Ind. Econ. Rev. (2017) 52:73–91
https://doi.org/10.1007/s41775-017-0003-x

ARTICLE

A combined estimator of regression models with measurement errors

Bai Huang[1] · **Tae-Hwy Lee**[2] · **Aman Ullah**[2]

Published online: 4 October 2017

Abstract When the regressors are observed with measurement errors, the OLS estimator is inconsistent and typically the use of the IV estimator is recommended. In this paper we place this recommendation under scrutiny, especially (1) when the instruments are weak and (2) when there are many instruments. Following Hansen (Econometric Reviews, http://www.ssc.wisc.edu/~bhansen/papers/shrinkiv.pdf, 2017), we use the Hausman (Econometrica 46(6):1251–1271, 1978) test for errors in variables to combine OLS and IV estimators. The combined estimator has the asymptotic risk strictly less than that of the IV estimator. Then we show some useful findings for small samples based on the Monte Carlo simulations. In terms of the mean squared error risk, we find that (a) typically OLS gets worse as the measurement error gets larger while IV is more robust and better than OLS, (b) OLS can be better than IV when the measurement error is small, and (c) the combined estimator outperforms IV as the asymptotic result predicts. (a) and (b) are true only when the instruments are not weak and when there are not many instruments. However, when the instruments are weak or when there are many instruments: (c) still holds as it is a theorem, but (a), (b) turn out to become quite the opposite, i.e., OLS can be much better than IV even when measurement error is large. This happens because IV is known to be inconsistent with weak instruments and many instruments (Staiger and Stock Econometrica (65(3), 557–5861997, Bekker

✉ Aman Ullah
aman.ullah@ucr.edu

Bai Huang
baihuang1988@gmail.com

Tae-Hwy Lee
tae.lee@ucr.edu

[1] School of Statistics and Mathematics, Central University of Finance and Economics, Beijing, China

[2] Department of Economics, University of California, Riverside, CA 92521, USA

Econometrica 62, 657–681, 1994)), and can be much worse than OLS, making the combined estimator close to OLS. In that case, the typical recommendation to use IV should be guided by the combined estimator, and IV and the combined estimator need to be regularized for weak instruments and many instruments.

Keywords OLS · IV · Measurement errors · Weak instruments · Many instruments · Regularized combined estimator

JEL Classification: C13 , C33 , C52

1 Introduction

When some of the regressors have errors in variables, the ordinary least squares (OLS) estimator is inconsistent and typically the instrumental variables (IV) estimator is recommended. In this paper we look into this recommendation closely and propose an estimator providing improvement over the IV as well as OLS estimators in terms of the total mean squared error risk. Motivated by Hansen (2017), we use the Hausman (1978) test for errors in variable to combine OLS and IV. Further, following Hansen (2017), the asymptotic distribution and the asymptotic risk of the combined estimator using a local asymptotic framework are presented. This shows that, if the regressor dimension exceeds two, the asymptotic risk of the combined estimator is strictly less than that of the IV estimator.

Our simulation result shows that the combined estimator can substantially reduce finite sample risk relatively to the IV estimator, as well as relative to the OLS estimator for moderate degrees of measurement errors. In terms of the mean squared error risk, we find that (a) typically OLS gets worse as the measurement error gets larger while IV is more robust and better than OLS, (b) OLS can be better than IV when the measurement error is small, and (c) the combined estimator outperforms IV as the asymptotic result predicts. (a) and (b) are true only when the instruments are not weak and when there are not many instruments.

However, when the instruments are weak or when there are many instruments, (c) still holds as it is a theorem, but (a), (b) turn out to become quite the opposite, i.e., OLS can be much better than IV even when measurement error is large. This happens because IV is known to be inconsistent with weak instruments and many instruments (Staiger and Stock 1997; Bekker 1994), and can be much worse than OLS, making the combined estimator become closer to OLS. In that case, the typical recommendation to use IV should be guided by the combined estimator and we need to regularize the OLS, IV, and the combined estimators to reduce the number of instruments and to shrink the effect of the weak instruments. The findings from this paper suggest that in the presence of many instruments and weak instruments, the regularization would be the first step to recover the consistency of IV before using it to construct a regularized combined estimator. Weak instruments and many instruments need to be treated before constructing the combined estimator.

The rest of the paper is organized as follows. Section 2 reviews issues in the regression model with measurement errors. Section 3 presents the combined estimator and summarizes its asymptotic distribution and the asymptotic risk. Monte Carlo simulation is provided in Sect. 4. Concluding remarks follows in Sect. 5.

2 Regression model with measurement errors

We consider the linear regression

$$y_i = x_i^{*\prime}\beta + \varepsilon_{1i}, \tag{1}$$

where y_i and ε_{1i} are scalars, x_i^* is a $q \times 1$ vector, and $\sigma_{\varepsilon_1}^2 = E\left(\varepsilon_{1i}^2\right)$. The covariates x_i^* is unobserved but measured with measurement error ε_{2i} such that

$$x_i = x_i^* + \varepsilon_{2i} \tag{2}$$

is observed and $\Sigma_{\varepsilon_2} = E\left(\varepsilon_{2i}\varepsilon_{2i}'\right)$. Our goal is to estimate β. Let z_i be an $\ell \times 1$ $(\ell \geq q)$ vector of instruments that are related to x_i^*

$$x_i^* = \Pi z_i + \varepsilon_{3i}, \tag{3}$$

where $x_i^*, \varepsilon_{2i}, \varepsilon_{3i}$ are $q \times 1$ vectors. For simplicity, we assume that all of $\varepsilon_{1i}, \varepsilon_{2i}, \varepsilon_{3i}$ are uncorrelated with each other, and we also assume that $\Sigma_{\varepsilon_2} = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \cdots, \sigma_q^2)$ is diagonal.

Then the model is written as

$$y_i = x_i^{*\prime}\beta + \varepsilon_{1i} = x_i'\beta + \left(\varepsilon_{1i} - \varepsilon_{2i}'\beta\right) \equiv x_i'\beta + u_i, \tag{4}$$

where $u_i = \varepsilon_{1i} - \varepsilon_{2i}'\beta$ and $\sigma_u^2 = E\left(u_i^2\right) = \sigma_{\varepsilon_1}^2 + \beta'\Sigma_{\varepsilon_2}\beta$. Also,

$$x_i = x_i^* + \varepsilon_{2i} = \Pi z_i + \varepsilon_{3i} + \varepsilon_{2i} \equiv \Pi z_i + v_i, \tag{5}$$

where $v_i = (\varepsilon_{3i} + \varepsilon_{2i})$ and $\Sigma_v = E\left(v_i v_i'\right) = \Sigma_{\varepsilon_2} + \Sigma_{\varepsilon_3}$.

The OLS estimator

$$\hat{\beta}_0 = (X'X)^{-1}(X'y) \tag{6}$$

may be inconsistent because

$$\begin{aligned} \hat{\beta}_0 - \beta \xrightarrow{p} \left[E\left(x_i x_i'\right)\right]^{-1}(Ex_i u_i) &= \left[E\left(x_i x_i'\right)\right]^{-1}\left[-E\left(\varepsilon_{2i}\varepsilon_{2i}'\right)\beta\right] \\ &= \left[E\left(x_i x_i'\right)\right]^{-1}\left[-\Sigma_{\varepsilon_2}\beta\right] \\ &= \left[E\left(x_i x_i'\right)\right]^{-1}\left[-\sigma_{\varepsilon_2}^2\beta\right]. \end{aligned} \tag{7}$$

The last line uses the assumption that the measurement errors to all regressors have the same degree (for simplicity), i.e.,

$$\Sigma_{\varepsilon_2} = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \cdots, \sigma_{\mathrm{q}}^2) = \sigma_{\varepsilon_2}^2 \mathrm{I}_{\mathrm{q}}.$$

When $\sigma^2_{\varepsilon_2} = 0$, the OLS estimator $\hat{\beta}_0$ is consistent and efficient. When $\sigma^2_{\varepsilon_2} \neq 0$, the OLS estimator $\hat{\beta}_0$ is inconsistent and a consistent estimator $\hat{\beta}_1$ using the instrument z can be obtained from

$$\hat{\beta}_1 = \left(\hat{X}'\hat{X}\right)^{-1}\left(\hat{X}'y\right), \tag{8}$$

where $P_Z = Z(Z'Z)^{-1}Z'$ and $\hat{X} = P_Z X$.

