



*The World's Largest Open Access Agricultural & Applied Economics Digital Library*

**This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.**

**Help ensure our sustainability.**

**Give to AgEcon Search**

AgEcon Search

<http://ageconsearch.umn.edu>

[aesearch@umn.edu](mailto:aesearch@umn.edu)

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

*No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.*

The Stata Journal (2016)  
16, Number 3, pp. 523–549

# The Keane and Runkle estimator for panel-data models with serial correlation and instruments that are not strictly exogenous

Michael Keane  
University of Oxford  
Oxford, UK  
michael.keane@nuffield.ox.ac.uk

Timothy Neal  
University of New South Wales  
Sydney, Australia  
Timothy.Neal@unsw.edu.au

**Abstract.** In this article, we introduce the new command `xtkr`, which implements the [Keane and Runkle \(1992a, \*Journal of Business and Economic Statistics\* 10: 1–9\)](#) approach for fitting linear panel-data models when the available instruments are predetermined but not strictly exogenous. This is a common case that includes dynamic panel-data models as a leading example. Monte Carlo simulations show that, in certain situations, this approach offers an improvement over the popular difference generalized method of moments and system generalized method of moments estimators in terms of bias and root mean squared error. An empirical application to cigarette demand also demonstrates its usefulness for applied researchers.

**Keywords:** `st0443`, `xtkr`, forward filtering, GMM, panel data, lagged dependent variable, endogeneity, strict exogeneity, predetermination

## 1 Introduction

Keane and Runkle (KR) ([1992a](#)) developed a new approach for fitting linear panel-data models when the available instruments are predetermined but not strictly exogenous. This is a common case that includes dynamic panel-data models as a leading example (a lagged dependent variable is predetermined but not strictly exogenous by construction). Unfortunately, until now, the KR estimator has not been available in Stata; this article seeks to remedy this with the new `xtkr` command.

The original motivation for the KR article ([1992a](#)) was the observation that the “within” transformation often used to eliminate fixed effects (FE) in panel-data models violates the orthogonality conditions implied by predetermination. The KR estimator avoids this problem by using “forward filtering” to eliminate serial correlation. Under the forward-filtering transformation, the orthogonality conditions implied by predetermination are maintained. In practice, a key feature of the KR approach was that they used only one or two lags of the predetermined variables as instruments rather than all available lags back to the first period.

[Chamberlain \(1992\)](#), [Hayashi \(1992\)](#), and Schmidt, Ahn, and Wyhowski ([1992](#)) noted that the KR estimator was not fully efficient because it failed to use all available instruments. But [Keane and Runkle \(1992b\)](#) responded that the use of many additional in-

struments would be unwise because, in small samples, it was likely to bias the instrumental variable results toward ordinary least squares (OLS) (that is, the “many-instrument problem”). Nevertheless, the development of more efficient panel-data estimators based on more instruments became a major research program in the 1990s. Examples are Arellano and Bond (1991), Ahn and Schmidt (1995), Arellano and Bover (1995), and Blundell and Bond (1998). For a review, see chapter 8 of Baltagi (2005).

Interestingly, Ziliak (1997) obtained Monte Carlo results that support the original argument of Keane and Runkle (1992b) that the use of additional instruments may lead to bias. In a nice summary of the results, Baltagi (2005, 151–152) states: “Ziliak (1997) performs an extensive set of Monte Carlo experiments for a dynamic panel-data model ... [and] finds that the downward bias of generalized method of moments (GMM) is quite severe as the number of moment conditions expands, outweighing the gains in efficiency. Interestingly, Ziliak finds that the forward filter two-stage least-squares (2SLS) estimator proposed by Keane and Runkle (1992a) performs best in terms of the bias/efficiency tradeoff and is recommended.” In this article, we provide additional Monte Carlo evidence that appears to confirm the conclusions of Ziliak (1997).

Recognizing the problem of too many instruments in GMM, later literature, including Koenker and Machado (1999), Windmeijer (2005), and Roodman (2009a), recommended restricting the instrument matrix to try to remove or reduce the bias. Possible solutions include restricting the number of lags used or “collapsing” the instrument matrix to prevent the proliferation of instruments as  $T$  expands. In this article, we test these restrictions using Monte Carlo analysis and find that there are situations where, in contrast to the KR approach, either technique cannot remove the bias.

Given that there is at least some evidence that the KR estimator may exhibit desirable small-sample properties relative to the alternatives, it seems prudent that applied researchers have easy access to their approach, at least as a robustness check relative to more efficient alternatives.

## 2 Dynamic panel-data models

Consider the linear panel model

$$y_{it} = \beta_0 + \beta_1 y_{it-1} + \beta_2 \mathbf{x}_{it} + \mu_i + \epsilon_{it} \quad (1)$$

where  $i = 1, 2, \dots, N$ ,  $t = 1, 2, \dots, T$ ,  $\mathbf{x}_{it} = (x_{1it}, x_{2it}, \dots, x_{Kit})$  is a  $K \times 1$  vector of regressors,  $\mu_i$  is an individual-specific FE, and  $\epsilon_{it}$  is a mean zero idiosyncratic error. We will assume that  $\mathbf{x}_{it}$  is predetermined, in the sense that  $E\mathbf{x}_{is}\epsilon_{it} = 0$  for  $t \geq s$ . This is a much weaker assumption than strict exogeneity ( $E\mathbf{x}_{is}\epsilon_{it} = 0$  for all  $s$  and  $t$ ). Notice that the lagged dependent variable  $y_{it-1}$  is correlated with the FE  $\mu_i$  by construction, so OLS will give inconsistent estimates of (1). This is true regardless of whether  $E\mathbf{x}_{it}\mu_i = 0$ .

## 2.1 FE estimator

The problem of individual FE in panel data is often addressed by demeaning the data (also known as the “within” transformation), which yields

$$(y_{it} - \bar{y}_i) = \beta_1(y_{it-1} - \bar{y}_i) + \beta_2(\mathbf{x}_{it} - \bar{\mathbf{x}}_i) + (\epsilon_{it} - \bar{\epsilon}_i) \quad (2)$$

The FE is eliminated because  $\mu_i - T^{-1} \sum_{t=1}^T \mu_i = 0$ . The FE estimator is then obtained by applying OLS to (2).

But the FE estimator is inconsistent (in  $N$ ) in the presence of a lagged dependent variable because  $y_{it-1}$  and  $\bar{y}_i$  are correlated with  $\bar{\epsilon}_i$  by construction.<sup>1</sup> Furthermore, even if we did not have a lagged dependent variable, the FE estimator would still be inconsistent because the predetermined regressors  $\mathbf{x}_{it}$  are also correlated with  $\bar{\epsilon}_i$ . The FE approach is valid only under the strong assumption that all covariates are strictly exogenous (in which case they will be uncorrelated with  $\bar{\epsilon}_i$ ).

## 2.2 First-difference estimator

Another transformation often used to eliminate individual-specific FE is first-differencing. First-differencing (1) yields

$$(y_{it} - y_{it-1}) = \beta_1(y_{it-1} - y_{it-2}) + \beta_2(\mathbf{x}_{it} - \mathbf{x}_{it-1}) + (\epsilon_{it} - \epsilon_{it-1}) \quad (3)$$

OLS applied to (3) is inconsistent for two reasons. First,  $y_{it-1}$  is correlated with  $\epsilon_{it-1}$ ; therefore, the covariate  $\Delta y_{it-1}$  is endogenous. Second,  $\mathbf{x}_{it}$  is correlated with  $\epsilon_{it-1}$  (because  $\mathbf{x}_{it}$  is predetermined but not strictly exogenous), which renders the covariate  $\Delta \mathbf{x}_{it}$  endogenous. However, one immediately sees that  $y_{it-2}$  and  $\mathbf{x}_{it-1}$  are potentially valid instruments for estimation of (3),<sup>2</sup> which leads us to consider 2SLS.<sup>3</sup>

## 2.3 2SLS estimator

2SLS applied to (1) or (3) will give consistent estimates provided the instruments are uncorrelated with the error term and satisfy the usual rank condition. To estimate (1) using 2SLS, we require instruments that are uncorrelated with both  $\epsilon_{it}$  and the individual-specific effects. Possibilities include  $\Delta y_{it-1}$  or further lagged differences (DIFF) of the dependent variable.<sup>4</sup> Other candidate instruments are any elements of  $\mathbf{x}_{it}$  that we are

1. The inconsistency of the FE estimator in dynamic panel models was first explained in Nickell (1981). It is worth noting that instrumenting  $y_{it-1}$  with further lags,  $y_{it-s}$  for  $s \geq 2$  does not solve the problem, because all observations of  $y_{it}$  are correlated with  $\bar{\epsilon}_i$ . However, because the strength of the correlation between a single observation and the cross-section average is a decreasing function in time, the FE estimator is consistent in  $T$ .
2. By “potentially valid”, we mean that  $y_{it-2}$  and  $\mathbf{x}_{it-1}$  are uncorrelated with  $\epsilon_{it}$  and  $\epsilon_{it-1}$  by our assumptions. Thus they will be valid instruments provided the rank condition is also satisfied.
3. Note that if we were to adopt a weaker version of the predetermination assumption on  $\mathbf{x}_{it}$ , and instead assume that  $E\mathbf{x}_{is}\epsilon_{it} = 0$  for  $t > s$ , then  $\mathbf{x}_{it-1}$  would no longer be a valid instrument for (3), but  $\mathbf{x}_{it-2}$  would be valid. That is, we would need to go back one more lag.
4. Note that  $\Delta y_{it-1}$  is only a valid instrument for  $y_{it-1}$  in (1) if the process is stationary. For example, if we start at  $y_{i0} = 0$ , then we will have  $E\mu_i \Delta y_{it-1} \neq 0$  for small  $t$ .

willing to assume are uncorrelated with the FE. In many cases, such an assumption is more palatable for first differences of the regressors (that is,  $\Delta x_{it}$ ).

Estimating (3) using 2SLS is possible using any lag of the dependent variable that is uncorrelated with  $\epsilon_{it-1}$ , such as  $y_{it-2}$  and further lags. The regressors in  $x_{it}$  will generally be correlated with  $\epsilon_{it-1}$  themselves (assuming we have only predetermination), but they can also be instrumented using lags.

