UvA ECONOMETRICS

Discussion Paper: 2007/08

On the optimal weighting matrix for the GMM system estimator in dynamic panel data models

Jan F. Kiviet

www.fee.uva.nl/ke/UvA-Econometrics

Amsterdam School of Economics
Department of Quantitative Economics
Roetersstraat 11
1018 WB AMSTERDAM
The Netherlands

UvA UNIVERSITEIT VAN AMSTERDAM

On the optimal weighting matrix for the GMM system estimator in dynamic panel data models

JAN F. KIVIET*
(Universiteit van Amsterdam & Tinbergen Institute)

preliminary version: 1 December 2007

JEL-code: C13; C15; C23

Keywords: effect stationarity, generalized method of moments, initial conditions, optimal weighting matrix, two-step estimation

Abstract

An analytical expression is obtained for the optimal weighting matrix for the generalized method of moments system estimator (GMMs) in the dynamic panel data model with predetermined regressors, individual effect stationarity of all variables, homoskedasticity and cross-section independence. As yet such an expression had not been obtained. Therefore GMMs has always been applied by first using some simple form of sub-optimal weighting matrix in 1-step GMMs, after which asymptotically efficient 2-step GMMs can be obtained in the usual way. The optimal weighting matrix is found to depend on unobservable model and data parameters, so asymptotically efficient 1-step GMMs is not operational. However, next to the few suboptimal naive 1-step GMMs weighting matrices presently in use, the expression for the optimal weighting matrix suggests alternative feasible more sophisticated GMMs implementations, including a point-optimal weighting matrix. In Monte Carlo experiments we analyse and compare some of these and find that 2-step GMMs is rather sensitive to the initial choice of weighting matrix in moderately large samples. Some guidelines are given for choosing weights in order to reduce the mean (squared) estimation errors.

* Department of Quantitative Economics, Amsterdam School of Economics, Universiteit van Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands (j.f.kiviet@uva.nl).

1 Introduction

To achieve asymptotic efficiency in GMM (generalized method of moments) estimation when there are more moment restrictions than parameters to be estimated, the sample moment conditions are minimized in a quadratic form that entails a weighting matrix that has a probability limit which is proportional to the inverse of the variance of the limiting distribution of the orthogonality condition vector. In many circumstances this variance is hard to derive and it may depend on nuisance parameters. Dynamic panel data models with unobserved individual effects are often estimated by GMM, and when the cross-sections are independent and homoskedastic the optimal weighting matrix for the linear moment conditions is well-known and not determined by nuisance parameters, see Arellano and Bond (1991). However, it has been observed that this estimator may suffer from serious bias in moderately large samples and it has been suggested that exploiting additional moment conditions that build on stationarity assumptions in a so-called GMM system estimator (GMMs) yields much better results. This estimator, put forward by Arellano and Bover (1995) and Blundell and Bond (1998), is called a system estimator because it simultaneously exploits moment conditions for the dynamic panel data model in levels and in its first-differenced form. Up to date the optimal weighting matrix for this estimator had not been derived so that even under very strict assumptions GMMs has to be employed in a 2-step procedure in which an empirical assessment of the optimal weights is used that is based on consistent 1-step GMMs estimates that are obtained from an operational but suboptimal weighting matrix.

In this paper we derive the optimal GMMs weighting matrix for the dynamic panel data model with both a lagged dependent variable regressor and an arbitrary number of further predetermined regressors, where both the dependent variable and the regression variables are effect stationary. This means that their correlation with the unobserved heterogeneity (the individual effects) is time-invariant. For many elements of the GMMs orthogonality condition vector it is extremely difficult to derive their covariance, which complicates the assessment of the optimal weighting matrix. However, we show that asymptotically equivalent results are obtained by deriving a conditional covariance, and that different conditioning sets may be used for the various elements of the variance matrix. By choosing for all the various types of elements in the variance matrix of the orthogonality conditions a conditioning set that leads to rather straightforward expressions for their conditional variance we are able to obtain an explicit expression for the asymptotically optimal weighting matrix. This matrix is shown to depend on all the observed instrumental variables and also on the unknown values of the covariance between the regressors and the individual effects, as well as on the unknown covariance between current regressors and lagged disturbances, i.e. on the pattern in the feedback mechanism of current disturbances into future observations of the regressors.

Due to the dependence on nuisance parameters asymptotically efficient 1-step GMMs is unfeasible, although a better founded initial choice of weighting matrix is possible now. At least a choice better than those currently in use, which are based either on the usually wrong assumption that the individual effects have zero variance, or on taking as weighting matrix the inverse of the sample moment matrix of the instrumental variables, i.e. wrongly assuming that the disturbances of the system are i.i.d. and therefore applying GIV (generalized instrumental variables), which in this model is clearly far from optimal. An alternative approach regarding improving the quality of weighting matrices can be found in Doran and Schmidt (2005).

In Monte Carlo experiments [in the present version of the paper we just simulated the autoregressive model without further predetermined regressors] we compare the results obtained by employing the initial weighting matrices presently in use, and by exploiting some new operational forms inspired by the optimal matrix. We find that 2-step GMMs is rather sensitive to the initial choice of weighting matrix in moderately large samples. As a yardstick we also present results obtained by the in practice unfeasible 1-step GMMs estimator that uses (or is inspired by) the optimal weights. Some guidelines, including a point-optimal version of the weighting matrix, are given for choosing the weights in order to reduce the mean (squared) estimation errors.

The paper is structured as follows. In Section 2 we introduce our notation, state the model assumptions, review some general GMM results for panels, present the GMMs estimator and the various forms of suboptimal weighting matrix that have been suggested in the literature. Section 3 contains our main analytical results regarding the derivation of the optimal weighting matrix. Next, in Section 4, we present the Monte Carlo design that we used and describe our simulation findings. Finally, Section 5 concludes.

2 Standard and system GMM for dynamic panels

In this section we introduce the linear first-order dynamic panel data model with random individual effects and an arbitrary number of further regressors. All regressors are assumed to be predetermined with respect to the homoskedastic and serially and contemporaneously uncorrelated idiosyncratic disturbance terms. From the various model assumptions the linear orthogonality conditions are derived that are exploited in what we will address as the standard GMM estimator, see Arellano and Bond (1991). We pay extra attention to an additional model assumption that we will call effect stationarity. This, when valid, gives rise to additional linear orthogonality conditions, which can be exploited in the system GMM estimator (GMMs), see Blundell and Bond (1998), who considered this estimator in the pure autoregressive case. Generic results are obtained for 1-step and 2-step estimation that cover both the standard and the system GMM panel data situation. This provides the setting from which we can embark on the derivation of the optimal weighting matrix for GMMs.

2.1 Model assumptions

We consider the linear dynamic panel data model

$$y_{it} = \gamma y_{i,t-1} + x_{it}'\beta + \eta_i + \varepsilon_{it}, \tag{1}$$

where x_{it} and β are $K \times 1$ vectors and $i = 1, ..., N$ refers to the cross-sections and $t = 1, ..., T$ to the time-periods. We suppose that we have a balanced panel data set; hence, if x_{it} happens to contain also the first lag of all current explanatory variables then both the dependent and the other individual explanatory variables should have been observed for the time-indices $\{0, ..., T\}$ for all cross-section units. Below we shall use the notation

$$\left.\begin{array}{l} y_i^{t-1} \equiv (y_{i,0}, ..., y_{i,t-1}) \\ X_i^t \equiv (x_{i1}', ..., x_{it}') \end{array}\right\} \; t = 1, ..., T,$$

where y_i^{t-1} is $1 \times t$ and X_i^t is $1 \times tK$. We also define

$$Y^{t-1} \equiv \begin{bmatrix} y_1^{t-1} \\ \vdots \\ y_N^{t-1} \end{bmatrix},\ X^t \equiv \begin{bmatrix} X_1^t \\ \vdots \\ X_N^t \end{bmatrix},\ \eta \equiv \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_N \end{pmatrix},$$

where η is $N \times 1$, Y^{t-1} is $N \times t$ and X^t is $N \times tK$.

Regarding the two random error components η_i and ε_{it} in model (1) we make the assumptions $(i, j = 1, ..., N;\ t, s = 1, ..., T)$:

$$\left.\begin{array}{l} E(\varepsilon_{it} \mid Y^{t-1}, X^t, \eta) = 0,\ \forall i, t \\ E(\varepsilon_{it}^2 \mid Y^{t-1}, X^t, \eta) = \sigma_\varepsilon^2 > 0,\ \forall i, t \\ E(\varepsilon_{it}\varepsilon_{js} \mid Y^{t-1}, X^t, \eta) = 0,\ \forall t < s,\ \forall i, j \text{ and } \forall i \neq j,\ \forall t \leq s \\ E(\eta_i) = 0,\ E(\eta_i^2) = \sigma_\eta^2 \geq 0,\ E(\eta_i\eta_j) = 0\ \forall i \neq j \end{array}\right\} \tag{2}$$

These assumptions entail that all regressors are predetermined with respect to the idiosyncratic disturbances ε_{it}, which are homoskedastic and serially and contemporaneously uncorrelated. In fact, we assume that all cross-section units are independent.

Additional assumptions that may be made — and which are crucial when the GMMs estimator is employed — involve

$$E(y_{it}\eta_i) = \sigma_{y\eta} \text{ and } E(x_{it}\eta_i) = \sigma_{x\eta},\ \forall i, t. \tag{3}$$

Here $\sigma_{y\eta}$ is a scalar and $\sigma_{x\eta}$ is a $K \times 1$ vector. Note that both $\sigma_{y\eta}$ and $\sigma_{x\eta}$ are assumed to be time-invariant. By multiplying the model equation (1) by η_i and taking expectations we find that (3) implies

$$\sigma_{y\eta} = \gamma\sigma_{y\eta} + \sigma_{x\eta}'\beta + \sigma_\eta^2$$

or

$$\sigma_{y\eta} = \frac{\sigma_{x\eta}'\beta + \sigma_\eta^2}{1 - \gamma}. \tag{4}$$

This condition we will call for obvious reasons *effect-stationarity*.

2.2 Moment conditions

By taking first-differences the model simplifies in the sense that only one unobservable error component remains. Estimating

$$\Delta y_{it} = \gamma\Delta y_{i,t-1} + (\Delta x_{it})'\beta + \Delta\varepsilon_{it} \tag{5}$$

by OLS would yield inconsistent estimators[1] because

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

Unless σ_ε^2 would be known this moment condition cannot directly be exploited in a method of moments estimator. Note, however, that it easily follows from the model assumptions (2) that for $i = 1, ..., N$ we have

$$\left.\begin{array}{l} E(Y^{t-2}\Delta\varepsilon_{it}) = O \\ E(X^{t-1}\Delta\varepsilon_{it}) = O \end{array}\right\} \qquad t = 2, ..., T,$$

[1]Note that in the (habitual) case where x_{it} contains a unit element and β an intercept this model has no longer $K + 1$ unknown coefficients. In the analysis that follows we neglect that situation, but it is reasonably straigtforward to adapt the results (mainly the dimensions of matrices) for models with an intercept.

which together provide $(K+1)NT(T-1)/2$ moment conditions[2] for estimating just $K+1$ coefficients. Especially when the cross-section units are independent many of these conditions will produce weak or completely ineffective instruments. Then it will be more appropriate to exploit just the $(K+1)T(T-1)/2$ moment conditions

$$\left.\begin{array}{r} E(y_i^{t-2}\Delta\varepsilon_{it}) = 0' \\ E(X_i^{t-1}\Delta\varepsilon_{it}) = 0' \end{array}\right\} \qquad t = 2, ..., T,$$

as is done in the Arellano-Bond (1991) estimator. Upon substituting (5) for $\Delta\varepsilon_{it}$ it is obvious that these moment conditions are linear in the unknown coefficients γ and β, i.e.

$$\left.\begin{array}{r} E\{y_i^{t-2}[\Delta y_{it} - \gamma\Delta y_{i,t-1} - (\Delta x_{it})'\beta]\} = 0' \\ E\{X_i^{t-1}[\Delta y_{it} - \gamma\Delta y_{i,t-1} - (\Delta x_{it})'\beta]\} = 0' \end{array}\right\} \qquad t = 2, ..., T. \tag{6}$$

Blundell and Bond (1998) argue that assumption (3), when valid, may yield relatively strong additional useful instruments for estimating the undifferenced equation (1). These additional instruments are the first-differenced variables. Defining

$$\left.\begin{array}{r} \Delta y_i^{t-1} \equiv (\Delta y_{i1}, ..., \Delta y_{i,t-1}) \\ \Delta X_i^{t} \equiv (\Delta x'_{i2}, ..., \Delta' x_{i,t}) \end{array}\right\} \; t = 2, ..., T,$$

which are $1 \times (t-1)$ and $1 \times (t-1)K$ respectively, it follows from (3) that (for $i = 1, ..., N$)

$$E(\Delta y_i^{t-1}\eta_i) = 0' \text{ and } E(\Delta X_i^{t}\eta_i) = 0'.$$

From (2) we find

$$E(\Delta y_i^{t-1}\varepsilon_{it}) = 0' \text{ and } E(\Delta X_i^{t}\varepsilon_{it}) = 0'.$$

Combining these and substituting $\eta_i + \varepsilon_{it} = y_{it} - \gamma y_{i,t-1} - x'_{it}\beta$ yields the $(K+1)T(T-1)/2$ linear moment conditions

$$\left.\begin{array}{r} E[\Delta y_i^{t-1}(y_{it} - \gamma y_{i,t-1} - x'_{it}\beta)] = 0' \\ E[\Delta X_i^{t}(y_{it} - \gamma y_{i,t-1} - x'_{it}\beta)] = 0' \end{array}\right\} \qquad t = 2, ..., T. \tag{7}$$

These can be transformed linearly into two subsets of $(K+1)(T-1)$ and $(K+1)(T-1)(T-2)/2$ conditions respectively, viz.

$$\left.\begin{array}{r} E[\Delta y_{i,t-1}(y_{it} - \gamma y_{i,t-1} - x'_{it}\beta)] = 0' \\ E[\Delta x'_{it}(y_{it} - \gamma y_{i,t-1} - x'_{it}\beta)] = 0' \end{array}\right\} \qquad t = 2, ..., T, \tag{8}$$

and

$$\left.\begin{array}{r} E\{\Delta y_i^{t-1}[\Delta y_{it} - \gamma\Delta y_{i,t-1} - (\Delta x_{it})'\beta]\} = 0' \\ E\{\Delta X_i^{t}[\Delta y_{it} - \gamma\Delta y_{i,t-1} - (\Delta x_{it})'\beta]\} = 0' \end{array}\right\} \qquad t = 3, ..., T. \tag{9}$$

The second subset (9), though, can also be obtained by a simple linear transformation of (6). Hence, effect stationarity only leads to the $(K+1)(T-1)$ additional linear moment conditions (8). These involve estimation of the undifferenced model (1) by employing all the first-differenced regressor variables as instruments.