The Hausman test for the null hypothesis that x_i has no measurement error against the alternative hypothesis that x_i has measurement errors is to compare the two estimators $\hat{\beta}_0$ and $\hat{\beta}_1$. Then the Hausman statistic is

$$H_n = n\left(\hat{\beta}_1 - \hat{\beta}_0\right)'\left(\hat{V}_1 - \hat{V}_0\right)^{-1}\left(\hat{\beta}_1 - \hat{\beta}_0\right), \tag{9}$$

where $\hat{V}_0$ and $\hat{V}_1$ are the consistent estimators for $V_0 = \sigma^2_u\left[E\left(x_i x_i'\right)\right]^{-1}$ and $V_1 = \sigma^2_u\left[E\left(x_i z_i'\right)\left[E\left(z_i z_i'\right)\right]^{-1}E\left(z_i x_i'\right)\right]^{-1}$ respectively. V_0 and V_1 are the asymptotic variances of $\hat{\beta}_0$ and $\hat{\beta}_1$. If H_n is larger than a certain critical value, $\hat{\beta}_1$ is considered to be the preferred estimator.

3 A combined estimator

When either $\sigma^2_{\varepsilon_2} = 0$ or $\sigma^2_{\varepsilon_2} \neq 0$, it is easy to select one of $\hat{\beta}_0$ and $\hat{\beta}_1$ from the Hausman statistic for measurement errors. But when $\sigma^2_{\varepsilon_2}$ is local to zero, it may not be clear which estimator to choose. In such a case a combined estimator of the following form is considered

$$\hat{\beta}_c = w\hat{\beta}_0 + (1 - w)\hat{\beta}_1 \tag{10}$$

where $w = \min\left(\frac{\tau}{H_n}, 1\right)$ and τ is a shrinkage parameter determined as suggested by James and Stein (1961). The asymptotic behavior of the three estimators $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_c$ has been elegantly derived in Hansen (2017) for the structural econometric model. As the same asymptotic theory of Hansen (2017) applies to the present cases with measurement errors in the regressors, as summarized below. The main goal of our paper is to examine the finite sample behavior to see how the asymptotic theory carries over to the finite sample cases. Not surprisingly, the finite sample behavior of the three estimators are quite similar in Hansen (2017) and in measurement error models. Interestingly though, we find that when q is large relative to the sample size n, the finite sample risks of $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_c$ are quite different than that when q is reasonably small.

The variable x_i is exogenous if there is no measurement error, i.e., $\sigma^2_{\varepsilon_2}$ is zero. Consider the case when $\sigma^2_{\varepsilon_2}$ is local to zero

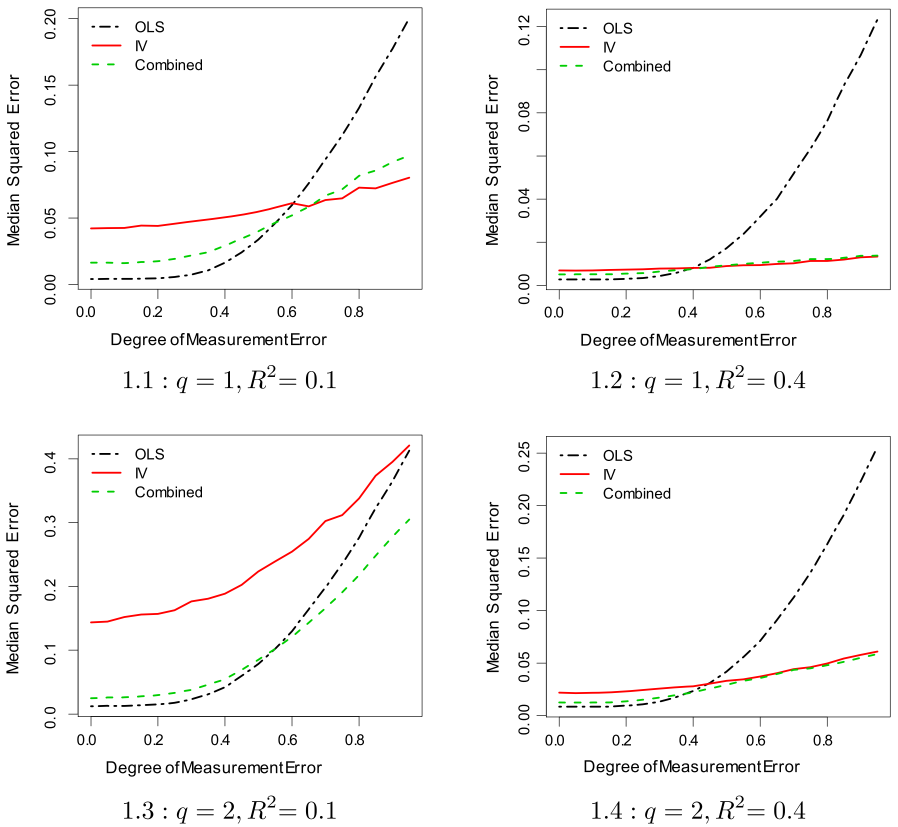


Fig. 1 $n = 100, W = I$

$$\sigma^2_{\varepsilon_2} = \frac{1}{\sqrt{n}}\delta_1 \tag{11}$$

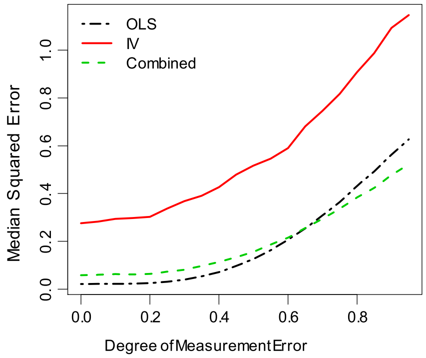
where δ_1 is the 1×1 localizing parameter, which is the degree of measurement error. Suppose the two error terms in (4) and (5) are linearly related and write the structural equation error u_i as a linear function of the reduced form equation error v_i and the orthogonal error η_i

$$\begin{aligned} u_i &= \rho' v_i + \eta_i \\ E(v_i \eta_i) &= 0. \end{aligned} \tag{12}$$

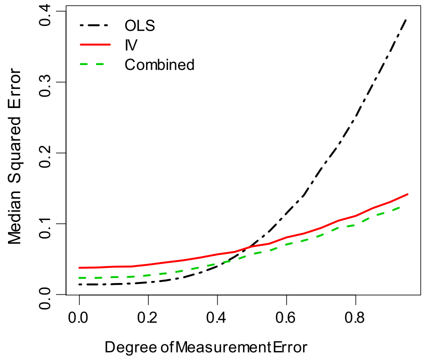
Hansen (2017) uses the local to zero exogeneity in the sense of

$$\rho = \frac{1}{\sqrt{n}}\delta_2 \tag{13}$$

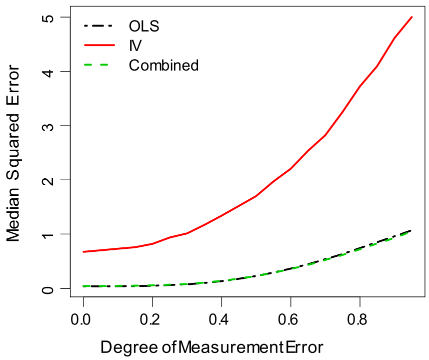
where δ_2 is the $q \times 1$ localizing parameter. To relate δ_1 and δ_2, recall $u_i = \varepsilon_{1i} - \varepsilon'_{2i}\beta$ and $v_i = \varepsilon_{3i} + \varepsilon_{2i}$, and thus