## 2.4 The KR (1992) estimator

While 2SLS is consistent for (3), the presence of serially correlated errors will lead to inefficient estimates. The KR-DIFF estimator introduced in KR (1992a) uses the idea of forward filtering from the time-series literature to increase efficiency (while still maintaining consistency) in this environment. The estimator first obtains a consistent estimate of  $\hat{\Sigma}_{FD} = 1/N \sum_{i=1}^N \hat{U}_{FD}^i \hat{U}_{FD}^{i'}$ , where  $\hat{U}_{FD}^i$  is the vector of residuals for individual  $i$  from 2SLS estimation of (3). Then, it calculates  $\hat{P}_{FD}$ , which is the upper-triangular Cholesky decomposition of  $\hat{\Sigma}_{FD}^{-1}$ . Finally, it premultiplies (3) with  $\hat{Q}_{FD} = (I_N \otimes \hat{P}_{FD})$  and estimates the transformed equation using 2SLS while retaining the original instruments. Accordingly, the  $(K+1) \times 1$  vector of slope coefficients will be

$$\hat{\beta}_{\text{KR-DIFF}} = \left\{ X' \hat{Q}_{FD}' Z (Z' Z)^{-1} Z' \hat{Q}_{FD} X \right\}^{-1} X' \hat{Q}_{FD}' Z (Z' Z)^{-1} \hat{Q}_{FD} Y$$

where  $X = (\Delta y_{t-1}, \Delta x_{1it}, \Delta x_{2it}, \dots, \Delta x_{Kit})$  and  $Z$  is the set of instruments. Note that the estimator can also be applied to a model in levels using a similar procedure, referred to as KR-LEV. First, estimate (1) through 2SLS, where the instruments  $Z$  are now typically lagged first DIFF in covariates, as we discussed in section 2.3. Then, calculate  $\hat{\Sigma}_{LEV} = 1/N \sum_{i=1}^N \hat{U}_{LEV}^i \hat{U}_{LEV}^{i'}$ , premultiply the equation with  $\hat{Q}_{LEV} = (I_N \otimes \hat{P}_{LEV})$ , where  $\hat{P}_{LEV}$  is the upper-triangular Cholesky decomposition of  $\hat{\Sigma}_{LEV}^{-1}$ , and finally estimate the transformed equation with 2SLS using the original instrument set. Accordingly, the  $(K+1) \times 1$  vector of slope coefficients will be

$$\hat{\beta}_{\text{KR-LEV}} = \left\{ X' \hat{Q}_{LEV}' Z (Z' Z)^{-1} Z' \hat{Q}_{LEV} X \right\}^{-1} X' \hat{Q}_{LEV}' Z (Z' Z)^{-1} \hat{Q}_{LEV} Y$$

where  $X = (y_{t-1}, x_{1it}, x_{2it}, \dots, x_{Kit})$  and  $Z$  is the set of instruments.

## 2.5 GMM estimators

GMM estimation is another way to improve the efficiency of 2SLS. Arellano and Bond (1991) and Blundell and Bond (1998) developed two types of GMM estimators for dynamic panel-data models, known as DIFF-GMM and system GMM (SYS-GMM), respectively. Extensive discussions of each method can be found in the original articles, as well as in Roodman (2009b). Accordingly, the treatment here will be brief.

The basic idea of DIFF-GMM is to estimate (3) using lagged instruments going all the way back to time  $t = 1$ . That is, at time  $t$ , we would use the instrument set consisting

of  $y_{it-2}, \dots, y_{i0}$  and  $x_{it-1}, \dots, x_{i1}$ . Note that the number of instruments grows with  $t$  and missing values are zeroed out, as in [Holtz-Eakin, Newey, and Rosen \(1988\)](#).<sup>5</sup>

The SYS-GMM estimator combines the moment conditions of the equation in levels and first DIFF,

$$\begin{pmatrix} \Delta y_{it} \\ y_{it} \end{pmatrix} = \begin{pmatrix} 0 \\ \beta_0 \end{pmatrix} + \beta_1 \begin{pmatrix} \Delta y_{it-1} \\ y_{it-1} \end{pmatrix} + \beta_2 \begin{pmatrix} \Delta x_{it} \\ x_{it} \end{pmatrix} + \begin{pmatrix} \Delta \epsilon_{it} \\ \mu_i + \epsilon_{it} \end{pmatrix}$$

which is possible because the same linear relationship is believed to apply to both the untransformed and transformed variables. It uses different instruments for each equation, with first-DIFF instruments applied to the equation in levels, and level instruments applied to the equation in first DIFF.<sup>6</sup>

Both DIFF-GMM and SYS-GMM suffer from the potential problem of proliferation of (potentially weak) instruments as  $T$  increases, which may induce bias. For both estimators, it is possible to restrict the instrument set to a specified number of lags, or to “collapse” the instrument matrix, to prevent an explosion in the number of instruments as  $T$  increases. The idea of limiting the number of lags used in the instrument set is explored in [Koenker and Machado \(1999\)](#), and both [Windmeijer \(2005\)](#) and [Roodman \(2009a\)](#) report that it is quite successful in correcting for the issue of weak or too many instruments in standard DIFF-GMM and SYS-GMM. [Roodman \(2009a\)](#) also reports that collapsing the instrument matrix, or a combination of the two techniques, can also reduce bias. Thus both of these options will be tested in the Monte Carlo simulations in section 4.

### 3 The xtkr command

#### 3.1 Syntax

`xtkr depvar [varlist1] (varlist2 = varlist3) [if] [in] [, nocons tdum]`

*varlist1* refers to any exogenous variables, *varlist2* refers to the endogenous variables, and *varlist3* refers to the instruments that are to be used (beyond the exogenous variables and constant term, which are added automatically to the instrument set).

5. By zeroing out any missing values, one avoids having to lose sample periods when allowing for more lagged instruments. Thus one can use as many lagged instruments as possible.

6. Recall that  $\Delta y_{it-1}$  is only a valid instrument for  $y_{it-1}$  in (1) if the process is stationary. Thus SYS-GMM relies on a stationarity assumption.

### 3.2 Options

`nocons` suppresses the constant term.

`tdum` demeans the data across the time dimension (that is, the average across  $i$  for a given  $t$ ). This is equivalent and preferable to adding time dummies to the regression because that can cause collinearity in the second stage.

### 3.3 Replicating Baltagi (2005)

Baltagi (2005) compares results from several methods of fitting dynamic panel-data models. Specifically, in table 8.1 of chapter 8, he estimates the dynamic demand for cigarettes in the United States from 1963 to 1992 using a variety of estimators.<sup>7</sup> The specified equation is

$$\ln C_{it} = \alpha + \beta_1 \ln(C_{i,t-1}) + \beta_2 \ln(P_{it}) + \beta_3 \ln(Y_{it}) + \beta_4 \ln(Pn_{it}) + u_{it}$$

Here  $C_{it}$  is the number of packs of cigarettes sold per person (over 14 years of age) in the  $i$ th state in year  $t$ , for  $i = 1, 2, \dots, N$  and  $t = 1, 2, \dots, T$ , where  $N = 46$  and  $T = 30$ .  $P_{it}$  is the average price (in real terms),  $Y_{it}$  is real per capita disposable income, and  $Pn_{it}$  is the minimum price in bordering states. These three variables in  $\mathbf{x}_{it}$  are assumed to be predetermined, and their lagged values are used to instrument for  $C_{i,t-1}$ . We now replicate Baltagi's (2005) results using the same dataset and the `xtkr` command:

```
use cigar.dta
generate logc = log(c/popgt16*pop)
generate logp = log(price/cpi)
generate logpn = log(pimin/cpi)
generate logy = log(ndi/cpi)

*! Without Time Dummies
regress logc l.logc logp logpn logy
xtabond2 logc l.logc logp logpn logy, gmmstyle(l.logc) ivstyle(logp logpn logy)
ivregress 2sls logc logp logpn logy (l.logc = l.logp l.logpn l.logy)
xtkr logc logp logpn logy (l.logc = l.logp l.logpn l.logy)
xtabond2 logc l.logc logp logpn logy, gmmstyle(l.logc) ///
    ivstyle(logp logpn logy) nolevel
ivregress 2sls d.logc d.logp d.logpn d.logy (d.l.logc = l.logp l.logpn l.logy)
xtkr d.logc d.logp d.logpn d.logy (d.l.logc = l.logp l.logpn l.logy)

*! With Time Dummies
regress logc l.logc logp logpn logy i.yr
xtabond2 logc l.logc logp logpn logy i.yr, gmmstyle(l.logc) ///
    ivstyle(logp logpn logy i.yr)
ivregress 2sls logc logp logpn logy i.yr (l.logc = l.logp l.logpn l.logy)
xtkr logc logp logpn logy (l.logc = l.logp l.logpn l.logy), tdum
xtabond2 logc l.logc logp logpn logy i.yr, gmmstyle(l.logc) ///
    ivstyle(logp logpn logy i.yr) nolevel
ivregress 2sls d.logc d.logp d.logpn d.logy i.yr ///
    (d.l.logc = l.logp l.logpn l.logy)
xtkr d.logc d.logp d.logpn d.logy (d.l.logc = l.logp l.logpn l.logy), tdum
```

7. The dataset is available at <http://www.wiley.com/legacy/wileychi/baltagi/supp/Cigar.txt>.

Table 1. Cigarette demand equation 1963–1992

Estimator	$\ln C_{it-1}$	$\ln P_{it}$	$\ln P n_{it}$	$\ln Y_{it}$	No. instruments
Without Time Dummies					
OLS	0.969 (157.7)	−0.090 (−6.2)	0.024 (1.8)	−0.031 (−5.1)	N/A
SYS-GMM	0.964 (133.73)	−0.105 (−6.52)	0.033 (2.29)	−0.028 (−4.6)	438
2SLS-LEV	0.850 (25.38)	−0.205 (−5.77)	0.052 (3.12)	−0.017 (−2.18)	6
KR-LEV	0.708 (22.71)	−0.311 (−13.88)	0.07 (3.67)	−0.015 (−1.5)	6
DIFF-GMM	0.891 (47.04)	−0.152 (−5.49)	0.050 (1.85)	−0.071 (−6.09)	409
2SLS-DIFF	0.521 (4.67)	−0.345 (−12.14)	0.116 (3.38)	0.175 (4.28)	6
KR-DIFF	0.536 (11.15)	−0.334 (−15.24)	0.088 (4.31)	0.196 (9.82)	6
With Time Dummies					
OLS	0.954 (148.5)	−0.137 (−8.67)	0.037 (2.72)	−0.009 (−1.13)	N/A
SYS-GMM	0.942 (123.85)	−0.172 (−9.51)	0.047 (3.19)	0.002 (0.29)	466
2SLS-LEV	0.611 (11.84)	−0.561 (−8.22)	0.087 (3.51)	0.158 (5.63)	6
KR-LEV	0.561 (15.93)	−0.543 (−15.32)	0.009 (0.15)	0.311 (4.83)	6
DIFF-GMM	0.843 (52.66)	−0.377 (−11.81)	−0.016 (−0.39)	0.139 (3.88)	437
2SLS-DIFF	0.645 (4.24)	−0.406 (−11.77)	0.038 (0.86)	0.156 (2.84)	6
KR-DIFF	0.703 (17.52)	−0.338 (−13.51)	0.075 (2.59)	0.225 (6.44)	6

Note:  $t$  statistics are in parentheses. Baltagi's (2005) table 8.1 contains OLS, 2SLS-LEV, KR-LEV, 2SLS-DIFF, and KR-DIFF, all without time dummies, and DIFF-GMM with time dummies. There are modest discrepancies in the results for 2SLS-DIFF and KR-DIFF, while the other four estimators can be replicated exactly. To avoid potential confusion, note that Baltagi (2005) refers to DIFF-GMM as "GMM-one-step" and to KR-LEV as "2SLS-KR." Baltagi (2005) does not report results for SYS-GMM.

A complication in replicating Baltagi's (2005) results is that he used time dummies for some estimators but not others.<sup>8</sup> So in table 1, we report results for all estimators with and without time dummies. Consider first the results without time dummies. Notice that the OLS coefficient on lagged consumption is very large (0.969). However, it is well known that such a high degree of state dependence may arise because of failure to deal with individual effects (see Keane [1997]). The negative income effect is also indicative of a failure to control for taste heterogeneity. The OLS estimates imply a short-run price elasticity of demand of only  $-0.09$  but a long-run elasticity of  $-0.090/0.031 = -2.9$ .

