A Reduced Bias GMM-like Estimator with Reduced Estimator Dispersion

Jerry Hausman, Konrad Menzel, Randall Lewis, and Whitney Newey*

Draft, September 14, 2007

2SLS is by far the most-used estimator for the simultaneous equation problem. However, it is now well-recognized that 2SLS can exhibit substantial finite sample (second-order) bias when the model is over-identified and the first stage partial R2 is low. The initial recommendation to solve this problem was to do LIML, e.g. Bekker (1994) or Staiger and Stock (1997). However, Hahn, Hausman, and Kuersteiner (HHK 2004) demonstrated that the "moment problem" of LIML led to undesirable estimates in this situation. Morimune (1983) analyzed both the bias in 2SLS and the lack of moments in LIML. While it was long known that LIML did not have finite sample moments, it was less known that this lack of moments led to the undesirable property of considerable dispersion in the estimates, e.g. the inter-quartile range was much larger than 2SLS. HHK developed a jackknife 2SLS (J2SLS) estimator that attenuated the 2SLS bias problem and had good dispersion properties. They found in their empirical results that the J2SLS estimator or the Fuller estimator, which modifies LIML to have moments, did well on both the bias and dispersion criteria. Since the Fuller estimator had smaller second order MSE, HHK recommended using the Fuller estimator.

However, Bekker and van der Ploeg (2005) and Hausman, Newey and Woutersen (HNW 2005) recognized that both Fuller and LIML are inconsistent with heteroscedasticity as the number of instruments becomes large in the Bekker (1994) sequence. Since econometricians recognize that heteroscedasticity is often present, this finding presents a problem. Hausman, Newey, Woutersen, Chao and Swanson (HNWCS 2007) solve this problem by proposing jackknife LIML (HLIML) and jackknife Fuller (HFull) estimators that are consistent in the presence of heteroscedasticity. HLIML does not have moments so HNWCS (2007) recommend using HFull, which does have moments. However, a problem remains. If serial correlation or clustering exists, neither HLIML nor HFull is consistent.

*jhausman@mit.edu, menzel@mit.edu, randalll@mit.edu, wnewey@mit.edu. Newey thanks the NSF for research support.

The continuous updating estimator, CUE, which is the GMM-like generalization of LIML, introduced by Hansen, Heaton, and Yaron (1996) would solve this problem. The CUE estimator also allows treatment of non-linear specifications which the above estimators need not allow for and also allows for general non-spherical disturbances. However, CUE suffers from the moment problem and exhibits wide dispersion.[1] GMM does not suffer from the no moments problem, but like 2SLS, GMM has finite sample bias that grows with the number of moments.

In this paper we modify CUE to solve the no moments/large dispersion problem. We consider the dual formulation of CUE and we modify the CUE first order conditions by adding a term of order T^{-1}. To first order the variance of the estimator is the same as GMM or CUE, so no large sample efficiency is lost. The resulting estimator has moments up to the degree of overidentification and demonstrates considerably reduced bias relative to GMM and reduced dispersion relative to CUE. Thus, we expect the new estimator will be useful for empirical research. We next consider a similar approach but use a class of functions which permits us to specify an estimator with all integral moments existing. Lastly, we demonstrate how this approach can be extended tothe entire family of Maximum Empirical Likelihood (MEL) Estimators, so these estimators will have integral moments of all orders.

In the next section we specify a general non-linear model of the type where GMM is often used. We then consider the finite-sample bias of GMM. We discuss how CUE partly removes the bias but evidence demonstrates the presence of the no moments/large dispersion problem. We next consider LIML and associated estimators under large instrument (moment) asymptotics and discuss why LIML, Fuller and HLIM and HFull will be inconsistent when time series or spatial correlation is present. In the next section we consider the no moment problem for CUE and demonstrate the source of the no moment condition. We then consider a general modification to the CUE objective function that solves the no moment problem. We then give a specific modification to CUE and specify the "regularized" CUE estimator, RCUE, and we prove that the estimator has moments. Lastly, we investigate or new estimators empirically. We find that they do not have either the bias arising from a large number of instruments (moments) or the no moment problem of CUE and LIML and perform well when heteroscedasticity and serial correlation are present. Thus, while further research is required to determine the optimal parameters of the modifications

[1] While a formal proof of absence of moments of CUE does not exist, empirical investigation e.g. Guggenberger (2005), demonstrate that CUE likely has not moments. In the spherical disturbance case, CUE is LIML, which is known to have no moments.

we present, the current estimators should allow improved empirical results in applied research.

1 Setup

Consider a sample of T observations of random vectors (W_t, Z_t) satisfying the conditional moment restriction

$$\mathbb{E}[\varrho(W_t, \beta)|Z_t] = 0$$

where β is a K-dimensional parameter. E.g. in the linear instrumental variables case with a correctly specified first stage,

$$\varrho(w_t, \beta) = y_t - x_t\beta \tag{1}$$

and

$$\mathbb{E}\left[\left.\frac{\partial}{\partial\beta}\varrho(W_t, \beta)\right| Z_t = z_t\right] = \mathbb{E}\left[X_t|Z_t = z_t\right] = z_t\pi \tag{2}$$

For purposes of estimation, we will consider an M-dimensional vector of unconditional moment restrictions

$$\mathbf{g}_t(\beta) = \mathbf{g}(w_t, \beta) = \mathbf{h}(z_t)\varrho(w_t, \beta) =: h_t\varrho(w_t, \beta)$$

where h_t is a K-dimensional vector of functions of z_t.

The two-step GMM estimator for this problem minimizes

$$Q(\beta) := T\bar{\mathbf{g}}(\beta)'\hat{\Omega}^{-1}\bar{\mathbf{g}}(\beta)$$

where

$$\bar{\mathbf{g}}(\beta) := \frac{1}{T}\sum_{t=1}^{T} g_t(\beta)$$

and $\hat{\Omega}$ is some consistent preliminary estimator of the variance covariance matrix of the moments. On the other hand, the CUE estimator in Hansen Heaton and Yaron (1996) minimizes the criterion

$$\tilde{Q}(\beta) := T\bar{\mathbf{g}}(\beta)'\hat{\mathbf{\Omega}}(\beta)^{-1}\bar{\mathbf{g}}(\beta)$$

where

$$\hat{\mathbf{\Omega}}(\beta) := \frac{1}{T}\sum_{s,t\leq T} w_T(s,t)\mathbf{g}_s(\beta)\mathbf{g}_t(\beta)' = \frac{1}{T}\sum_{s,t\leq T} w_T(s,t)h_s h_t' \varrho(w_s,\beta)\varrho(w_t,\beta)$$

is a consistent estimator of the variance-covariance matrix of the moment vector for an appropriate choice of $w_T(s,t)$. In the absence of autocorrelation, $w_T(s,t) = \mathbb{1}\{s=t\}$, for cluster sampling $w_T(s,t) = \mathbb{1}\{G_s = G_t\}$ is an indicator for observations s and t belonging to the same cluster, and under general autocorrelation, $w_T(s,t)$ are more general kernel weights (see Newey and West (1987) and Andrews (1991)).

For the following, denote

$$\begin{aligned}
\bar{\mathbf{G}}_k(\beta) &:= \frac{1}{T}\sum_{t\leq T} \frac{\partial}{\partial\beta_k}\mathbf{g}_t(\beta')\bigg|_{\beta'=\beta} \\
\bar{\mathbf{C}}_k(\beta) &:= \frac{1}{T}\sum_{s,t\leq T} w_T(s,t)\,\frac{\partial}{\partial\beta_k}\mathbf{g}_s(\beta')\bigg|_{\beta'=\beta}\mathbf{g}_t(\beta)'
\end{aligned}$$

where without loss of much generality, we suppose throughout that the parameter of interest β is a scalar. Also,

$$\begin{aligned}
\varrho_t &:= \varrho(w_t,\beta_0) \\
\nu_{kt} &:= \frac{\partial}{\partial\beta_k}\varrho(w_t,\beta_0) - \mathbb{E}\left[\frac{\partial}{\partial\beta_k}\varrho(w_t,\beta_0)\bigg| Z_t = z_t\right] \\
\zeta_{klt} &:= \frac{\partial^2}{\partial\beta_k\partial\beta_l}\varrho(w_t,\beta_0) - \mathbb{E}\left[\frac{\partial^2}{\partial\beta_k\partial\beta_l}\varrho(w_t,\beta_0)\bigg| Z_t = z_t\right] \\
\xi_{klmt} &:= \frac{\partial^3}{\partial\beta_k\partial\beta_l\partial\beta_m}\varrho(w_t,\beta_0) - \mathbb{E}\left[\frac{\partial^3}{\partial\beta_k\partial\beta_l\partial\beta_m}\varrho(w_t,\beta_0)\bigg| Z_t = z_t\right]
\end{aligned}$$

2 Finite-Sample Bias in GMM

Using the notation introduced above, efficient two-step GMM solves

$$\bar{\mathbf{G}}(\hat{\beta}_{GMM})'\hat{\Omega}^{-1}\bar{\mathbf{g}}(\hat{\beta}_{GMM}) = 0$$

where $\hat{\Omega}$ is a consistent preliminary estimator of the covariance matrix of $\bar{\mathbf{g}}(\beta_0)$. For the purposes of this section, we will in the following take β to be a scalar, and treat Ω as known so we can ignore the potential bias from using a consistent estimator of the optimal weighting matrix instead.[2]

Denoting

$$\begin{aligned}
\bar{\mathbf{G}}_2(\beta) &:= \frac{1}{T}\sum_{t\le T}\frac{\partial^2}{\partial\beta^2}\mathbf{g}_t(\beta')\bigg|_{\beta'=\beta} \\
\bar{\mathbf{G}}_3(\beta) &:= \frac{1}{T}\sum_{t\le T}\frac{\partial^3}{\partial\beta^3}\mathbf{g}_t(\beta')\bigg|_{\beta'=\beta}
\end{aligned}$$

from a second-order expansion of the first-order conditions about β_0, we get the approximation

$$\begin{aligned}
0 &= \bar{\mathbf{G}}(\beta_0)'\hat{\Omega}^{-1}\bar{\mathbf{g}}(\beta_0) + \left[\bar{\mathbf{G}}(\beta_0)'\hat{\Omega}^{-1}\bar{\mathbf{G}}(\beta_0) + \bar{\mathbf{G}}_2(\beta_0)'\hat{\Omega}^{-1}\bar{\mathbf{g}}(\beta_0)\right](\hat{\beta}_{GMM}-\beta_0) \\
&\quad + \frac{1}{2}\left[\bar{\mathbf{G}}_3(\beta_0)'\hat{\Omega}^{-1}\bar{\mathbf{g}}(\beta_0) + 2\bar{\mathbf{G}}_2(\beta_0)'\hat{\Omega}^{-1}\bar{\mathbf{G}}(\beta_0)\right](\hat{\beta}_{GMM}-\beta_0)^2 + o_p(|\hat{\beta}_{GMM}-\beta_0|^2)
\end{aligned}$$

Following the argument in Lemma 3.3 in Rilstone et al. (1996), we can replace $(\hat{\beta}_{GMM}-\beta_0)^2$ by the square of the first-order approximation

$$\hat{\beta}_{GMM}-\beta_0 = -\frac{\bar{\mathbf{G}}(\beta_0)'\hat{\Omega}^{-1}\bar{\mathbf{g}}(\beta_0)}{\bar{\mathbf{G}}(\beta_0)'\hat{\Omega}^{-1}\bar{\mathbf{G}}(\beta_0)} + O_p(T^{-1}) =: \psi + O_p(T^{-1})$$

and obtain an approximation up to terms of $O_p(T^{-1})$. Defining

$$\bar{m}_{lt}(\beta) := -\mathbb{E}\left[\frac{\partial^l}{\partial\beta^l}\varrho(W_t,\beta)\bigg| Z_t = z_t\right]$$

for $l = 1,2,3$, we can write

$$\sqrt{T}(\hat{\beta}_{GMM}-\beta_0) = \tilde{D}^{-1}\tilde{n} + o_p(T^{-1})$$

[2]For the general case, see Newey and Smith (2004)

where

$$
\begin{aligned}
\tilde{n} &= \frac{\bar{\mathbf{m}}_1(\beta_0)'\mathbf{H}\Omega^{-1}\mathbf{H}'\varrho}{\sqrt{T}} + \frac{1}{\sqrt{T}}\nu'\mathbf{H}\Omega^{-1}\mathbf{H}'\varrho + \frac{1}{\sqrt{T}}\left[\frac{1}{2}\bar{\mathbf{G}}_3(\beta_0)'\Omega^{-1}\bar{\mathbf{g}}(\beta_0) + \bar{\mathbf{G}}_2(\beta_0)'\Omega^{-1}\bar{\mathbf{G}}(\beta_0)\right]\psi^2 \\
&= \frac{\bar{\mathbf{m}}_1(\beta_0)'\mathbf{H}\Omega^{-1}\mathbf{H}'\varrho}{\sqrt{T}} + \frac{1}{\sqrt{T}}\nu'\mathbf{H}\Omega^{-1}\mathbf{H}'\varrho + \frac{1}{\sqrt{T}}\frac{\bar{\mathbf{m}}_2(\beta_0)'\mathbf{H}'\Omega^{-1}\mathbf{H}'\bar{\mathbf{m}}_1(\beta_0)}{\bar{\mathbf{m}}_1(\beta_0)'\mathbf{H}'\Omega^{-1}\mathbf{H}'\bar{\mathbf{m}}_1(\beta_0)} + O_p(T^{-1}) \\
\tilde{D} &= \frac{\bar{\mathbf{m}}_1(\beta_0)'\mathbf{H}\Omega^{-1}\mathbf{H}'\bar{\mathbf{m}}_1(\beta_0)}{T} + \frac{1}{T}\nu'\mathbf{H}\Omega^{-1}\mathbf{H}'\nu + \frac{2}{\sqrt{T}}\frac{\bar{\mathbf{m}}_1(\beta_0)'\mathbf{H}\Omega^{-1}\mathbf{H}'\nu}{\sqrt{T}} \\
&\quad + \frac{1}{\sqrt{T}}\frac{\bar{\mathbf{m}}_2(\beta)'\mathbf{H}\Omega^{-1}\mathbf{H}'\varrho}{\sqrt{T}} + \frac{1}{T}\zeta'\mathbf{H}\Omega^{-1}\mathbf{H}'\varrho
\end{aligned}
$$