Due to the i.i.d. assumption regarding ε_{it} further (non-linear) moment conditions do hold in the present dynamic panel data model, see Ahn and Schmidt (1995, 1997), but below we will stick to the linear conditions mentioned above.

[2] There would in fact be fewer unique moment conditions if the model includes an intercept, indicator or dymmy variables or particular regressors in combination with their lags. In what follows we will not take this into account, so that in actual applications of our results adaptations will have to be made.

2.3 Generic results for 1-step and 2-step panel GMM

Both the standard GMM and the system estimator GMMs fit into the following simple generic setup. After appropriate manipulation (transformation and stacking) of the panel data observations one has

$$y_i^* = X_i^* \beta^* + \varepsilon_i^*, \; i = 1, ..., N, \tag{10}$$

where y_i^* and ε_i^* are $T^* \times 1$ vectors and the $T^* \times K^*$ matrix X_i^* contains all regressor variables (including transformations of the lagged-dependent variable). The $K^* \times 1$ vector β^* of coefficients is estimated by employing $L^* \geq K^*$ moment conditions that hold for all $i = 1, ..., N$, viz.

$$E[Z_i^{*\prime}(y_i^* - X_i^* \beta^*)] = 0, \tag{11}$$

where Z_i^* is $T^* \times L^*$. The GMM estimator using the $L^* \times L^*$ semi-positive definite weighting matrix W^* is obtained by minimizing a quadratic form, viz.

$$\hat{\beta}^*_{W^*} = \arg\min_{\beta^*} \left(\sum_{i=1}^N Z_i^{*\prime}(y_i^* - X_i^* \beta^*) \right)^{\prime} W^* \left(\sum_{i=1}^N Z_i^{*\prime}(y_i^* - X_i^* \beta^*) \right). \tag{12}$$

Assuming that $X^{*\prime} Z^* W^* Z^{*\prime} X^*$ has rank K^* with probability 1 a unique minimum exists which yields

$$\hat{\beta}^*_{W^*} = (X^{*\prime} Z^* W^* Z^{*\prime} X^*)^{-1} X^{*\prime} Z^* W^* Z^{*\prime} y^*, \tag{13}$$

where $y^* = (y_1^{*\prime}, ..., y_N^{*\prime})^\prime$, $X^* = (X_1^{*\prime}, ..., X_N^{*\prime})^\prime$ and $Z^* = (Z_1^{*\prime}, ..., Z_N^{*\prime})^\prime$. It is obvious that $\hat{\beta}^*_{W^*}$ is invariant for the scale of W^*. Below, when useful, we will implicitly assume that scaling took place such that $W^* = O_p(1)$.

We only consider cases where convenient regularity conditions hold, including the existence with full column rank (for $N \to \infty$) of $\text{plim}\, N^{-1} Z^{*\prime} X^*$, $\text{plim}\, N^{-1} Z^{*\prime} Z^*$ and $\text{plim}\, N^{-2} X^{*\prime} Z^* W^* Z^{*\prime} X^*$. According to (11) the instruments are valid, thus

$$\underset{N\to\infty}{\text{plim}} \frac{1}{N} \textstyle\sum_{i=1}^N Z_i^{*\prime} \varepsilon_i^* = 0$$

and $\hat{\beta}^*_{W^*}$ is consistent. Since the cross-section units are assumed to be independent the CLT (central limit theorem) yields

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N Z_i^{*\prime} \varepsilon_i^* \to \text{N}\left(0, V_{Z^{*\prime}\varepsilon^*}\right), \text{ where } V_{Z^{*\prime}\varepsilon^*} \equiv \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \text{Var}(Z_i^{*\prime} \varepsilon_i^*), \tag{14}$$

from which we obtain

$$\begin{aligned} \sqrt{N}(\hat{\beta}^*_{W^*} - \beta_0^*) &\to \text{N}\left(0, V_{\hat{\beta}^*_{W^*}}\right), \text{ with} \\ V_{\hat{\beta}^*_{W^*}} &\equiv A^* V_{Z^{*\prime}\varepsilon^*} A^{*\prime}, \; A^* \equiv \text{plim}[N(X^{*\prime} Z^* W^* Z^{*\prime} X^*)^{-1} X^{*\prime} Z^* W^*]. \end{aligned} \tag{15}$$

The asymptotically efficient GMM estimator in the class of estimators exploiting instruments Z^* is obtained if W^* is chosen such[3] that, after appropriate scaling, it has probability limit proportional to the inverse of the variance matrix of the limiting distribution of $N^{-1/2} \sum_{i=1}^N Z_i^{*\prime} \varepsilon_i^*$.Hence, optimality is attained by using a weighting matrix W^{opt} such that

$$W^{opt} \propto \left(\sum_{i=1}^N \text{Var}(Z_i^{*\prime} \varepsilon_i^*) \right)^{-1}. \tag{16}$$

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

When W^{opt} is known the asymptotically efficient GMM estimator is self-evidently given by

$$\hat{\beta}^*_{W^{opt}} = (X^{*\prime}Z^*W^{opt}Z^{*\prime}X^*)^{-1}X^{*\prime}Z^*W^{opt}Z^{*\prime}y^*,$$

and

$$\sqrt{N}(\hat{\beta}^*_{W^{opt}} - \beta^*_0) \rightarrow \mathrm{N}\left(0, V_{\hat{\beta}^*_{W^{opt*}}}\right), \text{ with } V_{\hat{\beta}^*_{W^{opt*}}} \equiv [\operatorname{plim} N^{-2}(X^{*\prime}Z^*V^{-1}_{Z^{*\prime}\varepsilon^*}Z^{*\prime}X^*)]^{-1}. \tag{17}$$

For the special case $\varepsilon^*_{it} \mid Z^{*t}_i \sim$ i.i.d.$(0, \sigma^2_{\varepsilon^*})$, where Z^{*t}_i contains the first t rows of Z^*_i, one finds $\mathrm{Var}(Z^{*\prime}_i\varepsilon^*_i) = \sigma^2_{\varepsilon^*}E(Z^{*\prime}_iZ^*_i)$ and hence asymptotic efficiency is achieved by taking $W^{opt}_{idd} = (Z^{*\prime}Z^*)^{-1}$, which yields the familiar 2SLS or GIV result

$$\hat{\beta}^*_{GIV} = [X^{*\prime}P_{Z^*}X^*]^{-1}X^{*\prime}P_{Z^*}y^*, \tag{18}$$

where $P_{Z^*} \equiv Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime}$. In case $K^* = L^*$ this specializes to the simple instrumental variable estimator

$$\hat{\beta}^*_{IV} = (Z^{*\prime}X^*)^{-1}Z^{*\prime}y^*. \tag{19}$$

If matrix W^{opt} is not directly available then one may use some arbitrary initial weighting matrix W^* that produces a consistent (though inefficient) 1-step GMM estimator $\hat{\beta}^*_{W^*}$, and then exploit the 1-step residuals $\hat{\varepsilon}^*_i$ to construct the empirical weighting matrix $\widehat{W}^*$, where

$$\begin{aligned} \hat{\varepsilon}^*_i &= y^*_i - X^*_i\hat{\beta}^*_{W^*}, \\ \widehat{W}^* &= (\textstyle\sum_{i=1}^N Z^{*\prime}_i\hat{\varepsilon}^*_i\hat{\varepsilon}^{*\prime}_iZ^*_i)^{-1}, \end{aligned}$$

yielding the 2-step GMM estimator

$$\hat{\beta}^*_{\widehat{W}^*} = (X^{*\prime}Z^*\widehat{W}^*Z^{*\prime}X^*)^{-1}X^{*\prime}Z^*\widehat{W}^*Z^{*\prime}y^*. \tag{20}$$

Estimator $\hat{\beta}^*_{\widehat{W}^*}$ is asymptotically equivalent to $\hat{\beta}^*_{W^{opt}}$, and hence it is efficient in the class of estimators exploiting instruments Z^*. For inference purposes one uses the approximation

$$\hat{\beta}^*_{\widehat{W}^*} \overset{a}{\sim} \mathrm{N}\left(\beta^*, (X^{*\prime}Z^*\widehat{W}^*Z^{*\prime}X^*)^{-1}\right). \tag{21}$$

Note that, provided the moment conditions are valid, $\hat{\beta}^*_{GIV}$ is consistent thus could be employed as a 1-step GMM estimator. When $K^* = L^*$ using a weighting matrix for the orthogonality conditions is redundant, because the criterion function (12) will be zero for any W^*, as all moment conditions can then be imposed on the sample observations.

2.4 Implementation of ordinary and system GMM

In case of the Arellano-Bond ordinary GMM estimator the above generic setup involves

$$y^*_i = \begin{pmatrix} \Delta y_{i2} \\ \vdots \\ \Delta y_{iT} \end{pmatrix}, \ \varepsilon^*_i = \begin{pmatrix} \Delta \varepsilon_{i2} \\ \vdots \\ \Delta \varepsilon_{iT} \end{pmatrix} \text{ and } X^*_i = \begin{bmatrix} \Delta y_{i1} & \Delta x'_{i2} \\ \vdots & \vdots \\ \Delta y_{i,T-1} & \Delta x'_{iT} \end{bmatrix}, \tag{22}$$

hence $T^* = T - 1$. When there is an intercept in β and corresponding unit value in x_{it} (which vanishes after first-differencing) we have $K^* = K$, otherwise $K^* = K + 1$ and

$\beta^* = (\gamma, \beta')'$. From (6) it follows that $L^* = (K+1)T(T-1)/2$ and the $T^* \times L^*$ matrix Z_i^* is taken to be

$$Z_i^{AB} = \begin{bmatrix} y_i^0 & 0' & 0' & X_i^1 & 0' & 0' \\ 0 & \ddots & O & O & \ddots & O \\ 0 & 0' & y_i^{T-2} & 0' & 0' & X_i^{T-1} \end{bmatrix}. \tag{23}$$

Since $E(\Delta\varepsilon_{it}\Delta\varepsilon_{i,t-1}) \neq 0$ the ε_{it}^* are not i.i.d., thus the GIV estimator is not efficient. It can be derived that for the ordinary GMM estimator

$$W_{AB}^{opt} \propto \left(\sum_{i=1}^{N} (Z_i^{AB})' H Z_i^{AB} \right)^{-1} \tag{24}$$

with $(T-1) \times (T-1)$ matrix

$$H \equiv \begin{bmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & -1 & 2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{bmatrix}. \tag{25}$$

So, under our strong assumptions regarding homoskedasticity and independence of the idiosyncratic disturbances ε_{it} 1-step ordinary GMM employing (24) is asymptotically efficient within the class of estimators exploiting the instruments Z_i^{AB}.