The SYS-GMM results are very similar to OLS. We would strongly suspect that the use of 438 instruments biases the SYS-GMM results toward OLS. In contrast, the 2SLS-DIFF and KR-DIFF results (which rely on only 6 instruments) are very different from OLS but very similar to each other. The KR-DIFF standard errors are quite a bit smaller than those of 2SLS, suggesting a gain in efficiency from forward filtering. The KR-DIFF coefficient on lagged consumption is 0.536, implying a still strong but perhaps more plausible level of state dependence, while the income effect now has a theoretically correct positive sign. The KR-DIFF estimates imply a short-run price elasticity of demand of  $-0.33$  and a long-run elasticity of  $-0.334/0.464 = -0.72$ .

8. In the original table, the results for DIFF-GMM included time dummies, while the results for OLS, 2SLS, and KR did not.



The DIFF-GMM results, which rely on 409 instruments, are closer to OLS/SYS-GMM than to KR-DIFF/2SLS-DIFF. This may again be indicative of a many-instrument problem. The lag coefficient is 0.891, and the income effect is negative and significant. The estimates imply a short-run price elasticity of demand of only  $-0.15$  but a long-run elasticity of  $-0.090/0.109 = -0.83$ .

In summary, while DIFF-GMM and KR-DIFF/2SLS-DIFF all produce plausible short-run versus long-run elasticity patterns, we would prefer KR-DIFF/2SLS-DIFF based on having the theoretically correct positive income effect and then KR-DIFF based on giving more precise estimates.

The inclusion of time dummies accounts for aggregate taste shifts. With time dummies, the SYS-GMM results again appear problematic because they are nearly identical to OLS. All other estimators now produce theoretically correct positive signs on income. But DIFF-GMM generates a theoretically incorrect negative sign on the price in bordering states (and again produces a lag coefficient closer to OLS than to 2SLS). The remaining estimators, which are 2SLS and KR in either levels or DIFF, produce rather similar results. For instance, the KR-DIFF estimates imply a short-run price elasticity of demand of  $-0.34$  and a long-run elasticity of  $-0.338/0.297 = -1.14$ . Theoretically, one would expect long-run demand elasticities to be greater than one. Thus the inclusion of time effects appears to be important.

Finally, note that without time dummies, the 2SLS-DIFF and 2SLS-LEV estimates diverge noticeably. But with time dummies, they are quite close. This suggests that the stationarity assumption is violated in the levels equation unless time effects are included.

## 4 Monte Carlo simulation

### 4.1 Data-generating process

In this section, we present a set of Monte Carlo experiments that shed further light on the performance of the alternative panel-data estimators described in section 2. The data-generating process is

$$y_{it} = \beta_0 + \beta_1 y_{it-1} + \beta_2 x_{it} + \mu_i + \epsilon_{it} \quad (4)$$

where the regressor  $x_{it}$  follows the process

$$x_{it} = \rho x_{it-1} + \eta_i + \delta \mu_i + \kappa \epsilon_{it} + \theta \epsilon_{it-1} + \omega_{it} \quad (5)$$

In (4), we generate the individual effect  $\mu_i$  as independent and identically distributed (IID)  $N(0, \sigma_\mu^2)$  and the idiosyncratic error terms  $\epsilon_{it}$  as IID  $N(0, \sigma_\epsilon^2)$ . In (5), we generate the individual effect  $\eta_i$  as IID  $N(0, \sigma_\eta^2)$  and the idiosyncratic error term  $\omega_{it}$  as IID  $N(0, \sigma_\omega^2)$ .

Thus the parameters  $\delta$ ,  $\kappa$ , and  $\theta$  determine the extent of correlation of the regressor with the individual effects, the idiosyncratic error term, and the lagged idiosyncratic

error term, respectively. Scenarios where  $\kappa = 0$  yet  $\theta \neq 0$  produce a predetermined (or weakly exogenous) regressor that is correlated with the last period's error term. We will consider scenarios with different values of  $\delta$ ,  $\kappa$ , and  $\theta$ , as well as different values of the autoregressive-coefficient  $\rho$  in (5).

Parameter settings held invariant across scenarios are  $\beta_0 = 0.5$ ,  $\beta_1 = 0.5$ ,  $\beta_2 = 1$ ,  $\sigma_\mu^2 = \sigma_\epsilon^2 = 0.2$ , and  $\sigma_\eta^2 = \sigma_\omega^2 = 0.2$ . We felt these values were fairly realistic if one thinks of  $y_{it}$  as the log of wages or earnings.

We initialize the series (4) and (5) for  $t = -10, \dots, 0, 1, \dots, T$  with  $x_{i,-10} = y_{i,-10} = 0$ . The first 10 observations are discarded before estimation to ensure the process is (approximately) stationary, and so results are not sensitive to the initial value. Finally, we set  $N = 100$  in all cases, but we will consider cases of  $T = 5, 10, 15$ , or 20.

We consider four scenarios that vary the degree of endogeneity of the regressor  $x_{it}$ . The design is presented in table 2. In the first scenario, the regressor is strictly exogenous in the sense that  $Ex_{is}\epsilon_{it} = 0$  for all  $s$  and  $t$ . In the second scenario,  $\theta > 0$ , so the regressor is only predetermined, but we still have  $Ex_{it}\epsilon_{it} = 0$  for all  $t$ . In the third scenario, we allow the regressor to be correlated with the individual effect in (4) and to be contemporaneously correlated with the error term in (4). Finally, in the fourth scenario, we also introduce serial persistence of the first-order autoregressive form into the regressor. This means that  $x_{it}$  is correlated with both current and all lagged values of  $\epsilon_{it}$ .<sup>9</sup>

Table 2. Scenario design

Scenario	$\rho$	$\delta$	$\kappa$	$\theta$
1	0	0	0	0
2	0	0	0	0.5
3	0	0.5	0.5	0.5
4	0.5	0.5	0.5	0.5

The performance of each estimator under consideration will be measured using the mean bias and the root mean squared error (RMSE), which is defined as

$$\text{RMSE} = \sqrt{S^{-1} \sum_{s=1}^S (\hat{\beta}_s - \beta)^2}$$

where  $\beta$  is the true slope coefficient,  $s$  is the simulation number, and  $S$  is the total number of simulation repetitions, which was set to 2,000 for this study.

9. In section 4.3, we will also consider some sensitivity tests that vary this design. Specifically, we will consider cases where we increase persistence in the dependent variable by increasing  $\beta_1$  from 0.5 to 0.8. We will also consider effects of increasing or decreasing the sample size  $N$ .

## 4.2 Results

In table 3, we present results for the case where the regressor is strictly exogenous in the sense that  $Ex_{is}\epsilon_{it} = 0$  for all  $s$  and  $t$ . We report results for KR-DIFF, KR-LEV, DIFF-GMM, SYS-GMM, OLS, and FE.<sup>10</sup> For each estimator, the table reports the bias and RMSE across 2,000 Monte Carlo estimations. The top panel reports results for  $\beta_1$ , the coefficient on the lagged dependent variable, while the bottom panel reports results for  $\beta_2$ , the coefficient on the regressor.

For KR and GMM, we report results using a number of alternative instrument sets. The instrument sets are described in table 13. For example, for KR-DIFF, using 2 lags, we use  $\Delta x_{it}$  as an instrument for itself (as it is strictly exogenous), and we also use  $x_{it-1}$ ,  $x_{it-2}$  and  $y_{it-2}$ . We do not go beyond  $t - 2$ .

The GMM estimators adopt the Holtz-Eakin, Newey, and Rosen (1988)-type instrument sets, which are more difficult to explain as they vary over time. For example, consider DIFF-GMM when  $T = 5$ . As above,  $\Delta x_{it}$  is a valid instrument for itself because it is strictly exogenous. At  $t = 3$ , which is the first period that can be used, a valid instrument is  $y_{i1}$ . At  $t = 4$ , the valid instruments are  $y_{i2}$  and  $y_{i1}$ , and finally, at  $t = 5$ , they are  $y_{i3}$ ,  $y_{i2}$ , and  $y_{i1}$ . Thus we have seven instruments and seven moment conditions. Note how the number of instruments grows rapidly with  $T$ . For instance, as we see in table 3, it is 172 when  $T = 20$ .

One can visualize the GMM instrument matrix as a lower-triangular matrix where, with each period, we add a new row (leading to a quadratic growth rate in the number of instruments). Methods of “collapsing” the instrument set involve transforming these lower-triangular instrument matrices into row vectors, where the number of columns grows with  $t$ . One could also constrain the instrument set by restricting the number of lags to be used as instruments. Both techniques individually make the number of instruments grow linearly in  $t$ . Combined, they will make the number of instruments constant in  $t$  (as in 2SLS and KR).

As we would expect given the presence of an individual effect, OLS generates severely upward-biased estimates for the lag coefficient  $\beta_1$ . For instance, in the  $T = 5$  case, the mean of the bias is 0.227, implying a mean estimate of 0.727 compared with the true value of 0.50. When we implement FE, for  $T = 5$ , the bias is of similar magnitude but in the opposite direction (−0.224). Yet, as expected, the bias shrinks as  $T$  grows large, falling to a modest −0.0042 when  $T = 20$ .

The KR-DIFF, KR-LEV, and DIFF-GMM estimators all do a fine job of eliminating the OLS bias. SYS-GMM works well when  $T = 5$ , but it shows significant bias at  $T = 10$ , 15, and 20. Bias is still present when  $T = 15$  or 20 even once the instrument set is restricted to using 2 lags. This looks like evidence of a many-instrument problem. The bias vanishes if the instrument matrix is collapsed, and RMSE improves.

10. In all scenarios, the one-step estimator was used for DIFF-GMM and SYS-GMM. We found it performed better (on average) than the two-step estimator under this particular data-generating process.

At  $T = 20$ , the RMSEs for KR-DIFF (2 lags), KR-LEV (2 lags), DIFF-GMM (full-instrument set), and SYS-GMM (collapsed) are 2.65, 2.37, 2.54, and 2.56 for  $\beta_1$  and 3.00, 2.88, 2.71, and 3.37 for  $\beta_2$ , respectively. Thus there is little to choose between these estimators. It is interesting to note, however, that adding a third lag to the small-instrument sets used by the KR estimators does not lead to noticeable efficiency improvements.

In table 4, we consider the case where  $x_{it}$  is predetermined, in the sense that  $Ex_{is}\epsilon_{it} = 0$  for  $t \geq s$  but not for  $t < s$ . In this case, we can no longer use  $\Delta x_{it}$  as an instrument for itself, so the instrument sets change (see table 13). For instance, the instrument set for KR-DIFF using 2 lags is simply  $x_{it-1}$ ,  $x_{it-2}$ , and  $y_{it-2}$ . And the period  $t$  instrument set for DIFF-GMM consists of  $x_{is}$  for  $s = 1, \dots, t-1$  and  $y_{is}$  for  $s = 1, \dots, t-2$ .