Expanding the approximation from above in $\frac{1}{\sqrt{T}}$ around its limit and keeping terms of order up to $\frac{1}{\sqrt{T}}$, we obtain

$$
\begin{aligned}
\sqrt{T}(\hat{\beta}_{GMM} - \beta_0) &\approx \frac{1}{\sqrt{T}}\frac{\bar{\mathbf{m}}_1(\beta_0)'\mathbf{H}\Omega^{-1}\mathbf{H}'\varrho}{D} + \frac{1}{\sqrt{T}}\frac{\nu'\mathbf{H}\Omega^{-1}\mathbf{H}'\varrho}{D} + \frac{1}{\sqrt{T}}\frac{\bar{\mathbf{m}}_2(\beta_0)'\mathbf{H}'\Omega^{-1}\mathbf{H}'\bar{\mathbf{m}}_1(\beta_0)}{D^2} \\
&\quad - \frac{1}{\sqrt{T}}\frac{2\bar{\mathbf{m}}_1(\beta_0)'\mathbf{H}\Omega^{-1}\mathbf{H}'\varrho\nu'\mathbf{H}\Omega^{-1}\mathbf{H}'\bar{\mathbf{m}}_1(\beta_0) + \bar{\mathbf{m}}_1(\beta_0)'\mathbf{H}\Omega^{-1}\mathbf{H}'\varrho\varrho'\mathbf{H}\Omega^{-1}\mathbf{H}'\bar{\mathbf{m}}_2(\beta_0)}{D^2}
\end{aligned}
$$

where

$$
D := \frac{\bar{\mathbf{m}}_1(\beta_0)'\mathbf{H}\Omega^{-1}\mathbf{H}'\bar{\mathbf{m}}_1(\beta_0)}{T}
$$

Taking expectations,

$$
\mathbb{E}[\hat{\beta}_{GMM}] - \beta_0 \approx \frac{1}{T}\left[\frac{\text{tr }(\Omega^{-1}\Omega_{\varrho\nu})}{D} - 2\frac{\bar{\mathbf{m}}_1(\beta_0)'\mathbf{H}\Omega_{\varrho\varrho}^{-1}\Omega_{\varrho\nu}\Omega^{-1}\mathbf{H}'\bar{\mathbf{m}}_1(\beta_0)}{D^2}\right]
$$

where we denote $\Omega_{\varrho\varrho} := \mathbb{E}[\mathbf{H}'\varrho\varrho'\mathbf{H}|\mathbf{Z}] = \Omega$ and $\Omega_{\varrho\nu} := \mathbb{E}[\mathbf{H}'\varrho\nu'\mathbf{H}|\mathbf{Z}]$.

In the absence of autocorrelation, as in Theorem 4.5 in Newey and Smith,

$$
\begin{aligned}
\mathbb{E}[\hat{\beta}_{GMM}] - \beta_0 &\approx \frac{\mathbb{E}\left[\nu_t\varrho_t\mathbf{h}_t\Omega^{-1}\mathbf{h}_t'\right]}{TD} - 2\frac{\mathbb{E}\left[\nu_t\varrho_t\mathbf{h}_t\Omega^{-1}\mathbf{H}'\bar{\mathbf{m}}_1(\beta_0)\bar{\mathbf{m}}_1(\beta_0)'\mathbf{H}\Omega^{-1}\mathbf{h}_t'\right]}{TD^2} \\
&= \frac{\mathbb{E}\left[\nu_t\varrho_t\mathbf{h}_t\mathbf{P}_1\mathbf{h}_t'\right]}{TD} \geq (M-2)\frac{\kappa}{TD}
\end{aligned}
$$

where $\mathbf{P}_1 = \mathbf{\Omega}^{-1/2}(\mathbf{I} - 2\mathbf{P}_{\mathbf{\Omega^{-1}H'm_1}})\mathbf{\Omega}^{-1/2}$ and $\kappa := \min_t\left\{\frac{\mathbb{E}[\varrho_t\nu_t|Z_t]}{\sigma_t^2}\right\}$. Therefore the bias of one-step GMM using the efficient weighting matrix increases linearly in the number of moments.

The CUE estimator is known to remove this bias from the correlation between the moment functions and their Jacobians (see e.g. Donald and Newey (2000), Newey and Smith (2004)), but simulation studies suggest that like LIML in the case of the linear model with spherical errors, it may suffer from a no-moments problem in weakly identified moments (Guggenberger (2006)). Below, we will give a heuristic argument why CUE should be expected to have no integral moments in many common settings and suggest a modification of CUE to fix this problem.

3 Inconsistency of LIML with many Instruments and Clustering

In the following section, we show the inconsistency of LIML for clustered samples in order to illustrate the need for an estimator which remains higher-order unbiased in the presence of autocorrelation.

With many weak linear moment conditions, LIML can be thought of as the CUE estimator using an estimator of the covariance matrix which is only consistent under homoskedasticity,

$$\hat{\mathbf{\Omega}}(\beta) := \left[\frac{1}{T}\sum_{t=1}^{T}(y_t - x_t\beta)^2\right]\frac{1}{T}Z'Z$$

Therefore the expectation of the CUE objective function is equal to

$$\mathbb{E}[\tilde{Q}(\beta)] \approx \frac{(T-1)\mathbb{E}[\mathbf{g}_t(\beta)]'\mathbb{E}[Z_tZ_t']^{-1}\mathbb{E}[\mathbf{g}_t(\beta)]}{\mathbb{E}[(Y_t - X_t\beta)^2]} + \frac{\text{tr}\,\left(\mathbb{E}[Z_tZ_t']^{-1}\mathbf{\Omega}(\beta)\right)}{\mathbb{E}[(Y_t - X_t\beta)^2]} = \frac{T-1}{T}Q_0(\beta) + \frac{\text{tr}\,\left(\mathbb{E}[Z_tZ_t']^{-1}\mathbf{\Omega}(\beta)\right)}{\mathbb{E}[(Y_t - X_t\beta)^2]}$$

where

$$\mathbf{\Omega}(\beta) = \text{Var}\left(\frac{1}{T}Z'(y - X\beta)\right) \neq \mathbb{E}[(Y_t - X_t\beta)^2]\frac{1}{T}Z'Z$$

so that the expectation of the LIML objective function differs from the limiting objective function $Q_0(\beta)$ by a term which will in general depend on the parameter β except when (Y_t, X_t) are i.i.d. conditional on Z_t.

As an illustrative example, suppose we have a sample of clusters $g = 1, \ldots, G$ where cluster g contains the observations $i = 1, \ldots, n_g$, and we want to estimate the linear two-equation model

$$\begin{aligned} y_{ig} &= \alpha + x_{ig}\beta + \varepsilon_{ig} \\ x_{ig} &= z_{ig}\pi + \nu_{ig} \end{aligned}$$

using a k-dimensional vector of instrumental variables, z_{ig} satisfying

$$\mathbb{E}[\varepsilon_{ig}|Z_{ig}=z_{ig}]=\mathbb{E}[\nu_{ig}|Z_{ig}=z_{ig}]=0$$

Denoting

$$\sigma_{\varepsilon\varepsilon}:=\mathbb{E}[\varepsilon_{ig}^2|Z_{ig}] \qquad \sigma_{\varepsilon\nu}:=\mathbb{E}[\varepsilon_{ig}\nu_{ig}|Z_{ig}]$$

and for $i\neq j$,

$$r_{\varepsilon\varepsilon}\sigma_{\varepsilon\varepsilon}:=\mathbb{E}[\varepsilon_{ig}\varepsilon_{jg}|Z_{ig},Z_{jg}] \qquad r_{\varepsilon\nu}\sigma_{\varepsilon\varepsilon}:=\mathbb{E}[\varepsilon_{ig}\nu_{jg}|Z_{ig},Z_{jg}]$$

and assuming that observations are independent across clusters, i.e. for $g\neq h$,

$$\mathbb{E}[\varepsilon_{ig}\varepsilon_{jh}|Z_{ig},Z_{jh}]=\mathbb{E}[\varepsilon_{ig}\nu_{jh}|Z_{ig},Z_{jh}]=0$$

the first-order conditions for LIML with respect to β evaluated at the true parameter (β_0,π_0) can be written as

$$\begin{aligned}\sigma_{\varepsilon\varepsilon}\,\frac{\partial}{\partial\beta}\tilde{Q}(\beta)\Big|_{\beta=\beta_0} &= x'P_Z\varepsilon+\frac{\varepsilon'P_Z\varepsilon}{\varepsilon'\varepsilon}x'\varepsilon\\ &= \sum_{g,h\leq G}\sum_{\substack{i\leq n_g\\ j\leq n_h}}\varepsilon_{ig}(z_{jh}\pi_0+\nu_{jh})P_{(ig,jh)}-\frac{\sum_{g=1}^{G}\sum_{i=1}^{n_g}\varepsilon_{ig}(z_{ig}\pi_0+\nu_{ig})}{\sum_{g=1}^{G}\sum_{i=1}^{n_g}\varepsilon_{ig}^2}\sum_{g,h\leq G}\sum_{\substack{i\leq n_g\\ j\leq n_h}}\varepsilon_{ig}\varepsilon_{jh}P_{(ig,jh)}\end{aligned}$$

Taking expectations,

$$\frac{\partial}{\partial\beta}\tilde{Q}(\beta)\Big|_{\beta=\beta_0}=\left(r_{\varepsilon\nu}-\frac{r_{\varepsilon\varepsilon}\sigma_{\varepsilon\nu}}{\sigma_{\varepsilon\varepsilon}}\right)\sum_{g=1}^{G}\sum_{i\neq j}P_{(ig,jg)}$$

which will in general be different from zero unless the clustering coefficients for ε and that of its interaction with ν are the same. Under large-G/large n_g asymptotics LIML is of course still consistent since the minimal eigenvalue converges to zero in probability, but ignoring autocorrelation may lead to inconsistency of LIML under Bekker (1994) many-moment asymptotics.

Therefore estimators which implicitly rely on an independence assumption like LIML, JIVE or HLIM (see Hausman Newey and Woutersen, 2007) will likely fail to remove finite-sample bias in common empirical settings exhibiting time-series or spatial correlation and/or clustering since these methods do not remove the correlations between moments and their Jacobians across observations from the objective function. In contrast, CUE with a correctly specified weighting matrix has been shown to remove that component of the bias (Donald and Newey (2000), Newey and Smith (2004)).

4 The No-Moment Problem for CUE

The source of the no-moment problem is that, as we will now show, under weak identification the true parameter β_0 does not satisfy the second-order condition for a minimum, $\nabla_{\beta\beta}\tilde{Q}(\beta_0) > 0$ in the positive-definite matrix sense with probability 1. To see why this will result in poor performance of CUE, consider again the case of linear moment restrictions,

$$g_t(\beta) = z_t(y_t - x_t\beta)$$

for which

$$\hat{\beta}_{CUE} = -\left[\nabla_{\beta\beta}\tilde{Q}(\hat{\beta}_{CUE})\right]^{-1}\nabla_{\beta}\tilde{Q}(\hat{\beta}_{CUE}) + o_p(1)$$

If the second-order condition fails with probability greater than zero, $\nabla_{\beta\beta}\tilde{Q}(\hat{\beta})$ will be near singular with some probability,[3] resulting in "extreme" estimates for β.