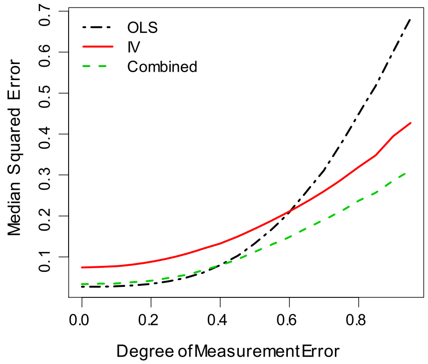
$1.5 : q = 3, R^2 = 0.1$



$1.6 : q = 3, R^2 = 0.4$



$1.7 : q = 5, R^2 = 0.1$



$1.8 : q = 5, R^2 = 0.4$

Fig. 1 continued

$$\rho = \left[E\left(v_i v_i'\right)\right]^{-1} E(v_i u_i) = \Sigma_v^{-1}\left(-\sigma_{\varepsilon_2}^2 \beta\right). \tag{14}$$

Therefore

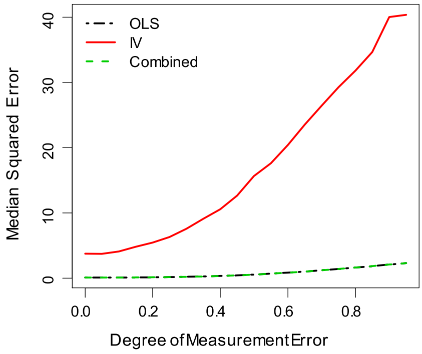
$$\delta_2 = \Sigma_v^{-1}(-\delta_1 \beta), \tag{15}$$

or

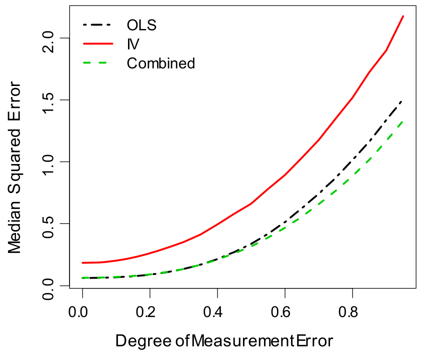
$$\Sigma_v \delta_2 = -\delta_1 \beta, \tag{16}$$

and thus the local to zero measurement error δ_1 is a linear function of the local to zero endogeneity δ_2. We state the following two theorems both in δ_1 (local degree of measurement error) and δ_2 (local degree of correlation between the structural error u_i and the reduced form error v_i).

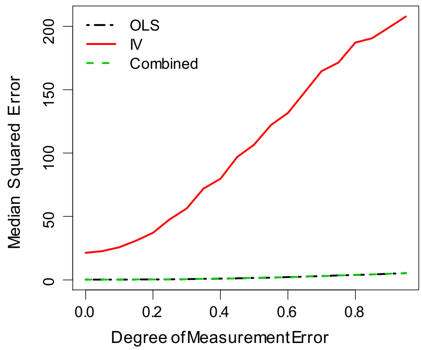
Theorem 1 (Hansen 2017). *Under the conditions that conventional central limit theory applies and that the error is conditionally homoskedastic*



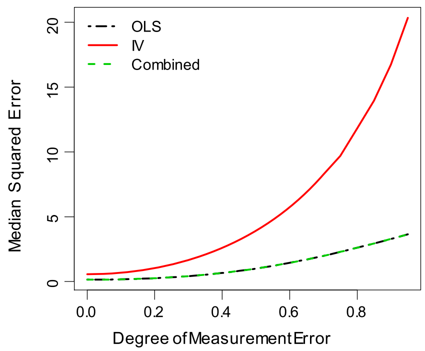
1.9 : $q = 10, R^2 = 0.1$



1.10 : $q = 10, R^2 = 0.4$



1.11 : $q = 20, R^2 = 0.1$



1.12 : $q = 20, R^2 = 0.4$

Fig. 1 continued

$$\sqrt{n}\begin{pmatrix} \hat{\beta}_0 - \beta \\ \hat{\beta}_1 - \beta \end{pmatrix} \xrightarrow{d} h + \xi \tag{17}$$

where

$$h = \begin{pmatrix} \sigma_u^{-2} V_0 \Sigma_v \delta_2 \\ 0 \end{pmatrix} = \begin{pmatrix} -\sigma_u^{-2} V_0 \beta \delta_1 \\ 0 \end{pmatrix} \tag{18}$$

$$\xi \sim N(0, V)$$

with

$$\begin{aligned} V &= \begin{pmatrix} V_0 & V_0 \\ V_0 & V_1 \end{pmatrix} \\ V_0 &= \sigma_u^2 \left[E\left(x_i x_i'\right)\right]^{-1} \\ V_1 &= \sigma_u^2 \left[E\left(x_i z_i'\right)\left[E\left(z_i z_i'\right)\right]^{-1} E\left(z_i x_i'\right)\right]^{-1}. \end{aligned}$$