The biases of OLS and FE are similar to the case where  $x_{it}$  was strictly exogenous (compare table 3 and table 4). As before, KR-DIFF and KR-LEV do a fine job of eliminating the OLS bias. However, now both SYS-GMM and DIFF-GMM show significant bias at  $T = 10, 15$ , and  $20$  for  $\beta_1$ . Restricting the instrument matrix to two lags lessens the bias in either case, while collapsing the matrix eliminates it entirely. RMSE generally improves with either technique.

At  $T = 20$ , the RMSEs for KR-DIFF (2 lags), KR-LEV (2 lags), DIFF-GMM (collapsed), and SYS-GMM (collapsed) are 2.65, 2.08, 2.35, and 2.01 for  $\beta_1$  and 2.99, 2.51, 2.65, and 2.53 for  $\beta_2$ , respectively. So as before, there is little to choose between these estimators. There is again little evidence that adding instruments beyond the small set used by KR leads to much gain in efficiency.

In table 5, we introduce correlation between the regressor and both the individual effect ( $Ex_{it}\mu_{it} \neq 0$ ) and the contemporaneous error term ( $Ex_{it}\epsilon_{it} \neq 0$ ).<sup>11</sup> In contrast to the previous case,  $x_{it-1}$  is no longer a valid instrument. So, for instance, the instrument set for KR-DIFF using 3 lags is now  $x_{it-2}$ ,  $x_{it-3}$ ,  $y_{it-2}$ , and  $y_{it-3}$ . And the period  $t$  instrument set for DIFF-GMM consists of  $x_{is}$  and  $y_{is}$  for  $s = 1, \dots, t-2$ .

Once again, KR-DIFF and KR-LEV do a fine job of eliminating the OLS bias. However, both SYS-GMM and DIFF-GMM exhibit substantial bias. In contrast to scenario 2, where this bias was primarily in the lagged dependent variable coefficient, now both DIFF-GMM and SYS-GMM exhibit substantial bias for the regressor coefficient as well. For instance, in the  $T = 20$  case, the mean upward bias for the estimate of  $\beta_2$  is 25% for DIFF-GMM and 19% for SYS-GMM. These are actually larger than the OLS bias (18%).

11. Note that in this case, the bias of the FE estimator no longer vanishes as  $T$  grows large. And, in contrast to table 4, the FE estimator is now severely biased for the regressor coefficient (not only the lagged dependent variable). This is because FE does nothing to deal with the contemporaneous correlation between the regressor and both the individual effect and the contemporaneous error term.

In contrast to scenarios 1 and 2, in scenario 3, collapsing the instrument matrix fails to completely eliminate the bias in the DIFF-GMM and SYS-GMM estimators (although it does substantially reduce it). For instance, at  $T = 20$ , the mean bias for the estimate of  $\beta_2$  is 11% for DIFF-GMM (collapsed) and 7.4% for SYS-GMM (collapsed).

Versions of GMM that restrict the number of lags to two also fail to eliminate the bias.<sup>12</sup> For instance, at  $T = 20$ , the mean bias for the estimate of  $\beta_2$  is 12.5% for DIFF-GMM (2 lags) and 11.6% for SYS-GMM (2 lags). A combination of both restricting lags and collapsing the instrument matrix does successfully eliminate the bias but at a high cost to efficiency (see below).

It is again interesting to compare the RMSE of the estimators. At  $T = 20$ , the RMSEs for KR-DIFF (3 lags), KR-LEV (3 lags), DIFF-GMM (collapsed), and SYS-GMM (collapsed) are 2.80, 3.00, 2.73, and 2.61 for  $\beta_1$  and 7.87, 8.17, 13.39, and 9.82 for  $\beta_2$ , respectively. Strikingly, the use of many instruments by DIFF-GMM and SYS-GMM leads to serious finite sample problems in estimating  $\beta_2$  in particular.

Similarly, the RMSEs for the two best-performing GMM estimators at  $T = 20$ , DIFF-GMM (2 lags, collapsed) and SYS-GMM (2 lags, collapsed), are 3.17 and 3.13 for  $\beta_1$  and 11.05 and 8.96 for  $\beta_2$ , respectively. Comparing these with the RMSEs of 2.80 and 7.87 for KR-DIFF, we see efficiency losses of 12% to 14% for SYS-GMM and up to 40% for DIFF-GMM.

Finally, table 6 reports results for scenario 4, where the regressor is correlated with the current and all lagged values of the idiosyncratic error. Results for this case are fairly similar to scenario 3. In particular, note that a combination of both lag reduction and collapsing the instrument matrix is again necessary to eliminate bias from the GMM estimators. This again comes at a high cost in terms of efficiency. For example, the RMSEs at  $T = 20$  for DIFF-GMM (2 lags, collapsed) and SYS-GMM (2 lags, collapsed) are 8.16 and 7.78 for  $\beta_2$ , compared with 6.13 for KR-DIFF. So we have efficiency losses of 33% for DIFF-GMM and 27% for SYS-GMM.

---

12. We also obtained results with three or four lags, and they were fairly similar. Some lag lengths performed slightly better than others in some cases but worse in others. We found no consistent pattern that would suggest a clearly preferred lag length. Obviously, an applied researcher would have no way of knowing the best lag length in any particular situation.

Table 3. Simulation results for scenario 1

$(N = 100, T)$		Bias ( $\times 100$ )				RMSE ( $\times 100$ )			
Instr. set	Instr. count	5	10	15	20	5	10	15	20
<b>Results for <math>\beta_1 = 0.5</math></b>									
<b>KR-DIFF</b>									
2 lags	4	-0.59	-0.01	0.02	0.05	7.36	4.11	3.14	2.65
3 lags	6	-1.63	-0.04	0.02	0.06	9.21	4.25	3.08	2.54
<b>KR-LEV</b>									
2 lags	4	-0.27	0.05	-0.01	0.08	4.78	3.43	2.68	2.37
3 lags	6	0.08	0.08	0.02	0.06	5.43	3.38	2.56	2.28
<b>DIFF-GMM</b>									
full	7/37/92/172	-4.73	-2.11	-1.49	-1.25	15.72	5.22	3.26	2.54
full (coll.)	4/9/14/19	-4.46	-1.07	-0.59	-0.29	22.53	7.13	4.40	3.28
2 lags	6/16/26/36	-5.39	-2.54	-1.46	-0.91	18.78	7.65	4.68	3.35
2 lags (coll.)	3	-1.47	-0.16	0.12	0.08	25.23	8.98	5.69	4.34
<b>SYS-GMM</b>									
full	11/46/106/191	1.19	4.33*	5.80*	7.08**	8.61	6.39	6.92	7.84
full (coll.)	6/11/16/21	-1.22	-0.11	-0.16	-0.04	8.96	4.51	3.18	2.56
2 lags	10/25/40/55	1.10	2.88	3.02*	3.11*	8.38	5.15	4.35	4.08
2 lags (coll.)	5	-1.15	0.07	0.02	0.08	9.02	4.70	3.39	2.77
<b>OLS</b>									
Pooled	N/A	22.74***	22.76***	22.80***	22.76***	22.89	22.87	22.90	22.85
FE	N/A	-22.40***	-9.29***	-5.76***	-4.14***	22.85	9.60	6.06	4.41
<b>Results for <math>\beta_2 = 1</math></b>									
<b>KR-DIFF</b>									
2 lags	4	-0.07	0.09	0.05	-0.04	7.03	4.51	3.54	3.00
3 lags	6	-0.38	0.12	0.07	-0.04	8.53	4.80	3.61	2.99
<b>KR-LEV</b>									
2 lags	4	-0.25	0.08	-0.02	-0.02	6.54	4.15	3.28	2.88
3 lags	6	-0.55	0.06	0.02	-0.01	8.17	4.37	3.32	2.90
<b>DIFF-GMM</b>									
full	7/37/92/172	-1.91	-0.61	-0.34	-0.34	9.49	4.54	3.33	2.71
full (coll.)	4/9/14/19	-1.95	-0.39	-0.19	-0.18	12.60	5.28	3.75	2.99
2 lags	6/16/26/36	-2.26	-0.95	-0.47	-0.33	10.75	5.26	3.75	2.96
2 lags (coll.)	3	-0.51	0.00	0.11	-0.03	13.77	5.90	4.18	3.28
<b>SYS-GMM</b>									
full	11/46/106/191	-0.90	-1.99	-2.98	-3.72*	6.26	4.68	4.73	5.08
full (coll.)	6/11/16/21	0.30	0.15	0.05	0.03	6.36	4.28	3.66	3.37
2 lags	10/25/40/55	-0.84	-1.16	-1.34	-1.33	6.21	4.30	3.78	3.54
2 lags (coll.)	5	0.26	0.07	-0.03	-0.02	6.38	4.30	3.67	3.39
<b>OLS</b>									
Pooled	N/A	-22.57***	-22.31***	-22.61***	-22.41***	23.29	22.80	23.00	22.76
FE	N/A	-7.88*	-1.74	-0.75	-0.42	9.83	3.98	2.88	2.35

Note: \*\*\*, \*\*, and \* denote significance at the 1%, 5%, and 10% confidence level, respectively. There are 2,000 Monte Carlo simulations with  $N = 100$  and variable  $T$ . "coll." refers to a collapsed instrument matrix in GMM. For further details on the instruments used, please see table 13.

Table 4. Simulation results for scenario 2

$(N = 100, T)$		Bias ( $\times 100$ )				RMSE ( $\times 100$ )			
Instr. set	Instr. count	5	10	15	20	5	10	15	20
<b>Results for <math>\beta_1 = 0.5</math></b>									
<b>KR-DIFF</b>									
2 lags	3	-0.85	-0.26	-0.15	-0.08	10.99	4.72	3.27	2.65
3 lags	5	-3.36	-0.34	-0.18	-0.09	13.32	4.74	3.21	2.57
<b>KR-LEV</b>									
2 lags	4	-0.27	-0.04	-0.05	0.04	4.49	3.04	2.38	2.08
3 lags	6	0.13	-0.03	-0.04	0.02	5.20	3.07	2.32	2.02
<b>DIFF-GMM</b>									
full	15/80/195/360	-6.71	-3.72*	-2.79*	-2.35*	11.00	5.06	3.60	2.92
full (coll.)	7/17/27/37	-3.99	-1.35	-0.80	-0.51	11.45	4.63	3.08	2.35
2 lags	11/31/51/71	-7.24	-3.85	-2.51	-1.77	12.88	6.36	4.27	3.18
2 lags (coll.)	4	-1.76	-0.26	-0.06	-0.02	12.46	5.98	4.11	3.28
<b>SYS-GMM</b>									
full	23/98/223/398	3.46	6.17*	8.00***	9.39***	7.16	7.24	8.58	9.79
full (coll.)	10/20/30/40	-1.10	-0.26	-0.27	-0.16	7.28	3.68	2.55	2.01
2 lags	19/49/79/109	2.91	3.93*	4.22*	4.27**	6.86	5.36	5.03	4.86
2 lags (coll.)	7	-0.87	0.10	0.07	0.10	7.48	3.96	2.85	2.33
<b>OLS</b>									
Pooled	N/A	22.17***	22.16***	22.21***	22.16***	22.33	22.27	22.31	22.25
FE	N/A	-23.14***	-9.96***	-6.20***	-4.48***	23.48	10.18	6.41	4.67
<b>Results for <math>\beta_2 = 1</math></b>									
<b>KR-DIFF</b>									
2 lags	3	-0.72	-0.13	-0.09	-0.13	12.60	5.30	3.77	2.99
3 lags	5	-3.15	-0.08	-0.08	-0.15	15.49	5.27	3.67	2.98
<b>KR-LEV</b>									
2 lags	4	-0.36	0.10	0.02	-0.02	6.13	3.66	2.86	2.51
3 lags	6	-0.86	0.05	0.05	-0.01	7.82	3.84	2.94	2.56
<b>DIFF-GMM</b>									
full	15/80/195/360	-6.05	-2.29	-1.37	-1.04	11.75	4.85	3.21	2.54
full (coll.)	7/17/27/37	-3.46	-0.89	-0.42	-0.36	12.43	5.04	3.31	2.65
2 lags	11/31/51/71	-6.07	-2.46	-1.40	-0.97	12.84	5.77	3.78	2.90
2 lags (coll.)	4	-1.32	-0.08	0.03	-0.09	13.00	5.82	3.92	3.15
<b>SYS-GMM</b>									
full	23/98/223/398	-0.22	-1.39	-3.11*	-4.71*	7.16	4.40	4.65	5.69
full (coll.)	10/20/30/40	-0.83	-0.07	-0.09	-0.12	7.79	4.09	3.02	2.53
2 lags	19/49/79/109	-0.09	0.03	-0.07	-0.18	7.13	3.96	3.00	2.62
2 lags (coll.)	7	-0.73	0.10	0.04	-0.02	7.83	4.19	3.15	2.70
<b>OLS</b>									
Pooled	N/A	-24.62***	-24.32***	-24.57***	-24.37***	25.19	24.69	24.87	24.65
FE	N/A	-10.62**	-2.75*	-1.34	-0.87	11.73	4.20	2.85	2.28