Denoting derivatives with subscripts, in the scalar parameter case the first-order conditions for CUE can be written as

$$\begin{aligned} 0 = \frac{\partial}{\partial\beta_k}\tilde{Q}(\beta) &= T\bar{\mathbf{g}}(\beta)'\hat{\mathbf{\Omega}}(\beta)^{-1}\bar{\mathbf{G}}_k(\beta) - T\bar{\mathbf{g}}(\beta)'\hat{\mathbf{\Omega}}(\beta)^{-1}\bar{\mathbf{C}}_k(\beta)\hat{\mathbf{\Omega}}(\beta)^{-1}\bar{\mathbf{g}}(\beta) \\ &= T\bar{\mathbf{g}}(\beta)'\hat{\mathbf{\Omega}}(\beta)^{-1}\bar{\mathbf{G}}_k(\beta) - T\mathrm{tr}\,\left\{\hat{\mathbf{\Omega}}(\beta)^{-1}\bar{\mathbf{C}}_k(\beta)\hat{\mathbf{\Omega}}(\beta)^{-1}\bar{\mathbf{g}}(\beta)\bar{\mathbf{g}}(\beta)'\right\} \\ &= \mathrm{tr}\,\left\{\hat{\mathbf{\Omega}}(\beta)^{-1}\left[T\bar{\mathbf{G}}_k(\beta)\bar{\mathbf{g}}(\beta)' - \bar{\mathbf{C}}_k(\beta)\right]\right\} + \mathrm{tr}\,\left\{\hat{\mathbf{\Omega}}(\beta)^{-1}\bar{\mathbf{C}}_k(\beta)\hat{\mathbf{\Omega}}(\beta)^{-1}\left[\hat{\mathbf{\Omega}}(\beta) - T\bar{\mathbf{g}}(\beta)\bar{\mathbf{g}}(\beta)'\right]\right\} \end{aligned}$$

for each $k = 1, \ldots K$.

In order to verify that the second-order condition for a minimum at β_0 doesn't hold with probability 1, it is sufficient to show that the expectation of the first derivative of $\nabla_{\beta\beta}\tilde{Q}(\beta)$ is singular, or has deficient

[3] By continuity of the eigenvalues and the intermediate value theorem.

rank. We can write

$$\tilde{Q}_{\beta_l\beta_k}(\beta_0) = \text{tr}\ \left(\hat{\Omega}(\beta_0)^{-1}[B_1 + B_2 + 2B_3] + B_4\right) \tag{3}$$

where

$$\begin{aligned}
B_1 &= T\bar{\mathbf{G}}_l(\beta_0)\bar{\mathbf{G}}_k(\beta_0)' - \frac{1}{T}\sum_{s,t\leq T} w_T(s,t)\ \frac{\partial}{\partial\beta_l}\mathbf{g}_s(\beta)\frac{\partial}{\partial\beta_k}\mathbf{g}_t(\beta)'\Big|_{\beta=\beta_0} \\
B_2 &= \left[\sum_{t=1}^{T}\frac{\partial^2}{\partial\beta_l\partial\beta_k}g_t(\beta)\Big|_{\beta=\beta_0}\right]\bar{\mathbf{g}}(\beta_0) - \frac{1}{T}\sum_{s,t\leq T} w_T(s,t)\ \frac{\partial^2}{\partial\beta_l\partial\beta_k}\mathbf{g}_s(\beta)\Big|_{\beta=\beta_0}\mathbf{g}_t(\beta_0) \\
B_3 &= \bar{\mathbf{C}}_l(\beta_0)\hat{\mathbf{\Omega}}(\beta_0)^{-1}\left[T\bar{\mathbf{G}}_k(\beta_0)\bar{\mathbf{g}}(\beta_0)' - \bar{\mathbf{C}}_k(\beta_0)\right] - \bar{\mathbf{C}}_k(\beta_0)\hat{\mathbf{\Omega}}(\beta_0)^{-1}\left[T\bar{\mathbf{G}}_l(\beta_0)\bar{\mathbf{g}}(\beta_0)' - \bar{\mathbf{C}}_l(\beta_0)\right] \\
B_4 &= \frac{\partial}{\partial\beta_l}\left[\hat{\mathbf{\Omega}}(\beta)^{-1}\bar{\mathbf{C}}_k(\beta)\hat{\mathbf{\Omega}}(\beta)^{-1}\right]\left[\hat{\mathbf{\Omega}}(\beta_0) - T\bar{\mathbf{g}}(\beta_0)\bar{\mathbf{g}}(\beta_0)'\right] = o_p(1)
\end{aligned}$$

where B_4 converges in probability to zero by the Slutsky theorem and the assumption that $\hat{\Omega}(\beta)$ is a consistent estimator of the GMM variance covariance matrix. In the following we will make the usual GMM assumptions (see e.g. Gallant and White, 1988), in particular the moment functions and their first derivatives are assumed to have bounded $4+\delta$th moments, and the moment functions evaluated at the true parameter have mean zero. Furthermore we will consider two different forms of autocorrelation.

4.1 Case 1: Finitely Correlated Errors

Assumption 1 *There is $\tau_0 << T$ such that for all $\tau > \tau_0$ and $k, l, m, n = 1, \ldots, K$,*

$$\mathbb{E}[\varrho_t\varrho_{t+\tau}|Z_t, Z_{t+\tau}] = \mathbb{E}[\nu_{kt}\nu_{l,t+\tau}|Z_t, Z_{t+\tau}] = \mathbb{E}[\zeta_{klt}\zeta_{mn,t+\tau}|Z_t, Z_{t+\tau}] = 0$$

and

$$\mathbb{E}[\varrho_t\nu_{k,t\pm\tau}|Z_t, Z_{t\pm\tau}] = \mathbb{E}[\varrho_t\zeta_{kl,t\pm\tau}|Z_t, Z_{t\pm\tau}] = \mathbb{E}[\nu_{kt}\zeta_{lm,t\pm\tau}|Z_t, Z_{t\pm\tau}] = 0$$

This assumption covers both the no autocorrelation case and clustered sampling. In this case, we can choose weights $w_T(s,t) = \mathbb{1}\{|s-t| \leq \tau_0\}$ so that

$$\begin{aligned}
\mathbb{E}\left[B_1|Z\right] &= \mathbb{E}\left[\frac{1}{T}\sum_{s,t\leq T}(1-w_T(s,t))\left.\frac{\partial}{\partial\beta_l}\mathbf{g}_s(\beta)\frac{\partial}{\partial\beta_k}\mathbf{g}_t(\beta)'\right|_{\beta=\beta_0}\middle| Z\right] \\
&= \frac{1}{T}\sum_{|s-t|>\tau_0} h_s h_t' \mathbb{E}\left[\frac{\partial}{\partial\beta_l}\varrho(w_s,\beta_0)\middle| Z_s = z_s\right]\mathbb{E}\left[\frac{\partial}{\partial\beta_k}\varrho(w_t,\beta_0)\middle| Z_t = z_t\right] \\
&\quad + \mathbb{E}\left[\frac{1}{T}\sum_{|s-t|>\tau_0} h_s h_t' \nu_{ls}\nu_{kt}\middle| Z\right] \\
&= \frac{1}{T}\sum_{|s-t|>\tau_0} h_s h_t' \mathbb{E}\left[\frac{\partial}{\partial\beta_l}\varrho(w_s,\beta_0)\middle| Z_s = z_s\right]\mathbb{E}\left[\frac{\partial}{\partial\beta_k}\varrho(w_t,\beta_0)\middle| Z_t = z_t\right]
\end{aligned}$$

Similarly,

$$\mathbb{E}\left[B_2|Z\right] = \mathbb{E}\left[\frac{1}{T}\sum_{|s-t|>\tau_0} h_s h_t' \left(\zeta_{kls}\varrho_t + \mathbb{E}\left[\frac{\partial^2}{\partial\beta_l\partial_k}\varrho(w_s,\beta_0)\middle| Z_s = z_s\right]\varrho_t\right)\middle| Z\right] = 0$$

and

$$\begin{aligned}
\mathbb{E}\left[B_3|Z\right] &= \bar{\mathbf{C}}_l(\beta_0)\hat{\mathbf{\Omega}}(\beta_0)^{-1}\mathbb{E}\left[\frac{1}{T^2}\sum_{|s-t|>\tau_0} h_s h_t' \left(\nu_{ks}\varrho_t + \mathbb{E}\left[\frac{\partial}{\partial\beta_k}\varrho(w_s,\beta_0)\middle| Z_s = z_s\right]\varrho_t\right)\middle| Z\right] \\
&\quad -\bar{\mathbf{C}}_k(\beta_0)\hat{\mathbf{\Omega}}(\beta_0)^{-1}\mathbb{E}\left[\frac{1}{T^2}\sum_{|s-t|>\tau_0} h_s h_t' \left(\nu_{ls}\varrho_t + \mathbb{E}\left[\frac{\partial}{\partial\beta_l}\varrho(w_s,\beta_0)\middle| Z_s = z_s\right]\varrho_t\right)\middle| Z\right] = 0
\end{aligned}$$

Hence,

$$\mathbb{E}[\tilde{Q}_{\beta_l\beta_k}(\beta_0)|Z] = \mathrm{tr}\left(\hat{\mathbf{\Omega}}(\beta_0)^{-1}\frac{1}{T}\sum_{|s-t|>\tau_0} h_s h_t' \mathbb{E}\left[\frac{\partial}{\partial\beta_l}\varrho(w_s,\beta_0)\middle| Z_s = z_s\right]\mathbb{E}\left[\frac{\partial}{\partial\beta_k}\varrho(w_t,\beta_0)\middle| Z_t = z_t\right]\right)$$

Weak identification means that

$$\mathbb{E}\left[H_t\frac{\partial}{\partial\beta}\varrho(W_t,\beta_0)\middle| Z_t = z_t\right]$$

is near singular. As an example, the Staiger and Stock (1997) weak instruments asymptotics for the familiar linear specification (1) correspond to

$$\mathbb{E}[X_t|Z_t = z_t] = z_t \frac{\Pi}{\sqrt{T}}$$

which is the fastest rate in the first stage for which CUE will be well-posed in the limit.

Therefore, under weak identification the expectation of $\nabla_{\beta\beta}\tilde{Q}(\beta_0) := \left\{\tilde{Q}_{\beta_l\beta_k}(\beta_0)\right\}_{k,l<K}$ in (3) becomes near singular, and the inverse in the approximation (2) for $\hat{\beta}_{CUE}$ is ill-defined, so that we would expect CUE to suffer from the no-moments problem in that case.

4.2 Case 2: Mixing Sequences

Assumption 2 *$(\varrho_t, \nu_t, \zeta_t)$ is a uniform mixing sequence with $\phi(m) = O\left(m^{-\frac{r}{r-1}}\right)$ for some $r > 2$.*

In particular it is possible to apply a mixing law of large numbers. Also for every $s, t \in \{1, 2, \dots\}$, the kernel weights have to satisfy $\lim_T w_T(s,t) = 1$ and $w_T(t, t \pm m_T) = 0$ for some sequence $m_T \longrightarrow \infty$, $\frac{m_T}{T} \longrightarrow 0$ in order for $\hat{\mathbf{\Omega}}(\beta)$ to be a consistent estimator of $\mathbf{\Omega}(\beta)$. Under these conditions, following an argument completely analogous to the proof of Lemma 6.6 in Gallant and White (1988),

$$\mathbb{E}[\tilde{Q}_{\beta_l\beta_k}(\beta_0)|Z] = \text{tr}\left(\hat{\mathbf{\Omega}}(\beta_0)^{-1}\frac{1}{T}\sum_{s,t\leq T}(1 - w_T(s,t))h_s h_t' \mathbb{E}\left[\frac{\partial}{\partial\beta_l}\varrho(w_s,\beta_0)\middle| Z_s = z_s\right]\mathbb{E}\left[\frac{\partial}{\partial\beta_k}\varrho(w_t,\beta_0)\middle| Z_t = z_t\right]\right)$$

This suggests that *in the limit* we should expect CUE to have no integral moments under weak identification if we allow for the moments and their derivatives to be mixing sequences.

5 A Modification of CUE with Finite Moments

With possibly dependent data, we can derive CUE from the dual maximization problem

$$\max_{\pi,\beta}\sum_{t=1}^{T}\varrho(\pi_t) \qquad \text{s.t.} \quad \sum_{t=1}^{T}\pi_t\hat{g}_t(\beta) = 0, \text{ and } \sum_{t=1}^{T}\pi_t = 1 \tag{4}$$

where for CUE

$$\varrho(v) = -\frac{(v-1)^2}{2}$$

and for a positive definite matrix $W^{1/2}$ and the $T \times k$ matrix $\tilde{\mathbf{g}} := \tilde{\mathbf{g}}(\beta)$ containing the moment vector $g_t(\beta)$ in its t-th column,

$$\hat{g}_t := \mathbf{W}_t^{1/2}\tilde{\mathbf{g}}$$

As for the blockwise empirical likelihood estimator described in Kitamura and Stutzer (1997) possible choices for $W^{1/2}$ are e.g. $W^{1/2} := \frac{1}{T-\tau}(\mathbb{1}\{|s-t| < \tau\})_{s,t \le T}$ which corresponds to a matrix with the Bartlett kernel function used for the Newey-West variance estimator, or$W^{1/2} := \left(\frac{1}{n_g(t)}\mathbb{1}\{s,t \in g(t)\}\right)_{s,t\le T}$ which yields the robust variance matrix in the presence of one-way clustering.