In case of GMMs we have $K^* = K+1$, $\beta^* = (\gamma, \beta')$ and $T^* = 2(T-1)$ with

$$y_i^* = \begin{pmatrix} \Delta y_{i2} \\ \vdots \\ \Delta y_{iT} \\ y_{i2} \\ \vdots \\ y_{iT} \end{pmatrix}, \; \varepsilon_i^* = \begin{pmatrix} \Delta\varepsilon_{i2} \\ \vdots \\ \Delta\varepsilon_{iT} \\ \eta_i + \varepsilon_{i2} \\ \vdots \\ \eta_i + \varepsilon_{iT} \end{pmatrix} \text{ and } X_i^* = \begin{bmatrix} \Delta y_{i1} & \Delta x_{i2}' \\ \vdots & \vdots \\ \Delta y_{i,T-1} & \Delta x_{iT}' \\ y_{i1} & x_{i2}' \\ \vdots & \vdots \\ y_{i,T-1} & x_{iT}' \end{bmatrix}, \tag{26}$$

and from (6) and (**??**) it follows that $L^* = (K+1)(T-1)(T+2)/2$ whereas the $T^* \times L^*$ matrix $Z_i^* = Z_i^{BB}$ is

$$Z_i^{BB} = \begin{bmatrix} Z_i^{AB} & 0 & \cdots & 0 & O & \cdots & O \\ 0' & \Delta y_{i1} & 0' & 0 & \Delta x_{i2}' & 0' & 0' \\ \vdots & 0 & \ddots & 0 & 0' & \ddots & O \\ 0' & 0 & 0' & \Delta y_{i,T-1} & 0' & 0' & \Delta x_{iT}' \end{bmatrix}. \tag{27}$$

Note that $E[(\Delta\varepsilon_{it})^2] = 2\sigma_\varepsilon^2$ differs from $E[(\varepsilon_{it} + \eta_i)^2] = \sigma_\varepsilon^2 + \sigma_\eta^2$ when $\sigma_\eta^2 \neq \sigma_\varepsilon^2$, and for $t \neq s$ we find $E[(\varepsilon_{it} + \eta_i)(\varepsilon_{is} + \eta_i)] = \sigma_\eta^2 \geq 0$. Thus, again it does not hold here that $\varepsilon_{it}^* \mid Z_i^t \sim \text{i.i.d.}(0, \sigma_{\varepsilon^*}^2)$, so GIV is not efficient. However, here an appropriate weighting matrix is not readily available due to the complexity of $\text{Var}(Z_i^* \varepsilon_i^*)$.

2.5 Weighting matrices in use for 1-step GMMs

To date the GMMs optimal weighting matrix has only been obtained for the no individual effects case $\sigma_\eta^2 = 0$ by Windmeijer (2000), who presents

$$W_{BB}^{FW} \propto \left(\sum_{i=1}^{N} (Z_i^{BB})' D^{FW} Z_i^{BB} \right)^{-1} \tag{28}$$

with

$$D^{FW} = \begin{pmatrix} H & C_1 \\ C_1' & I_{T-1} \end{pmatrix}, \tag{29}$$

where C_1 is the $(T-1) \times (T-1)$ matrix

$$C_1 = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ -1 & 1 & 0 & \ddots & \vdots \\ 0 & -1 & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & -1 & 1 \end{bmatrix}. \tag{30}$$

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 weighting matrix

$$W_{BB}^{DPD} \propto \left(\sum_{i=1}^{N} (Z_i^{BB})' D^{DPD} Z_i^{BB} \right)^{-1}, \tag{31}$$

with

$$D^{DPD} = \begin{pmatrix} H & O \\ O & I_{T-1} \end{pmatrix}. \tag{32}$$

The motivation for that is unclear. The block diagonality of D^{DPD} does not lead to an interesting reduction of computational requirements, and there is no special situation (not even $\sigma_\eta^2 = 0$) 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

$$D^{GIV} = \begin{pmatrix} I_{T-1} & O \\ O & I_{T-1} \end{pmatrix} = I_{2T-2}, \tag{33}$$

which yields the simple GIV estimator. This, as we already indicated, is certainly not optimal either, but is easy and could perhaps suit well as first step in a 2-step procedure. Blundell and Bond (1998) mention that, in most of the cases they examined, 2-step and 1-step GMMs gave similar results, suggesting that the weighting matrix to be used in combination with Z_i^{BB} seems of minor concern under homoskedasticity. However in Kiviet (2007), in less restrained simulation experiments, it is demonstrated that different initial weighting matrices can lead to huge differences in the performance of 1-step and 2-step GMMs. There it is argued that a better (though yet suboptimal and unfeasible) weighting matrix would be

$$W_{BB}^{\sigma_\eta^2/\sigma_\varepsilon^2} \propto \left(\sum_{i=1}^{N} (Z_i^{BB})' D^{\sigma_\eta^2/\sigma_\varepsilon^2} Z_i^{BB} \right)^{-1}, \tag{34}$$

with

$$D^{\sigma_\eta^2/\sigma_\varepsilon^2} = \begin{pmatrix} H & C \\ C' & I_{T-1} + \frac{\sigma_\eta^2}{\sigma_\varepsilon^2}\iota_{T-1}\iota_{T-1}' \end{pmatrix}. \tag{35}$$

This can be made operational by choosing some value for $\sigma_\eta^2/\sigma_\varepsilon^2$. From the simulations it follows that this value should not be chosen too low, and reasonably satisfying results were obtained by choosing $\sigma_\eta^2/\sigma_\varepsilon^2 = 10$.

3 The optimal weighting matrix for GMMs

The obtain the optimal weighting matrix W^{opt} of (16) for the generic model (10) one has to consider the matrices

$$Q_i \equiv Z_i^{*\prime}\varepsilon_i^*\varepsilon_i^{*\prime}Z_i^*$$

and, ideally, evaluate

$$E(Q_i) = E(Z_i^{*\prime}\varepsilon_i^*\varepsilon_i^{*\prime}Z_i^*) = \mathrm{Var}(Z_i^{*\prime}\varepsilon_i^*).$$

Denoting the typical element of the $L^* \times L^*$ matrix Q_i by $q_{i,lm}$, where $l, m = 1, ..., L^*$, and defining

$$\bar{q}_{i,lm}^s \equiv E_s(q_{i,lm}),$$

where E_s denotes expectation conditional on information available at time-period s, we have, by the LIE (law of iterated expectations),

$$E(\bar{q}_{i,lm}^s) = E(q_{i,lm}) = [\mathrm{Var}(Z_i^{*\prime}\varepsilon_i^*)]_{l,m} \tag{36}$$

and by the LLN (law of large numbers)

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\bar{q}_{i,lm}^s = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}E(\bar{q}_{i,lm}^s) = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}[\mathrm{Var}(Z_i^{*\prime}\varepsilon_i^*)]_{lm}. \tag{37}$$

Hence, $N^{-1}\sum_{i=1}^N \bar{q}_{i,lm}^s$ has probability limit equal to the corresponding element of the covariance matrix of the limiting distribution of $N^{-1/2}\sum_{i=1}^N Z_i^{*\prime}\varepsilon_i^*$. So, we do not necessarily have to evaluate $E(q_{i,lm})$; finding $\bar{q}_{i,lm}^s$ for a convenient s and calculating $N^{-1}\sum_{i=1}^N \bar{q}_{i,lm}^s$ for all $l, m = 1, ..., L^*$ suffices for the present purpose. Of course, other types of conditioning sets are allowed too, and different conditioning sets may be used for the individual elements $q_{i,lm}$ of Q_i, and even for separate components of these individual elements. Since a GMM estimator is invariant to the scale of the weighting matrix we may multiply all elements by N/σ_ε^2, and thus W^{opt} can be obtained by inverting the matrix that has elements $\sigma_\varepsilon^{-2}\sum_{i=1}^N \bar{q}_{i,lm}^s$.

3.1 Ordinary GMM

We first derive formally (most published earlier derivations are rather sketchy) the well-known optimal weighting matrix W_{AB}^{opt}, given in (23), to be used when the instruments Z_i^{AB}, given in (23), are employed in model (5). In this case $Z_i^{*\prime}\varepsilon_i^*\varepsilon_i^{*\prime}Z_i^*$ consists of components which are of the following four forms: $(y_i^{t-2})'\Delta\varepsilon_{it}\Delta\varepsilon_{is}y_i^{s-2}$, $(y_i^{t-2})'\Delta\varepsilon_{it}\Delta\varepsilon_{is}X_i^{s-1}$, $(X_i^{t-1})'\Delta\varepsilon_{it}\Delta\varepsilon_{is}y_i^{s-2}$, or $(X_i^{t-1})'\Delta\varepsilon_{it}\Delta\varepsilon_{is}X_i^{s-1}$, where $t, s = 2, ..., T$. Because of the symmetry of $Z_i^{*\prime}\varepsilon_i^*\varepsilon_i^{*\prime}Z_i^*$ we only have to consider components for $t \geq s$.

We find for $t > s+1$

$$\begin{aligned}
&E_{t-2}[(y_i^{t-2})'\Delta\varepsilon_{it}\Delta\varepsilon_{is}y_i^{s-2}] = (y_i^{t-2})'y_i^{s-2}\Delta\varepsilon_{is}E_{t-2}(\Delta\varepsilon_{it}) = O,\\
&E_{t-2}[(y_i^{t-2})'\Delta\varepsilon_{it}\Delta\varepsilon_{is}X_i^{s-1}] = (y_i^{t-2})'X_i^{s-1}\Delta\varepsilon_{is}E_{t-2}(\Delta\varepsilon_{it}) = O,\\
&E_s[(X_i^{t-1})'\Delta\varepsilon_{it}\Delta\varepsilon_{is}y_i^{s-2}] = \Delta\varepsilon_{is}E_s[(X_i^{t-1})'\Delta\varepsilon_{it}]y_i^{s-2} = O,\\
&E_s[(X_i^{t-1})'\Delta\varepsilon_{it}\Delta\varepsilon_{is}X_i^{s-1}] = \Delta\varepsilon_{is}E_s[(X_i^{t-1})'\Delta\varepsilon_{it}]X_i^{s-1} = O.
\end{aligned}$$

For $s = t-1$ we obtain

$$\begin{aligned}
&E_{t-2}[(y_i^{t-2})'\Delta\varepsilon_{it}\Delta\varepsilon_{i,t-1}y_i^{t-3}] = (y_i^{t-2})'y_i^{t-3}[E_{t-2}(\Delta\varepsilon_{it}\varepsilon_{i,t-1}) - \varepsilon_{i,t-2}E_{t-2}(\Delta\varepsilon_{it})] = -\sigma_\varepsilon^2(y_i^{t-2})'y_i^{t-3},\\
&E_{t-2}[(y_i^{t-2})'\Delta\varepsilon_{it}\Delta\varepsilon_{i,t-1}X_i^{t-2}] = (y_i^{t-2})'X_i^{t-2}E_{t-2}(\Delta\varepsilon_{it}\Delta\varepsilon_{i,t-1}) = -\sigma_\varepsilon^2(y_i^{t-2})'X_i^{t-2},\\
&E_{t-3}[(X_i^{t-1})'\Delta\varepsilon_{it}\Delta\varepsilon_{i,t-1}y_i^{t-3} \mid X_i^{t-1}] = (X_i^{t-1})'y_i^{t-3}E[\Delta\varepsilon_{it}\Delta\varepsilon_{i,t-1} \mid X_i^{t-1}] = -\sigma_\varepsilon^2(X_i^{t-1})'y_i^{t-3},\\
&E[(X_i^{t-1})'\Delta\varepsilon_{it}\Delta\varepsilon_{i,t-1}X_i^{t-2} \mid X_i^{t-1}] = (X_i^{t-1})'X_i^{t-2}E(\Delta\varepsilon_{it}\Delta\varepsilon_{i,t-1} \mid X_i^{t-1}) = -\sigma_\varepsilon^2(X_i^{t-1})'X_i^{t-2},
\end{aligned}$$

because $E(\Delta\varepsilon_{it}\Delta\varepsilon_{i,t-1} \mid X_i^{t-1}) = E(\varepsilon_{it}\varepsilon_{i,t-1} - \varepsilon_{i,t-1}^2 - \varepsilon_{it}\varepsilon_{i,t-2} + \varepsilon_{i,t-1}\varepsilon_{i,t-2} \mid X_i^{t-1}) = -\sigma_\varepsilon^2$, since for $j = 1, 2$ we find $E(\varepsilon_{it}\varepsilon_{i,t-j} \mid X_i^{t-1}) = E[E_{t-1}(\varepsilon_{it}\varepsilon_{i,t-j} \mid X_i^{t-1})] = E[\varepsilon_{i,t-j}E_{t-1}(\varepsilon_{it} \mid X_i^{t-1})] = 0$ and $E(\varepsilon_{i,t-1}\varepsilon_{i,t-2} \mid X_i^{t-1}) = E_{t-2}[E(\varepsilon_{i,t-1}\varepsilon_{i,t-2} \mid X_i^{t-1})] = E_{t-2}\varepsilon_{i,t-2}[E(\varepsilon_{i,t-1} \mid X_i^{t-1})] = 0$, whereas $E(\varepsilon_{i,t-1}^2 \mid X_i^{t-1}) = \sigma_\varepsilon^2$. Finally, for $s = t = 2, ..., T$, we get

$$\begin{aligned}
&E_{t-2}[(y_i^{t-2})'(\Delta\varepsilon_{it})^2y_i^{t-2}] = (y_i^{t-2})'y_i^{t-2}E_{t-2}[(\Delta\varepsilon_{it})^2] = 2\sigma_\varepsilon^2(y_i^{t-2})'y_i^{t-2},\\
&E_{t-2}[(y_i^{t-2})'(\Delta\varepsilon_{it})^2X_i^{t-1}] = 2\sigma_\varepsilon^2(y_i^{t-2})'X_i^{t-1},\\
&E_{t-2}[(X_i^{t-1})'(\Delta\varepsilon_{it})^2y_i^{t-2} \mid X_i^{t-1}] = (X_i^{t-1})'y_i^{t-2}E_{t-2}[(\Delta\varepsilon_{it})^2 \mid X_i^{t-1}] = 2\sigma_\varepsilon^2(X_i^{t-1})'y_i^{t-2},\\
&E[(X_i^{t-1})'(\Delta\varepsilon_{it})^2X_i^{t-1} \mid X_i^{t-1}] = 2\sigma_\varepsilon^2(X_i^{t-1})'X_i^{t-1}.
\end{aligned}$$

Collecting from the above all elements $\sigma_\varepsilon^{-2}\sum_{i=1}^N \bar{q}_{i,lm}^s$, we obtain $\sum_{i=1}^N Z_i^{AB\prime}HZ_i^{AB}$. Hence, its inverse (24) is proportional to the optimal weighting matrix W_{AB}^{opt} when $\varepsilon_i \sim$ i.i.d.$(0, \sigma_\varepsilon^2 I_T)$ for all i.