Note: \*\*\*, \*\*, and \* denote significance at the 1%, 5%, and 10% confidence level, respectively. There are 2,000 Monte Carlo simulations with  $N = 100$  and variable  $T$ . "coll." refers to a collapsed instrument matrix in GMM. For further details on the instruments used, please see table 13.

Table 5. Simulation results for scenario 3

$(N = 100, T)$		Bias ( $\times 100$ )				RMSE ( $\times 100$ )			
Instr. set	Instr. count	5	10	15	20	5	10	15	20
<b>Results for <math>\beta_1 = 0.5</math></b>									
<b>KR-DIFF</b>									
3 lags	4	-3.09	-0.53	-0.33	-0.20	15.22	5.15	3.51	2.80
4 lags	6	-9.11	-1.13	-0.51	-0.34	23.56	5.18	3.36	2.70
<b>KR-LEV</b>									
3 lags	4	-1.40	-0.58	-0.22	-0.17	11.70	5.12	3.58	3.00
4 lags	6	1.50	-0.91	-0.39	-0.24	13.10	4.70	3.15	2.54
<b>DIFF-GMM</b>									
full	12/72/182/342	-6.78	-5.85**	-5.56***	-5.46***	10.65	6.54	5.87	5.63
full (coll.)	6/16/26/36	-6.81	-3.27	-2.33*	-1.86*	12.99	5.20	3.48	2.73
2 lags	10/30/50/70	-5.94	-4.03*	-3.58*	-3.40*	10.46	5.59	4.52	4.07
2 lags (coll.)	4	-5.08	-0.71	-0.27	-0.13	15.83	6.37	4.07	3.17
<b>SYS-GMM</b>									
full	19/89/209/379	0.21	2.53	3.74*	4.66**	6.44	4.23	4.58	5.21
full (coll.)	9/19/29/39	-5.63	-2.82	-2.10	-1.67	11.10	4.86	3.36	2.61
2 lags	17/47/77/107	0.20	1.25	1.57	1.66	6.65	3.78	3.13	2.86
2 lags (coll.)	7	-4.40	-0.58	-0.28	-0.08	11.88	5.49	3.96	3.13
<b>OLS</b>									
Pooled	N/A	11.85***	11.76***	11.82***	11.75***	12.05	11.91	11.96	11.87
FE	N/A	-19.33***	-10.71***	-8.42***	-7.39***	19.56	10.82	8.50	7.45
<b>Results for <math>\beta_2 = 1</math></b>									
<b>KR-DIFF</b>									
3 lags	4	1.53	0.99	0.66	0.39	23.40	11.92	9.21	7.87
4 lags	6	4.75	2.68	1.53	0.86	31.66	12.12	8.96	7.66
<b>KR-LEV</b>									
3 lags	4	2.46	1.41	0.79	0.48	29.50	12.78	9.55	8.17
4 lags	6	9.08	2.94	1.65	0.99	45.41	12.74	9.30	7.78
<b>DIFF-GMM</b>									
full	12/72/182/342	11.95	18.79***	22.62***	24.99***	20.72	20.05	23.13	25.24
full (coll.)	6/16/26/36	20.74	15.15*	12.74*	10.93*	40.03	21.07	16.14	13.39
2 lags	10/30/50/70	10.79	12.37*	12.52**	12.54**	21.36	15.58	14.44	13.91
2 lags (coll.)	4	18.08	3.20	1.53	0.71	56.80	24.90	14.77	11.05
<b>SYS-GMM</b>									
full	19/89/209/379	11.97	16.01**	17.90***	18.65***	20.36	18.12	19.06	19.49
full (coll.)	9/19/29/39	14.74	9.74*	8.29*	7.37*	28.95	14.95	11.58	9.82
2 lags	17/47/77/107	10.82	11.74*	11.66*	11.59**	20.27	15.11	13.86	13.16
2 lags (coll.)	7	11.35	2.36	1.33	0.69	31.76	15.39	11.13	8.96
<b>OLS</b>									
Pooled	N/A	17.98***	18.37***	18.09***	18.28***	18.67	18.80	18.44	18.61
FE	N/A	25.03***	32.61***	34.04***	34.60***	25.43	32.73	34.11	34.65

Note: \*\*\*, \*\*, and \* denote significance at the 1%, 5%, and 10% confidence level, respectively. There are 2,000 Monte Carlo simulations with  $N = 100$  and variable  $T$ . "coll." refers to a collapsed instrument matrix in GMM. For further details on the instruments used, please see table 13.



Table 6. Simulation results for scenario 4

$(N = 100, T)$		Bias ( $\times 100$ )				RMSE ( $\times 100$ )			
Instr. set	Instr. count	5	10	15	20	5	10	15	20
<b>Results for <math>\beta_1 = 0.5</math></b>									
<b>KR-DIFF</b>									
3 lags	4	-4.08	-0.80	-0.49	-0.25	23.05	7.43	4.73	3.75
4 lags	6	-13.66	-1.92	-0.92	-0.51	31.13	7.55	4.69	3.64
<b>KR-LEV</b>									
3 lags	4	0.19	-0.94	-0.45	-0.24	15.31	7.22	4.96	3.97
4 lags	6	3.39	-1.35	-0.70	-0.33	17.22	6.41	4.31	3.39
<b>DIFF-GMM</b>									
full	12/72/182/342	-12.01*	-10.15***	-9.78***	-9.75***	16.71	10.84	10.05	9.89
full (coll.)	6/16/26/36	-14.13	-5.82*	-4.09*	-3.16*	24.22	8.69	5.63	4.24
2 lags	10/30/50/70	-11.34*	-8.10*	-6.94**	-6.24**	16.99	9.71	7.95	6.95
2 lags (coll.)	4	-9.00	-1.12	-0.48	-0.16	27.50	10.32	6.71	5.03
<b>SYS-GMM</b>									
full	19/89/209/379	1.71	2.76	3.52*	3.96*	8.09	5.05	4.76	4.87
full (coll.)	9/19/29/39	-6.69	-4.80	-3.76*	-2.99*	15.78	7.79	5.36	4.13
2 lags	17/47/77/107	1.90	2.43	2.81	2.93*	8.50	5.29	4.55	4.26
2 lags (coll.)	7	-4.49	-0.84	-0.30	-0.07	16.76	8.75	6.08	4.72
<b>OLS</b>									
Pooled	N/A	8.18***	8.02***	8.11***	8.01***	8.55	8.29	8.35	8.23
FE	N/A	-24.93***	-16.37***	-14.01***	-12.93***	25.11	16.45	14.05	12.96
<b>Results for <math>\beta_2 = 1</math></b>									
<b>KR-DIFF</b>									
3 lags	4	-0.96	0.51	0.39	0.17	23.75	9.42	7.04	6.13
4 lags	6	1.61	1.81	1.09	0.53	32.77	10.02	7.18	6.16
<b>KR-LEV</b>									
3 lags	4	-1.80	0.91	0.63	0.32	26.24	10.64	7.73	6.49
4 lags	6	-1.04	2.04	1.27	0.66	38.55	10.85	7.59	6.38
<b>DIFF-GMM</b>									
full	12/72/182/342	8.80	16.03***	19.97***	22.48***	17.92	17.19	20.40	22.70
full (coll.)	6/16/26/36	13.76	9.32*	8.08*	6.91*	28.99	14.20	10.92	9.05
2 lags	10/30/50/70	8.08	10.68*	10.88**	10.71**	18.74	13.40	12.50	11.82
2 lags (coll.)	4	9.60	1.59	0.81	0.30	35.00	15.46	10.58	8.16
<b>SYS-GMM</b>									
full	19/89/209/379	2.30	4.52	5.01	5.42*	15.11	9.58	8.30	8.09
full (coll.)	9/19/29/39	9.46	8.48	7.23*	6.25*	24.25	13.37	10.20	8.41
2 lags	17/47/77/107	1.51	2.30	1.96	1.82	15.64	9.31	7.13	6.00
2 lags (coll.)	7	6.03	1.61	0.73	0.36	26.09	13.81	9.93	7.78
<b>OLS</b>									
Pooled	N/A	3.79	4.29*	3.96*	4.24*	6.72	6.35	5.83	5.91
FE	N/A	26.81***	32.53***	33.48***	33.80***	27.20	32.64	33.54	33.84

Note: \*\*\*, \*\*, and \* denote significance at the 1%, 5%, and 10% confidence level, respectively. There are 2,000 Monte Carlo simulations with  $N = 100$  and variable  $T$ . "coll." refers to a collapsed instrument matrix in GMM. For further details on the instruments used, please see table 13.

### 4.3 Sensitivity tests

In this section, we consider the sensitivity of our results to two aspects of our experiments: 1) the degree of persistence in the dependent variable ( $\beta_1$ ) and 2) the sample size ( $N$ ). First, we consider increasing ( $\beta_1$ ) from 0.50 to 0.80. All other parameters are fixed at their values for scenarios 3 or 4.

In table 8, we report results for scenario 3 with  $\beta_1 = 0.8$ . Recall that in this case, we have correlation between the regressor and both the individual effect and the contemporaneous error term. Note that  $\beta_1 = 0.8$  implies a very high degree of persistence in the dependent variable, given that we also have individual effects.

As in table 5, KR-DIFF and KR-LEV do a fine job of eliminating the OLS bias. Also, both SYS-GMM and DIFF-GMM continue to exhibit substantial bias for both the lagged dependent variable coefficient and the regressor coefficient. For instance, in the  $T = 20$  case, the mean upward bias for the estimate of  $\beta_2$  is 21% for DIFF-GMM and 12% for SYS-GMM.