In the dual formulation of the corresponding maximization problem, CUE solves the first-order conditions[4]

$$\left[\sum_{t=1}^{T}\hat{\pi}_t \frac{\partial}{\partial\beta}\hat{\mathbf{g}}_t(\beta)\right]\Omega(\beta)^{-1}\bar{\hat{\mathbf{g}}}(\beta) = 0$$

where

$$\hat{\pi}_t := \frac{1 - \bar{\hat{\mathbf{g}}}(\hat{\beta}_{CUE})\Omega(\hat{\beta}_{CUE})^{-1}\hat{\mathbf{g}}_t(\hat{\beta}_{CUE})}{\sum_{s=1}^{T} 1 - \bar{\hat{\mathbf{g}}}(\hat{\beta}_{CUE})\Omega(\hat{\beta}_{CUE})^{-1}\hat{\mathbf{g}}_s(\hat{\beta}_{CUE})}$$

which have an interpretation as "empirical probabilities" in the Generalized Empirical Likelihood framework. We propose to solve a modification of the first-order conditions in which we replace $\hat{\pi}_t$ with

$$\tilde{\pi}_t^{\alpha} := \frac{1 - \bar{\hat{\mathbf{g}}}(\hat{\beta}_{CUE})\Omega(\hat{\beta}_{CUE})^{-1}\hat{\mathbf{g}}_t(\hat{\beta}_{CUE}) + \alpha_T}{\sum_{s=1}^{T}[1 + \alpha_T - \bar{\hat{\mathbf{g}}}(\hat{\beta}_{CUE})\Omega(\hat{\beta}_{CUE})^{-1}\hat{\mathbf{g}}_s(\hat{\beta}_{CUE})]}$$

for some sequence $\alpha_T > 0$ with $\alpha_T = O\left(\frac{1}{T}\right)$, e.g. $\alpha_T = \frac{a}{T}$ for some constant $a > 0$, and hold all terms except $\bar{g}(\beta)$ fixed at $\hat{\beta}_{CUE}$, i.e. solve

$$\begin{aligned} &= \left[\sum_{t=1}^{T}\tilde{\pi}_t^{\alpha}\frac{\partial}{\partial\beta}\hat{g}_t(\hat{\beta}_{CUE})\right]\Omega(\hat{\beta}_{CUE})^{-1}\bar{\hat{\mathbf{g}}}(\beta) \\ &\propto \left[\sum_{t=1}^{T}\left(1 - \bar{\hat{\mathbf{g}}}(\hat{\beta}_{CUE})\Omega(\hat{\beta}_{CUE})^{-1}\hat{\mathbf{g}}_t(\hat{\beta}_{CUE})' + \alpha_T\right)\frac{\partial}{\partial\beta}\hat{\mathbf{g}}_t(\hat{\beta}_{CUE})\right]\Omega(\hat{\beta}_{CUE})^{-1}\bar{\hat{\mathbf{g}}}(\beta) \qquad (5) \end{aligned}$$

Therefore we "bias" the estimates of the "empirical probabilities" toward $\frac{1}{T}$ and skews the first-order conditions towards those for the two-step optimal GMM estimator. This modification could be interpreted as a regularization in the that it penalizes variation in the "empirical distribution" (in the GEL sense).

[4]see Newey and Smith (2004)

We can rewrite these first-order conditions in vector form as

$$0 = (\mathbf{I_K} \otimes \iota_T)' \left\{ \mathbf{I}_K \otimes \left[(1+\alpha_T)\, \mathbf{I_T} - \mathbf{W^{1/2}} \tilde{\mathbf{g}} (\tilde{\mathbf{g}}' \mathbf{W} \tilde{\mathbf{g}})^{-1} \tilde{\mathbf{g}}' \mathbf{W^{1/2}} \right] \mathbf{W^{1/2}} \right\} \tilde{\mathbf{G}} (\tilde{\mathbf{g}}' \mathbf{W} \tilde{\mathbf{g}})^{-1} \tilde{\mathbf{g}}' \mathbf{W}^{1/2} \iota_{\mathbf{T}}$$

where ι_T is a T-dimensional column vector of ones, $\tilde{\mathbf{g}} := \tilde{\mathbf{g}}(\beta)$ is a $T \times M$ matrix consisting of the moment functions for each observation, and $\tilde{\mathbf{G}} = \left[\tilde{\mathbf{G}}_1(\beta)', \ldots, \tilde{\mathbf{G}}_K(\beta)' \right]'$ is a $KT \times M$ matrix containing the stacked first derivatives. For the remaining part of the argument, we assume w.l.o.g. $\mathbf{W} = \mathbf{I}_T$, the T-dimensional identity matrix.

If $\mathbf{g}_t(\beta)$ is linear in β, we can solve for

$$\begin{aligned} \hat{\beta}_\alpha &= \left[(\mathbf{I_K} \otimes \iota_T)' \left\{ \mathbf{I_K} \otimes \left[(1+\alpha_T)\, \mathbf{I_T} - \tilde{\mathbf{g}} \hat{\mathbf{\Omega}}^{-1} \tilde{\mathbf{g}}' \right] \right\} \tilde{\mathbf{G}} \hat{\mathbf{\Omega}}^{-1} \tilde{\mathbf{G}}' \iota_{TK} \right]^{-1} \\ &\quad \times \left[(\mathbf{I_K} \otimes \iota_T)' \left\{ \mathbf{I_K} \otimes \left[(1+\alpha_T)\, \mathbf{I_T} - \tilde{\mathbf{g}} \hat{\mathbf{\Omega}}^{-1} \tilde{\mathbf{g}}' \right] \right\} \tilde{\mathbf{G}} \hat{\mathbf{\Omega}}^{-1} \tilde{\mathbf{g}}(\mathbf{0})' \iota \right] \end{aligned}$$

The denominator of the new estimator is equal to

$$\tilde{\Delta}_\alpha = (\mathbf{I_K} \otimes \iota_T)' \left[(1+\alpha_T)\, \mathbf{I_T} - \tilde{\mathbf{g}} \mathbf{\Omega}^{-1} \tilde{\mathbf{g}}' \right] \tilde{\mathbf{G}} \mathbf{\Omega}^{-1} \tilde{\mathbf{G}}' (\mathbf{I_K} \otimes \iota_T) =: \Delta_{CUE} + \alpha_T (\mathbf{I_K} \otimes \iota_T)' \tilde{\mathbf{G}} \mathbf{\Omega}^{-1} \tilde{\mathbf{G}}' (\mathbf{I_K} \otimes \iota_T)$$

Since the first summand

$$\Delta_{CUE} := (\mathbf{I_K} \otimes \iota_T)' \left[\mathbf{I_T} - \tilde{\mathbf{g}} \mathbf{\Omega}^{-1} \tilde{\mathbf{g}}' \right] \tilde{\mathbf{G}} \mathbf{\Omega}^{-1} \tilde{\mathbf{G}}' (\mathbf{I_K} \otimes \iota_T)$$

is almost surely positive semidefinite,[5] the eigenvalues of the denominator

$$T\tilde{\Delta}_\alpha = (\mathbf{I_K} \otimes \iota_T)' \left[\mathbf{I_T} - \tilde{\mathbf{g}} \mathbf{\Omega}^{-1} \tilde{\mathbf{g}}' \right] \tilde{\mathbf{G}} \mathbf{\Omega}^{-1} \tilde{\mathbf{G}}' (\mathbf{I_K} \otimes \iota_T) + \alpha_T (\mathbf{I_K} \otimes \iota_T)' \tilde{\mathbf{G}} \mathbf{\Omega}^{-1} \tilde{\mathbf{G}}' (\mathbf{I_K} \otimes \iota_T)$$

[5] Since the matrix in the middle of this expression has the structure $\mathbf{PA}$ where $\mathbf{P}$ is a projection matrix, and $\mathbf{A}$ is symmetric and positive semi-definite, there exists a decomposition $\mathbf{A} = \mathbf{X \Lambda X}'$ where $\mathbf{X}$ has full rank and is orthonormal and symmetric, and Λ is a non-negative diagonal matrix. Therefore, we can write

$$\mathbf{PA} = \mathbf{PX}'\mathbf{\Lambda X} = \mathbf{X}^{-1}\mathbf{XPX}'\mathbf{\Lambda X} = \mathbf{X}^{-1}(\mathbf{XPX\Lambda})\mathbf{X}$$

This matrix has the same eigenvalues as $\mathbf{XPX\Lambda}$, which in turn is positive semi-definite since $\mathbf{P}$ is an orthogonal projector, and all elements of Λ are non-negative. Since Δ is a quadratic form corresponding to a positive semi-definite matrix, $\Delta \geq 0$ almost surely.

are bounded from below by $\min \operatorname{eig}(T\alpha_T \mathbf{V}) > 0$ for any choice of $\alpha_T > 0$, where

$$\mathbf{V} := (\mathbf{I_K} \otimes \iota_T)' \, \tilde{\mathbf{G}} \boldsymbol{\Omega}^{-1} \tilde{\mathbf{G}} \, (\mathbf{I_K} \otimes \iota_T)$$

Following a similar line of reasoning as in Lemma 2 in Fuller (1977), we can apply Lemma B from Desgracie and Fuller (1972) to show that for any sequence $\alpha_T = O\left(\frac{1}{T}\right)$, the modification of CUE in (5) has all moments up to the degree of overidentification for T sufficiently large. Also, since this estimator solves a convex combination of the first-order conditions for two-step GMM and CUE, it is asymptotically equivalent to both estimators to first order. If $\sqrt{T}\alpha_T \to 0$, the modified estimator will furthermore be second-order equivalent to CUE.

6 Extensions

From this argument, it can be seen that we can alternatively add any function $J_\beta(\beta)$ which is strictly increasing in β at β_0 to the first order condition and solve

$$\left[\sum_{t=1}^{T} \hat{\pi}_t \frac{\partial}{\partial \beta} \hat{\mathbf{g}}_t(\hat{\beta}_{CUE})\right] \Omega(\hat{\beta}_{CUE})^{-1} \bar{\tilde{\mathbf{g}}}(\beta) + \frac{\alpha}{T} J_\beta(\beta) = 0 \tag{6}$$

for β, where now $\alpha > 0$ is held fixed. In the simulations we implement

$$J_\beta^{(1)}(\beta) = \frac{\tilde{\mathbf{G}}(\hat{\beta}_{CUE})' \tilde{\mathbf{g}}(\beta)}{T \sigma_{\varrho\varrho}}, \qquad \text{and} \qquad J_\beta^{(2)}(\beta) = \beta$$

where the first choice mimics the modification proposed by Fuller (1977). By an argument parallel to that in the preceding section it can be seen that any integral moment of the resulting estimator exists for a sufficiently large sample.

The idea of regularizing the "empirical probabilities" for CUE can also be extended to the entire family of Generalized Empirical Likelihood (GEL) Estimators, which solve the problem (4) for a general concave function $\varrho(v)$. As shown in Newey and Smith (2004), the GEL estimator solves

$$\left[\sum_{t=1}^{T} \hat{\pi}_t \hat{G}_t(\beta)\right]' \left[\sum_{t=1}^{T} k_t \hat{g}_t(\beta) \hat{g}(\beta)'\right]^{-1} \bar{\hat{g}}(\beta) = 0$$

where

$$\hat{\pi}_t = \frac{\varrho'(\hat{\lambda}'\hat{g}_t(\beta))}{\sum_{s=1}^T \varrho'(\hat{\lambda}'\hat{g}_s(\beta))},$$

$\hat{\lambda} = -\left(\frac{1}{T}\sum_{t=1}^T k_t \hat{g}_t(\beta)\hat{g}_t(\beta)'\right)^{-1}\bar{\hat{g}}(\hat{\beta})$, and $k_t = \frac{\varrho'(\lambda'\hat{g}_t)+1}{\frac{\lambda'\hat{g}_t}{\hat{\lambda}'\hat{g}_s}[\varrho'(\hat{\lambda}'\hat{g}_s)+1]}$. Under regularity conditions, $\hat{\lambda} = O_p\left(\frac{1}{\sqrt{T}}\right)$, so that we can expand $\hat{\pi}$ around $\lambda_0 = 0$ and obtain

$$\hat{\pi}_t \propto \varrho'(0) + \varrho''(0)\hat{\lambda}'\hat{g}_t(\beta) + O_p\left(\frac{1}{T}\right) = -\left[1 - \bar{\hat{g}}(\beta)'\left(\frac{1}{T}\sum_{t=1}^T k_t \hat{g}_t(\beta)\hat{g}_t(\beta)'\right)^{-1}\hat{g}_t(\beta)\right] + O_p\left(\frac{1}{T}\right)$$

where without loss of generality we impose the normalizations $\varrho'(0) = \varrho''(0) = -1$. Also, by L'Hospital's rule

$$\begin{aligned} k_t &= \lim_{v_t, v_s \to 0} \frac{\varrho'(v_t)+1}{\sum_{s=1}^T \frac{v_t}{v_s}[\varrho'(v_s)+1]} \\ &+ \lim_{v_t, v_s \to 0} \frac{\frac{\varrho''(v_t)v_t - (\varrho'(v_t)+1)}{v_t^2}\left[\sum_{s=1}^T \frac{\varrho'(v_s)+1}{v_s}\right]\hat{\lambda}'\hat{g}_t - \frac{\varrho'(v_t)+1}{v_t}\left[\sum_{s=1}^T \frac{\varrho''(v_s)v_s - (\varrho'(v_s)+1)}{v_s^2}\hat{\lambda' g_s}\right]}{\left[\sum_{s=1}^T \frac{\varrho'(v_s)+1}{v_s}\right]^2} + O_p\left(\frac{1}{T^2}\right) \\ &= \frac{1}{T} + \frac{\varrho'''(0)}{2T}\hat{\lambda}'(\bar{\hat{g}} - \hat{g}_t) + O_p\left(\frac{1}{T^2}\right) \end{aligned}$$

Plugging this back into the expansion for $\hat{\pi}_t$, it can be seen that $\hat{\pi}_t^{GEL} = \hat{\pi}_t^{CUE} + O_p\left(\frac{1}{T}\right)$, so that in large samples the projection argument made for CUE will be valid for all GEL estimators. It should also be noted that if the third moments of $\hat{g}_t(\beta_0)$ are zero and the fourth moments are bounded, all GEL estimators are the same as CUE up to second order.

