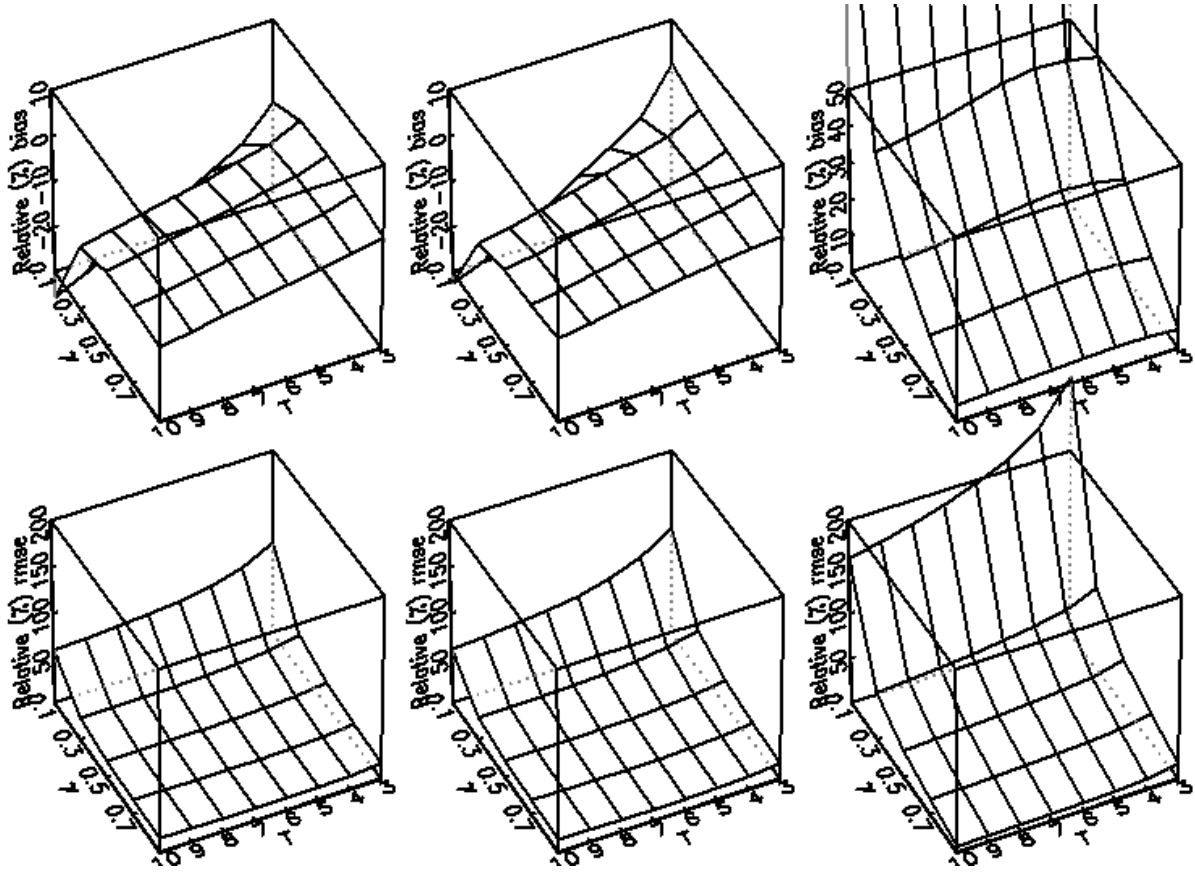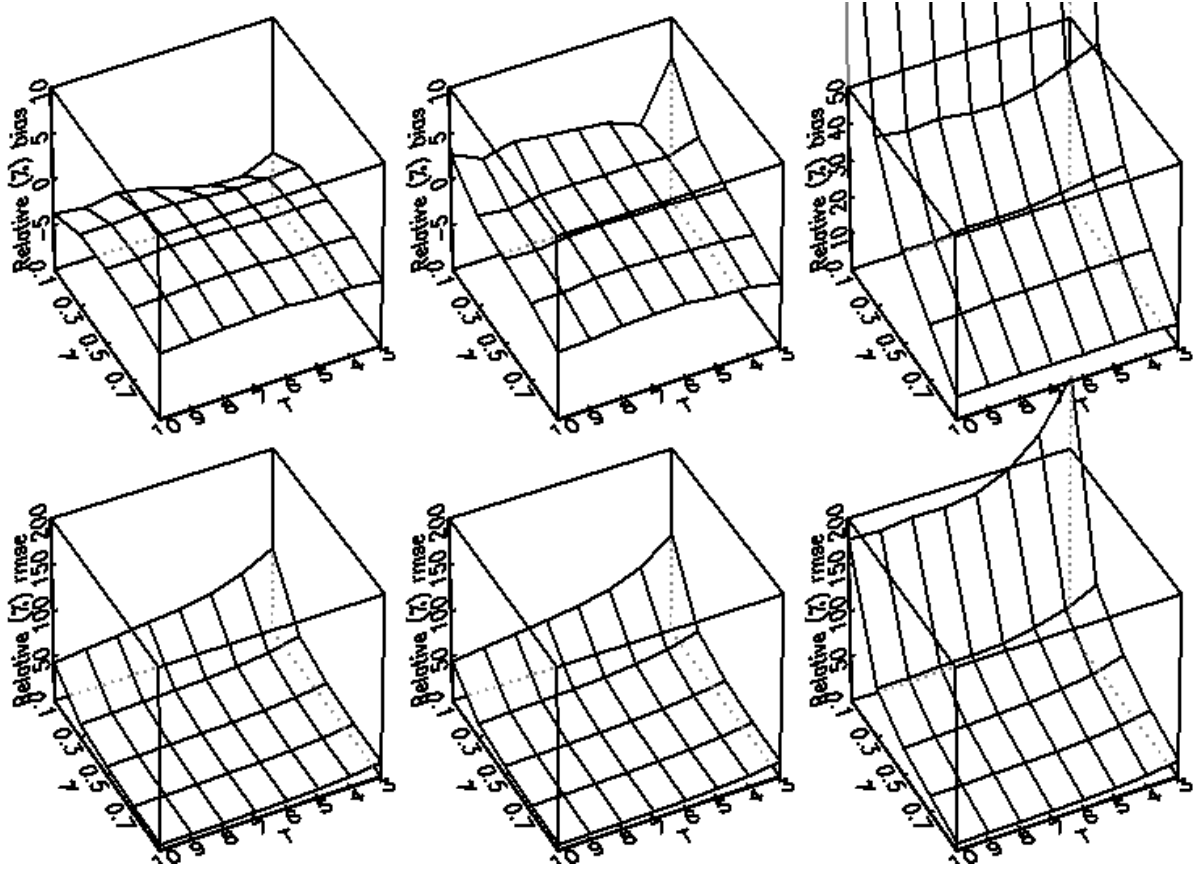Figure 3: GMMs$_1^{DPD}$ at $N = 100$; $\phi = 