Also, as in table 5, collapsing the instrument matrix fails to eliminate bias in the DIFF-GMM and SYS-GMM estimators. For instance, at  $T = 20$ , the mean bias for the estimate of  $\beta_2$  is 13% for DIFF-GMM (collapsed) and 14% for SYS-GMM (collapsed).

Versions of GMM that restrict the number of lags to two also fail to eliminate the bias, although this leads to improvements. For instance, at  $T = 20$ , the mean bias for the estimate of  $\beta_2$  is 7.3% for DIFF-GMM (2 lags) and 6.7% for SYS-GMM (2 lags).

A combination of both restricting lags and collapsing the instrument matrix does successfully eliminate the bias of GMM but at a high cost to efficiency. For instance, at  $T = 20$ , the RMSEs for KR-DIFF (3 lags), KR-LEV (3 lags), DIFF-GMM (2 lags, collapsed), and SYS-GMM (2 lags, collapsed) are 3.86, 2.83, 6.90, and 4.31 for  $\beta_1$  and 6.46, 6.67, 13.25, and 12.92 for  $\beta_2$ , respectively.

Next, table 9 reports results for scenario 4 but with  $\beta_1$  increased from 0.50 to 0.80. Recall that in scenario 4, the regressor is correlated with the current and all lagged values of the idiosyncratic error.

In this scenario, the performance of the GMM estimators noticeably improves. While GMM results based on all lagged instruments still show significant bias, collapsing the instrument set is now sufficient to eliminate this bias. And an efficiency comparison between the KR estimators and the GMM (collapsed) estimators is now somewhat ambiguous. For instance, at  $T = 20$ , the RMSEs for KR-DIFF (3 lags), KR-LEV (3 lags), DIFF-GMM (collapsed), and SYS-GMM (collapsed) are 4.23, 2.20, 2.94, and 1.71 for  $\beta_1$  and 4.85, 3.55, 4.38, and 4.47 for  $\beta_2$ , respectively. Also, unlike previous cases we have looked at, the RMSEs for the KR estimators drop quite substantially if one goes from three lags to four lags.

A striking aspect of this scenario is that the biases in pooled OLS are actually quite small. The OLS bias for  $\beta_2$  is very small and insignificant. The OLS bias for  $\beta_1$  is significant but quantitatively modest (that is, it is only about 5% of the true parameter

value, compared with about 16% when  $\beta_1 = 0.5$ ). Thus one conjecture about why GMM performs relatively well in this case is that the many-instrument bias is not very severe when the OLS bias is not very severe to begin with.

A second conjecture is that if the persistence of the dependent variable or the regressor is increased, then lagged instruments become more informative, so the weak- or many-instrument problem is less severe. This point is illustrated in table 7. This conjecture is also consistent with the fact that it is only in the  $\beta_1 = 0.8$  and  $\rho = 0.80$  case that the RMSEs of the KR estimators drop quite substantially if one goes from 3 lags to 4 lags. We leave it to future research to explore these conjectures more carefully.

Table 7. RMSE for SYS-GMM (full) under varying assumptions when  $T = 10$

Scenario	Parameters	RMSE for $\beta_2$	RMSE for $\beta_1$
Scenario 3	$\beta_1 = 0.5, \rho = 0.0$	18.12	4.23
Scenario 4	$\beta_1 = 0.5, \rho = 0.5$	9.58	5.05
Scenario 3	$\beta_1 = 0.8, \rho = 0.0$	12.51	4.83
Scenario 4	$\beta_1 = 0.8, \rho = 0.5$	5.71	3.41

Next, we consider the sensitivity of our results to the sample size. We use scenario 3 as the baseline in these experiments.<sup>13</sup> In the baseline results in table 5, we had  $N = 100$ . In tables 10, 11, and 12, we report results for  $N = 50, 250$ , and  $500$ , respectively.

The results for  $N = 50$  in table 10 are similar to those for  $N = 100$  in table 5. The KR estimators show no significant bias. The GMM estimators show significant and quantitatively large biases unless one both collapses the instrument matrix and reduces the number of lags to two. In that case, the GMM estimators suffer from a loss of efficiency primarily for the regressor coefficient. For instance, at  $T = 20$ , the RMSEs for KR-DIFF (3 lags), KR-LEV (3 lags), DIFF-GMM (2 lags, collapsed), and SYS-GMM (2 lags, collapsed) are 11.12, 11.99, 16.32, and 12.45 for  $\beta_2$ , respectively.

Increasing the sample size to  $N = 250$ , as in table 11, somewhat improves the performance of GMM. In most instances, collapsing the instrument matrix is now sufficient to eliminate significant bias. The DIFF-GMM estimator still exhibits an efficiency loss relative to the KR estimators, but SYS-GMM is now fairly comparable. For instance, at  $T = 20$ , the RMSEs for KR-DIFF (3 lags), KR-LEV (3 lags), DIFF-GMM (collapsed), and SYS-GMM (collapsed) are 4.92, 5.07, 8.07, and 5.85 for  $\beta_2$ , respectively.

Finally, in table 12, we increase the sample size to  $N = 500$ . Again there is no evidence of bias for the KR estimators. There is significant bias for the GMM estimators using the full-instrument set, but collapsing the instrument matrix eliminates the bias. In terms of efficiency, SYS-GMM (collapsed) and the KR estimators are very similar. But DIFF-GMM (collapsed) generates a substantially higher RMSE for  $\beta_2$ .

13. Scenario 3 seems more interesting than scenario 4 because the OLS bias is greater.

To summarize, we see that the KR estimators show no significant bias for any of the sample sizes we consider, while the GMM estimators continue to show significant bias even with  $N = 500$ . The performance of the GMM estimators improves with the sample size in the sense that reducing the size of the instrument matrix becomes more effective as a way to reduce bias and at a lower cost in efficiency. But at no point do the GMM estimators appear to deliver notable efficiency gains over the KR estimators.

Table 8. Simulation results for scenario 3 when  $\beta_1 = 0.8$ 

$(N = 100, T)$		Bias ( $\times 100$ )				RMSE ( $\times 100$ )			
Instr. set	Instr. count	5	10	15	20	5	10	15	20
<b>Results for <math>\beta_1 = 0.8</math></b>									
<b>KR-DIFF</b>									
3 lags	4	-12.76	-2.09	-0.79	-0.43	45.40	11.69	5.74	3.86
4 lags	6	-23.08	-2.62	-0.98	-0.64	44.29	9.86	5.17	3.66
<b>KR-LEV</b>									
3 lags	4	3.73	1.34	0.63	0.29	6.90	4.61	3.49	2.83
4 lags	6	5.34	1.61	0.67	0.35	8.86	4.08	3.07	2.42
<b>DIFF-GMM</b>									
full	12/72/182/342	-12.43*	-7.34**	-5.57***	-4.75***	18.60	8.34	5.97	4.97
full (coll.)	6/16/26/36	-11.79	-5	-3.08	-2.22*	23.81	8.29	4.80	3.39
2 lags	10/30/50/70	-12	-7.6*	-6.24*	-5.52*	19.64	10.25	7.91	6.70
2 lags (coll.)	4	-8.7	-1.77	-0.79	-0.33	27.13	12.19	8.74	6.90
<b>SYS-GMM</b>									
full	19/89/209/379	3.91*	4.51***	4.81***	4.96***	5.10	4.83	5.00	5.10
full (coll.)	9/19/29/39	0.05	-1.03	-1.51	-1.52	7.32	4.14	3.18	2.66
2 lags	17/47/77/107	3.84*	3.79**	3.85***	3.83***	5.15	4.29	4.17	4.08
2 lags (coll.)	7	0.74	1.83	1.71	1.63	8.59	6.05	5.06	4.31
<b>OLS</b>									
Pooled	N/A	6.84***	6.67***	6.62***	6.56***	6.91	6.72	6.67	6.60
FE	N/A	-20.2***	-9.59***	-6.67***	-5.32***	20.42	9.69	6.74	5.37
<b>Results for <math>\beta_2 = 1</math></b>									
<b>KR-DIFF</b>									
3 lags	4	-6.96	-0.26	0.2	0.06	37.62	11.08	7.59	6.46
4 lags	6	-9.37	0.8	0.79	0.34	39.68	11.01	7.41	6.35
<b>KR-LEV</b>									
3 lags	4	0.72	1.69	0.97	0.53	23.88	10.94	7.92	6.67
4 lags	6	4.28	2.64	1.52	0.82	35.26	11.11	7.76	6.51
<b>DIFF-GMM</b>									
full	12/72/182/342	3.44	13.48**	18.38***	21.33***	18.45	15.07	18.93	21.59
full (coll.)	6/16/26/36	11.05	14.13*	14.22*	13.13*	35.85	20.55	17.77	15.69
2 lags	10/30/50/70	1.78	5.71	6.63*	7.25*	19.90	11.21	9.73	9.28
2 lags (coll.)	4	10.34	3.67	2.21	1.1	47.92	23.94	16.96	13.25
<b>SYS-GMM</b>									
full	19/89/209/379	8.02	10.12*	11.36**	12.09***	16.03	12.51	12.67	13.04
full (coll.)	9/19/29/39	16.2	17.11*	15.93**	13.86**	30.89	21.49	18.40	15.64
2 lags	17/47/77/107	7.4	7.71*	7.09*	6.74*	16.21	11.33	9.54	8.63
2 lags (coll.)	7	13.19	5.73	3.5	1.91	34.88	21.24	16.27	12.92
<b>OLS</b>									
Pooled	N/A	14.46***	14.71***	14.35***	14.48***	15.25	15.21	14.76	14.85
FE	N/A	20.55***	29.73***	31.86***	32.74***	21.03	29.85	31.93	32.78

Note: \*\*\*, \*\*, and \* denote significance at the 1%, 5%, and 10% confidence level, respectively. There are 2,000 Monte Carlo simulations with  $N = 100$  and variable  $T$ . "coll." refers to a collapsed instrument matrix in GMM. For further details on the instruments used, please see table 13.