3.2 System GMM

For GMMs, where the instruments Z_i^{BB} given in (27) are applied to the system (10) with hybrid disturbances as in (26), the derivation of an optimal weighting matrix W_{BB}^{opt} is much more involved, because the various components of $Z_i^{*\prime}\varepsilon_i^*\varepsilon_i^{*\prime}Z_i^*$ are of a very different nature. The elements associated with $\Delta\varepsilon_{it}\Delta\varepsilon_{is}$ yield results equivalent to those just obtained. Of the additional components we first examine those that involve $\Delta\varepsilon_{it}(\eta_i+\varepsilon_{is})$, viz. $(y_i^{t-2})'\Delta\varepsilon_{it}(\eta_i+\varepsilon_{is})\Delta y_{i,s-1}$, $(X_i^{t-1})'\Delta\varepsilon_{it}(\eta_i+\varepsilon_{is})\Delta y_{i,s-1}$, $(y_i^{t-2})'\Delta\varepsilon_{it}(\eta_i+\varepsilon_{is})\Delta x_{is}'$ and $(X_i^{t-1})'\Delta\varepsilon_{it}(\eta_i+\varepsilon_{is})\Delta x_{is}'$. Next we will examine components that involve $(\eta_i+\varepsilon_{it})(\eta_i+\varepsilon_{is})$, viz. $\Delta y_{i,t-1}(\eta_i+\varepsilon_{it})(\eta_i+\varepsilon_{is})\Delta y_{i,s-1}$, $\Delta y_{i,t-1}(\eta_i+\varepsilon_{it})(\eta_i+\varepsilon_{is})\Delta x_{is}'$, the transpose of the latter, and finally $\Delta x_{it}(\eta_i+\varepsilon_{it})(\eta_i+\varepsilon_{is})\Delta x_{is}'$. We examine these eight types of components successively, for $t, s = 2, ..., T$, upon assuming effect stationarity, which implies (3).

The first component can be decomposed into two contributions, viz.

$$(y_i^{t-2})'\Delta\varepsilon_{it}(\eta_i+\varepsilon_{is})\Delta y_{i,s-1} = \eta_i(y_i^{t-2})'\Delta y_{i,s-1}\Delta\varepsilon_{it} + (y_i^{t-2})'\Delta y_{i,s-1}(\Delta\varepsilon_{it})\varepsilon_{is}.$$

For $t = s$ we find for these

$$\begin{aligned}
&E[\eta_i(y_i^{t-2})'\Delta y_{i,t-1}\Delta\varepsilon_{it}] = -E[\eta_i(y_i^{t-2})'\Delta y_{i,t-1}\varepsilon_{i,t-1}] = -\sigma_\varepsilon^2\sigma_{y\eta}\iota_{t-1}\\
&E[(y_i^{t-2})'\Delta y_{i,t-1}(\Delta\varepsilon_{it})\varepsilon_{it}] = E_{t-1}[(y_i^{t-2})'\Delta y_{i,t-1}(\Delta\varepsilon_{it})\varepsilon_{it}]\\
&\qquad = \sigma_\varepsilon^2(y_i^{t-2})'\Delta y_{i,t-1}[E_{t-1}(\varepsilon_t^2) - \varepsilon_{t-1}E_{t-1}(\varepsilon_t)] = \sigma_\varepsilon^2(y_i^{t-2})'\Delta y_{i,t-1},
\end{aligned}$$

for $s = t-1$ they give

$$\begin{array}{l} E_{t-2}[\eta_i(y_i^{t-2})'\Delta y_{i,t-2}\Delta\varepsilon_{it}] = 0 \\ E_{t-2}[(y_i^{t-2})'\Delta y_{i,t-2}(\Delta\varepsilon_{it})\varepsilon_{i,t-1}] = -\sigma_\varepsilon^2(y_i^{t-2})'\Delta y_{i,t-2}, \end{array}$$

for $s < t-1$ we find

$$E_s[(y_i^{t-2})'\Delta\varepsilon_{it}(\eta_i + \varepsilon_{is})\Delta y_{i,s-1}] = 0,$$

and for $s = t+1$ we obtain

$$\begin{array}{l} E[\eta_i(y_i^{t-2})'\Delta y_{it}\Delta\varepsilon_{it}] = E(\Delta y_{it}\Delta\varepsilon_{it})\sigma_{y\eta}\iota_{t-1} \\ E_t[(y_i^{t-2})'\Delta y_{it}(\Delta\varepsilon_{it})\varepsilon_{i,t+1}] = 0. \end{array}$$

Substituting (5) and using the notation

$$\sigma_{x\varepsilon,j} \equiv E(x_{it}\varepsilon_{t-j}),\ \text{for } j = ..., -1, 0, 1, 2, ..., \tag{38}$$

where $\sigma_{x\varepsilon,j} = 0$ for $j \leq 0$, we find

$$\begin{aligned} E(\Delta y_{it}\Delta\varepsilon_{it}) &= \gamma E(\Delta y_{i,t-1}\Delta\varepsilon_{it}) + E[\Delta\varepsilon_{it}(\Delta x_{it})']\beta + E[(\Delta\varepsilon_{it})^2] \\ &= -\gamma E(y_{i,t-1}\varepsilon_{i,t-1}) - \sigma'_{x\varepsilon,1}\beta + 2\sigma_\varepsilon^2 \\ &= (2-\gamma)\sigma_\varepsilon^2 - \sigma'_{x\varepsilon,1}\beta, \end{aligned}$$

thus

$$E[\eta_i(y_i^{t-2})'\Delta y_{it}\Delta\varepsilon_{it}] = \sigma_{y\eta}[(2-\gamma)\sigma_\varepsilon^2 - \sigma'_{x\varepsilon,1}\beta]\iota_{t-1}.$$

Now for $s \geq t+1$ we find

$$\begin{aligned} E[(y_i^{t-2})'\Delta\varepsilon_{it}(\eta_i + \varepsilon_{is})\Delta y_{i,s-1}] &= E[\eta_i(y_i^{t-2})'\Delta\varepsilon_{it}\Delta y_{i,s-1}] \\ &= \sigma_{y\eta}\iota_{t-1}E(\Delta y_{i,s-1}\Delta\varepsilon_{it}) \\ &= \sigma_{y\eta}\pi_{s-t-1}\iota_{t-1}, \end{aligned}$$

where

$$\pi_j \equiv E(\Delta y_{i,t+j}\Delta\varepsilon_{it}),\ \text{for } j \geq 0. \tag{39}$$

We already found $\pi_0 = (2-\gamma)\sigma_\varepsilon^2 - \sigma'_{x\varepsilon,1}\beta$, and further obtain

$$\begin{aligned} \pi_1 &= E(\Delta y_{i,t+1}\Delta\varepsilon_{it}) \\ &= \gamma E(\Delta y_{i,t}\Delta\varepsilon_{it}) + E[\Delta\varepsilon_{it}(\Delta x_{i,t+1})']\beta + E[(\Delta\varepsilon_{i,t+1})(\Delta\varepsilon_{it})] \\ &= \gamma\pi_0 + 2\sigma'_{x\varepsilon,1}\beta - \sigma'_{x\varepsilon,2}\beta - \sigma_\varepsilon^2 \\ &= [(2-\gamma)\sigma'_{x\varepsilon,1} - \sigma'_{x\varepsilon,2}]\beta - (1-\gamma)^2\sigma_\varepsilon^2, \end{aligned}$$

whereas for $j > 1$

$$\begin{aligned} \pi_j &= E(\Delta y_{i,t+j}\Delta\varepsilon_{it}) \\ &= \gamma E(\Delta y_{i,t+j-1}\Delta\varepsilon_{it}) + E[\Delta\varepsilon_{it}(\Delta x_{i,t+j})']\beta + E[(\Delta\varepsilon_{i,t+j})(\Delta\varepsilon_{it})] \\ &= \gamma\pi_{j-1} + (2\sigma'_{x\varepsilon,j} - \sigma'_{x\varepsilon,j+1} - \sigma'_{x\varepsilon,j-1})\beta. \end{aligned}$$

For the second component we obtain

$$(X_i^{t-1})'\Delta\varepsilon_{it}(\eta_i + \varepsilon_{is})\Delta y_{i,s-1} = \eta_i(X_i^{t-1})'\Delta\varepsilon_{it}\Delta y_{i,s-1} + (X_i^{t-1})'(\Delta\varepsilon_{it})\varepsilon_{is}\Delta y_{i,s-1}.$$

For $s = t$ we find

$$\begin{array}{l} E[\eta_i (X_i^{t-1})' \Delta\varepsilon_{it} \Delta y_{i,t-1}] = -\sigma_\varepsilon^2 \iota_{t-1} \otimes \sigma_{x\eta} \\ E_{t-1}[(X_i^{t-1})' (\Delta\varepsilon_{it}) \varepsilon_{it} \Delta y_{i,t-1}] = \sigma_\varepsilon^2 (X_i^{t-1})' \Delta y_{i,t-1}, \end{array}$$

for $s = t - 1$

$$\begin{array}{l} E[\eta_i (X_i^{t-1})' \Delta\varepsilon_{it} \Delta y_{i,t-2}] = 0 \\ E_{t-2}[(X_i^{t-1})' (\Delta\varepsilon_{it}) \varepsilon_{i,t-1} \Delta y_{i,t-2} \mid X_i^{t-1}] = -\sigma_\varepsilon^2 (X_i^{t-1})' \Delta y_{i,t-2}, \end{array}$$

for $s < t - 1$

$$E[(X_i^{t-1})' \Delta\varepsilon_{it} (\eta_i + \varepsilon_{is}) \Delta y_{i,s-1}] = 0,$$

for $s = t + 1$

$$\begin{array}{l} E[\eta_i (X_i^{t-1})' \Delta\varepsilon_{it} \Delta y_{i,t}] = [(2-\gamma)\sigma_\varepsilon^2 - \sigma'_{x\varepsilon,1} \beta] \iota_{t-1} \otimes \sigma_{x\eta} \\ E_t[(X_i^{t-1})' (\Delta\varepsilon_{it}) \varepsilon_{i,t+1} \Delta y_{i,t}] = 0, \end{array}$$

and for $s > t + 1$

$$\begin{aligned} E[(X_i^{t-1})' \Delta\varepsilon_{it} (\eta_i + \varepsilon_{is}) \Delta y_{i,s-1}] &= E[\eta_i (X_i^{t-1})' \Delta\varepsilon_{it} \Delta y_{i,s-1}] \\ &= E(\Delta\varepsilon_{it} \Delta y_{i,s-1}) \iota_{t-1} \otimes \sigma_{x\eta} \\ &= \pi_{s-t-1} \iota_{t-1} \otimes \sigma_{x\eta}. \end{aligned}$$

The third component is

$$(y_i^{t-2})' \Delta\varepsilon_{it} (\eta_i + \varepsilon_{is}) \Delta x'_{is} = \eta_i (y_i^{t-2})' \Delta\varepsilon_{it} \Delta x'_{is} + (y_i^{t-2})' (\Delta\varepsilon_{it}) \varepsilon_{is} \Delta x'_{is}.$$

For $s = t$ we find for the first term

$$\eta_i (y_i^{t-2})' \Delta\varepsilon_{it} \Delta x'_{it} = \eta_i (y_i^{t-2})' \Delta\varepsilon_{it} x'_{it} - \eta_i (y_i^{t-2})' \Delta\varepsilon_{it} x'_{i,t-1},$$

where the second sub-term has expectation zero and the first yields

$$E[\eta_i (y_i^{t-2})' \Delta\varepsilon_{it} x'_{it}] = -\sigma_{y\eta} \iota_{t-1} \sigma'_{x\varepsilon,1},$$

and for the second term we find

$$(y_i^{t-2})' (\Delta\varepsilon_{it}) \varepsilon_{it} \Delta x'_{it} = (y_i^{t-2})' \varepsilon_{it}^2 \Delta x'_{it} - (y_i^{t-2})' \varepsilon_{it} \varepsilon_{i,t-1} \Delta x'_{i,t-1},$$

where the second sub-term has expectation zero and the first yields

$$E_{t-1}[(y_i^{t-2})' \varepsilon_{it}^2 \Delta x'_{it} \mid x_{it}] = \sigma_\varepsilon^2 (y_i^{t-2})' \Delta x'_{it}.$$