The Monte Carlo results in the next section demonstrate that CUE still has some bias for a large degree of over-identification most of which arises from the estimation of the optimal weighting matrix $\hat{\Omega}$. Newey and Smith (2004) showed that Maximum Empirical Likelihood (MEL) does not have this bias, so that applying the proposed modification to the MEL estimator should be expected to have even better finite-sample properties than the modified CUE. We leave this topic for future research.

7 Monte Carlo Study

We performed a simulation study to investigate the properties of the proposed estimators, "Regularized" CUE (RCUE) defined by the first-order conditions in equation (5), and RCUE2 (specified in equation (6)), and compare their performance with GMM, LIML, and CUE. In particular, we find that RCUE fixes the no-moments problem present in CUE while significantly reducing the overidentification bias of GMM in both the linear and nonlinear cases.

7.1 Baseline Design: Linear Model without Autocorrelation

For the simulation study we followed the heteroskedastic design of Hausman, Newey, Woutersen, Chao, Swanson (2007) in order to provide empirically relevant parameter choices for the severity of the endogeneity and heteroskedasticity. We then introduced autocorrelation into the model in the form of groupwise clustering and further generalized the design to allow for nonlinear models. The baseline design is as follows:

$$
\begin{aligned}
y_i &= x_i\beta + \varepsilon_i && (7)\\
x &= z_{1i}\pi + \nu_i && (8)\\
\varepsilon_i &= \rho\nu_i + \sqrt{1-\rho^2}\left(\phi\theta_{1i} + \sqrt{1-\phi^2}\theta_{2i}\right)
\end{aligned}
$$

$\theta_{1i} \overset{iid}{\sim} N(0, z_{1i}^2)$, and $\nu_i, \theta_{2i} \overset{iid}{\sim} N(0,1)$ in order to allow for heteroskedasticity.

The estimators we chose to compare for the linear model simulations are 2-step GMM, LIML, Fuller(c=1), CUE, and the two RCUE estimators. For RCUE, we found that the adjustment term c=1 (labeled $RCUE_1$ and $RCUE2_1$ in the tables) performed well in the simulations, although the choice of the optimal constant will be a topic of future research. We also report the results for c=4 (labeled $RCUE_4$ and $RCUE2_4$) to illustrate a small range of the bias and variance that the proposed estimators can have. These values for c are the most widely used values for the Fuller estimator since respectively they lead to a mean unbiased or a minimum feasible MSE estimator, respectively. In the nonlinear model, we restricted our attention to GMM, CUE, and the two RCUE estimators with c=1 and c=4.

In Table 1, we compare the median bias of the several estimators. All of the estimators have biases which increase with the degree of overidentification, but the bias of GMM is much more pronounced as theory predicts, even approaching that of the OLS bias (unreported) for small values of the concentration parameter, CP, (more weakly identified models) and higher degrees of overidentification, $M-K$. On the other hand, the bias of CUE and LIML (under homoskedasticity) is much smaller, though also increasing in $M-K$ and decreasing in CP. The RCUE estimators perform similarly to CUE, although they do exhibit a slightly larger median bias.

In Tables 2 and 3, we see why it is possible for the RCUE estimators to restore the moments of CUE and lead to lower dispersion of the estimation results. In Table 2, for each set of model parameters, the interquartile range for CUE is greater than for the RCUE estimators. In Table 3, the result is qualitatively similar for the nine decile range (5lower CP and higher $M-K$. Thus, these results favor the use of RCUE over CUE.

In Table 4, we see that GMM's mean bias is increasing in the degree of overidentification, $M-K$, and is in all cases greater than the bias for the RCUE estimators. In addition, the mean bias for the RCUE estimators is also well defined and stable for all simulation parameters while the inconsistent estimates (using the standard sample mean and variance estimators) of the biases of CUE and LIML demonstrate the instability of their first moment, which arises from the "moments problem" of non-existence of moments for these estimators. The simulated variance and MSE of the RCUE estimators in Tables 5 and 6 further support the claim that RCUE adjusts CUE in such a way to restore its moments and decrease estimator dispersion. Further, the first two moments of the RCUE estimators are comparable to those of the HFUL estimator in HNWCS (2007) (unreported). Other than the apparent inconsistency of LIML as evidenced by its erratic performance, no other substantive differences in relative performance emerge when testing the estimators under heteroskedasticity.

7.2 Linear Model with Clustering

The next set of simulations was generated by drawing G clusters based on the design in equations (7)-(8) where the disturbances (ε_i, ν_i) are generated as

$$\nu_{gi} = \gamma_\nu v_g + \sqrt{1-\gamma_\nu^2} v_{gi} \qquad v_g, v_{gi} \overset{iid}{\sim} N(0,1)$$

$$\varepsilon_{gi} = \rho \nu_{gi} + \sqrt{1-\rho^2}\theta_{gi} \qquad \theta_{gi} \overset{iid}{\sim} N(0,1)$$

for a cluster correlation coefficient $\gamma_\nu = 0, 0.5$, $g = 1, \ldots, G$, and $i = 1, \ldots, n_g$, the number of observations in the gth cluster. We also allow for clustering in the instrumental variables,

$$Z_{gi} = \gamma_z \zeta_g + \sqrt{1 - \gamma_z^2} \zeta_{gi}$$

choosing parameter values $\gamma_z = 0, 0.5$.

Once we allow for clustering in the first stage with $G = 20, 50$ clusters given the total sample size of $T = 400$ in Tables 7 and 10, we first note the large median and mean biases arising from many moments persist in GMM, even after correcting for clustering. Also, LIML, which is typically suggested as a solution to the many moments problem has substantial bias under various conditions. On the other hand, in terms of median bias in Table 7, CUE performs quite well, with much less bias than GMM, although it also exhibits a bias which increases in the degree of overidentification. The RCUE estimators have larger bias, but it still remains much smaller than that of GMM. The same results arise in Tables 8 and 9 in terms of the quantiles. Quantiles of the RCUE estimators are smaller than quantiles of CUE in all cases and provide preliminary evidence for the existence of the first two moments under autocorrelation in Tables 10-12. There we find the mean, variance, and MSE of the estimators generally[6] existing for the RCUE estimators, but not for LIML or CUE, whose performance further suggests the moments problem. We also tested HLIM and HFUL from HNWCS (2007) as well as CUE with a heteroskedastic consistent covariance matrix, but all three estimators performed similar to LIML and FULL1 in demonstrating sizable bias.

7.3 Nonlinear Specification

For the Monte Carlo study in Tables 13-18, we changed the specification of the second stage of the baseline model to

$$\begin{aligned} y_i &= \exp\{x_i\beta\} + \varepsilon_i \\ x_i &= z_{1i}\pi + \nu_i \end{aligned}$$

[6] c=1 and c=4 are not guaranteed to solve the moments problem, and for a few cases, the evidence suggests that c=1 is not optimal for the RCUE2 estimator. Yet, as previously mentioned the optimal choice of c will be a topic of future research.

which has a similar structure as the CCAPM specification in Stock and Wright (2000).

The results resemble those in the linear case. In Tables 13 and 16, GMM shows substantial median and mean bias increasing with the degree of overidentification while CUE and the RCUE estimators exhibit much less median bias in all cases. In addition, in Tables 16 and 17, the first two moments of the RCUE estimator exist under all cases, while they do not appear to exist for CUE.[7] Thus, the nonlinear simulation results suggest the wide array of the problems to which the RCUE estimators can be applied as they are capable of removing the bias present in GMM under quite general conditions while decreasing the dispersion of the CUE estimator to permit the existence of moments.

8 Conclusion

In this paper we attempt to solve two problems: the bias in GMM which increases with the number of instruments (moment condition) and the "no moment" problem of CUE that leads to wide dispersion of the estimates. To date, much of the discussion of these problems has been under the assumption of homoscedasticity and absence of serial correlation, which limits their use in empirical research. HNWCS (2005) propose a solution to the homoscedasticity problem for the linear model. In this paper we propose a solution to the more general problem for linear and non-linear specifications when both heteroscedasticity and serial correlation may be present. We propose a class of estimators "regularized CUE" or RCUE that reduce the moment problem and reduce the dispersion of the CUE estimator. The Monte Carlo results accord well with the theory and demonstrate that the new RCUE estimator performs better than existing estimators. As topics for future research, we recommend investigation of the optimal value of the parameter in the RCUE estimators and application of the estimators to the class of Maximum Empirical Likelihood estimators, which is likely to decrease estimator bias even more.

References

[1] Andrews, D. **(1991):** Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation, *Econometrica* **59(4)**, 817-58

[7] The RCUE2 estimator appears to behave contrary to what the theory would predict as evidenced by its erratic variances in Table 17. The choice of the optimal c will be further investigated to test this estimator.

[2] BEKKER, P. **(1994):** Alternative Approximations to the Distributions of Instrumental Variables Estimators, *Econometrica* **62(3)**, 657-81

[3] BEKKER, P., AND J. VAN DER PLOEG **(2005):** Instrumental Variable Estimation Based on Grouped Data, *Statistica Neerlandica* **59**, 239-67

[4] DEGRACIE, J., AND W. FULLER **(1972):** Estimation of the Slope and Analysis of Covariance when the Concomitant Variable is Measured with Error, *Journal of the American Statistical Association* **67** No.340, 930-7

[5] DONALD, S., AND W.NEWEY **(2000):** A Jackknife Interpretation of the Continuous Updating Estimator, *Economics Letters* **67**, 239-43

[6] FULLER, W. **(1977):** Some Properties of a Modification of the Limited Information Estimator, *Econometrica* **45(4)**, 939-54

[7] GALLANT, A., AND H. WHITE **(1988):** A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models, Oxford

[8] GUGGENBERGER, P. **(2005):** Monte-Carlo Evidence Suggesting a No Moment Problem of the Continuous Updating Estimator, *Economics Bulletin* **3(13)**, 1-6

[9] HAHN, J., J. HAUSMAN, AND G. KUERSTEINER **(2004):** Estimation with Weak Instruments: Accuracy of higher-order bias and MSE approximations, *Econometrics Journal* **7**, 272-306

[10] HAN, CH., AND P.PHILLIPS **(2006):** GMM with Many Moment Conditions, *Econometrica* **74(1)**, 147-92

[11] HANSEN, L., J. HEATON, AND A. YARON **(1996):** Finite-Sample Properties of Some Alternative GMM Estimators, *Journal of Business and Economic Statistics* **14(3)**, 262-80

[12] HAUSMAN, J., W. NEWEY, AND T. WOUTERSEN **(2005):** IV Estimation with Heteroskedasticity and Many Instruments, working paper MIT

[13] HAUSMAN, J., W. NEWEY, T. WOUTERSEN, J. CHAO, AND N. SWANSON **(2007):** Instrumental Variable Estimation with Heteroskedasticity and Many Instruments, working paper MIT

[14] Kitamura, Y., and M. Stutzer **(1997):** An Information-Theoretic Alternative to Generalized Method of Moments Estimation, *Econometrica* **65(4)**, 861-74

[15] Morimune, K. **(1983):** Approximate Distributions of k-Class Estimators when the Degree of Overidentifiability is Large Compared with the Sample Size, *Econometrica* **51(3)**, 821-41

[16] Nagar, A. **(1959):** The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations, *Econometrica* **27(4)**, 575-95

[17] Newey, W., and R. Smith **(2004):** Higher Order Propertiew of GMM and Generalized Empirical Likelihood Estimators, *Econometrica* **72(1)**, 219-55

[18] Newey, W., and K. West **(1987):** A Simple Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica* **55(3)**, 703-08

[19] Rilstone, P., V. Srivastava, and A. Ullah **(1996):** The Second-Order Bias and Mean-Squared Error of Nonlinear Estimators, *Journal of Econometrics* **75**, 369-95

[20] Staiger, D., and J. Stock **(1997):** Instrumental Variables Regression with Weak Instruments, *Econometrica* **65(3)**, 557-86

[21] Stock, J., and J. Wright **(2000):** GMM with Weak Identification, *Econometrica* **68(5)**, 1055-96

Table 1: Linear Model under Homoskedasticity and Heteroskedasticity - Median Bias $\rho = 0.3$, $T = 400$