Table 9. Simulation results for scenario 4 when  $\beta_1 = 0.8$ 

$(N = 100, T)$		Bias ( $\times 100$ )				RMSE ( $\times 100$ )			
Instr. set	Instr. count	5	10	15	20	5	10	15	20
<b>Results for <math>\beta_1 = 0.8</math></b>									
<b>KR-DIFF</b>									
3 lags	4	-8.21	-1.58	-0.64	-0.24	49.39	16.14	6.04	4.23
4 lags	6	-14.02	-2.71	-0.95	-0.50	35.66	11.22	5.06	3.37
<b>KR-LEV</b>									
3 lags	4	2.89	1.27	0.59	0.32	4.84	3.29	2.69	2.20
4 lags	6	3.59	1.48	0.66	0.36	6.28	2.98	2.40	1.93
<b>DIFF-GMM</b>									
full	12/72/182/342	-9.64	-6.13**	-5.03***	-4.55***	15.82	7.02	5.35	4.73
full (coll.)	6/16/26/36	-9.49	-3.94	-2.35	-1.58	22.55	7.71	4.36	2.94
2 lags	10/30/50/70	-9.98	-7.08*	-5.49*	-4.47*	17.36	9.37	6.97	5.52
2 lags (coll.)	4	-5.63	-1.03	-0.27	-0.09	27.96	14.94	9.17	6.75
<b>SYS-GMM</b>									
full	19/89/209/379	2.88*	3.11**	3.29***	3.35***	3.85	3.41	3.48	3.49
full (coll.)	9/19/29/39	2.23	1.64	1.26	0.94	4.23	2.59	2.07	1.71
2 lags	17/47/77/107	2.86*	2.73**	2.79**	2.79***	3.88	3.13	3.05	2.99
2 lags (coll.)	7	2.42	2.19*	1.92*	1.66*	4.38	3.07	2.62	2.30
<b>OLS</b>									
Pooled	N/A	4.44***	4.26***	4.23***	4.17***	4.55	4.34	4.30	4.23
FE	N/A	-19.52***	-10.11***	-7.49***	-6.28***	19.69	10.18	7.53	6.31
<b>Results for <math>\beta_2 = 1</math></b>									
<b>KR-DIFF</b>									
3 lags	4	-11.61	-1.64	-0.55	-0.26	74.97	22.95	7.24	4.85
4 lags	6	-15.10	-2.33	-0.40	-0.28	53.96	15.68	6.02	4.24
<b>KR-LEV</b>									
3 lags	4	0.32	1.52	0.84	0.42	14.18	6.42	4.33	3.55
4 lags	6	-1.03	1.65	0.98	0.49	21.44	6.96	4.53	3.66
<b>DIFF-GMM</b>									
full	12/72/182/342	-7.03	3.39	8.08***	10.98***	21.44	6.38	8.72	11.25
full (coll.)	6/16/26/36	-8.29	-0.39	1.41	1.81	32.47	9.47	5.57	4.38
2 lags	10/30/50/70	-8.67	-2.66	-0.27	1.02	24.46	9.62	5.81	4.46
2 lags (coll.)	4	-5.55	-0.36	0.22	0.12	42.04	17.83	8.71	6.08
<b>SYS-GMM</b>									
full	19/89/209/379	1.98	1.59	1.30	1.38	9.59	5.71	4.46	4.04
full (coll.)	9/19/29/39	3.07	3.32	3.04	2.86	11.79	6.69	5.25	4.47
2 lags	17/47/77/107	1.91	1.68	1.02	0.66	9.75	5.84	4.31	3.48
2 lags (coll.)	7	2.27	1.36	0.84	0.60	11.80	6.45	4.80	3.89
<b>OLS</b>									
Pooled	N/A	1.90	2.21	1.88	2.06	5.30	4.62	4.08	4.03
FE	N/A	18.56***	24.38***	25.61***	26.11***	19.08	24.49	25.67	26.15

Note: \*\*\*, \*\*, and \* denote significance at the 1%, 5%, and 10% confidence level, respectively. There are 2,000 Monte Carlo simulations with  $N = 100$  and variable  $T$ . "coll." refers to a collapsed instrument matrix in GMM. For further details on the instruments used, please see table 13.

Table 10. Simulation results for scenario 3 when  $N = 50$ 

$(N = 50, T)$		Bias ( $\times 100$ )				RMSE ( $\times 100$ )			
Instr. set	Instr. count	5	10	15	20	5	10	15	20
<b>Results for <math>\beta_1 = 0.5</math></b>									
<b>KR-DIFF</b>									
3 lags	4	-5.71	-1.48	-0.43	-0.53	24.54	7.70	5.24	4.05
4 lags	6	-15.04	-2.24	-0.95	-0.82	33.08	7.58	5.24	4.08
<b>KR-LEV</b>									
3 lags	4	-0.71	-0.79	-0.46	-0.42	15.89	7.41	5.44	4.30
4 lags	6	4.05	-1.44	-0.54	-0.61	18.63	6.68	4.61	3.80
<b>DIFF-GMM</b>									
full	12/72/182/342	-10.18*	-8.15**	-7.36***	-7.11***	14.89	8.98	7.72	7.34
full (coll.)	6/16/26/36	-9.02	-4.94*	-3.55*	-3.18*	17.55	7.22	4.94	4.17
2 lags	10/30/50/70	-9.23	-6.46*	-5.71*	-5.4**	15.00	8.26	6.81	6.16
2 lags (coll.)	4	-6.73	-1.5	-0.5	-0.49	20.44	8.75	6.09	4.55
<b>SYS-GMM</b>									
full	19/89/209/379	1.77	4.74*	6.46**	7.57***	8.15	6.22	7.26	8.10
full (coll.)	9/19/29/39	-5.7	-4.43*	-3.29*	-3*	12.95	6.69	4.78	4.07
2 lags	17/47/77/107	1.52	2.79	3.13*	3.16*	8.34	5.15	4.73	4.42
2 lags (coll.)	7	-4.44	-1.27	-0.45	-0.45	14.26	7.78	5.76	4.36
<b>OLS</b>									
Pooled	N/A	11.66***	11.62***	11.63***	11.61***	12.07	11.92	11.91	11.87
FE	N/A	-19.21***	-10.73***	-8.39***	-7.49***	19.66	10.96	8.55	7.61
<b>Results for <math>\beta_2 = 1</math></b>									
<b>KR-DIFF</b>									
3 lags	4	4.43	1.91	0.65	0.82	37.87	17.12	13.30	11.12
4 lags	6	9.62	4.3	2.23	1.72	40.37	17.22	13.25	11.09
<b>KR-LEV</b>									
3 lags	4	3.35	1.82	1.09	0.9	45.24	19.30	14.86	11.99
4 lags	6	13.91	4.92	2.28	2.07	54.84	18.75	13.75	11.32
<b>DIFF-GMM</b>									
full	12/72/182/342	16.36	23.98***	27.6***	29.92***	25.90	25.29	28.09	30.17
full (coll.)	6/16/26/36	23.39	20.42*	18.61*	17.43**	47.01	26.60	22.22	19.86
2 lags	10/30/50/70	14.98	17.25*	18.11**	18.18***	26.87	20.74	19.99	19.48
2 lags (coll.)	4	21.42	5.83	2.5	1.66	62.19	33.02	22.28	16.32
<b>SYS-GMM</b>									
full	19/89/209/379	16	19.04**	19.46***	19.52***	25.46	21.40	20.85	20.48
full (coll.)	9/19/29/39	16.99	15.33*	13.34*	12.89*	34.50	20.65	16.76	15.30
2 lags	17/47/77/107	14.91	15.75*	15.74**	15.93**	25.56	19.45	18.10	17.63
2 lags (coll.)	7	13.26	4.8	2.31	1.94	38.99	22.21	16.05	12.45
<b>OLS</b>									
Pooled	N/A	18.48***	18.68***	18.49***	18.64***	19.81	19.49	19.21	19.22
FE	N/A	25.36***	32.5***	33.97***	34.58***	26.19	32.72	34.11	34.68

Note: \*\*\*, \*\*, and \* denote significance at the 1%, 5%, and 10% confidence level, respectively. There are 2,000 Monte Carlo simulations with  $N = 50$  and variable  $T$ . "coll." refers to a collapsed instrument matrix in GMM. For further details on the instruments used, please see table 13.

Table 11. Simulation results for scenario 3 when  $N = 250$ 

$(N = 250, T)$		Bias ( $\times 100$ )				RMSE ( $\times 100$ )			
Instr. set	Instr. count	5	10	15	20	5	10	15	20
<b>Results for <math>\beta_1 = 0.5</math></b>									
<b>KR-DIFF</b>									
3 lags	4	-0.66	-0.3	-0.14	-0.13	9.37	3.30	2.21	1.70
4 lags	6	-2.67	-0.44	-0.25	-0.18	13.64	3.24	2.10	1.61
<b>KR-LEV</b>									
3 lags	4	-0.12	-0.14	-0.14	-0.13	6.65	3.23	2.33	1.87
4 lags	6	0.96	-0.28	-0.19	-0.17	8.73	2.93	2.04	1.66
<b>DIFF-GMM</b>									
full	12/72/182/342	-2.96	-3.29*	-3.35***	-3.49***	6.37	3.87	3.59	3.63
full (coll.)	6/16/26/36	-3.4	-1.78	-1.16	-0.97	8.91	3.30	2.11	1.70
2 lags	10/30/50/70	-2.4	-1.98	-1.76*	-1.72*	6.33	3.22	2.58	2.31
2 lags (coll.)	4	-1.81	-0.34	-0.16	-0.16	11.17	3.78	2.56	2.01
<b>SYS-GMM</b>									
full	19/89/209/379	0.0	0.84	1.42	1.93*	4.62	2.59	2.33	2.52
full (coll.)	9/19/29/39	-2.46	-1.45	-0.99	-0.86	7.75	3.09	2.02	1.63
2 lags	17/47/77/107	0.07	0.42	0.51	0.58	4.69	2.44	1.93	1.65
2 lags (coll.)	7	-1.4	-0.28	-0.14	-0.16	8.12	3.54	2.52	1.98
<b>OLS</b>									
Pooled	N/A	12.02***	11.96***	11.95***	11.96***	12.10	12.01	12.00	12.01
FE	N/A	-19.12***	-10.69***	-8.39***	-7.4***	19.22	10.73	8.42	7.43
<b>Results for <math>\beta_2 = 1</math></b>									
<b>KR-DIFF</b>									
3 lags	4	0.72	0.38	0.09	0.23	13.81	7.71	5.83	4.92
4 lags	6	2.17	0.98	0.49	0.41	19.56	7.77	5.78	4.80
<b>KR-LEV</b>									
3 lags	4	0.59	0.35	0.2	0.21	17.01	8.08	6.14	5.07
4 lags	6	4.07	1.07	0.52	0.45	29.55	8.00	5.95	4.88
<b>DIFF-GMM</b>									
full	12/72/182/342	6.66	11.25**	14.39***	16.86***	13.96	12.49	14.89	17.13
full (coll.)	6/16/26/36	12.66	8.2	6.3*	5.58*	30.86	13.89	9.66	8.07
2 lags	10/30/50/70	5.8	6.24*	6.38*	6.52*	14.47	9.36	8.29	7.89
2 lags (coll.)	4	8.11	1.13	0.35	0.5	41.06	14.57	9.46	7.36
<b>SYS-GMM</b>									
full	19/89/209/379	6.76	10.52*	12.86***	14.58***	13.71	12.33	13.84	15.22
full (coll.)	9/19/29/39	7.52	4.75	3.76	3.57	20.83	9.41	6.72	5.85
2 lags	17/47/77/107	5.9	6.33*	6.43*	6.42*	13.69	9.30	8.41	7.85
2 lags (coll.)	7	4.6	0.94	0.47	0.55	21.77	9.85	7.06	5.73
<b>OLS</b>									
Pooled	N/A	17.84***	17.93***	17.96***	17.96***	18.11	18.10	18.10	18.09
FE	N/A	25.23***	32.62***	34.01***	34.62***	25.40	32.66	34.04	34.64

Note: \*\*\*, \*\*, and \* denote significance at the 1%, 5%, and 10% confidence level, respectively. There are 2,000 Monte Carlo simulations with  $N = 250$  and variable  $T$ . "coll." refers to a collapsed instrument matrix in GMM. For further details on the instruments used, please see table 13.