1$; $\psi = 0, 1, 5$.

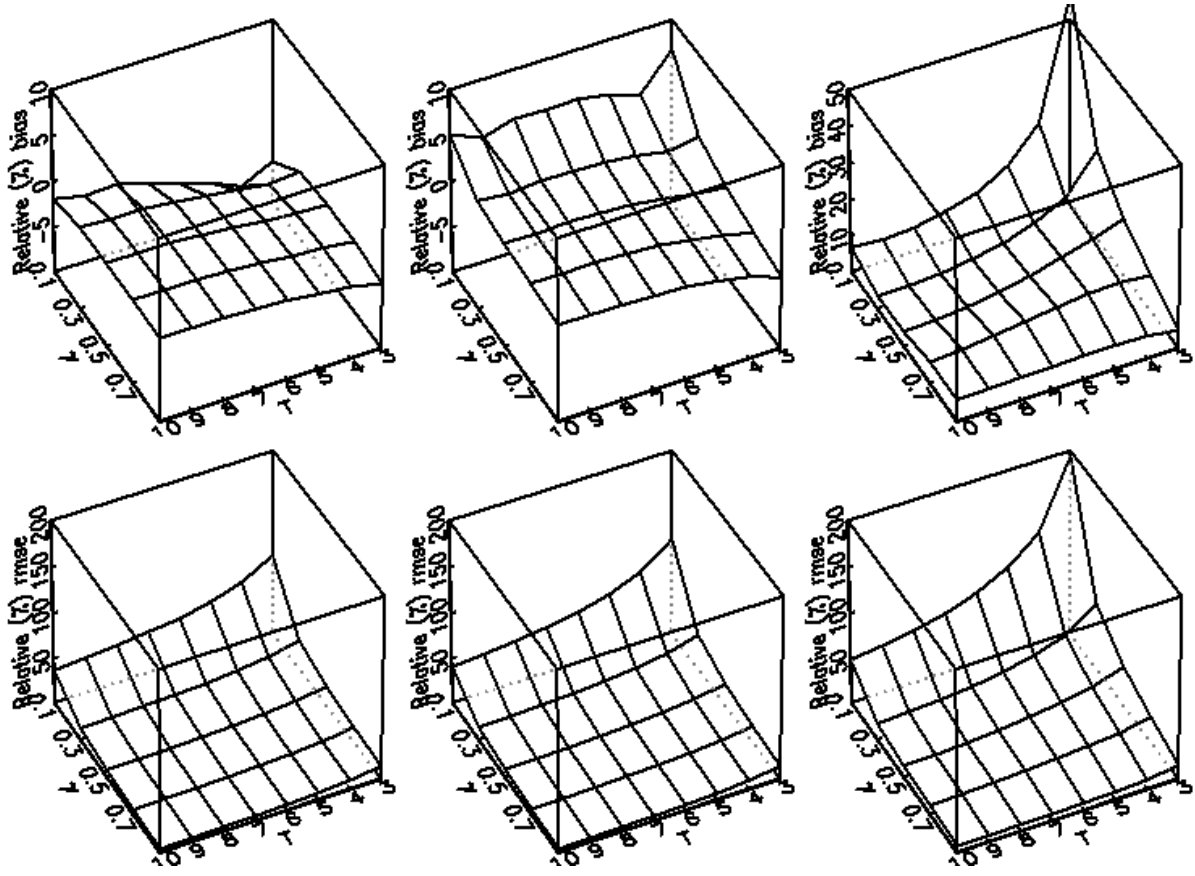Figure 4: GMMs$_2^{DPD}$ at $N = 100$; $\phi = 1$; $\psi = 0, 1, 5$.