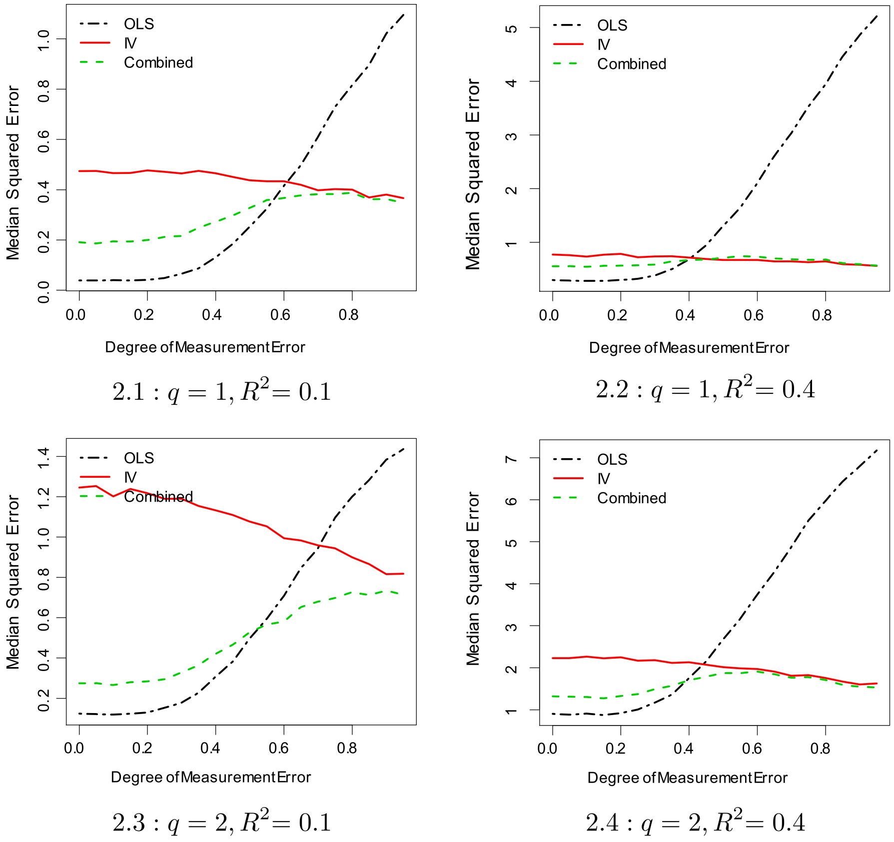


Fig. 2 $n = 100, W = (\hat{V}_1 - \hat{V}_0)^{-1}$

Furthermore,

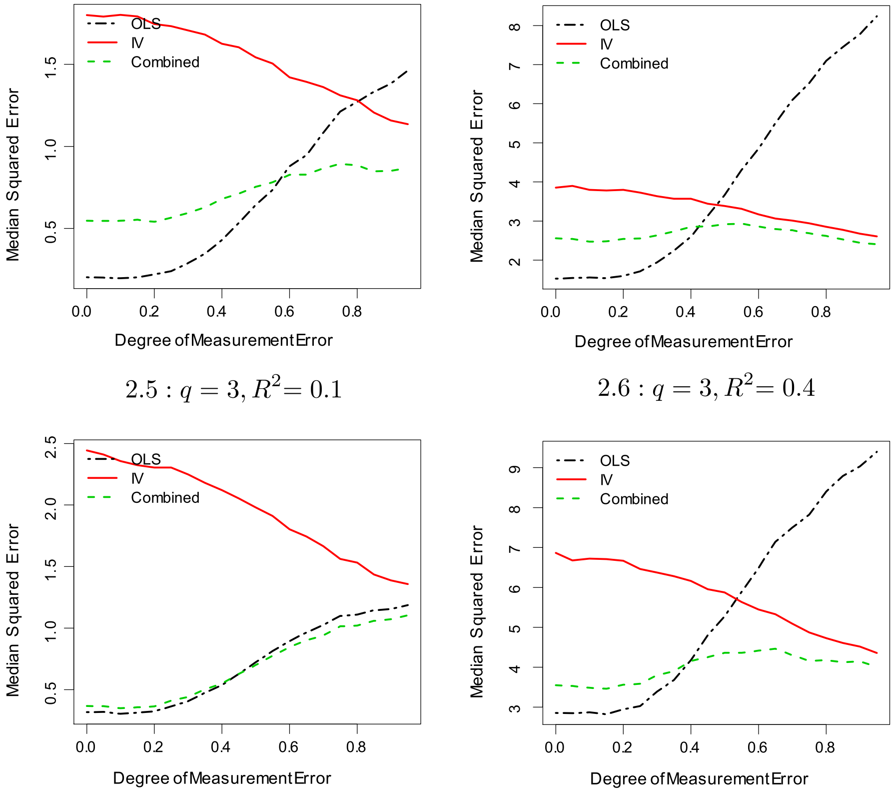
$$H_n \xrightarrow{d} (h + \xi)' P (h + \xi) \tag{19}$$

and

$$\sqrt{n}\left(\hat{\beta}_c - \beta\right) \xrightarrow{d} G_1' \xi - \left(\frac{\tau}{(h + \xi)' P (h + \xi)}\right)_1 G'(h + \xi) \tag{20}$$

where $P = G(V_1 - V_0)^{-1} G'$, $G = (-I \quad I)'$, $G_1 = (0 \quad I)'$, *and* $(a)_1 = \min[1, a]$.

Remark 1 Theorem 1 presents the joint asymptotic distribution of the OLS and IV estimators, the Hausman statistic, and the combined estimator under the local to zero measurement error assumption. The joint asymptotic distribution of the OLS and IV estimators is normal with a classic covariance matrix. The OLS estimator has an asymptotic bias when $\delta_1 \neq 0$ (and thus $\delta_2 \neq 0$), but the IV estimator does not have any asymptotic bias and is consistent. However, we emphasize that the



2.5 : $q = 3, R^2 = 0.1$ 2.6 : $q = 3, R^2 = 0.4$

2.7 : $q = 5, R^2 = 0.1$ 2.8 : $q = 5, R^2 = 0.4$

Fig. 2 continued

theorem holds *under some conditions that conventional central limit theory applies.* As is well known (Staiger and Stock 1997; Bekker 1994), our Monte Carlo results show that when there are many instruments (ℓ is large) or when there are weak instruments, the IV estimator is inconsistent. In that case we need to regularize the IV estimator to reduce the number of instruments and to shrink the weak instruments. Based on the regularized IV, we can then recover the theorem and construct a *regularized* combined estimator. We leave this for our next research agenda.

The asymptotic risk of an estimator $\hat{\beta}$ of β is defined as

$$R\left(\hat{\beta}, \beta, W\right) = \lim_{n\to\infty} E\left[n\left(\hat{\beta} - \beta\right)' W\left(\hat{\beta} - \beta\right)\right] \tag{21}$$

so long as the estimator has an asymptotic distribution

$$\sqrt{n}\left(\hat{\beta} - \beta\right) \xrightarrow{d} \psi \tag{22}$$

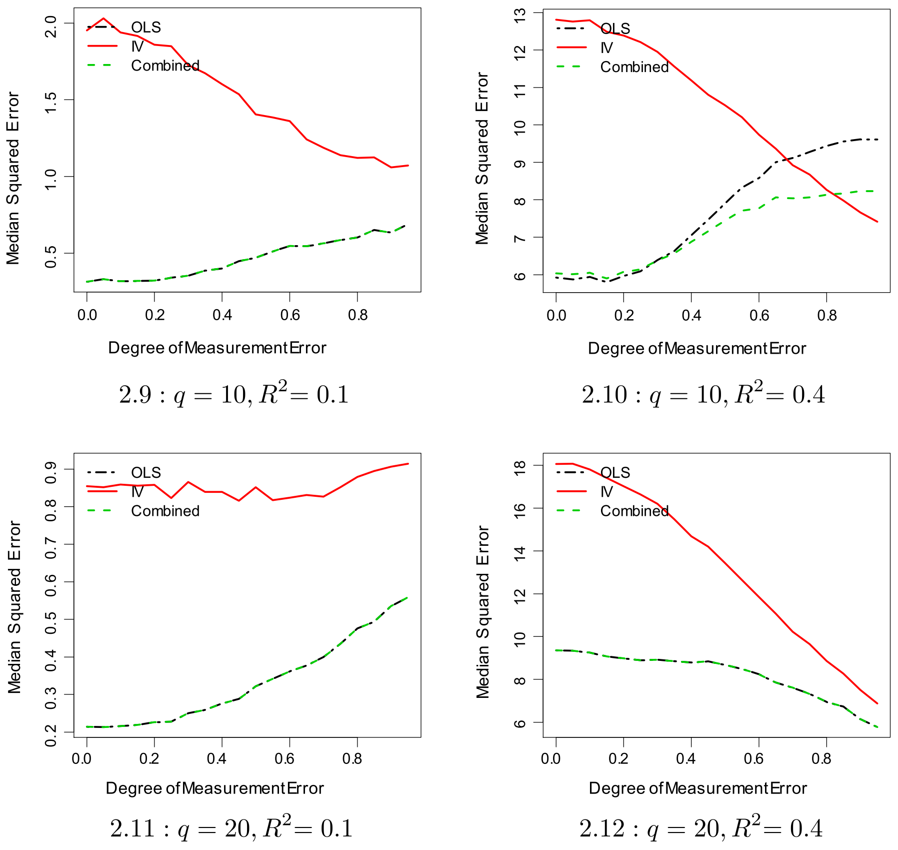


Fig. 2 continued

for some random variable ψ, the asymptotic risk can be calculated using

$$R\left(\hat{\beta}, \beta, W\right) = E(\psi' W \psi) = \text{tr}(\text{WE}(\psi\psi')). \tag{23}$$

Define the largest eigenvalue of the matrix $W(V_1 - V_0)$

$$\lambda_1 = \lambda_{\max}(W(V_1 - V_0)), \tag{24}$$

and the ratio

$$d = \frac{\text{tr}(\text{W}(\text{V}_1 - \text{V}_0))}{\lambda_1}. \tag{25}$$

Note that $1 \le d \le q$. In the case $W = (V_1 - V_0)^{-1}$, $\lambda_1 = 1$ and we have the simplification $d = q$. Hansen (2017) shows that the asymptotic risk of the combined estimator is smaller than that of the IV estimator for all values of β.