For $s = t - 1$ we find

$$\begin{array}{l} E[\eta_i (y_i^{t-2})' \Delta\varepsilon_{it} \Delta x'_{i,t-1}] = 0 \\ E_{t-2}[(y_i^{t-2})' (\Delta\varepsilon_{it}) \varepsilon_{i,t-1} \Delta x'_{i,t-1} \mid x_{i,t-1}] = -\sigma_\varepsilon^2 (y_i^{t-2})' \Delta x'_{i,t-1}, \end{array}$$

for $s < t - 1$

$$E[(y_i^{t-2})' \Delta\varepsilon_{it} (\eta_i + \varepsilon_{is}) \Delta x'_{is}] = 0,$$

for $s = t + 1$

$$\begin{aligned} E[\eta_i (y_i^{t-2})' \Delta\varepsilon_{it} \Delta x'_{i,t+1}] &= \sigma_{y\eta} \iota_{t-1} E(\Delta\varepsilon_{it} \Delta x'_{i,t+1}) \\ &= \sigma_{y\eta} \iota_{t-1} E(\varepsilon_{it} x'_{i,t+1} - \varepsilon_{i,t-1} x'_{i,t+1} - \varepsilon_{it} x'_{it} + \varepsilon_{i,t-1} x'_{it}) \\ &= \sigma_{y\eta} \iota_{t-1} (2\sigma'_{x\varepsilon,1} - \sigma'_{x\varepsilon,2}) \\ E_t[(y_i^{t-2})'(\Delta\varepsilon_{it}) \varepsilon_{i,t+1} \Delta x'_{i,t+1}] &= 0, \end{aligned}$$

and for $s > t + 1$

$$\begin{aligned} E[\eta_i (y_i^{t-2})' \Delta\varepsilon_{it} \Delta x'_{is}] &= \sigma_{y\eta} \iota_{t-1} E(\varepsilon_{it} x'_{is} - \varepsilon_{i,t-1} x'_{is} - \varepsilon_{it} x'_{i,s-1} + \varepsilon_{i,t-1} x'_{i,s-1}) \\ &= \sigma_{y\eta} \iota_{t-1} (2\sigma'_{x\varepsilon,s-t} - \sigma'_{x\varepsilon,s-t+1} - \sigma'_{x\varepsilon,s-t-1}) \\ E[(y_i^{t-2})' \Delta\varepsilon_{it} (\varepsilon_{is}) \Delta x'_{is}] &= 0. \end{aligned}$$

The fourth component is

$$(X_i^{t-1})' \Delta\varepsilon_{it} (\eta_i + \varepsilon_{is}) \Delta x'_{is} = \eta_i (X_i^{t-1})' \Delta\varepsilon_{it} \Delta x'_{is} + (X_i^{t-1})' (\Delta\varepsilon_{it}) \varepsilon_{is} \Delta x'_{is}.$$

For $t = s$ we find

$$E[\eta_i (X_i^{t-1})' \Delta\varepsilon_{it} \Delta x'_{it}] = -(\iota_{t-1} \otimes \sigma_{x\eta}) \sigma'_{x\varepsilon,1},$$

and since

$$(X_i^{t-1})' (\Delta\varepsilon_{it}) \varepsilon_{it} \Delta x'_{it} = (X_i^{t-1})' \varepsilon_{it}^2 \Delta x'_{it} - (X_i^{t-1})' \varepsilon_{it} \varepsilon_{i,t-1} \Delta x'_{it},$$

we also have

$$\begin{aligned} &E[(X_i^{t-1})' \varepsilon_{it}^2 \Delta x'_{it} \mid X_i^t] = \sigma_\varepsilon^2 (X_i^{t-1})' \Delta x'_{it} \\ &E_{t-1}[(X_i^{t-1})' \varepsilon_{it} \varepsilon_{i,t-1} \Delta x'_{it}] = O. \end{aligned}$$

For $s = t - 1$

$$\begin{aligned} &E[\eta_i (X_i^{t-1})' \Delta\varepsilon_{it} \Delta x'_{i,t-1}] = O \\ &E[(X_i^{t-1})' (\Delta\varepsilon_{it}) \varepsilon_{i,t-1} \Delta x'_{i,t-1} \mid X_i^{t-1}] = -\sigma_\varepsilon^2 (X_i^{t-1})' \Delta x'_{i,t-1}, \end{aligned}$$

for $s < t - 1$

$$\begin{aligned} &E[\eta_i (X_i^{t-1})' \Delta\varepsilon_{it} \Delta x'_{is}] = O \\ &E[(X_i^{t-1})' (\Delta\varepsilon_{it}) \varepsilon_{is} \Delta x'_{is}] = O, \end{aligned}$$

for $s = t + 1$

$$\begin{aligned} E[\eta_i (X_i^{t-1})' \Delta\varepsilon_{it} \Delta x'_{i,t+1}] &= (\iota_{t-1} \otimes \sigma_{x\eta}) E(\Delta\varepsilon_{it} \Delta x'_{i,t+1}) \\ &= (2\sigma'_{x\varepsilon,1} - \sigma'_{x\varepsilon,2}) \iota_{t-1} \otimes \sigma_{x\eta} \\ E_t[(X_i^{t-1})' (\Delta\varepsilon_{it}) \varepsilon_{i,t+1} \Delta x'_{i,t+1}] &= O, \end{aligned}$$

and for $s > t + 1$

$$\begin{aligned} &E[\eta_i (X_i^{t-1})' \Delta\varepsilon_{it} \Delta x'_{is}] = (2\sigma'_{x\varepsilon,s-t} - \sigma'_{x\varepsilon,s-t+1} - \sigma'_{x\varepsilon,s-t-1}) \iota_{t-1} \otimes \sigma_{x\eta} \\ &E[(X_i^{t-1})' \Delta\varepsilon_{it} (\varepsilon_{is}) \Delta x'_{is}] = O. \end{aligned}$$

Next we commence with the second group of four components. The fifth,

$$\Delta y_{i,t-1} (\eta_i + \varepsilon_{it})(\eta_i + \varepsilon_{is}) \Delta y_{i,s-1} = \Delta y_{i,t-1} \Delta y_{i,s-1} [\eta_i^2 + \eta_i (\varepsilon_{it} + \varepsilon_{is}) + \varepsilon_{it} \varepsilon_{is}],$$

is symmetric in s and t. Just examining $t \geq s$ we find

$$\begin{aligned} &E[\Delta y_{i,t-1} \Delta y_{i,s-1} \eta_i^2 \mid \Delta y_{i,t-1}, \Delta y_{i,s-1}] = \sigma_\eta^2 \Delta y_{i,t-1} \Delta y_{i,s-1} \\ &E[\Delta y_{i,t-1} \Delta y_{i,s-1} \eta_i (\varepsilon_{it} + \varepsilon_{is})] = 0. \end{aligned}$$

For $\Delta y_{i,t-1}\Delta y_{i,s-1}\varepsilon_{it}\varepsilon_{is}$ we find when $s=t$

$$E_{t-1}[\Delta y_{i,t-1}\Delta y_{i,t-1}\varepsilon_{it}^2]=\sigma_\varepsilon^2(\Delta y_{i,t-1})^2$$

and for $s<t$

$$E_{t-1}[\Delta y_{i,t-1}\varepsilon_{it}\varepsilon_{is}\Delta y_{i,s-1}]=0.$$

The sixth and seventh component follow from

$$\Delta y_{i,t-1}\Delta x_{is}'[\eta_i^2+\eta_i(\varepsilon_{it}+\varepsilon_{is})+\varepsilon_{it}\varepsilon_{is}],$$

where

$$\begin{aligned}&E[\eta_i^2\Delta y_{i,t-1}\Delta x_{is}'\mid \Delta y_{i,t-1},\Delta x_{is}']=\sigma_\eta^2\Delta y_{i,t-1}\Delta x_{is}'\\&E[\eta_i(\varepsilon_{it}+\varepsilon_{is})\Delta y_{i,t-1}\Delta x_{is}']=0',\end{aligned}$$

and regarding $\varepsilon_{it}\varepsilon_{is}\Delta y_{i,t-1}\Delta x_{is}'$ we find, for $t=s$

$$E_{t-1}[\varepsilon_{it}^2\Delta y_{i,t-1}\Delta x_{it}'\mid \Delta x_{it}]=\sigma_\varepsilon^2\Delta y_{i,t-1}\Delta x_{it}',$$

for $t>s$

$$E_{t-1}[\varepsilon_{it}\varepsilon_{is}\Delta y_{i,t-1}\Delta x_{is}']=0',$$

and for $s>t$

$$E_{s-1}[\varepsilon_{it}\varepsilon_{is}\Delta y_{i,t-1}\Delta x_{is}'\mid \Delta x_{is}]=0'.$$

The eighth and final component is again symmetric in t and s. From

$$\Delta x_{it}(\varepsilon_{it}+\eta_i)(\varepsilon_{is}+\eta_i)\Delta x_{is}'=\Delta x_{it}\Delta x_{is}'[\eta_i^2+\eta_i(\varepsilon_{it}+\varepsilon_{is})+\varepsilon_{it}\varepsilon_{is}],$$

we find

$$\begin{aligned}&E[\Delta x_{it}\Delta x_{is}'\eta_i^2\mid \Delta x_{it},\Delta x_{is}']=\sigma_\eta^2\Delta x_{it}\Delta x_{is}'\\&E[\Delta x_{it}\Delta x_{is}'\eta_i(\varepsilon_{it}+\varepsilon_{is})]=O,\end{aligned}$$

and for $t=s$

$$E[\Delta x_{it}\Delta x_{it}'\varepsilon_{it}^2\mid \Delta x_{it}]=\sigma_\varepsilon^2\Delta x_{it}\Delta x_{it}',$$

while for $s<t$

$$E_{t-1}[\Delta x_{it}\Delta x_{is}'\varepsilon_{it}\varepsilon_{is}]=O.$$

Collecting all the above results and taking for all elements $\sigma_\varepsilon^{-2}\sum_{i=1}^N\bar{q}_{i,lm}^s$, we find for the GMMs estimator that exploits $2N(T-1)\times L^*$ instrument matrix Z^{BB} and has the disturbances given in (26) that the optimal weighting matrix is given by

$$W_{BB}^{opt}\propto[Z^{BB\prime}(I_N\otimes D_1^{opt})Z^{BB}+ND_2^{opt}]^{-1}, \tag{40}$$

where D_1^{opt} is the $2(T-1)\times 2(T-1)$ matrix

$$D_1^{opt}=\left(\begin{array}{cc}H & C_1\\ C_1' & I_{T-1}+\frac{\sigma_\eta^2}{\sigma_\varepsilon^2}\iota_{T-1}\iota_{T-1}'\end{array}\right), \tag{41}$$

with C_1 the $(T-1)\times(T-1)$ matrix given in (30), and hence $D_1^{opt}=D^{\sigma_\eta^2/\sigma_\varepsilon^2}$ given in (35), and D_2^{opt} is the $L^*\times L^*$ matrix

$$D_2^{opt}=\frac{1}{\sigma_\varepsilon^2}\left(\begin{array}{cc}O & C_2\\ C_2{}' & O\end{array}\right), \tag{42}$$

with C_2 the $(K+1)T(T-1)/2 \times (K+1)(T-1)$ matrix

$$C_2 = \begin{pmatrix} C_{21} & C_{22} \end{pmatrix}, \tag{43}$$

with C_{21} the $(K+1)T(T-1)/2 \times (T-1)$ matrix

$$C_{21} = \begin{bmatrix} -\sigma_\varepsilon^2 \iota_1 \sigma_{y\eta} & \pi_0 \iota_1 \sigma_{y\eta} & \pi_1 \iota_1 \sigma_{y\eta} & \cdots & \cdots & \pi_{T-3} \iota_1 \sigma_{y\eta} \\ 0 & -\sigma_\varepsilon^2 \iota_2 \sigma_{y\eta} & \pi_0 \iota_2 \sigma_{y\eta} & \cdots & \cdots & \pi_{T-4} \iota_2 \sigma_{y\eta} \\ 0 & 0 & -\sigma_\varepsilon^2 \iota_3 \sigma_{y\eta} & & & \cdot \\ \vdots & \ddots & 0 & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \pi_0 \iota_{T-2} \sigma_{y\eta} \\ 0 & 0 & 0 & \cdots & 0 & -\sigma_\varepsilon^2 \iota_{T-1} \sigma_{y\eta} \\ -\sigma_\varepsilon^2 \iota_1 \otimes \sigma_{x\eta} & \pi_0 \iota_1 \otimes \sigma_{x\eta} & \pi_1 \iota_1 \otimes \sigma_{x\eta} & & \cdots & \pi_{T-3} \iota_1 \otimes \sigma_{x\eta} \\ 0 & -\sigma_\varepsilon^2 \iota_2 \otimes \sigma_{x\eta} & \pi_0 \iota_2 \otimes \sigma_{x\eta} & & \cdots & \pi_{T-4} \iota_2 \otimes \sigma_{x\eta} \\ 0 & 0 & -\sigma_\varepsilon^2 \iota_3 \otimes \sigma_{x\eta} & & & \cdot \\ \vdots & \ddots & 0 & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \pi_0 \iota_{T-2} \otimes \sigma_{x\eta} \\ 0 & 0 & 0 & \cdots & 0 & -\sigma_\varepsilon^2 \iota_{T-1} \otimes \sigma_{x\eta} \end{bmatrix}, \tag{44}$$