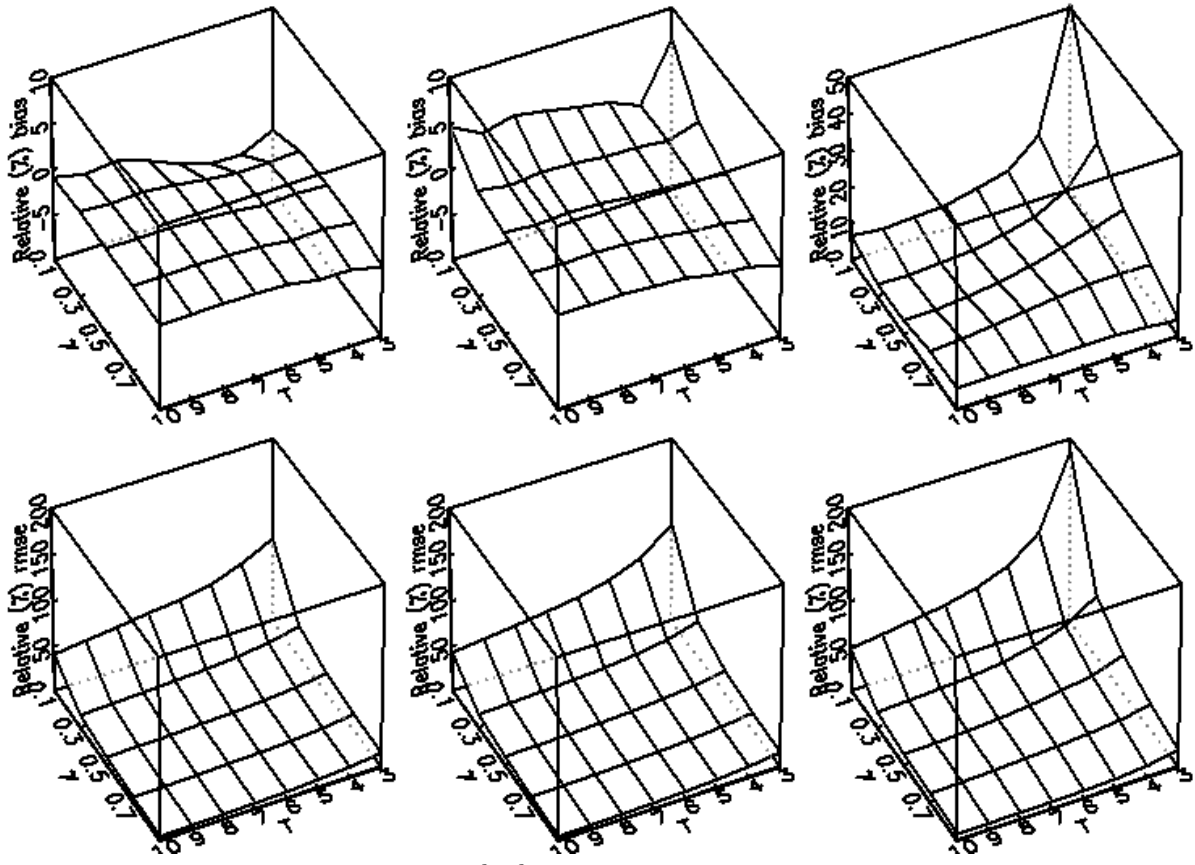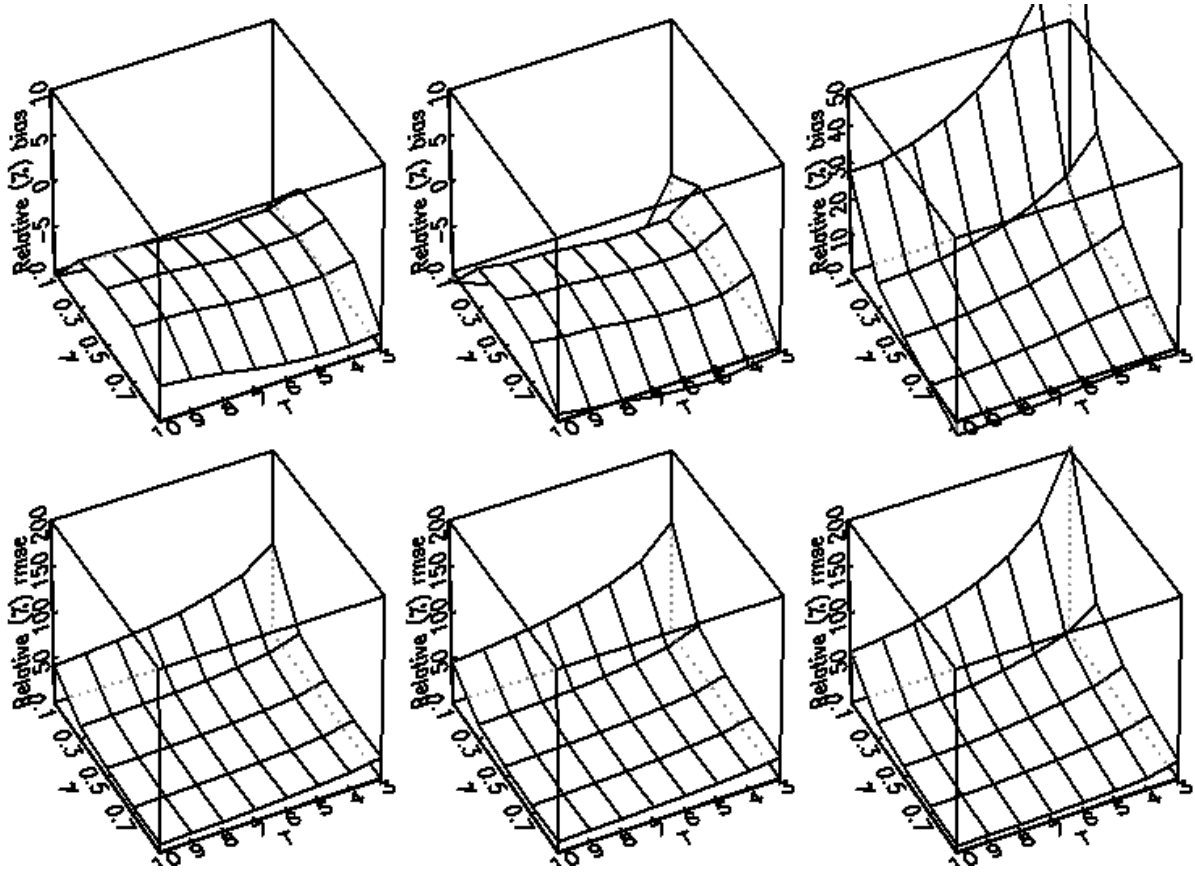