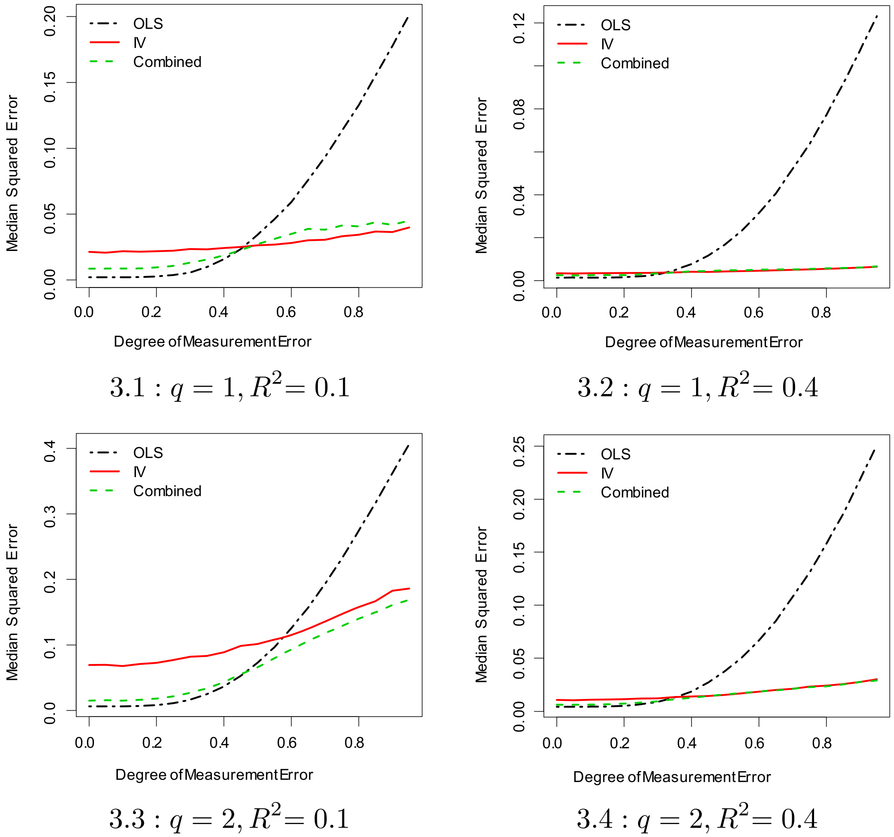


Fig. 3 $n = 200, W = I$

Theorem 2 (Hansen 2017). *Under some assumptions, if* $q > 2$ *and* $0 < \tau \leq 2(q-2)$, *then*

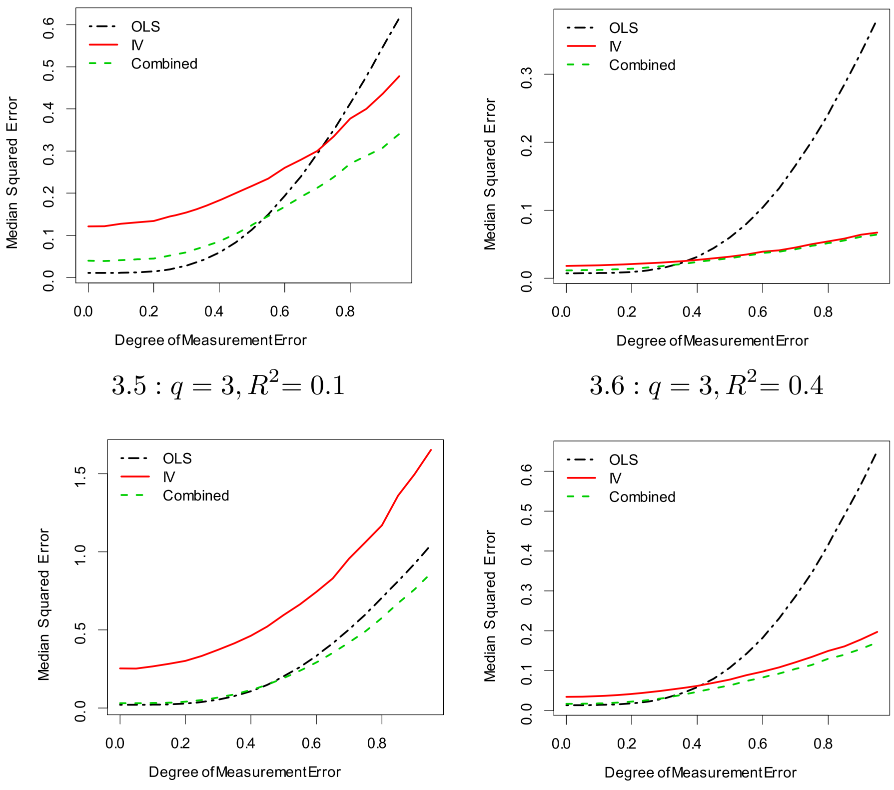
$$R\left(\hat{\beta}_1, \beta, W\right) = \mathrm{tr}(\mathrm{WV}_1), \tag{26}$$

$$R\left(\hat{\beta}_c, \beta, W\right) < R\left(\hat{\beta}_1, \beta, W\right) - \frac{\tau[2(q-2)-\tau]}{\sigma_u^{-4}\delta_2'\Sigma_v V_0 (V_1 - V_0)^{-1} V_0 \Sigma_v \delta_2 + q}, \tag{27}$$

where $\Sigma_v = E\left(v_i v_i'\right)$.

Remark 2 Theorem 2 can be stated in terms of δ_1 using the relationship $\Sigma_v \delta_2 = -\delta_1 \beta$ in (16),

$$R\left(\hat{\beta}_c, \beta, W\right) < R\left(\hat{\beta}_1, \beta, W\right) - \frac{\tau[2(q-2)-\tau]}{\sigma_u^{-4}\delta_1^2 \beta' V_0 (V_1 - V_0)^{-1} V_0 \beta + q}. \tag{28}$$



3.5 : $q = 3, R^2 = 0.1$

3.6 : $q = 3, R^2 = 0.4$

3.7 : $q = 5, R^2 = 0.1$

3.8 : $q = 5, R^2 = 0.4$

Fig. 3 continued

The improvement in the asymptotic risk by the combined estimator $\hat{\beta}_c$ over the IV estimator $\hat{\beta}_1$ depends on the measurement error bias of the OLS estimator $\hat{\beta}_0$, as the OLS bias in (7) depends on $-\sigma^2_{\varepsilon_2}\beta = -\delta_1\beta/\sqrt{n}$.