Table 12. Simulation results for scenario 3 when  $N = 500$ 

$(N = 500, T)$		Bias ( $\times 100$ )				RMSE ( $\times 100$ )			
Instr. set	Instr. count	5	10	15	20	5	10	15	20
<b>Results for <math>\beta_1 = 0.5</math></b>									
<b>KR-DIFF</b>									
3 lags	4	-0.4	-0.05	-0.01	-0.05	6.50	2.31	1.47	1.23
4 lags	6	-1.4	-0.15	-0.11	-0.07	9.66	2.28	1.41	1.12
<b>KR-LEV</b>									
3 lags	4	-0.02	-0.06	-0.09	-0.03	4.60	2.26	1.64	1.25
4 lags	6	0.42	-0.15	-0.09	-0.05	5.95	2.06	1.46	1.09
<b>DIFF-GMM</b>									
full	12/72/182/342	-1.6	-1.82*	-2.02**	-2.18***	4.52	2.40	2.23	2.31
full (coll.)	6/16/26/36	-2.04	-0.85	-0.59	-0.48	6.50	2.24	1.41	1.11
2 lags	10/30/50/70	-1.29	-0.97	-0.92	-0.9	4.55	2.16	1.62	1.41
2 lags (coll.)	4	-1.18	-0.1	-0.04	-0.01	6.78	2.67	1.79	1.39
<b>SYS-GMM</b>									
full	19/89/209/379	-0.01	0.38	0.66	0.88	3.57	1.86	1.53	1.43
full (coll.)	9/19/29/39	-1.31	-0.67	-0.47	-0.41	5.82	2.13	1.34	1.07
2 lags	17/47/77/107	0.05	0.21	0.22	0.28	3.58	1.77	1.34	1.09
2 lags (coll.)	7	-0.7	-0.09	-0.03	-0.01	5.95	2.55	1.76	1.35
<b>OLS</b>									
Pooled	N/A	11.97***	11.98***	11.98***	12.01***	12.01	12.00	12.01	12.03
FE	N/A	-19.17***	-10.68***	-8.42***	-7.38***	19.21	10.70	8.44	7.40
<b>Results for <math>\beta_2 = 1</math></b>									
<b>KR-DIFF</b>									
3 lags	4	0.63	0.21	0.03	0.11	10.20	5.23	4.03	3.45
4 lags	6	1.81	0.57	0.27	0.17	14.23	5.39	4.00	3.27
<b>KR-LEV</b>									
3 lags	4	0.34	0.26	0.17	0.05	12.10	5.65	4.25	3.43
4 lags	6	1.71	0.58	0.31	0.17	21.70	5.64	4.19	3.25
<b>DIFF-GMM</b>									
full	12/72/182/342	3.78	6.76*	8.91***	10.85***	9.88	7.93	9.40	11.11
full (coll.)	6/16/26/36	7.88	4.62	3.44	2.86	22.97	9.41	6.45	5.15
2 lags	10/30/50/70	3.14	3.57	3.6*	3.55*	10.24	6.28	5.31	4.82
2 lags (coll.)	4	5.09	0.69	0.21	0.03	25.19	10.07	6.67	5.10
<b>SYS-GMM</b>									
full	19/89/209/379	3.38	6.44*	8.44**	10.12***	9.62	8.06	9.28	10.67
full (coll.)	9/19/29/39	3.84	2.38	1.89	1.79	15.22	6.47	4.50	3.73
2 lags	17/47/77/107	2.92	3.62	3.68*	3.7*	9.54	6.21	5.35	4.91
2 lags (coll.)	7	2.18	0.48	0.17	0.13	15.71	7.09	4.85	3.92
<b>OLS</b>									
Pooled	N/A	17.84***	17.89***	17.84***	17.83***	17.98	17.97	17.92	17.90
FE	N/A	25.28***	32.61***	34.05***	34.61***	25.36	32.64	34.07	34.62

Note: \*\*\*, \*\*, and \* denote significance at the 1%, 5%, and 10% confidence level, respectively. There are 2,000 Monte Carlo simulations with  $N = 500$  and variable  $T$ . "coll." refers to a collapsed instrument matrix in GMM. For further details on the instruments used, please see table 13.

Table 13. Instrument list

Estimator	Instrument set	Scenario	Instrument list
KR-DIFF	2 lags	1	$\Delta x_{it}, x_{it-1}, x_{it-2}, y_{it-2}$
KR-DIFF	2 lags	2	$x_{it-1}, x_{it-2}, y_{it-2}$
KR-DIFF	3 lags	1	$\Delta x_{it}, x_{it-1}, \dots, x_{it-3}, y_{it-2}, y_{it-3}$
KR-DIFF	3 lags	2	$x_{it-1}, \dots, x_{it-3}, y_{it-2}, y_{it-3}$
KR-DIFF	3 lags	3/4	$x_{it-2}, x_{it-3}, y_{it-2}, y_{it-3}$
KR-DIFF	4 lags	3/4	$x_{it-2}, \dots, x_{it-4}, y_{it-2}, \dots, y_{it-4}$
KR-LEV	2 lags	1/2	$x_{it}, \Delta x_{it}, \Delta x_{it-1}, \Delta y_{it-1}$
KR-LEV	3 lags	1/2	$x_{it}, \Delta x_{it}, \dots, \Delta x_{it-2}, \Delta y_{it-1}, \Delta y_{it-2}$
KR-LEV	3 lags	3/4	$\Delta x_{it-1}, \Delta x_{it-2}, \Delta y_{it-1}, \Delta y_{it-2}$
KR-LEV	4 lags	3/4	$\Delta x_{it-1}, \dots, \Delta x_{it-3}, \Delta y_{it-1}, \dots, \Delta y_{it-3}$
DIFF-GMM	Full	1	$\Delta x_{it}, y_{it-2}, \dots, y_{it-(T-2)}$
DIFF-GMM	Full	2	$x_{it-1}, \dots, x_{it-(T-2)}, y_{it-2}, \dots, y_{it-(T-2)}$
DIFF-GMM	Full	3/4	$x_{it-2}, \dots, x_{it-(T-2)}, y_{it-2}, \dots, y_{it-(T-2)}$
DIFF-GMM	2 lags	1	$\Delta x_{it}, y_{it-2}, y_{it-3}$
DIFF-GMM	2 lags	2	$x_{it-1}, x_{it-2}, y_{it-2}, y_{it-3}$
DIFF-GMM	2 lags	3/4	$x_{it-2}, x_{it-3}, y_{it-2}, y_{it-3}$
SYS-GMM	Full	1	$x_{it}, \Delta x_{it}, \Delta y_{it-1}, y_{it-2}, \dots, y_{it-(T-1)}$
SYS-GMM	Full	2	$\Delta x_{it}, \Delta y_{it-1}, x_{it-1}, \dots, x_{it-(T-1)}, y_{it-2}, \dots, y_{it-(T-1)}$
SYS-GMM	Full	3/4	$\Delta x_{it}, \Delta y_{it-1}, x_{it-2}, \dots, x_{it-(T-1)}, y_{it-2}, \dots, y_{it-(T-1)}$
SYS-GMM	2 lags	1	$x_{it}, \Delta x_{it}, \Delta y_{it-1}, y_{it-2}, y_{it-3}$
SYS-GMM	2 lags	2	$\Delta x_{it}, \Delta y_{it-1}, x_{it-1}, x_{it-2}, y_{it-2}, y_{it-3}$
SYS-GMM	2 lags	3/4	$\Delta x_{it}, \Delta y_{it-1}, x_{it-2}, x_{it-3}, y_{it-2}, y_{it-3}$

## 5 Conclusion

We provide a command, `xtkr`, for implementing the KR (1992a) estimator for dynamic panel-data models in Stata. The two key features of the KR approach relative to other popular panel-data estimators (for example, Arellano and Bond [1991] and Blundell and Bond [1998]) are 1) that KR rely on small-instrument sets (typically one or two lags of the predetermined variables) and 2) that KR rely on forward filtering to eliminate serial correlation.

We show in an application to a dataset taken from Baltagi (2005) that the KR estimator generates estimates that are arguably more theoretically plausible than alternative methods. Indeed, the popular “SYS-GMM” estimator yields results almost identical to OLS, implying that its use of many instruments induces severe bias.

We also report on a Monte Carlo exercise where the KR estimator is compared with a variety of popular methods, including SYS-GMM and DIFF-GMM. We consider both cases where the GMM estimators use all lagged instruments, as well as cases where the instrument set is restricted or collapsed (as suggested by Roodman [2009a], among others). We find that the KR estimator typically performs better than these popular alternatives in terms of both bias and RMSE. This finding is consistent with Monte Carlo results for a different context reported by Ziliak (1997).

Of course, our results are for particular cases, and one can likely find other data-generating processes where GMM estimators that rely on many instruments would perform as well as or better than the KR estimator. Nevertheless, our results suggest that the KR estimator should definitely be in the toolkit of applied researchers.

## 6 Acknowledgments

Keane’s work on this project has been supported by Australian Research Council grants FF0561843 and FL110100247, and the Australian Research Council Centre of Excellence in Population Ageing Research (grant CE110001029) at the University of New South Wales. But the views expressed are entirely our own.

## 7 References

- Ahn, S. C., and P. Schmidt. 1995. Efficient estimation of models for dynamic panel data. *Journal of Econometrics* 68: 5–27.
- Arellano, M., and S. Bond. 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., and O. Bover. 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*. 3rd ed. New York: Wiley.

- Blundell, R., and S. Bond. 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87: 115–143.
- Chamberlain, G. 1992. Comment: Sequential moment restrictions in panel data. *Journal of Business and Economic Statistics* 10: 20–26.
- Hayashi, F. 1992. Comment. *Journal of Business and Economic Statistics* 10: 15–17.
- Holtz-Eakin, D., W. Newey, and H. S. Rosen. 1988. Estimating vector autoregressions with panel data. *Econometrica* 56: 1371–1395.
- Keane, M. P. 1997. Modeling heterogeneity and state dependence in consumer choice behavior. *Journal of Business and Economic Statistics* 15: 310–327.
- Keane, M. P., and D. E. Runkle. 1992a. On the estimation of panel-data models with serial correlation when instruments are not strictly exogenous. *Journal of Business and Economic Statistics* 10: 1–9.
- . 1992b. Reply. *Journal of Business and Economic Statistics* 10: 26–29.
- Koenker, R., and J. A. F. Machado. 1999. GMM inference when the number of moment conditions is large. *Journal of Econometrics* 93: 327–344.
- Nickell, S. 1981. Biases in dynamic models with fixed effects. *Econometrica* 49: 1417–1426.
- Roodman, D. 2009a. A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics* 71: 135–158.
- . 2009b. How to do xtabond2: An introduction to difference and system GMM in Stata. *Stata Journal* 9: 86–136.
- Schmidt, P., S. C. Ahn, and D. Wyhowski. 1992. Comment. *Journal of Business and Economic Statistics* 10: 10–14.
- Windmeijer, F. 2005. A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics* 126: 25–51.
- Ziliak, J. P. 1997. Efficient estimation with panel data when instruments are predetermined: An empirical comparison of moment-condition estimators. *Journal of Business and Economic Statistics* 14: 419–431.

#### About the authors

Michael Keane is the Nuffield Professor of Economics at the University of Oxford, an Australian Laureate Fellow at the University of New South Wales, and a chief investigator of the Australian Research Council Centre of Excellence in Population Ageing Research.

Timothy Neal is currently a PhD candidate at the University of New South Wales and a research fellow at the Australian Research Council Centre of Excellence in Population Ageing Research.