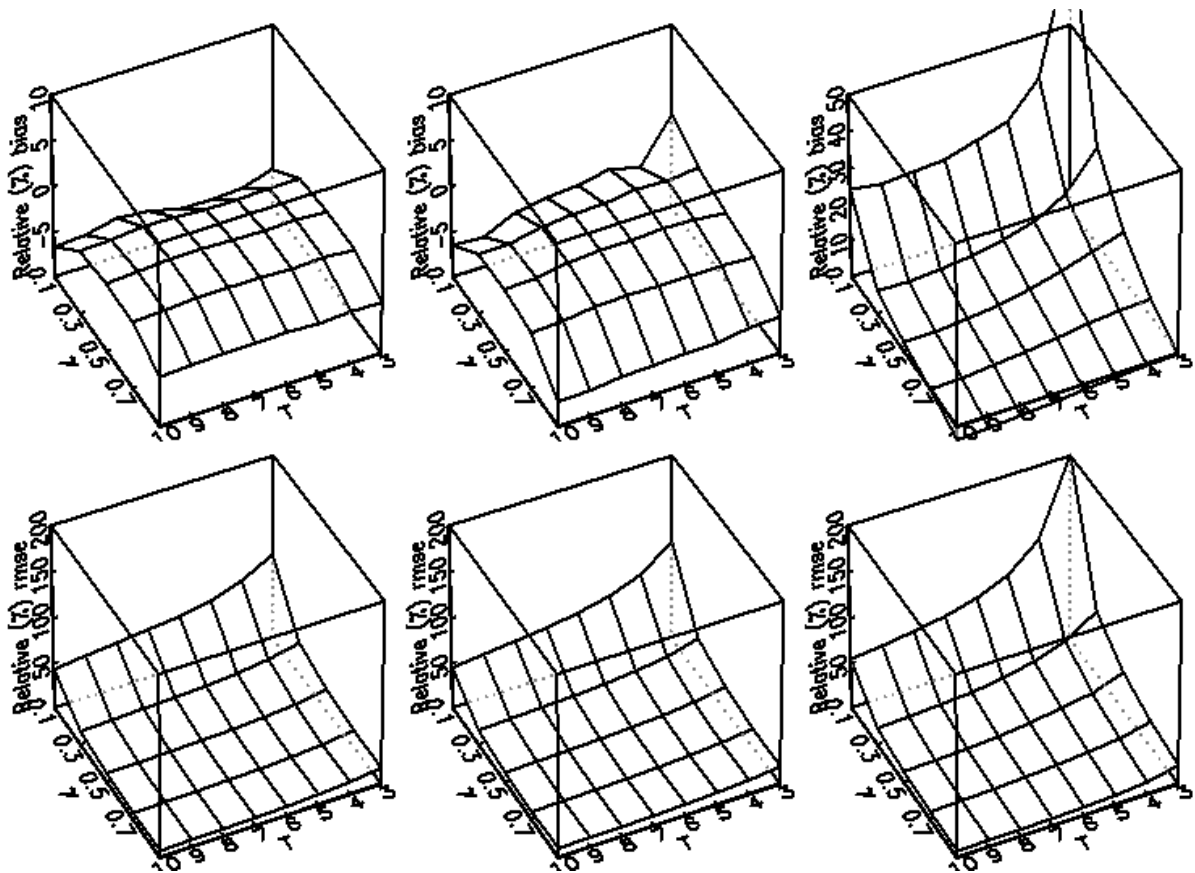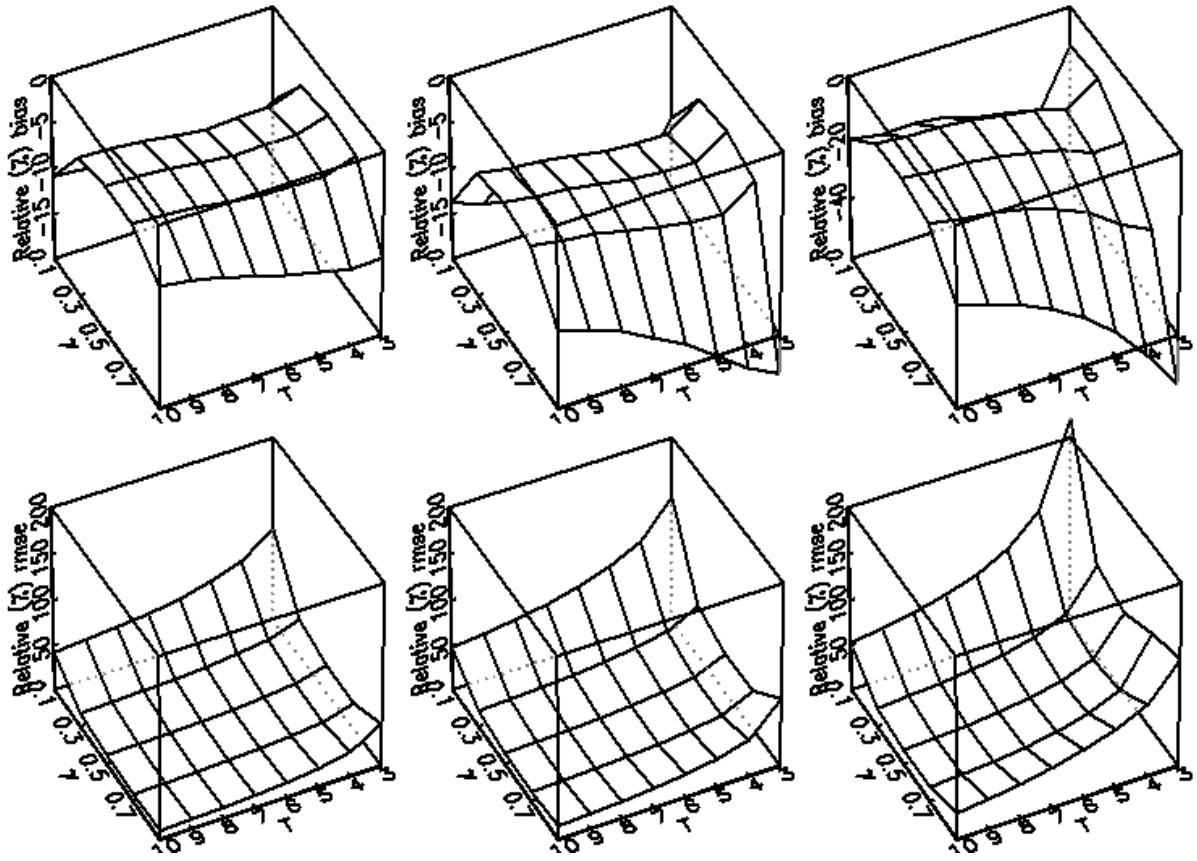