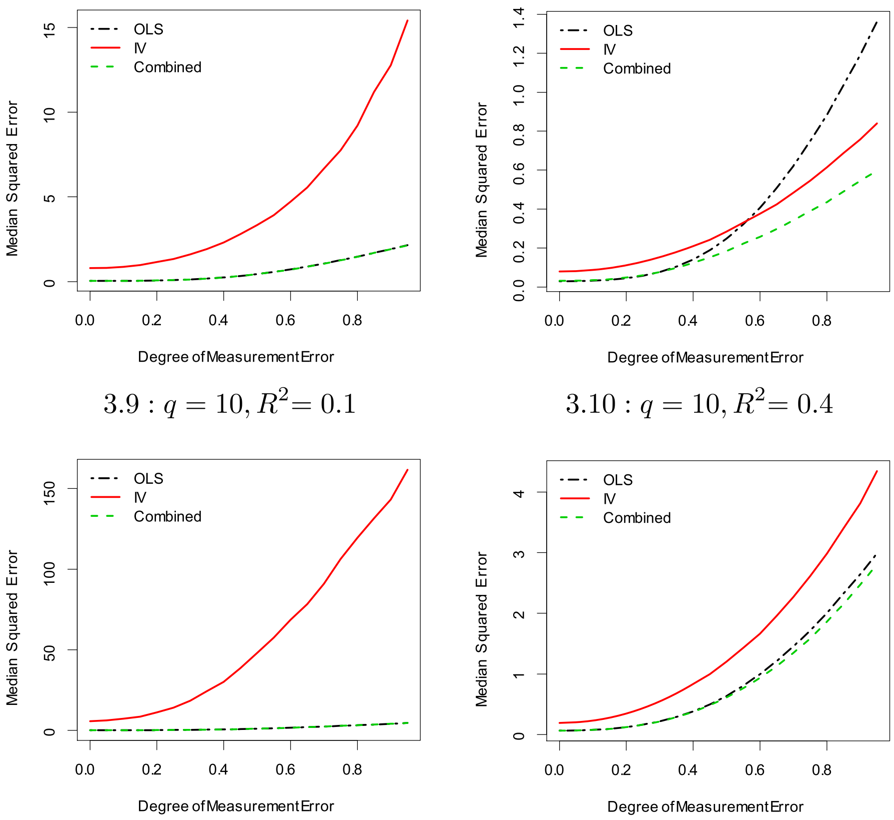
4 Monte Carlo

We thank Bruce Hansen for his R code, which we downloaded from his website. For our measurement model we made a minor modification of his code. The observations $\{y_i, x_i, z_i\}_{i=1}^n$ are generated by the process

$$y_i = x_i^{*\prime}\beta + \varepsilon_{1i} \tag{29}$$

$$x_i = x_i^* + \varepsilon_{2i} \tag{30}$$

$$x_i^* = \Pi z_i + v_i \tag{31}$$



3.9 : $q = 10, R^2 = 0.1$ 3.10 : $q = 10, R^2 = 0.4$

3.11 : $q = 20, R^2 = 0.1$ 3.12 : $q = 20, R^2 = 0.4$

Fig. 3 continued

where $\varepsilon_{1i} \sim N(0,1)$, the measurement errors ε_{2i} are uncorrelated with each other and follows $N(0, \Sigma_{\varepsilon_2})$, $\Sigma_{\varepsilon_2} = \text{diag}\left(\sigma_1^2, \sigma_2^2, \ldots, \sigma_q^2\right)$ is a $q \times q$ diagonal matrix of measurement error variances, v_i is a $q \times 1$ vector following $N\left(0, I_q\right)$, z_i is an $\ell \times 1$ vector of instruments with $N(0, I_\ell)$ and $\ell \geq q$. We report the results only for $\ell = q$.

We set $\sigma_1 = \sigma_2 = \cdots = \sigma_q = \sigma_{\varepsilon_2}$ so that $\Sigma_{\varepsilon_2} = \sigma_{\varepsilon_2}^2 I_q$. We have experimented with $\sigma_{\varepsilon_2} \in [0, 3]$ but report results with σ_{ε_2} only on [0, 1]. Note that our experiment sets the dimension of z_i equal to that of x_i, so the IV estimates are just-identified $(\ell = q)$. We set the $q \times q$ reduced form matrix as $\Pi = I_q d$ and the scale d set as $d = \sqrt{R^2/(1 - R^2)}$ so that R^2 is the reduced form population R^2 for each x_{ji}, $j = 1, \ldots, q$. We vary $n \in \{100, 200\}$, $q \in \{1, 2, 3, 5, 10, 20\}$, $R^2 \in \{0.10, 0.40\}$ and σ_{ε_2} on a 40-point grid on [0, 0.975]. The parameter R^2 controls the strength of the instruments (small R^2 is for the case of weak instruments) and the parameter σ_{ε_2} controls the size of measurement errors and thus the degree of endogeneity ($\sigma_{\varepsilon_2} = 0$

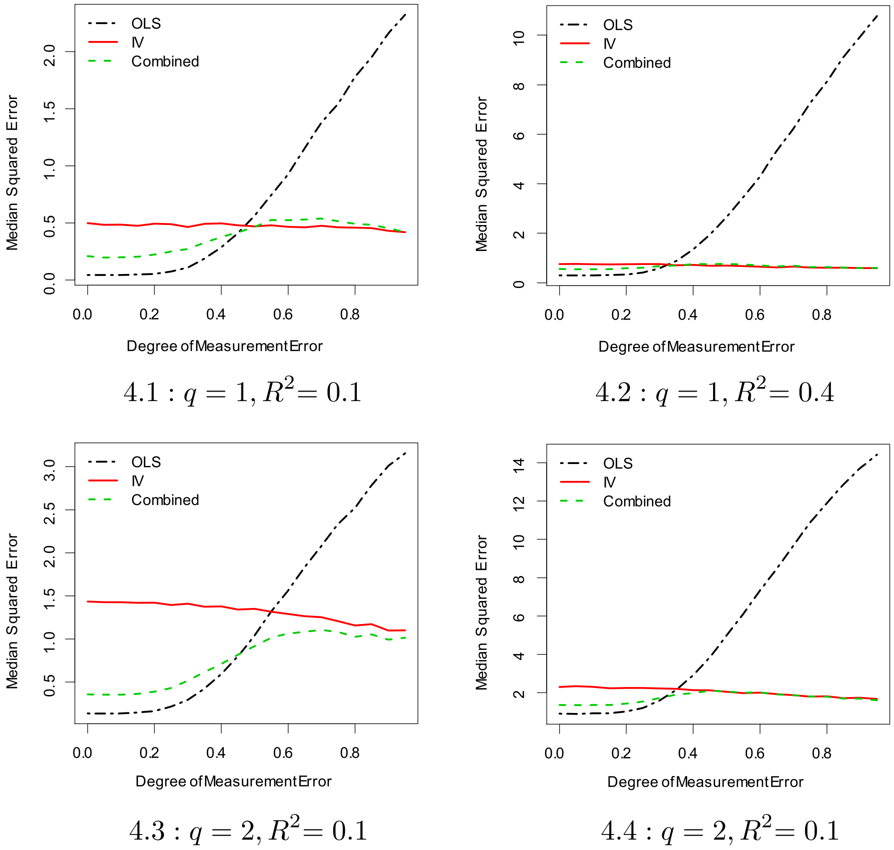


Fig. 4 $n = 200, W = (\hat{V}_1 - \hat{V}_0)^{-1}$

is for the case of no measurement error and large σ_{ε_2} is for the case of large measurement error).

We generated 10,000 samples for each configuration, and on each calculated $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_c$. For $\hat{\beta}_c$, as in Hansen (2017), we set $\tau = \frac{1}{4}$ for $q = 1$, $\tau = 1$ for $q = 2$, and $\tau = q - 2$ for $q > 2$. Our focus will be with a large q, such as $q = 3, 5, 10, 20$, for consideration of using *many* instruments. The measurement error bias in the OLS estimator $\hat{\beta}_0$ depends on $\sigma^2_{\varepsilon_2}$ and β. We fix $\beta = (1 \ldots 1)'$ such that the bias is controlled only by $\sigma^2_{\varepsilon_2}$, the size of the measurement error.