$\mathcal{R}^2_{\varepsilon^2\|z_1^2}$	CP	M	$RCUE_1$	$RCUE_4$	$RCUE2_1$	$RCUE2_4$	GMM	CUE	$LIML$	$FULL1$
0.00	8	5	0.046	0.064	0.059	0.038	0.090	0.026	0.017	0.052
0.00	8	10	0.070	0.097	0.075	0.062	0.156	0.044	0.024	0.058
0.00	8	30	0.130	0.151	0.148	0.135	0.237	0.097	0.067	0.090
0.00	8	50	0.166	0.182	0.182	0.174	0.260	0.131	0.098	0.115
0.00	16	5	0.022	0.034	0.023	0.012	0.053	0.009	0.006	0.025
0.00	16	10	0.031	0.056	0.025	0.015	0.103	0.010	0.009	0.026
0.00	16	30	0.055	0.084	0.053	0.043	0.194	0.030	0.016	0.034
0.00	16	50	0.097	0.119	0.101	0.096	0.227	0.068	0.028	0.046
0.00	32	5	0.011	0.019	0.013	0.007	0.030	0.005	0.005	0.014
0.00	32	10	0.017	0.033	0.010	0.005	0.065	0.004	0.005	0.014
0.00	32	30	0.026	0.050	0.015	0.012	0.143	0.010	0.005	0.015
0.00	32	50	0.036	0.057	0.028	0.025	0.182	0.022	0.005	0.014
0.20	8	5	0.029	0.045	0.070	0.036	0.062	0.015	−0.222	−0.122
0.20	8	10	0.054	0.072	0.075	0.050	0.134	0.029	−0.620	−0.320
0.20	8	30	0.123	0.142	0.147	0.135	0.228	0.099	−1.806	−0.837
0.20	8	50	0.164	0.179	0.190	0.181	0.256	0.143	−2.579	−1.289
0.20	16	5	0.008	0.020	0.026	0.006	0.027	−0.001	−0.177	−0.124
0.20	16	10	0.019	0.040	0.026	0.011	0.085	0.004	−0.426	−0.320
0.20	16	30	0.051	0.076	0.051	0.042	0.185	0.029	−1.629	−1.035
0.20	16	50	0.087	0.106	0.089	0.085	0.224	0.068	−2.524	−1.602
0.20	32	5	0.004	0.010	0.013	0.002	0.013	−0.002	−0.093	−0.077
0.20	32	10	0.006	0.022	0.007	−0.002	0.048	−0.005	−0.211	−0.185
0.20	32	30	0.013	0.039	0.006	0.001	0.135	−0.002	−1.009	−0.814
0.20	32	50	0.027	0.053	0.021	0.018	0.180	0.013	−1.881	−1.444

***Simulation results based on 6250 replications.

Table 2: Linear Model under Homoskedasticity and Heteroskedasticity - Interquartile Range $\rho = 0.3$, $T = 400$

$\mathcal{R}^2_{\varepsilon^2\|z_1^2}$	CP	M	$RCUE_1$	$RCUE_4$	$RCUE2_1$	$RCUE2_4$	GMM	CUE	$LIML$	$FULL1$
0.00	8	5	0.521	0.464	0.482	0.548	0.410	0.581	0.575	0.481
0.00	8	10	0.607	0.502	0.587	0.651	0.350	0.717	0.692	0.582
0.00	8	30	0.862	0.676	0.855	0.915	0.237	1.034	0.922	0.792
0.00	8	50	0.951	0.774	0.929	0.994	0.193	1.156	1.060	0.920
0.00	16	5	0.363	0.340	0.359	0.380	0.315	0.387	0.374	0.346
0.00	16	10	0.419	0.364	0.432	0.453	0.288	0.462	0.431	0.399
0.00	16	30	0.625	0.512	0.649	0.674	0.216	0.709	0.575	0.528
0.00	16	50	0.756	0.627	0.778	0.806	0.182	0.878	0.682	0.625
0.00	32	5	0.250	0.243	0.250	0.257	0.232	0.260	0.251	0.242
0.00	32	10	0.278	0.257	0.288	0.295	0.223	0.298	0.273	0.261
0.00	32	30	0.419	0.361	0.447	0.455	0.186	0.459	0.336	0.325
0.00	32	50	0.541	0.461	0.573	0.584	0.164	0.592	0.394	0.377
0.20	8	5	0.707	0.647	0.558	0.689	0.589	0.763	2.015	1.334
0.20	8	10	0.824	0.693	0.701	0.829	0.481	0.936	3.734	2.040
0.20	8	30	1.118	0.906	1.054	1.165	0.312	1.313	9.974	4.222
0.20	8	50	1.211	1.011	1.159	1.250	0.240	1.400	13.901	5.602
0.20	16	5	0.469	0.449	0.430	0.475	0.433	0.493	1.038	0.852
0.20	16	10	0.530	0.475	0.514	0.556	0.375	0.576	1.729	1.292
0.20	16	30	0.788	0.659	0.791	0.831	0.284	0.886	6.297	3.350
0.20	16	50	0.944	0.799	0.947	0.994	0.224	1.054	10.368	4.853
0.20	32	5	0.323	0.314	0.312	0.327	0.310	0.332	0.528	0.497
0.20	32	10	0.341	0.321	0.343	0.356	0.280	0.360	0.746	0.679
0.20	32	30	0.502	0.437	0.524	0.540	0.240	0.546	2.292	1.625
0.20	32	50	0.631	0.550	0.656	0.670	0.202	0.682	3.910	2.392

***Simulation results based on 6250 replications.

Table 3: Linear Model under Homoskedasticity and Heteroskedasticity - Nine Decile Range $\rho = 0.3$, $T = 400$

$\mathcal{R}^2_{\varepsilon^2 \mid z_1^2}$	CP	M	$RCUE_1$	$RCUE_4$	$RCUE2_1$	$RCUE2_4$	GMM	CUE	$LIML$	$FULL1$
0.00	8	5	1.478	1.280	1.194	1.527	1.093	1.930	1.945	1.328
0.00	8	10	1.737	1.287	1.517	1.945	0.893	2.708	2.808	1.664
0.00	8	30	2.418	1.488	2.203	2.913	0.587	4.886	5.115	2.394
0.00	8	50	2.726	1.670	2.575	3.224	0.467	5.734	6.310	2.872
0.00	16	5	0.969	0.879	0.919	1.041	0.805	1.094	1.030	0.907
0.00	16	10	1.137	0.942	1.137	1.282	0.725	1.383	1.275	1.093
0.00	16	30	1.813	1.252	1.807	2.131	0.534	2.581	2.203	1.630
0.00	16	50	2.263	1.499	2.240	2.610	0.440	3.580	3.062	2.062
0.00	32	5	0.637	0.610	0.644	0.669	0.586	0.676	0.641	0.613
0.00	32	10	0.730	0.655	0.759	0.793	0.553	0.803	0.701	0.666
0.00	32	30	1.128	0.886	1.214	1.279	0.457	1.315	0.918	0.861
0.00	32	50	1.524	1.153	1.605	1.718	0.398	1.816	1.172	1.071
0.20	8	5	2.362	2.021	1.427	2.076	1.693	3.019	14.835	3.737
0.20	8	10	2.885	2.038	1.812	2.670	1.276	4.636	26.192	4.608
0.20	8	30	3.534	2.254	2.680	3.780	0.776	6.289	59.017	6.834
0.20	8	50	3.862	2.431	3.124	4.187	0.613	6.553	78.679	8.326
0.20	16	5	1.325	1.238	1.123	1.331	1.168	1.443	5.671	2.971
0.20	16	10	1.671	1.360	1.447	1.798	0.997	2.049	12.118	3.895
0.20	16	30	2.666	1.879	2.371	2.965	0.699	3.706	37.617	6.340
0.20	16	50	3.214	2.172	2.855	3.536	0.575	4.465	60.802	7.898
0.20	32	5	0.824	0.803	0.792	0.841	0.791	0.862	1.821	1.605
0.20	32	10	0.929	0.850	0.921	0.974	0.738	1.002	3.096	2.290
0.20	32	30	1.523	1.229	1.548	1.671	0.601	1.736	16.167	5.150
0.20	32	50	2.020	1.589	2.090	2.253	0.515	2.370	33.146	7.011

***Simulation results based on 6250 replications.

Table 4: Linear Model under Homoskedasticity and Heteroskedasticity - Mean Bias $\rho = 0.3$, $T = 400$

$\mathcal{R}^2_{\varepsilon^2\|z_1^2}$	CP	M	$RCUE_1$	$RCUE_4$	$RCUE2_1$	$RCUE2_4$	GMM	CUE	$LIML$	$FULL1$
0.00	8	5	0.024	0.050	0.066	0.022	0.078	1.3E+011	−0.008	0.043
0.00	8	10	0.066	0.107	0.086	0.048	0.155	-2.0E+011	1.309	0.066
0.00	8	30	0.132	0.173	0.142	0.116	0.238	-4.7E+011	−0.522	0.102
0.00	8	50	0.155	0.192	0.161	0.138	0.260	1.1E+011	0.142	0.133
0.00	16	5	0.002	0.020	0.015	−0.010	0.041	7.2E+009	−0.028	0.009
0.00	16	10	0.017	0.053	0.021	−0.005	0.101	4.0E+010	0.097	0.014
0.00	16	30	0.045	0.098	0.045	0.017	0.195	-5.3E+010	−0.310	0.023
0.00	16	50	0.077	0.126	0.079	0.052	0.228	-3.3E+011	−0.183	0.043
0.00	32	5	0.001	0.010	0.002	−0.007	0.021	−0.010	−0.008	0.003
0.00	32	10	0.005	0.026	−0.001	−0.011	0.059	−0.015	−0.008	0.003
0.00	32	30	0.008	0.048	−0.004	−0.018	0.143	-1.3E+011	0.016	0.001
0.00	32	50	0.011	0.059	0.004	−0.016	0.182	-9.1E+010	0.020	−0.003
0.20	8	5	0.012	0.033	0.094	0.039	0.047	-9.4E+010	1.047	−0.060
0.20	8	10	0.041	0.082	0.099	0.050	0.133	-1.2E+011	8.845	−0.123
0.20	8	30	0.120	0.160	0.145	0.113	0.231	7.5E+011	−1.698	−0.189
0.20	8	50	0.184	0.204	0.188	0.176	0.258	-3.0E+011	−1.619	−0.231
0.20	16	5	−0.010	0.005	0.030	−0.006	0.016	5.2E+011	−0.534	−0.132
0.20	16	10	0.001	0.034	0.027	−0.009	0.080	7.1E+011	−11.442	−0.230
0.20	16	30	0.036	0.086	0.053	0.019	0.187	-1.4E+011	2.021	−0.470
0.20	16	50	0.083	0.124	0.091	0.066	0.226	-3.0E+010	−8.330	−0.587
0.20	32	5	−0.008	−0.001	0.005	−0.010	0.005	6.6E+010	2.291	−0.106
0.20	32	10	−0.007	0.012	−0.003	−0.018	0.043	4.4E+010	−1.058	−0.200
0.20	32	30	−0.015	0.029	−0.014	−0.035	0.135	7.1E+009	−2.383	−0.603
0.20	32	50	0.002	0.047	−0.001	−0.020	0.181	1.0E+010	−7.190	−0.888

***Simulation results based on 6250 replications.

Table 5: Linear Model under Homoskedasticity and Heteroskedasticity - Variance of Estimates $\rho = 0.3$, $T = 400$

$\mathcal{R}^2_{\varepsilon^2\|z_1^2}$	CP	M	$RCUE_1$	$RCUE_4$	$RCUE2_1$	$RCUE2_4$	GMM	CUE	$LIML$	$FULL1$
0.00	8	5	0.223	0.162	0.130	0.210	0.120	3.8E+026	5.997	0.163
0.00	8	10	0.277	0.155	0.198	0.321	0.075	3.3E+026	9.2E+003	0.249
0.00	8	30	0.487	0.215	0.410	0.650	0.032	3.6E+026	2.2E+003	0.492
0.00	8	50	0.611	0.270	0.530	0.803	0.020	4.7E+026	267.805	0.685
0.00	16	5	0.090	0.074	0.078	0.103	0.062	3.2E+023	0.309	0.080
0.00	16	10	0.125	0.083	0.117	0.158	0.049	1.5E+025	135.774	0.117
0.00	16	30	0.290	0.145	0.280	0.399	0.026	6.7E+026	107.529	0.254
0.00	16	50	0.432	0.206	0.406	0.587	0.018	1.8E+026	320.255	0.388
0.00	32	5	0.039	0.035	0.039	0.043	0.032	0.045	0.039	0.035
0.00	32	10	0.050	0.040	0.054	0.061	0.029	0.068	0.069	0.043
0.00	32	30	0.124	0.075	0.138	0.167	0.019	4.7E+025	3.596	0.081
0.00	32	50	0.221	0.123	0.233	0.299	0.015	3.5E+025	19.187	0.132
0.20	8	5	0.686	0.485	0.195	0.386	0.305	1.2E+027	1.7E+003	1.260
0.20	8	10	0.755	0.408	0.294	0.570	0.161	1.5E+027	7.5E+005	2.045
0.20	8	30	1.044	0.474	0.624	1.078	0.058	7.3E+026	2.6E+005	5.461
0.20	8	50	1.234	0.557	0.812	1.311	0.034	7.4E+026	2.9E+005	8.709
0.20	16	5	0.208	0.168	0.123	0.192	0.137	1.4E+027	722.773	0.764
0.20	16	10	0.309	0.193	0.194	0.313	0.096	4.5E+026	7.6E+005	1.325
0.20	16	30	0.615	0.312	0.453	0.697	0.047	3.8E+026	1.3E+005	4.214
0.20	16	50	0.849	0.415	0.645	0.973	0.030	2.9E+026	1.8E+005	7.226
0.20	32	5	0.070	0.065	0.061	0.074	0.060	2.7E+025	7.0E+004	0.309
0.20	32	10	0.092	0.072	0.083	0.103	0.052	6.5E+024	1.9E+003	0.541
0.20	32	30	0.248	0.148	0.230	0.300	0.034	1.3E+026	7.5E+003	2.238
0.20	32	50	0.394	0.225	0.364	0.484	0.024	6.9E+025	2.3E+005	4.594