and C_{22} the $(K+1)T(T-1)/2 \times K(T-1)$ matrix which is

$$\begin{bmatrix} -\iota_1 \sigma_{y\eta} \sigma'_{x\varepsilon,1} & \pi_0 \iota_1 \sigma_{y\eta} \bar{\sigma}'_{x\varepsilon,1} & \pi_1 \iota_1 \sigma_{y\eta} \bar{\sigma}'_{x\varepsilon,2} & \cdots & \cdots & \pi_{T-3} \iota_1 \sigma_{y\eta} \bar{\sigma}'_{x\varepsilon,T-1} \\ O & -\iota_2 \sigma_{y\eta} \sigma'_{x\varepsilon,1} & \pi_0 \iota_2 \sigma_{y\eta} \bar{\sigma}'_{x\varepsilon,1} & \cdots & \cdots & \pi_{T-4} \iota_2 \sigma_{y\eta} \bar{\sigma}'_{x\varepsilon,T-2} \\ O & O & -\iota_3 \sigma_{y\eta} \sigma'_{x\varepsilon,1} & & & \cdot \\ \vdots & \ddots & O & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \pi_0 \iota_{T-2} \sigma_{y\eta} \bar{\sigma}'_{x\varepsilon,1} \\ O & O & O & \cdots & O & -\iota_{T-1} \sigma_{y\eta} \sigma'_{x\varepsilon,1} \\ -(\iota_1 \otimes \sigma_{x\eta}) \sigma'_{x\varepsilon,1} & (\iota_1 \otimes \sigma_{x\eta}) \bar{\sigma}'_{x\varepsilon,1} & (\iota_1 \otimes \sigma_{x\eta}) \bar{\sigma}'_{x\varepsilon,2} & \cdots & \cdots & (\iota_1 \otimes \sigma_{x\eta}) \bar{\sigma}'_{x\varepsilon,T-1} \\ O & -(\iota_2 \otimes \sigma_{x\eta}) \sigma'_{x\varepsilon,1} & (\iota_2 \otimes \sigma_{x\eta}) \bar{\sigma}'_{x\varepsilon,1} & \cdots & \cdots & (\iota_2 \otimes \sigma_{x\eta}) \bar{\sigma}'_{x\varepsilon,T-2} \\ O & O & -(\iota_3 \otimes \sigma_{x\eta}) \sigma'_{x\varepsilon,1} & & & \cdot \\ \vdots & \ddots & O & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & (\iota_{T-2} \otimes \sigma_{x\eta}) \bar{\sigma}'_{x\varepsilon,1} \\ O & O & O & \cdots & O & -(\iota_{T-1} \otimes \sigma_{x\eta}) \sigma'_{x\varepsilon,1} \end{bmatrix}, \tag{45}$$

where we used the short-hand notation

$$\begin{aligned} \bar{\sigma}'_{x\varepsilon,1} &\equiv 2\sigma'_{x\varepsilon,1} - \sigma'_{x\varepsilon,2} \\ \bar{\sigma}'_{x\varepsilon,t} &\equiv 2\sigma'_{x\varepsilon,t} - \sigma'_{x\varepsilon,t+1} - \sigma'_{x\varepsilon,t-1}, \text{ for } t = 2, ..., T-1. \end{aligned} \tag{46}$$

Note that the weighting matrix (34) explored in Kiviet (2007) is suboptimal because it sets $C_2 = O$, which would only be true if $\sigma_\eta^2 = 0$.

When the model is pure autoregressive, i.e. $K = 0$, the optimal weighting matrix is much simpler. Then it is obvious that C_{22} is void. It leads to the following simplifications

$$\begin{aligned} C_2 &= C_{21} \\ \pi_0 &= (2-\gamma)\sigma_\varepsilon^2 \\ \pi_1 &= -(1-\gamma)^2 \sigma_\varepsilon^2 \\ \pi_j &= \gamma \pi_{j-1}, \text{ for } j \geq 2 \\ \sigma_{y\eta} &= \sigma_\eta^2/(1-\gamma). \end{aligned}$$

Substituting these we find that $\frac{1}{\sigma_\varepsilon^2}C_2$ simplifies to the $T(T-1)/2 \times (T-1)$ matrix

$$\frac{1}{\sigma_\varepsilon^2}C_2 = \frac{\sigma_\eta^2}{1-\gamma}C_3(\gamma),$$

with $C_3(\gamma)$ given by:

$$\begin{bmatrix} -1 & 2-\gamma & -(1-\gamma)^2 & -\gamma(1-\gamma)^2 & \dots & \dots & -\gamma^{T-5}(1-\gamma)^2 & -\gamma^{T-4}(1-\gamma)^2 \\ 0 & -\iota_2 & (2-\gamma)\iota_2 & -(1-\gamma)^2\iota_2 & \dots & \dots & -\gamma^{T-6}(1-\gamma)^2\iota_2 & -\gamma^{T-5}(1-\gamma)^2\iota_2 \\ 0 & 0 & -\iota_3 & (2-\gamma)\iota_3 & \dots & \dots & -\gamma^{T-7}(1-\gamma)^2\iota_3 & -\gamma^{T-6}(1-\gamma)^2\iota_3 \\ 0 & 0 & 0 & -\iota_4 & \dots & \dots & \gamma^{T-8}(1-\gamma)^2\iota_4 & -\gamma^{T-7}(1-\gamma)^2\iota_4 \\ 0 & 0 & 0 & 0 & \ddots & & & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & & \vdots \\ \vdots & \vdots & \vdots & \vdots & & \ddots & -\iota_{T-2} & (2-\gamma)\iota_{T-2} \\ 0 & 0 & 0 & 0 & \dots & \dots & 0 & -\iota_{T-1} \end{bmatrix}. \tag{47}$$

A more promising operational alternative to D^{DPD} or D^{GIV} in the pure first-order autoregressive model could be the following. Instead of capitalizing on the assumption $\sigma_\eta^2 = 0$, as in D^W, one could proceed as follows: capitalize on particular chosen values $\sigma_\eta^2 = \bar{\sigma}_\eta^2$, $\sigma_\varepsilon^2 = \bar{\sigma}_\varepsilon^2$ and $\gamma = \bar{\gamma}$ that seem relevant for the situation at hand and then use the operational point-optimal weighting matrix

$$W_{BB}^{popt}(\bar{\gamma}, \bar{\sigma}_\eta^2, \bar{\sigma}_\varepsilon^2) \propto \left\{ Z^{BB\prime}[I_N \otimes D^{\bar{\sigma}_\eta^2/\bar{\sigma}_\varepsilon^2}]Z^{BB} + N\frac{\bar{\sigma}_\eta^2}{1-\bar{\gamma}} \begin{pmatrix} O & C_3(\bar{\gamma}) \\ C_3(\bar{\gamma})' & O \end{pmatrix} \right\}^{-1}. \tag{48}$$

This is optimal for the chosen values of $\bar{\gamma}$, $\bar{\sigma}_\eta^2$ and $\bar{\sigma}_\varepsilon^2$.

Note that the optimal weighting matrix derived above allows a different approach in obtaining a 2-step GMMs estimator (under effects stationarity and homoskedasticity) than the usually employed $\hat{\theta}_{\widehat{W}^*}$ of (20) based on first step residuals, viz. by simply substituting in D_1^{opt} and D_2^{opt} the first-stage estimators of γ, σ_η^2 and σ_ε^2. Below, by GMMs$_1^{opt}$ we indicate the unfeasible GMMs estimator using in the optimal weighting matrix the true values of γ, σ_η^2 and σ_ε^2, whereas GMMs$_1^{popt}$ is the feasible point-optimal 1-step GMMs estimator using particular values of $\bar{\gamma}$, $\bar{\sigma}_\eta^2$ and $\bar{\sigma}_\varepsilon^2$.

4 Monte Carlo study

In the simulation experiments below we have severely restricted ourselves regarding the generality of the Monte Carlo design. At this stage we only focussed on the case $K = 0$, i.e. like Blundell and Bond (1998) we just examine the fully stationary pure first-order autoregressive case. This implies that the panel data have to be generated such that for all $i = 1, ..., N$ first start-up values are obtained according to

$$y_{i0} = \frac{1}{1-\gamma}\eta_i + \sqrt{\frac{1}{1-\gamma^2}}\varepsilon_{i0},$$

and next in a recursion the T observations

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

from the $N \times 1$ white-noise series $\eta_i \sim$ i.i.d.$(0, \sigma_\eta^2)$ and the N distinct $(T+1) \times 1$ white-noise series $\varepsilon_{it} \sim$ i.i.d.$(0, \sigma_\varepsilon^2)$, $t = 0, ..., T$. All these white-noise series have to be mutually independent from each other. Note that this yields $\mathrm{Var}(y_{it}) = \sigma_\eta^2/(1-\gamma)^2 + \sigma_\varepsilon^2/(1-\gamma^2)$, so there is both effect stationarity and stationarity with respect to the idiosyncratic disturbances.

We will investigate just a few particular combinations of values for the design parameters. Without loss of generality we may scale with respect to σ_ε and choose $\sigma_\varepsilon = 1$. We will examine only positive stable values of γ, viz. $\gamma = 0.1(+0.2)0.9$, and just a few different values of the ratio $\sigma_\eta^2/\sigma_\varepsilon^2$. This variance ratio we vary, for reasons explained in Kiviet (1995, 2007), by examining $\mu \in \{0.5, 1, 2, 4\}$, where

$$\mu = \frac{1}{1-\gamma}\frac{\sigma_\eta}{\sigma_\varepsilon}. \tag{49}$$

Hence, we fix $\sigma_\eta = (1-\gamma)\mu$. For the sample sizes N and T we choose $N = 100$ and $T = 4(+2)10$. For all cells we ran 1000 replications and all the results for all separate cells are obtained by precisely the same series of random numbers.

All results are presented graphically. In 6 Figures we give results for 6 different weighting matrices, viz. D^{GIV}, D^{DPD}, $D^{\bar{\sigma}_\eta^2/\bar{\sigma}_\varepsilon^2}$ for $\bar{\sigma}_\eta^2/\bar{\sigma}_\varepsilon^2 = 0$, $D^{\bar{\sigma}_\eta^2/\bar{\sigma}_\varepsilon^2}$ for $\bar{\sigma}_\eta^2/\bar{\sigma}_\varepsilon^2 = 10$, $D^{\sigma_\eta^2/\sigma_\varepsilon^2}$ for the true value of $\sigma_\eta^2/\sigma_\varepsilon^2$ (which is not feasible in practice) and finally the non-operational optimal W_{BB}^{opt} matrix. Each figure contains 16 diagrams. The top 8 contain 1-step GMMs results and the bottom 8 the related 2-step GMMs results, where residuals have been used obtained from the 1-step GMMs procedure in that same figure. The figures have four columns of diagrams, which correspond from left to right to the four examined values of $\mu = \{0.5, 1, 2, 4\}$. Each group of 8 diagrams consists of two rows of 4 diagrams: the first row presents bias, i.e. the Monte Carlo estimate of $E(\hat{\gamma} - \gamma)$; the second row depicts relative precision, which is expressed as the Monte Carlo estimate of $\mathrm{RMSE}(\hat{\gamma})/\gamma$ in %.

In Figure 1 we investigate the operational but sub-optimal simple weighting matrix based on $D^{GIV} = I_{2T-2}$. We find negative bias for moderate μ, but positive bias for μ larger and γ small. The bias does not change much between 1-step and 2-step estimation, as already noticed by Blundell and Bond (1998). However, the precision (measured by relative root mean squared error) improves slightly by 2-step estimation (which is in agreement with asymptotic theory). Whether these are general phenomena or is special for the GIV weighting matrix will follow when examining other weighting devices. We note that the precision gets poor when γ is small (which is natural when one divides by γ), but is especially poor when γ is small while μ large.

Figure 2 shows that D^{DPD} yields bias figures shifted in upward direction (especially for large γ values), when compared with those in Figure 1. Therefore the bias is smaller and thus the precision higher for μ moderate. Now 2-step is hardly better than 1-step.

Figure 3 is based on the weighting matrix presented in Windmeijer (2000), which assumes $\mu = 0$. For $\mu \leq 1$ the results are slightly better than those of Figures 1 and 2. But for larger actual values of μ very poor results are obtained when γ is small or moderate.

From Figure 4 we learn that overstating the true value of $\sigma_\eta^2/\sigma_\varepsilon^2$ seems less harmful than understating it. Aiming at point-optimality at $\bar{\sigma}_\eta^2/\bar{\sigma}_\varepsilon^2 = 10$, while neglecting the D_2^{opt} contribution given in (42) to the weighting matrix, is better than the methods in Figures 1 through 4 when γ is small. The method examined in Figure 5 differs from that of Figure 4 in just one respect, viz. that the true value of $\sigma_\eta^2/\sigma_\varepsilon^2$ has been used in weighting matrix (35). This non-operational method removes almost all bias and clearly shows the best precision results obtained at this stage.

Finally we examine the qualities of the optimal weighting matrix W_{BB}^{opt} given in (40). In performing the calculations we experienced a serious problem, viz. that this matrix is not always positive definite. Although we found W_{BB}^{opt} to be non-singular in all our experiments, it was not always positive definite due to the contribution of D_2^{opt} in (42). For $\mu = 0$, when $D_2^{opt} = O$, the optimal weighting matrix was always positive definite, but not for $\sigma_\eta^2 > 0$. We found that, although the matrix is asymptotically proportional to a covariance matrix which is certainly positive definite, in finite sample it may not be, apparently due to the hybrid nature of W_{BB}^{opt}, which is partly determined by data moments and partly by parameters. Whenever W_{BB}^{opt} was not positive definite in our simulations, we removed the D_2^{opt} contribution from the weighting matrix, which then proved to be positive definite. For $\mu = 1$ we could retain D_2^{opt} for the larger γ and smaller T values, but we had to remove it in 60% of the replications for $\gamma = 0.1$ and $T = 10$. However, at $\mu = 4$ this problem emerged much more frequently: in fact in 99% of the replications when T large and γ small, though hardly ever for $\gamma = 0.9$. Hence, strictly speaking we were only able for the case $\mu = 0$ to fully examine GMM with optimal weight matrix. Figure 6 presents results for the adapted procedure. The results do not differ much from those in Figure 5.