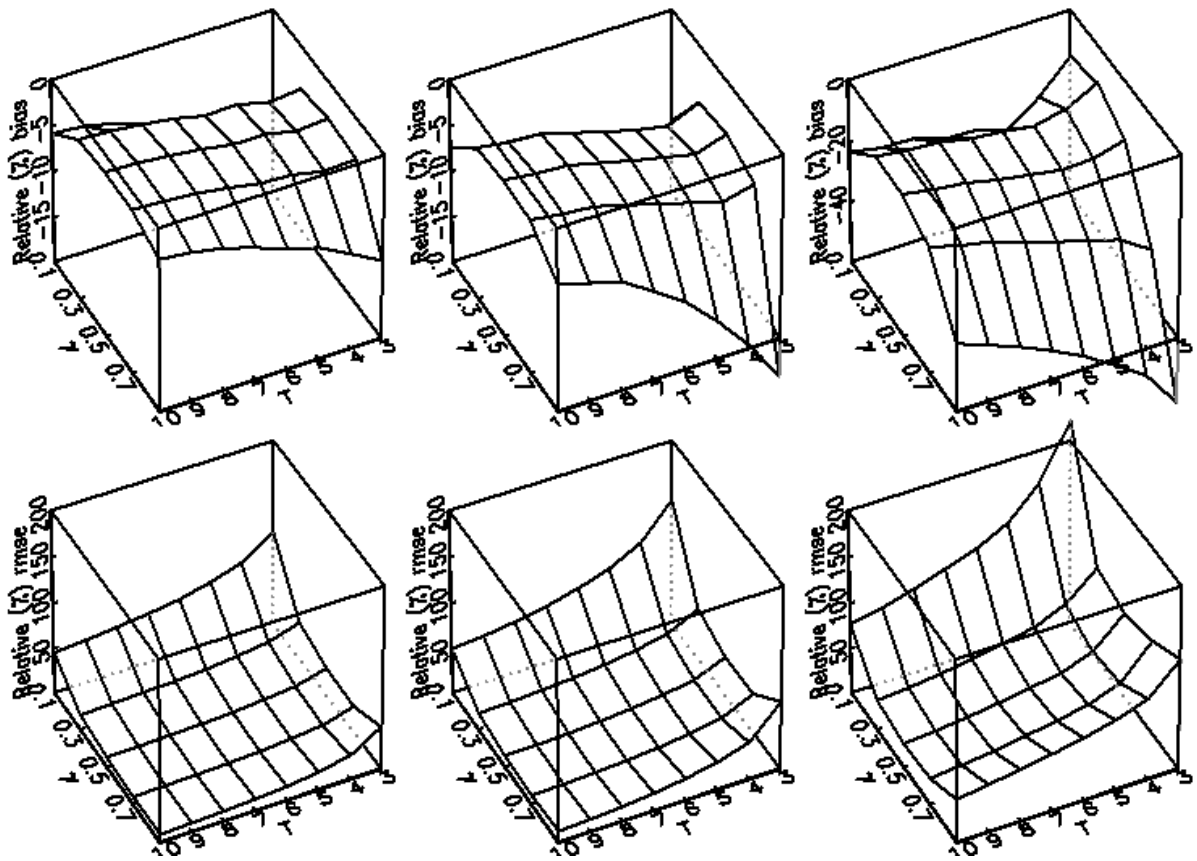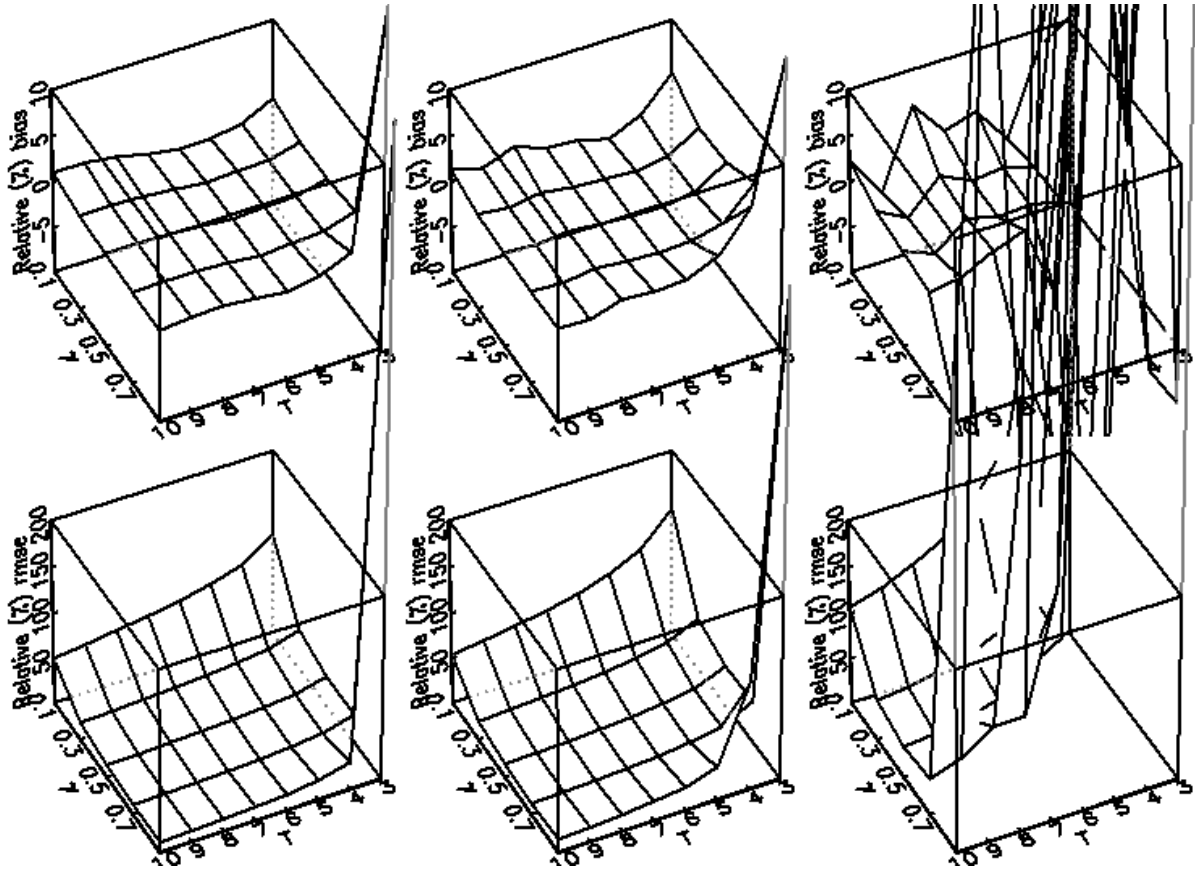