***Simulation results based on 6250 replications.

Table 6: Linear Model under Homoskedasticity and Heteroskedasticity - Mean Square Error $\rho = 0.3$, $T = 400$

$\mathcal{R}^2_{\varepsilon^2 \vert z_1^2}$	CP	M	$RCUE_1$	$RCUE_4$	$RCUE2_1$	$RCUE2_4$	GMM	CUE	$LIML$	$FULL1$
0.00	8	5	0.224	0.164	0.135	0.211	0.126	3.8E+026	5.997	0.165
0.00	8	10	0.282	0.166	0.206	0.323	0.099	3.3E+026	9.2E+003	0.253
0.00	8	30	0.504	0.245	0.430	0.664	0.088	3.6E+026	2.2E+003	0.502
0.00	8	50	0.635	0.307	0.556	0.822	0.088	4.7E+026	267.825	0.702
0.00	16	5	0.090	0.075	0.078	0.103	0.064	3.2E+023	0.310	0.080
0.00	16	10	0.125	0.085	0.118	0.158	0.059	1.5E+025	135.784	0.117
0.00	16	30	0.292	0.154	0.282	0.400	0.064	6.7E+026	107.625	0.255
0.00	16	50	0.438	0.222	0.412	0.589	0.070	1.8E+026	320.288	0.390
0.00	32	5	0.039	0.035	0.039	0.043	0.033	0.046	0.039	0.035
0.00	32	10	0.050	0.040	0.054	0.061	0.032	0.068	0.069	0.043
0.00	32	30	0.124	0.077	0.138	0.167	0.040	4.7E+025	3.596	0.081
0.00	32	50	0.221	0.126	0.233	0.300	0.048	3.5E+025	19.187	0.132
0.20	8	5	0.687	0.486	0.204	0.387	0.308	1.2E+027	1.7E+003	1.264
0.20	8	10	0.756	0.414	0.304	0.572	0.179	1.5E+027	7.5E+005	2.061
0.20	8	30	1.058	0.500	0.645	1.091	0.111	7.3E+026	2.6E+005	5.497
0.20	8	50	1.268	0.598	0.848	1.342	0.101	7.4E+026	2.9E+005	8.762
0.20	16	5	0.209	0.168	0.124	0.192	0.137	1.4E+027	723.059	0.782
0.20	16	10	0.309	0.194	0.195	0.313	0.102	4.5E+026	7.6E+005	1.377
0.20	16	30	0.617	0.319	0.456	0.697	0.082	3.8E+026	1.3E+005	4.435
0.20	16	50	0.856	0.431	0.653	0.978	0.081	2.9E+026	1.8E+005	7.571
0.20	32	5	0.071	0.065	0.061	0.074	0.060	2.7E+025	7.0E+004	0.321
0.20	32	10	0.092	0.072	0.083	0.103	0.054	6.5E+024	1.9E+003	0.581
0.20	32	30	0.248	0.148	0.230	0.301	0.052	1.3E+026	7.5E+003	2.601
0.20	32	50	0.394	0.228	0.364	0.484	0.057	6.9E+025	2.3E+005	5.383

***Simulation results based on 6250 replications.

Table 7: Linear Model under Groupwise Clustering - Median Bias $\rho = 0.3$, $T = 400$, $\gamma_\nu = 0.5$

CP	M	Z Clustering	$Clusters$	$RCUE_1$	$RCUE_4$	$RCUE2_1$	$RCUE2_4$	GMM	CUE	$LIML$	$FULL1$
8	5	No	20	0.036	0.053	0.019	0.018	0.087	0.004	0.011	0.045
8	10	No	20	0.040	0.064	0.020	0.018	0.151	−0.002	0.017	0.050
8	15	No	20	0.047	0.063	0.023	0.022	0.191	0.014	0.027	0.056
32	5	No	20	0.003	0.010	−0.005	−0.005	0.024	−0.005	−0.001	0.009
32	10	No	20	0.008	0.018	−0.003	−0.003	0.059	−0.003	−0.003	0.006
32	15	No	20	0.016	0.023	0.008	0.008	0.091	0.006	0.004	0.014
8	5	Yes	20	0.102	0.131	0.086	0.081	0.179	0.007	0.151	0.162
8	10	Yes	20	0.087	0.122	0.063	0.058	0.230	0.012	0.206	0.210
8	15	Yes	20	0.057	0.080	0.026	0.023	0.249	0.009	0.228	0.230
32	5	Yes	20	0.025	0.043	0.008	0.007	0.075	0.000	0.054	0.061
32	10	Yes	20	0.032	0.053	0.015	0.013	0.132	0.002	0.102	0.106
32	15	Yes	20	0.030	0.046	0.013	0.012	0.166	0.007	0.132	0.134
8	5	No	50	0.034	0.057	0.024	0.020	0.089	0.005	0.014	0.051
8	10	No	50	0.053	0.086	0.037	0.032	0.156	0.002	0.021	0.053
8	15	No	50	0.050	0.085	0.034	0.030	0.187	0.001	0.021	0.050
32	5	No	50	0.010	0.017	0.003	0.002	0.028	0.002	0.003	0.012
32	10	No	50	0.016	0.029	0.004	0.003	0.064	0.003	0.002	0.011
32	15	No	50	0.014	0.033	−0.001	−0.002	0.089	−0.002	−0.000	0.009
8	5	Yes	50	0.071	0.100	0.062	0.056	0.137	0.008	0.088	0.108
8	10	Yes	50	0.081	0.120	0.065	0.059	0.197	0.008	0.139	0.150
8	15	Yes	50	0.084	0.122	0.074	0.065	0.228	0.016	0.178	0.185
32	5	Yes	50	0.017	0.030	0.005	0.004	0.050	0.003	0.021	0.030
32	10	Yes	50	0.022	0.044	0.006	0.005	0.099	0.002	0.054	0.061
32	15	Yes	50	0.025	0.052	0.007	0.006	0.130	0.001	0.072	0.078

***Simulation results based on 12000 replications.

Table 8: Linear Model under Groupwise Clustering - Interquartile Range $\rho = 0.3$, $T = 400$, $\gamma_\nu = 0.5$

CP	M	Z Clustering	$Clusters$	$RCUE_1$	$RCUE_4$	$RCUE2_1$	$RCUE2_4$	GMM	CUE	$LIML$	$FULL1$
8	5	No	20	0.531	0.485	0.584	0.588	0.421	0.606	0.569	0.478
8	10	No	20	0.590	0.525	0.665	0.672	0.366	0.708	0.676	0.572
8	15	No	20	0.553	0.506	0.616	0.621	0.310	0.632	0.760	0.636
32	5	No	20	0.271	0.262	0.281	0.281	0.242	0.281	0.254	0.243
32	10	No	20	0.340	0.318	0.367	0.367	0.233	0.368	0.271	0.260
32	15	No	20	0.390	0.369	0.427	0.428	0.218	0.429	0.288	0.276
8	5	Yes	20	0.477	0.416	0.545	0.562	0.349	0.698	0.417	0.379
8	10	Yes	20	0.454	0.378	0.545	0.559	0.256	0.637	0.324	0.309
8	15	Yes	20	0.403	0.353	0.469	0.473	0.214	0.487	0.283	0.273
32	5	Yes	20	0.297	0.271	0.335	0.338	0.236	0.344	0.251	0.240
32	10	Yes	20	0.344	0.302	0.403	0.406	0.201	0.415	0.221	0.215
32	15	Yes	20	0.342	0.310	0.389	0.391	0.180	0.397	0.213	0.209
8	5	No	50	0.527	0.471	0.563	0.575	0.418	0.598	0.578	0.486
8	10	No	50	0.595	0.491	0.648	0.659	0.350	0.709	0.683	0.578
8	15	No	50	0.640	0.518	0.704	0.717	0.309	0.773	0.758	0.641
32	5	No	50	0.250	0.243	0.257	0.258	0.233	0.259	0.249	0.239
32	10	No	50	0.294	0.271	0.312	0.313	0.227	0.314	0.267	0.256
32	15	No	50	0.329	0.295	0.355	0.355	0.215	0.356	0.286	0.274
8	5	Yes	50	0.503	0.435	0.548	0.566	0.371	0.649	0.496	0.437
8	10	Yes	50	0.544	0.426	0.607	0.628	0.297	0.734	0.454	0.416
8	15	Yes	50	0.564	0.441	0.620	0.636	0.254	0.730	0.415	0.389
32	5	Yes	50	0.269	0.251	0.288	0.290	0.232	0.291	0.251	0.242
32	10	Yes	50	0.313	0.274	0.349	0.351	0.213	0.352	0.251	0.243
32	15	Yes	50	0.356	0.303	0.400	0.403	0.197	0.408	0.249	0.242

***Simulation results based on 12000 replications.

Table 9: Linear Model under Groupwise Clustering - Nine Decile Range $\rho = 0.3$, $T = 400$, $\gamma_\nu = 0.5$

CP	M	Z Clustering	$Clusters$	$RCUE_1$	$RCUE_4$	$RCUE2_1$	$RCUE2_4$	GMM	CUE	$LIML$	$FULL1$
8	5	No	20	1.484	1.309	1.736	1.837	1.112	2.129	1.961	1.299
8	10	No	20	1.684	1.421	2.061	2.198	0.924	2.663	2.728	1.648
8	15	No	20	1.616	1.420	1.959	2.034	0.792	2.198	3.689	1.912
32	5	No	20	0.703	0.672	0.745	0.746	0.615	0.747	0.648	0.618
32	10	No	20	0.893	0.819	0.997	0.999	0.585	1.003	0.706	0.672
32	15	No	20	1.041	0.966	1.171	1.176	0.539	1.191	0.755	0.718
8	5	Yes	20	1.405	1.192	1.664	1.929	1.005	1.7E+013	1.587	1.166
8	10	Yes	20	1.324	1.083	1.773	1.970	0.701	6.795	1.032	0.926
8	15	Yes	20	1.260	1.052	1.572	1.673	0.549	2.176	0.817	0.767
32	5	Yes	20	0.818	0.730	1.022	1.068	0.634	1.189	0.706	0.665
32	10	Yes	20	0.959	0.802	1.247	1.298	0.534	1.528	0.633	0.607
32	15	Yes	20	1.016	0.890	1.229	1.267	0.454	1.340	0.564	0.548
8	5	No	50	1.468	1.273	1.597	1.763	1.122	2.098	1.992	1.319
8	10	No	50	1.624	1.266	1.949	2.159	0.901	2.863	2.723	1.665
8	15	No	50	1.753	1.313	2.157	2.365	0.801	3.242	3.418	1.930
32	5	No	50	0.650	0.625	0.681	0.685	0.593	0.686	0.649	0.619
32	10	No	50	0.742	0.668	0.809	0.813	0.559	0.816	0.697	0.664
32	15	No	50	0.866	0.749	0.968	0.976	0.534	0.982	0.745	0.707
8	5	Yes	50	1.488	1.239	1.605	1.889	1.060	5.156	1.893	1.275
8	10	Yes	50	1.473	1.100	1.835	2.146	0.772	7.541	1.632	1.246
8	15	Yes	50	1.555	1.126	1.989	2.285	0.654	5.678	1.435	1.189
32	5	Yes	50	0.728	0.664	0.804	0.819	0.607	0.831	0.690	0.651
32	10	Yes	50	0.855	0.710	1.042	1.066	0.539	1.107	0.671	0.639
32	15	Yes	50	0.973	0.779	1.204	1.233	0.497	1.304	0.662	0.635

***Simulation results based on 12000 replications.

Table 10: Linear Model under Groupwise Clustering - Mean Bias $\rho = 0.3$, $T = 400$, $\gamma_\nu = 0.5$