In a next version of the paper we will also simulate panel data models with further predetermined regressors, cf. Bun and Kiviet (2006).

5 Conclusions

The GMM system estimator for effect stationary dynamic panel data models exploits two hybrid sets of orthogonality conditions. The covariance matrix of all individual orthogonality condition expressions determines the optimal weighting matrix, and due to the hybrid nature of the orthogonality conditions deriving those covariances is not straightforward. We show that obtaining a conditional covariance suffices, and that different conditioning sets may be chosen for all separate elements of this variance matrix. By choosing for all individual covariances conditioning sets that simplify the analytical derivation of them we find an expression for the optimal weighting matrix for the model that contains the first lag of the dependent variable as an explanatory variable and an arbitrary number of further predetermined regressors. The resulting weighting matrix depends on the in practice unknown parameters of the model and on further characteristics of the data series, viz. the covariance between the regressors and the unobserved individual effects and the autocovariance between predetermined regressors and idiosyncratic disturbances. Hence, it cannot be used to obtain optimal 1-step GMM estimators, but it can serve as a guide in finding close to optimal or point-optimal operational weighting matrices for GMMs.

We performed a number of Monte Carlo experiments in the context of the very specific but simple stationary zero-mean panel data AR(1) model and examined and compared the results of a few implementations of 1 and 2-step GMMs, which differ in the weighting matrix employed in 1-step estimation. We found that the results of 2-step estimation hardly improve upon the corresponding 1-step GMMs results. Hence, the quality of the weighting matrix used in 1-step estimation is found to be very important. Apparently, the theoretical result of asymptotic optimality of 2-step estimation, irrespective of the weighting matrix used in 1-step estimation, has limited relevance in this model for the sample sizes examined. For reasons that still require further study we were not able yet to exploit in a really satisfactory way the derived optimal weighting matrix for GMMs. However, it did lead to an operational though suboptimal weighting matrix which seems superior in many cases to those presently in use (the one that yields the GIV estimator and the weighting matrix used in the DPD program). For this new weighting matrix one has to make a reasonable prior guess of the error-component variance ratio $\sigma_\eta^2/\sigma_\varepsilon^2$.

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.

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-components models. *Journal of Econometrics* 68, 29-51.

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

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.

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

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

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

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

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

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).

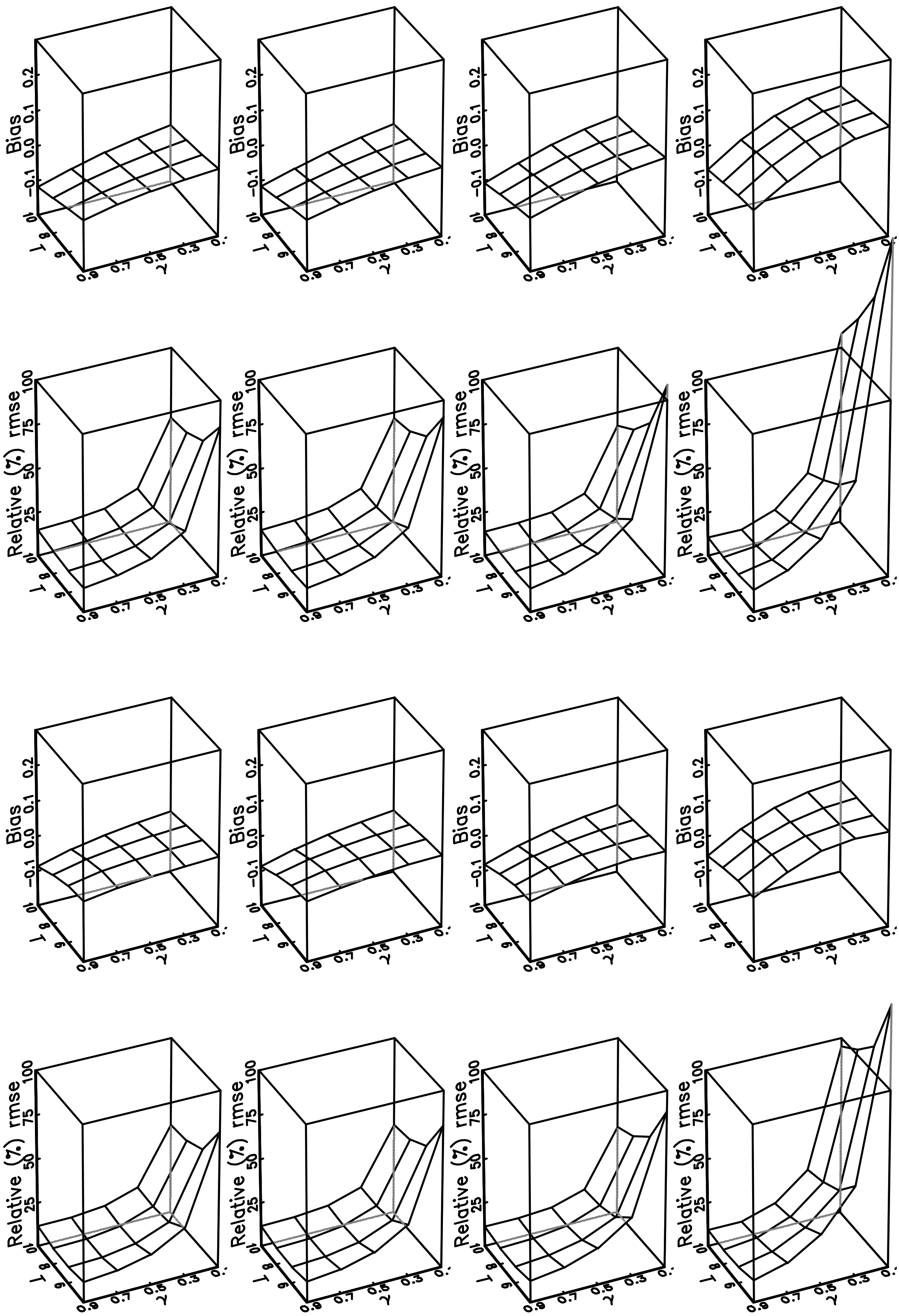

Figure 1. GMMs1 and GMMs2 employing D^{GIV}, for $\mu = 0.5, 1, 2, 4$.

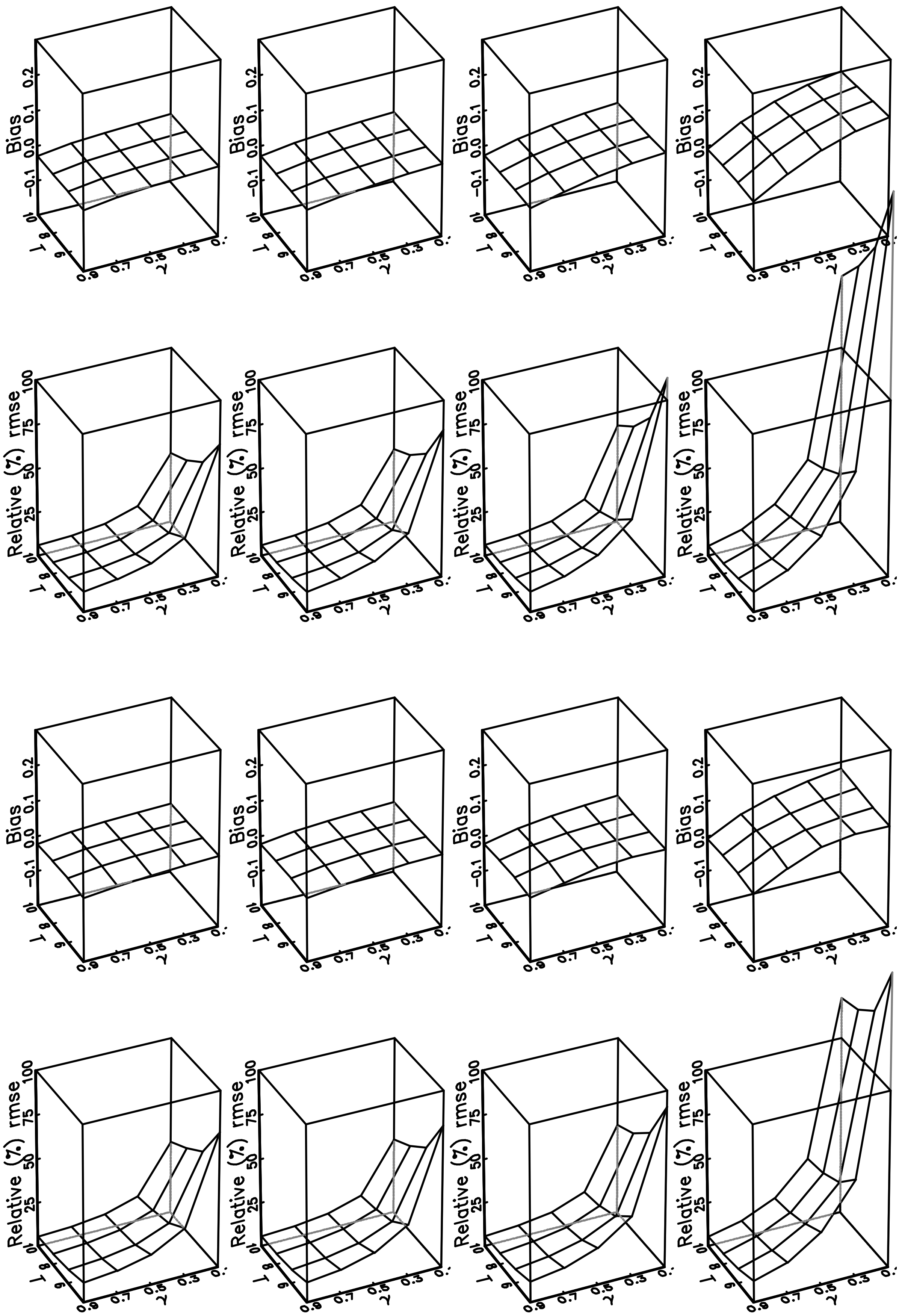

Figure 2. GMMs1 and GMMs2 employing D^{DPD}, $\mu = 0.5, 1, 2, 4$.

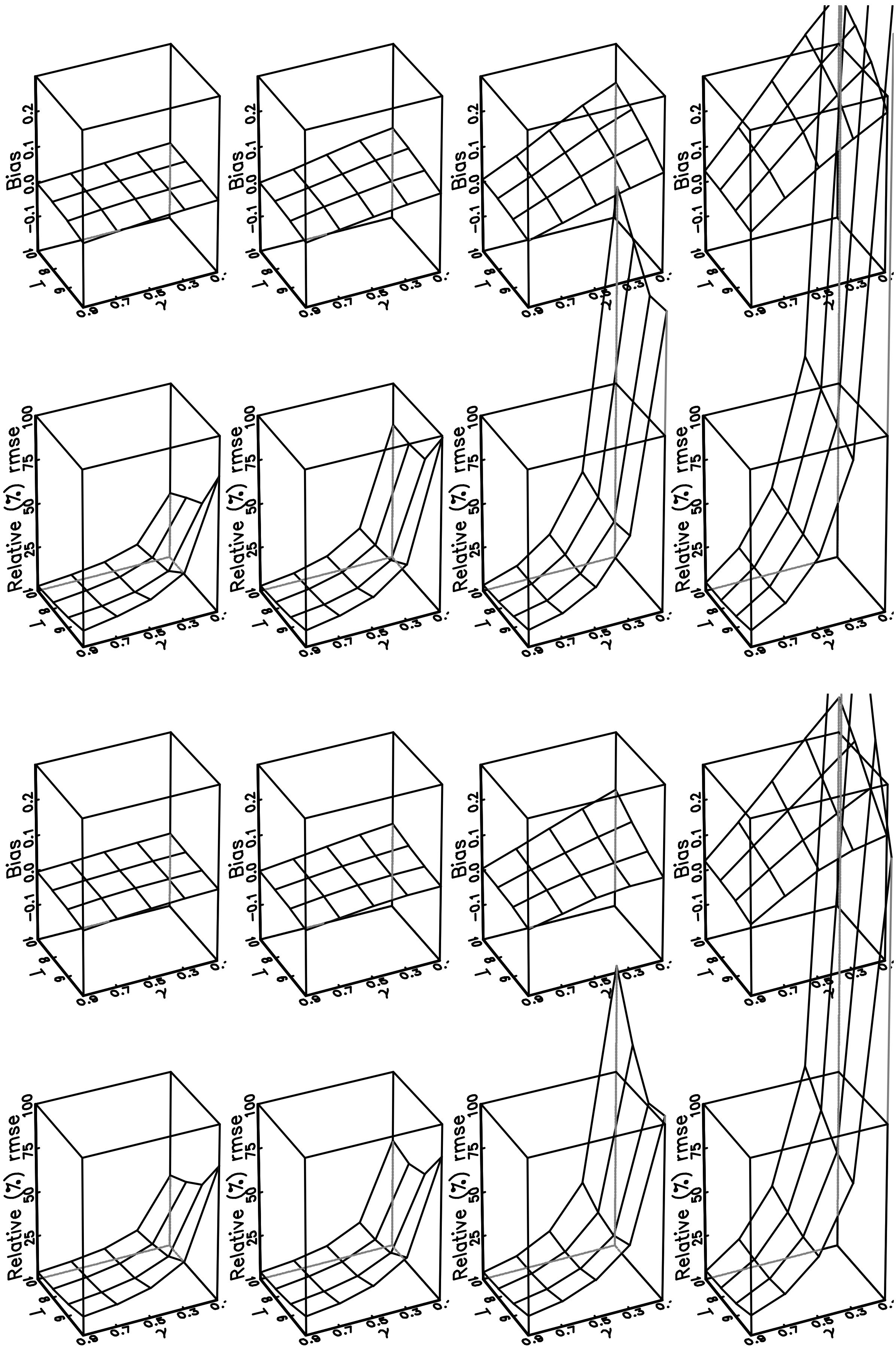

Figure 3. GMMs1 and GMMs2 employing $D^{\bar{\sigma}_\eta^2/\bar{\sigma}_\varepsilon^2}$ with $\bar{\sigma}_\eta^2/\bar{\sigma}_\varepsilon^2 = 0$, for $\mu = 0.5, 1, 2, 4$.