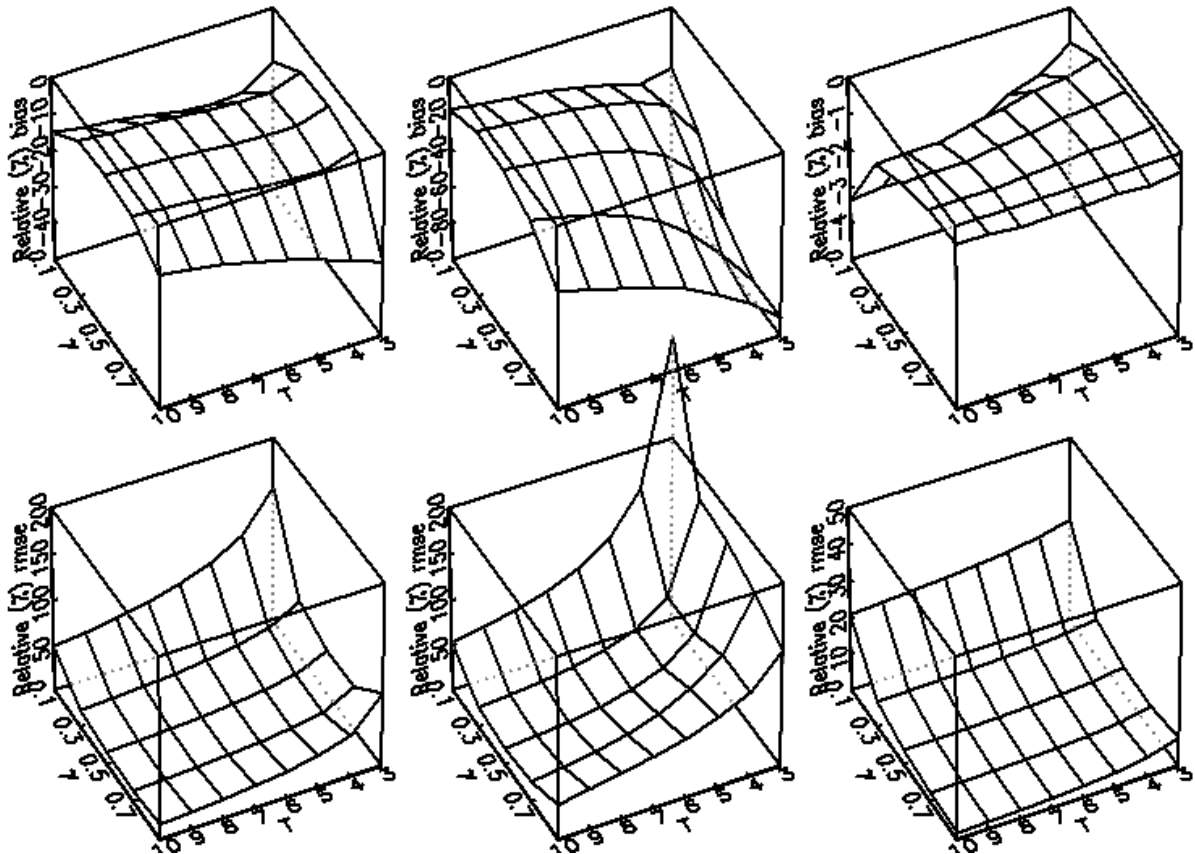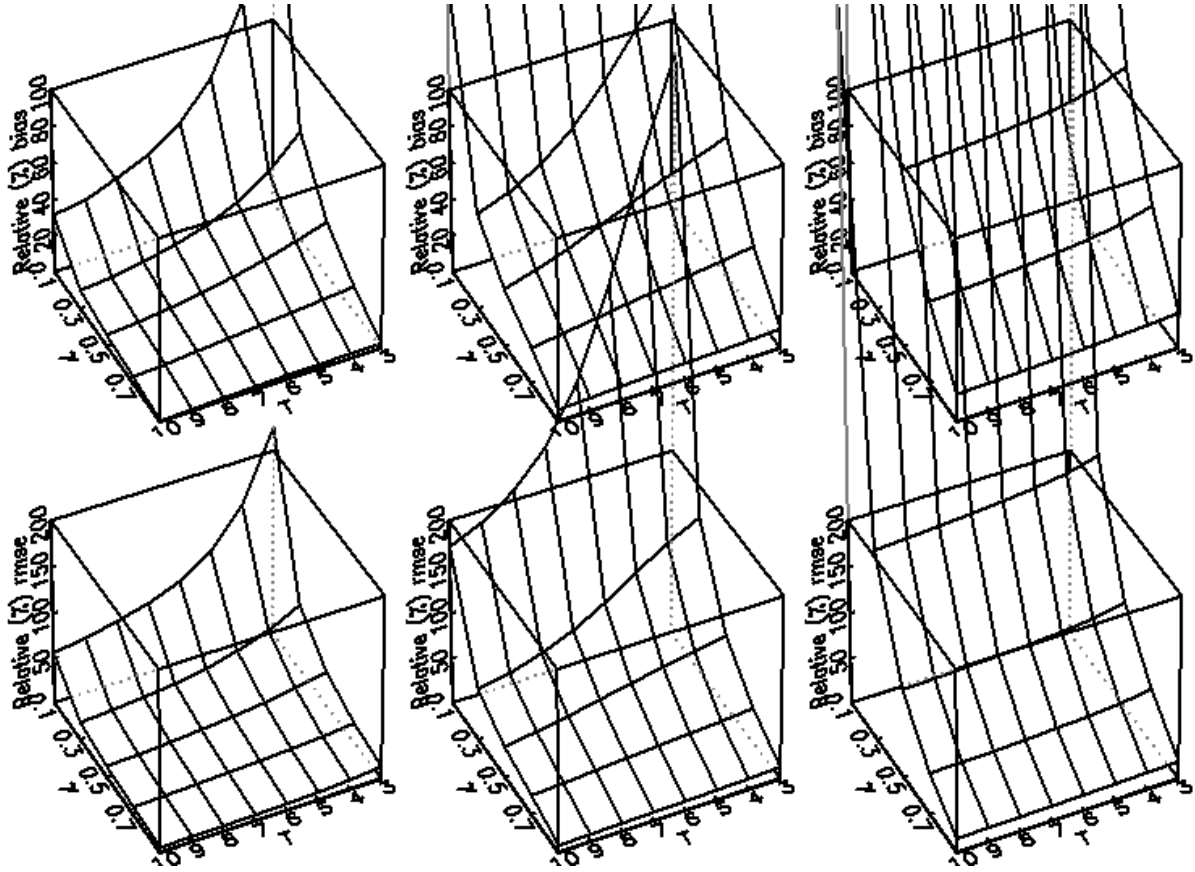Figure 5: $\text{GMMs}_1^{\sigma_\eta^2/\sigma_\varepsilon^2}$ at $N = 100$; $\phi = 1$; $\psi = 0, 1, 5$.