CP	M	Z Clustering	$Clusters$	$RCUE_1$	$RCUE_4$	$RCUE2_1$	$RCUE2_4$	GMM	CUE	$LIML$	$FULL1$
8	5	No	20	0.008	0.037	−0.027	−0.046	0.074	-2.2E+012	−0.028	0.035
8	10	No	20	0.014	0.055	−0.036	−0.058	0.147	-2.6E+012	0.136	0.047
8	15	No	20	0.014	0.034	−0.366	−0.073	0.188	-9.8E+011	−0.036	0.061
32	5	No	20	−0.008	0.001	−0.019	−0.020	0.016	−0.021	−0.016	−0.003
32	10	No	20	−0.007	0.008	−0.026	−0.032	0.054	-2.0E+010	−0.015	−0.002
32	15	No	20	−0.006	0.007	−0.028	−0.031	0.085	-6.5E+010	0.052	−0.002
8	5	Yes	20	0.081	0.120	0.021	−0.010	0.170	-1.3E+013	0.215	0.152
8	10	Yes	20	0.069	0.115	−0.017	−0.048	0.229	-6.2E+012	0.227	0.206
8	15	Yes	20	0.018	0.057	−0.046	−0.081	0.249	-1.7E+012	0.141	0.230
32	5	Yes	20	−0.003	0.021	−0.043	−0.054	0.060	-1.2E+012	0.012	0.039
32	10	Yes	20	0.004	0.038	−0.048	−0.064	0.128	-8.7E+011	0.082	0.095
32	15	Yes	20	−0.001	0.024	−0.047	−0.058	0.163	-3.8E+011	0.123	0.128
8	5	No	50	0.008	0.041	−0.011	−0.034	0.071	-2.3E+012	−0.244	0.034
8	10	No	50	0.032	0.083	−0.005	−0.030	0.149	-3.4E+012	−0.059	0.049
8	15	No	50	0.028	0.089	−0.019	−0.047	0.184	-3.6E+012	−0.103	0.058
32	5	No	50	−0.001	0.008	−0.011	−0.013	0.019	1.6E+010	−0.011	0.001
32	10	No	50	0.006	0.025	−0.012	−0.014	0.059	1.8E+010	−0.012	0.002
32	15	No	50	−0.000	0.026	−0.024	−0.027	0.084	-9.1E+010	−0.015	−0.004
8	5	Yes	50	0.056	0.094	0.031	−0.001	0.128	-1.0E+013	−0.292	0.102
8	10	Yes	50	0.066	0.122	0.014	−0.021	0.194	-7.3E+012	0.139	0.142
8	15	Yes	50	0.066	0.128	0.011	−0.024	0.226	-6.0E+012	0.085	0.174
32	5	Yes	50	−0.002	0.016	−0.022	−0.026	0.036	-1.0E+011	−0.003	0.014
32	10	Yes	50	0.004	0.038	−0.031	−0.040	0.092	-2.4E+011	0.037	0.046
32	15	Yes	50	0.006	0.046	−0.032	−0.041	0.126	-5.6E+011	0.049	0.069

***Simulation results based on 12000 replications.

Table 11: Linear Model under Groupwise Clustering - Variance of Estimates $\rho = 0.3$, $T = 400$, $\gamma_\nu = 0.5$

CP	M	Z Clustering	$Clusters$	$RCUE_1$	$RCUE_4$	$RCUE2_1$	$RCUE2_4$	GMM	CUE	$LIML$	$FULL1$
8	5	No	20	0.232	0.175	0.288	0.367	0.132	2.6E+027	182.247	0.162
8	10	No	20	0.282	0.192	0.408	0.527	0.084	3.2E+027	1.8E+003	0.248
8	15	No	20	2.152	0.195	1.2E+003	0.493	0.060	6.1E+026	66.550	0.317
32	5	No	20	0.047	0.043	0.056	0.057	0.036	0.064	0.065	0.036
32	10	No	20	0.080	0.066	0.107	0.191	0.032	1.2E+025	0.052	0.044
32	15	No	20	0.113	0.093	0.154	0.164	0.027	1.7E+025	61.775	0.051
8	5	Yes	20	0.216	0.156	0.257	0.365	0.116	1.0E+028	36.471	0.130
8	10	Yes	20	0.339	0.113	0.297	0.411	0.046	4.3E+027	7.891	0.094
8	15	Yes	20	0.452	0.113	2.105	0.363	0.028	6.6E+026	39.801	0.064
32	5	Yes	20	0.072	0.055	0.112	0.140	0.041	8.6E+026	2.451	0.048
32	10	Yes	20	0.093	0.064	0.162	0.212	0.026	2.8E+026	0.382	0.038
32	15	Yes	20	0.123	0.079	0.169	0.209	0.020	1.2E+026	0.038	0.031
8	5	No	50	0.224	0.161	0.235	0.313	0.125	5.4E+027	251.221	0.164
8	10	No	50	0.252	0.149	0.342	0.461	0.078	7.2E+027	64.200	0.248
8	15	No	50	0.288	0.160	0.411	0.551	0.060	3.7E+027	311.093	0.324
32	5	No	50	0.040	0.037	0.045	0.047	0.033	3.2E+024	0.041	0.036
32	10	No	50	0.054	0.043	0.067	0.071	0.029	1.7E+024	0.071	0.043
32	15	No	50	0.074	0.054	0.100	0.111	0.026	2.9E+025	0.153	0.051
8	5	Yes	50	0.222	0.155	0.227	0.325	0.117	1.3E+028	834.605	0.152
8	10	Yes	50	0.201	0.113	0.299	0.423	0.057	5.3E+027	54.942	0.152
8	15	Yes	50	0.222	0.117	0.345	0.485	0.040	3.3E+027	38.744	0.143
32	5	Yes	50	0.052	0.043	0.065	0.074	0.036	1.9E+025	0.139	0.041
32	10	Yes	50	0.072	0.048	0.111	0.135	0.027	9.0E+025	0.065	0.042
32	15	Yes	50	0.092	0.057	0.147	0.179	0.023	2.1E+026	2.581	0.044

***Simulation results based on 12000 replications.

Table 12: Linear Model under Groupwise Clustering - Mean Square Error $\rho = 0.3$, $T = 400$, $\gamma_\nu = 0.5$

CP	M	Z Clustering	$Clusters$	$RCUE_1$	$RCUE_4$	$RCUE2_1$	$RCUE2_4$	GMM	CUE	$LIML$	$FULL1$
8	5	No	20	0.233	0.177	0.289	0.369	0.137	2.6E+027	182.248	0.163
8	10	No	20	0.282	0.195	0.409	0.530	0.105	3.2E+027	1.8E+003	0.250
8	15	No	20	2.152	0.196	1.2E+003	0.498	0.095	6.1E+026	66.551	0.321
32	5	No	20	0.047	0.043	0.056	0.058	0.036	0.064	0.066	0.036
32	10	No	20	0.080	0.066	0.108	0.192	0.035	1.2E+025	0.052	0.044
32	15	No	20	0.113	0.093	0.155	0.165	0.035	1.7E+025	61.778	0.051
8	5	Yes	20	0.223	0.171	0.258	0.365	0.145	1.0E+028	36.517	0.153
8	10	Yes	20	0.343	0.126	0.298	0.413	0.098	4.3E+027	7.943	0.136
8	15	Yes	20	0.453	0.116	2.107	0.369	0.090	6.6E+026	39.821	0.117
32	5	Yes	20	0.072	0.056	0.113	0.143	0.045	8.6E+026	2.452	0.050
32	10	Yes	20	0.093	0.065	0.165	0.216	0.043	2.8E+026	0.389	0.047
32	15	Yes	20	0.123	0.079	0.171	0.213	0.046	1.2E+026	0.053	0.048
8	5	No	50	0.224	0.162	0.235	0.314	0.130	5.4E+027	251.280	0.165
8	10	No	50	0.253	0.156	0.342	0.462	0.100	7.2E+027	64.203	0.250
8	15	No	50	0.289	0.168	0.411	0.553	0.094	3.7E+027	311.103	0.327
32	5	No	50	0.040	0.037	0.046	0.047	0.034	3.2E+024	0.041	0.036
32	10	No	50	0.054	0.044	0.068	0.071	0.033	1.7E+024	0.071	0.043
32	15	No	50	0.074	0.054	0.101	0.112	0.033	2.9E+025	0.153	0.051
8	5	Yes	50	0.225	0.164	0.228	0.325	0.134	1.3E+028	834.690	0.162
8	10	Yes	50	0.205	0.128	0.299	0.423	0.095	5.3E+027	54.961	0.173
8	15	Yes	50	0.226	0.133	0.345	0.486	0.091	3.3E+027	38.752	0.173
32	5	Yes	50	0.052	0.043	0.066	0.075	0.037	1.9E+025	0.139	0.042
32	10	Yes	50	0.073	0.050	0.112	0.137	0.036	9.0E+025	0.066	0.044
32	15	Yes	50	0.092	0.059	0.148	0.181	0.039	2.2E+026	2.583	0.048

***Simulation results based on 12000 replications.

Table 13: Nonlinear Model under Homoskedasticity - Median Bias $T = 400$

$\hat{\rho}$	CP	M	GMM	CUE	$RCUE_1$	$RCUE_4$	$RCUE2_1$	$RCUE2_4$
0.4	12.50	5	0.050	0.006	0.023	0.036	0.023	0.012
0.4	12.50	10	0.083	0.008	0.030	0.053	0.021	0.017
0.4	12.50	30	0.115	0.015	0.047	0.061	0.035	0.040
0.4	12.50	50	0.124	0.019	0.045	0.058	0.038	0.041
0.4	25.00	5	0.029	0.005	0.014	0.021	0.013	0.007
0.4	25.00	10	0.056	0.005	0.017	0.032	0.012	0.008
0.4	25.00	30	0.096	0.011	0.026	0.043	0.017	0.017
0.4	25.00	50	0.108	0.017	0.034	0.046	0.027	0.027

***Simulation results based on 7000 replications.

Table 14: Nonlinear Model under Homoskedasticity - Interquartile Range $T = 400$

$\hat{\rho}$	CP	M	GMM	CUE	$RCUE_1$	$RCUE_4$	$RCUE2_1$	$RCUE2_4$
0.4	12.50	5	0.165	0.239	0.205	0.184	0.188	0.222
0.4	12.50	10	0.127	0.272	0.216	0.170	0.223	0.252
0.4	12.50	30	0.080	0.304	0.236	0.169	0.259	0.272
0.4	12.50	50	0.067	0.298	0.228	0.165	0.255	0.262
0.4	25.00	5	0.126	0.161	0.149	0.139	0.145	0.157
0.4	25.00	10	0.105	0.183	0.157	0.134	0.166	0.178
0.4	25.00	30	0.073	0.226	0.186	0.141	0.208	0.217
0.4	25.00	50	0.062	0.237	0.195	0.147	0.216	0.221

***Simulation results based on 7000 replications.

Table 15: Nonlinear Model under Homoskedasticity - Nine Decile Range $T = 400$

$\hat{\rho}$	CP	M	GMM	CUE	$RCUE_1$	$RCUE_4$	$RCUE2_1$	$RCUE2_4$
0.4	12.50	5	0.467	0.800	0.638	0.572	0.483	0.656
0.4	12.50	10	0.341	0.954	0.695	0.552	0.583	0.774
0.4	12.50	30	0.205	1.369	0.860	0.572	0.806	0.991
0.4	12.50	50	0.167	1.388	0.883	0.606	0.915	1.040
0.4	25.00	5	0.342	0.467	0.421	0.388	0.385	0.440
0.4	25.00	10	0.280	0.543	0.456	0.385	0.454	0.518
0.4	25.00	30	0.186	0.786	0.611	0.443	0.649	0.731
0.4	25.00	50	0.154	0.941	0.730	0.513	0.766	0.837

***Simulation results based on 7000 replications.

Table 16: Nonlinear Model under Homoskedasticity - Mean Bias $T = 400$

$\hat{\rho}$	CP	M	GMM	CUE	$RCUE_1$	$RCUE_4$	$RCUE2_1$	$RCUE2_4$
0.4	12.50	5	0.043	-0.024	0.015	0.029	0.023	0.001
0.4	12.50	10	0.083	-0.028	0.034	0.057	0.044	0.034
0.4	12.50	30	0.118	-0.074	0.054	0.073	0.015	0.025
0.4	12.50	50	0.127	-0.094	0.048	0.066	0.038	0.042
0.4	25.00	5	0.022	-0.007	0.003	0.012	0.006	-0.006
0.4	25.00	10	0.055	-0.009	0.011	0.029	0.005	-0.004
0.4	25.00	30	0.098	-0.042	0.017	0.043	0.000	-0.003
0.4	25.00	50	0.111	-0.057	0.028	0.047	0.015	0.013

***Simulation results based on 7000 replications.

Table 17: Nonlinear Model under Homoskedasticity - Variance of Estimates $T = 400$

$\hat{\rho}$	CP	M	GMM	CUE	$RCUE_1$	$RCUE_4$	$RCUE2_1$	$RCUE2_4$
0.4	12.50	5	0.023	0.245	0.048	0.042	0.022	0.040
0.4	12.50	10	0.011	0.328	0.051	0.036	2.245	2.901
0.4	12.50	30	0.004	0.271	0.068	0.036	1.013	0.986
0.4	12.50	50	0.003	0.675	0.070	0.035	0.408	0.229
0.4	25.00	5	0.012	0.144	0.020	0.017	0.015	0.020
0.4	25.00	10	0.008	0.167	0.024	0.017	0.020	0.028
0.4	25.00	30	0.003	0.107	0.041	0.021	0.042	0.052
0.4	25.00	50	0.002	0.495	0.055	0.026	0.364	0.154

***Simulation results based on 7000 replications.

Table 18: Nonlinear Model under Homoskedasticity - Mean Square Error $T = 400$

$\hat{\rho}$	CP	M	GMM	CUE	$RCUE_1$	$RCUE_4$	$RCUE2_1$	$RCUE2_4$
0.4	12.50	5	0.025	0.245	0.048	0.043	0.023	0.040
0.4	12.50	10	0.018	0.329	0.052	0.039	2.246	2.903
0.4	12.50	30	0.018	0.276	0.071	0.041	1.013	0.986
0.4	12.50	50	0.019	0.684	0.073	0.039	0.410	0.231
0.4	25.00	5	0.012	0.144	0.020	0.017	0.015	0.020
0.4	25.00	10	0.011	0.167	0.025	0.018	0.020	0.028
0.4	25.00	30	0.013	0.109	0.041	0.023	0.042	0.052
0.4	25.00	50	0.015	0.498	0.056	0.028	0.364	0.154

***Simulation results based on 7000 replications.