To compare the estimators we calculated the weighted median squared error risk

$$\text{median}\left(\left(\hat{\beta} - \beta\right)' W \left(\hat{\beta} - \beta\right)\right) \tag{32}$$

for each of the three estimators $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_c$. The weight is either $W = I$ or $W = \left(\hat{V}_1 - \hat{V}_0\right)^{-1}$. As in Hansen (2017), we report the median squared error rather

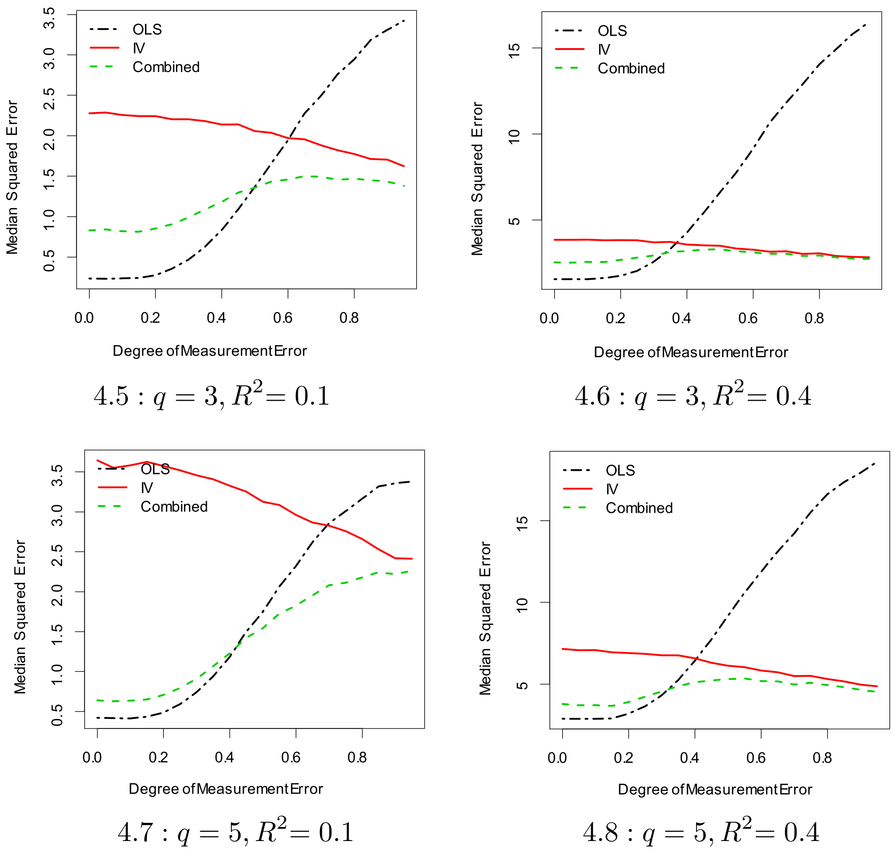
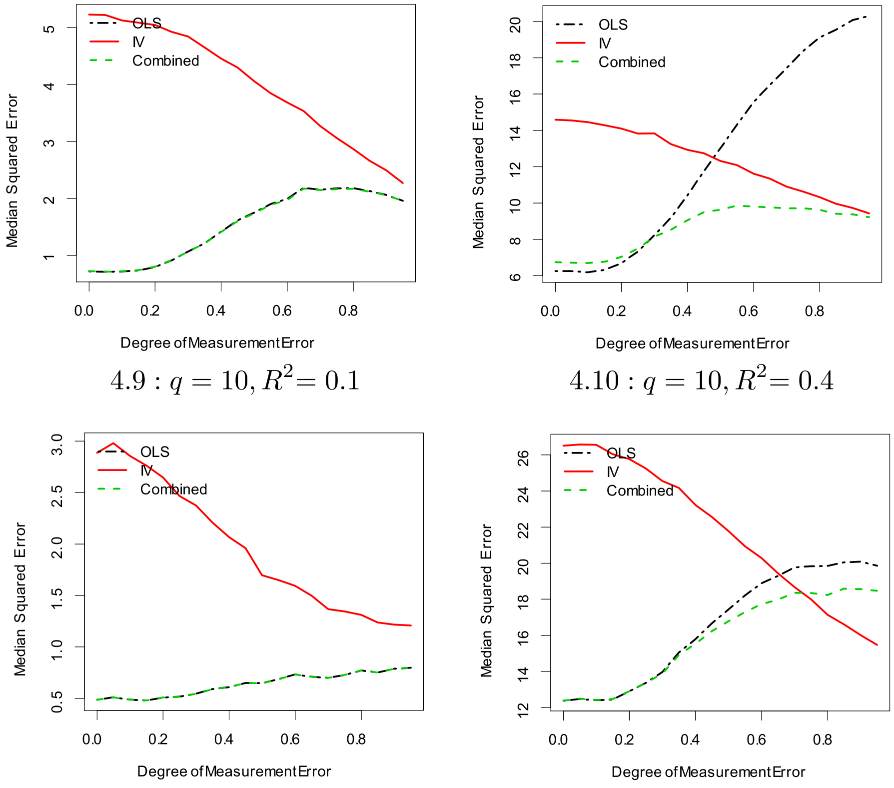


Fig. 4 continued

than the mean squared error because IV estimators for the just-identified model ($\ell = q$) may not have finite moments.

We produced many figures for various configurations. For space, we report only Figs. 1, 2, 3, 4 in the paper. Figure 1 is for $n = 100$, $W = I$. Figure 2 is for $n = 100$, $W = \left(\hat{V}_1 - \hat{V}_0\right)^{-1}$. Figure 3 is for $n = 200$, $W = I$. Figure 4 is for $n = 200$, $W = \left(\hat{V}_1 - \hat{V}_0\right)^{-1}$. For all of Figs. 1, 2, 3, 4, we set $\sigma_1 = \sigma_2 = \cdots = \sigma_q = \sigma_{\varepsilon_2}$ so that $\Sigma_{\varepsilon_2} = \sigma^2_{\varepsilon_2} I_q$.[1] There are 12 subfigures in each figure. Subfigures 1 and 2 are with $q = 1$. Subfigures 3 and 4 are for $q = 2$. Subfigures 5 and 6 are the with $q = 3$. Subfigures 7 and 8 are with $q = 5$. Subfigures 9 and 10 are using $q = 10$. Subfigures 11 and 12 have $q = 20$. The subfigures with odd numbers are the case

[1] We have also considered cases where only one (the first) regressor has measurement error, i.e., $\sigma_1 = \sigma_{\varepsilon_2}$ and $\sigma_2 = \cdots = \sigma_q = 0$ so that $\Sigma_{\varepsilon_2} = \text{diag}\left(\sigma^2_{\varepsilon_2}, \mathbf{0}\right)$. These are Figs. 5, 6, 7, 8, in the supplemental appendix, available in our website.



4.9 : $q = 10, R^2 = 0.1$

4.10 : $q = 10, R^2 = 0.4$

4.10 : $q = 20, R^2 = 0.1$

4.12 : $q = 20, R^2 = 0.4$

Fig. 4 continued

$R^2 = 0.1$ (weaker instruments). The subfigures with even numbers are the case $R^2 = 0.4$ (stronger instruments).