Figure 6: $\text{GMMs}_2^{\sigma_\eta^2/\sigma_\varepsilon^2}$ at $N = 100$; $\phi = 1$; $\psi = 0, 1, 5$.

Figure 7: GMMs_1^{10} at $N = 100$; $\phi = 1$; $\psi = 0, 1, 5$.

Figure 8: GMMs_2^{10} at $N = 100$; $\phi = 1$; $\psi = 0, 1, 5$.

Figure 9: GMM_1^{AB} at $N = 100$; $\phi = 1$; $\psi = 0, 1, 5$.

Figure 10: GMM_1^{BK} at $N = 100$; $\phi = 1$; $\psi = 0, 1, 5$.

Figure 11: MM^{AHl} at $N = 100$; $\phi = 1$; $\psi = 0, 1, 5$.

Figure 12: GMM$_1^{AB}$ at $N = 100$; $\phi = 0.5$; $\psi = 0.5, 1, 5$.

Figure 13: Inconsistent GMMs$_2^{DPD}$ at $N = 100$; $\phi = 0.5$; $\psi = 0.5, 1, 5$.
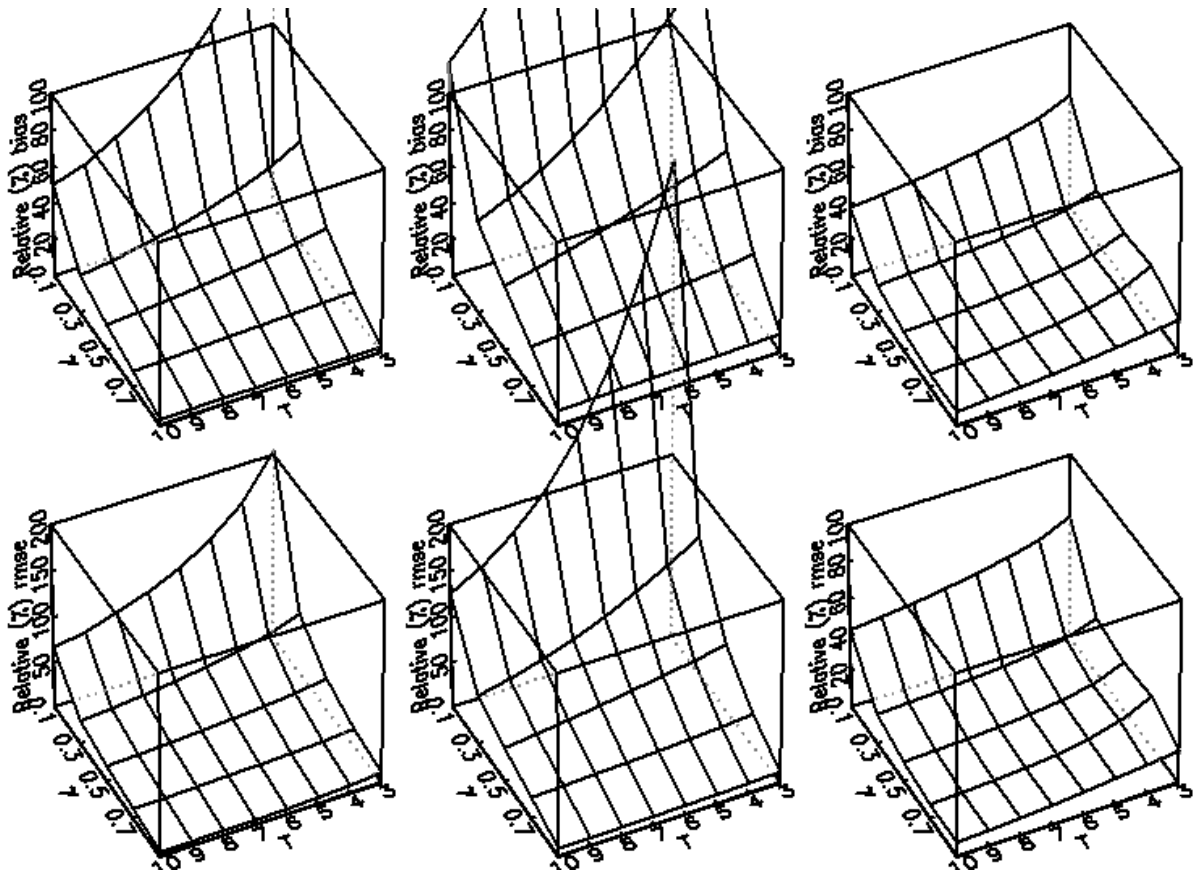
Figure 14: Inconsistent GMMs$_1^{\sigma_\eta^2/\sigma_\varepsilon^2}$ at $N = 100$; $\phi = 0.5$; $\psi = 0.5, 1, 5$.