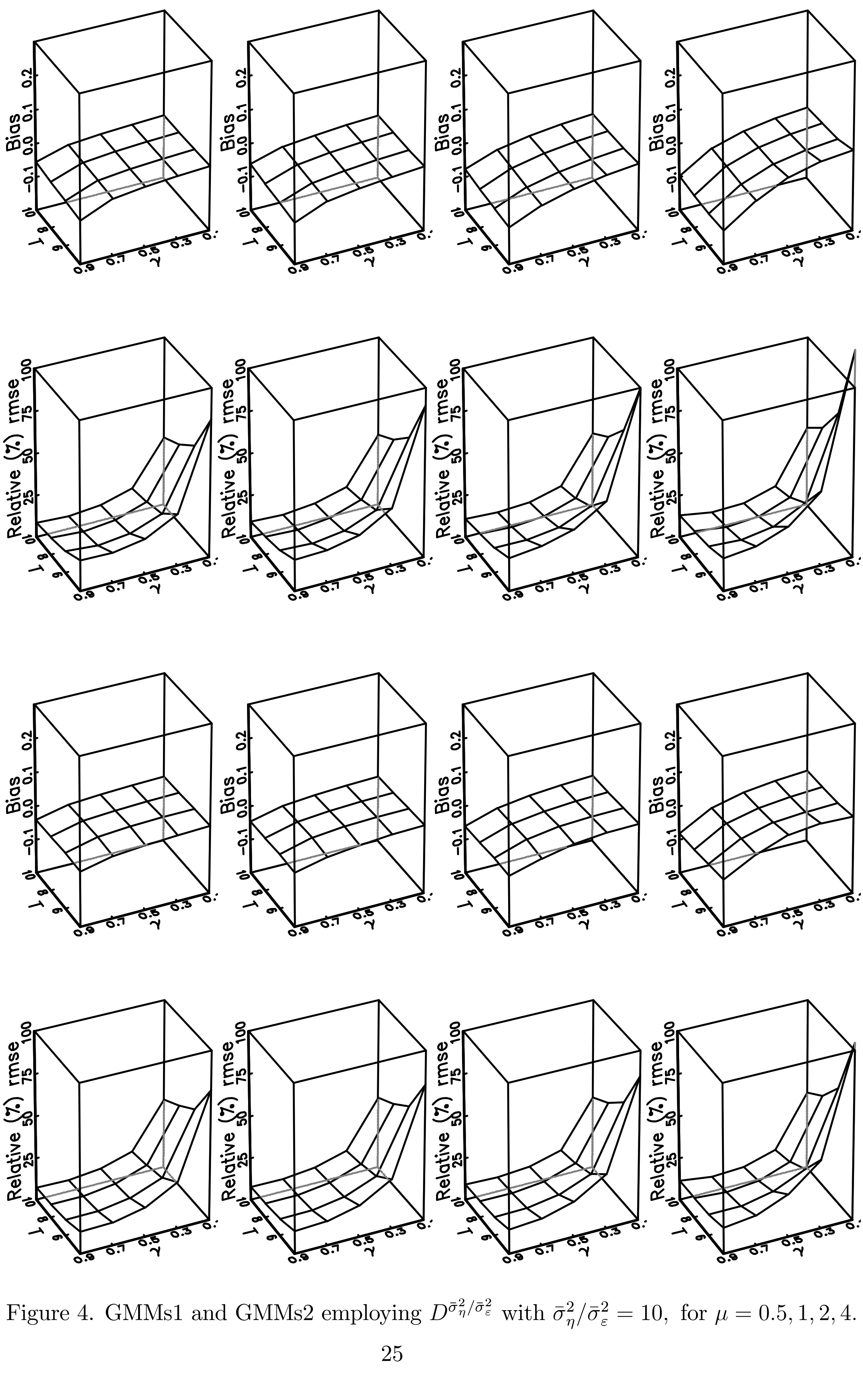

Figure 4. GMMs1 and GMMs2 employing $D^{\bar{\sigma}_\eta^2/\bar{\sigma}_\varepsilon^2}$ with $\bar{\sigma}_\eta^2/\bar{\sigma}_\varepsilon^2 = 10$, for $\mu = 0.5, 1, 2, 4$.

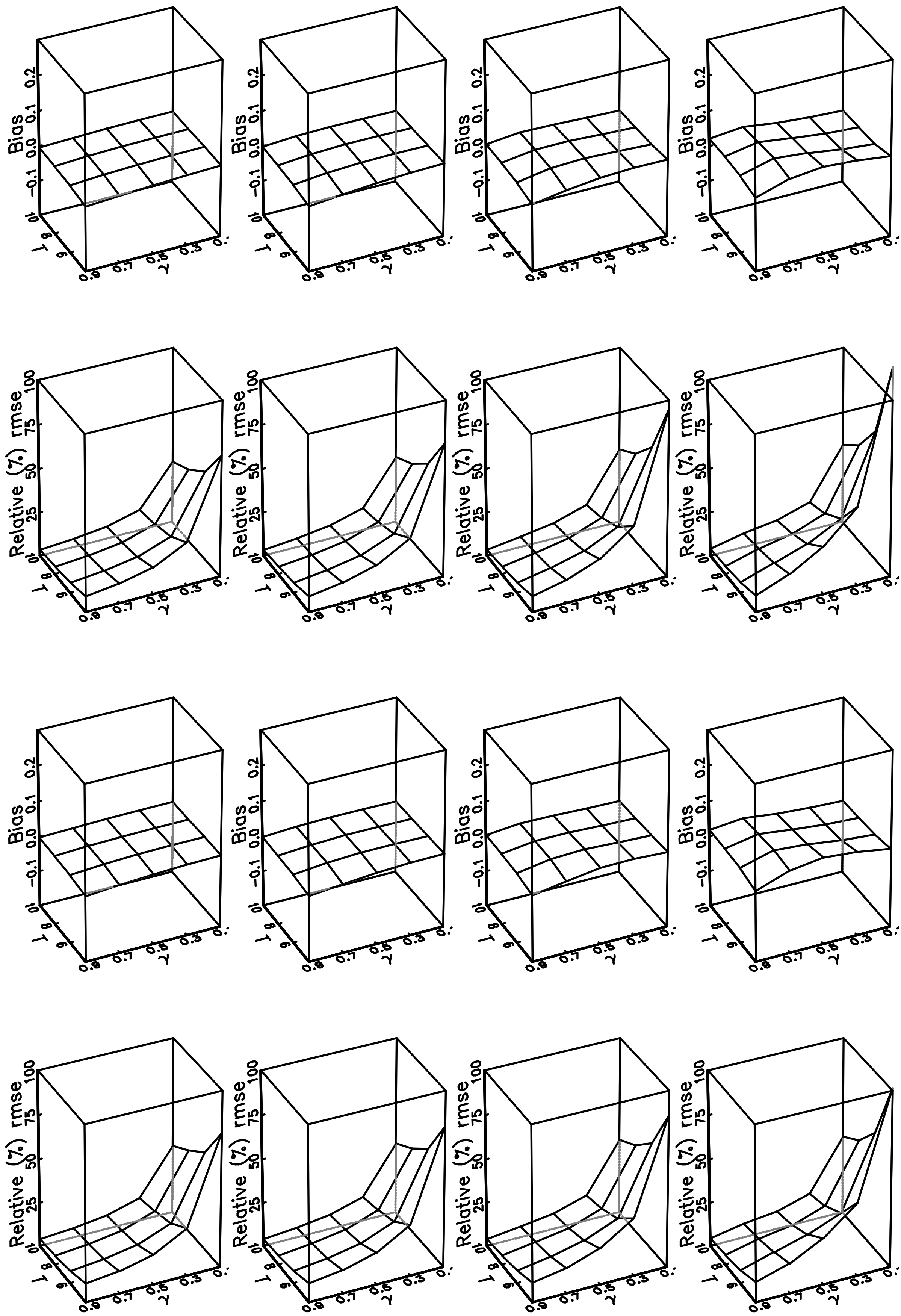

Figure 5. GMMs1 and GMMs2 employing $D^{\sigma_\eta^2/\sigma_\varepsilon^2}$ (non-operational), for $\mu = 0.5, 1, 2, 4$.

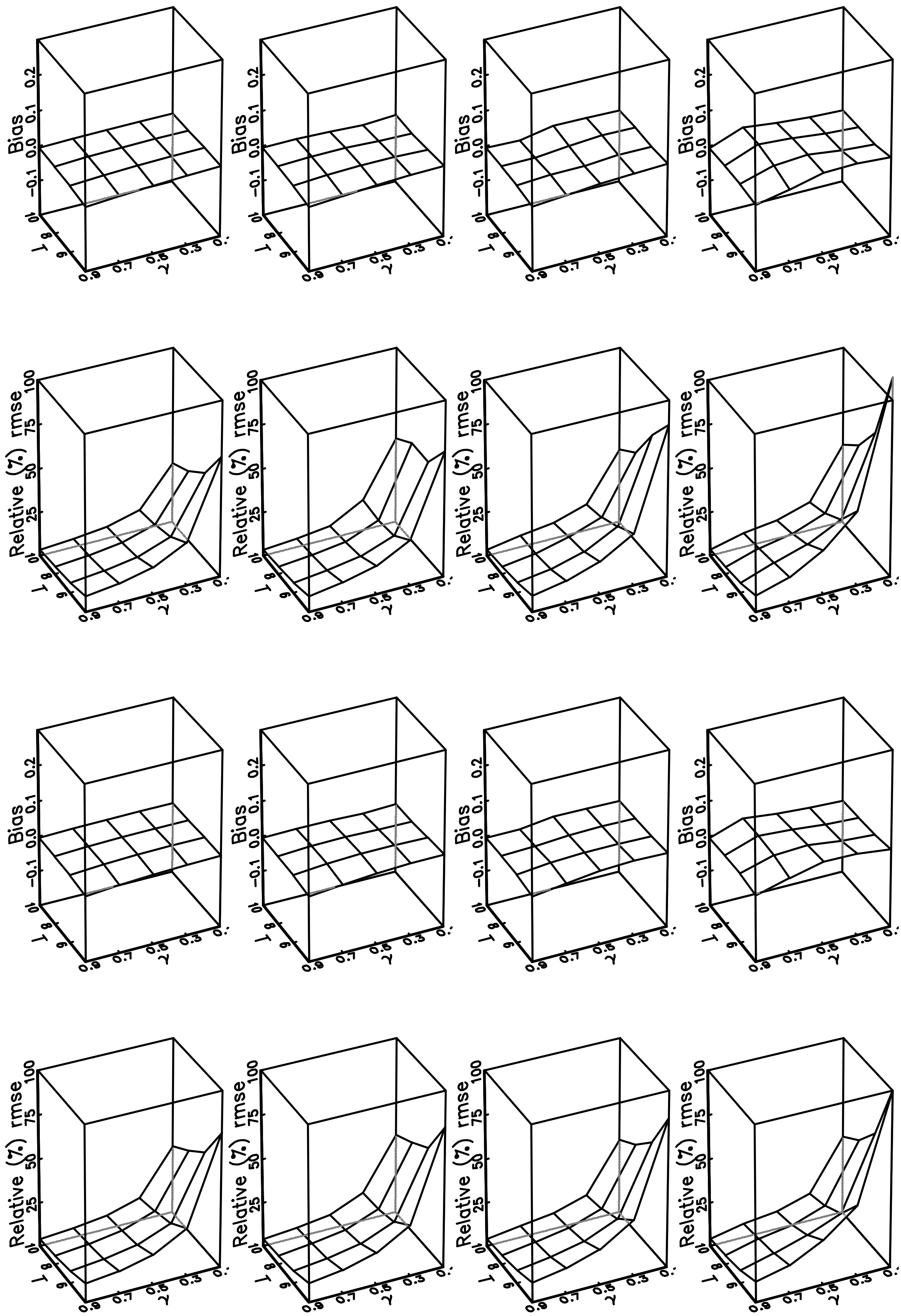

Figure 6. GMMs1 and GMMs2 employing W_{BB}^{opt}-adapted (non-operational), $\mu = 0.5, 1, 2, 4$.