Some observations from these results are summarized here. Theorem 2 shows that the asymptotic risk of the combined estimator is uniformly smaller than that of the IV estimator when $q > 2$. Subfigures for $q > 2$ show that the asymptotic uniform ranking holds in finite samples. When q is too large ($q = 10, 20$), subfigures 1.9, 1.10, 1.11 and 1.12 show that both OLS and IV are inconsistent, little or no improvements can be achieved by the combined estimator as the combined estimator puts (almost) all weight on OLS and little weight on IV as q gets larger relative to n. It is interesting that IV is worse than OLS when q is large, for a quite large size of measurement errors over a wide range of σ_{ε_2} . It should be note that, when IV loses consistency due to many instruments, the Hausman statistic may not be used for the construction of the combined estimator.

Finally, in order to see if this would also happen with other form of endogeneity, we also replicated Hansen (2017) monte carlo results for his setup of a structural model using large $q = 5, 10, 20$.[2] In Hansen (2017), it is reported only for small values of $q \in \{1, 2, 3, 4\}$. As expected, for larger values of q, the findings from the structural model are similar to what we see from Figs. 1, 2, 3, 4 for the measurement error model: namely, when q is large, both OLS and IV are inconsistent, IV is worse than OLS, and little or no improvements can be achieved by the combined estimator.

5 Conclusions

Given the findings from this paper, a next question is when we may use IV. When does IV do better than OLS? We find that's when R^2 is not small (when instruments are not weak), $\sigma^2_{\varepsilon_2}$ is not small (when measurement errors are large), and q is not large relative to n (when there are not too many instruments). Otherwise, IV can be bad, possibly much worse than OLS. When we do not know how small R^2 or $\sigma^2_{\varepsilon_2}$ is too small or how large q is too large, a safe guard would be to use the combined estimator. Although these findings are shown for the errors in variables model they may apply to other econometric models using IV estimators.

Recently, many papers have been written to deal with the endogeneity in high dimensional models with large q. The basic idea is to reduce the dimension q by selecting some of the instruments. The lasso type methods pioneered by Tibshirani (1996) reduce the number of variables in the high dimensional models by shrinking some of the estimators toward zero. With the rapid advance of statistical learning and machine learning in big data environment, the regularization of the standard econometric methods such as OLS, IV or GMM have been very actively studied in recent econometrics community, for example, Friedman (2001), Bühlmann and Yu (2003), Chao and Swanson (2007), Hansen et al. (2008), Ng and Bai (2008), Newey and Windmeijer (2009), Caner (2009), Donald et al. (2008), Carrasco (2012), Belloni et al. (2012), Belloni and Chernozhukov (2011; 2013), Liao (2013), Fan and Liao (2014), Chernozhukov and Hansen (2004; 2005; 2006; 2013), Chernozhukov et al. (2015), Cheng and Liao (2015), DiTraglia (2016), Harding et al. (2016), Chudik et al. (2016), Caner et al. (2017a, b), and Lee and Xu (2017), among others. They show various methods such as lasso, adaptive lasso, SCAD, elastic net, boosting, to mitigate the effect of high dimension on consistent estimation of the econometric models such as the conditional mean and conditional quantiles. Many of these papers have studied various regularization methods to select instruments to improve IV. It would be interesting to see how these methods for high dimensional models could also improve the combined estimator. As noted in Remark 1 above, we are currently investigating this in relation to the current paper and Hansen (2017).

[2] These are Figs. 9 and 10 in the supplemental appendix, available in our website.

References

Bekker, P. A. (1994). Alternative approximations to the distributions of instrumental variable estimators. *Econometrica*, *62*, 657–681.

Belloni, A., Chen, D., Chernozhukov, V., & Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, *80*, 2369–2429.

Belloni, A., & Chernozhukov, V. (2011). ℓ_1 -penalized quantile regression in high-dimensional sparse models. *Annals of Statististics*, *39*(1), 82–130.

Belloni, A., & Chernozhukov, V. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli*, *19*(2), 521–547.

Bühlmann, P., & Yu, B. (2003). Boosting with the L_2 Loss. *Journal of the American Statistical Association*, *98*, 324–339.

Caner, M. (2009). Lasso-type GMM estimator. *Econometric Theory*, *25*(1), 270–290.

Caner, M., Han, X., & Lee, Y. (2017a). Adaptive elastic net GMM estimation with many invalid moment conditions: simultaneous model and moment selection. *Journal of Business and Economic Statistics*, forthcoming.

Caner, M., Maasoumi, E., & Riquelme, J. (2017b). Moment and IV selection approaches: a comparative simulation study. *Econometric Reviews*, forthcoming.

Carrasco, M. (2012). A regularization approach to the many instruments problem. *Journal of Econometrics*, *170*, 383–398.

Chao, J., & Swanson, N. (2007). Alternative specifications of the bias and MSE of the IV estimator under weak identification with an application to bias correction. *Journal of Econometrics*, *137*, 515–555.

Chernozhukov, V., & Hansen, C. (2004). The effects of 401(k) participation on the wealth distribution: an instrumental quantile regression analysis. *The Review of Economics and Statistics*, *86*(3), 735–751.

Chernozhukov, V., & Hansen, C. (2005). An IV model of quantile treatment effects. *Econometrica*, *73*(1), 245–261.

Chernozhukov, V., & Hansen, C. (2006). Instrumental quantile regression inference for structural and treatment effects models. *Journal of Econometrics*, *132*, 491–525.

Chernozhukov, V., & Hansen, C. (2013). Quantile models with endogeneity. *Annual Review of Economics*, *5*(1), 57–81.

Chernozhukov, V., Hansen, C., Spindler, M. (2015). Post-Selection and Post-Regularization Inference in Linear Models with Many Controls and Instruments. MIT Department of Economics, Working Paper No. 15-02.

Cheng, X., & Liao, Z. (2015). Select the valid and relevant moments: an information-based LASSO for GMM with many moments. *Journal of Econometrics*, *186*, 443–464.

Chudik, A., Kapetanios, G., Pesaran, M.H. (2016). Big data analytics: a new perspective. USC Globalization and Monetary Policy Institute Working Paper 268.

DiTraglia, F. (2016). Using invalid instruments on purpose: focused moment selection and averaging for GMM. *Journal of Econometrics*, *195*(2), 187–208.

Donald, S., Imbens, G., & Newey, W. (2008). Choosing instrumental variables in conditional moment restriction models. *Journal of Econometrics*, *152*, 28–36.

Fan, J., & Li, Runze. (2001). Variable selection via nonconcave penalized likelihood and its Oracle properties. *Journal of the American Statistical Association*, *96*(456), 1348–1360.

Fan, J., & Liao, Y. (2014). Endogeneity in high dimension. *The Annals of Statistics*, *42*(3), 872–917.

Friedman, J. (2001). Greedy function approximation: a gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232.

Hansen, B.E. (2017). Stein-like 2SLS estimator. *Econometric Reviews*, *36*, 840–852.

Hansen, C., Hausman, J., & Newey, W. (2008). Estimation with many instrumental variables. *Journal of Business & Economic Statistics*, *26*(4), 398–422. http://www.ssc.wisc.edu/~bhansen/papers/shrinkiv.pdf

Harding, M., Hausman, J., & Palmer, C. J. (2016). Finite sample bias corrected IV estimation for weak and many instruments. *Advances in Econometrics*, *36*, 245–273.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, *46*(6), 1251–1271.

James, W., & Stein, C. M. (1961). Estimation with quadratic loss. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, *1*, 361–379.

Lee, T.-H. & Xu, H. (2017). Boosting penalized GMM estimation with many instruments that are possibly irrelevant or invalid. UC Riverside.

Newey, W. K., & Windmeijer, F. (2009). Generalized method of moments with many weak moment conditions. *Econometrica*, *77*(3), 687–719.

Ng, S. & Bai, J. (2008). Selecting instrumental variables in a data rich environment. Journal of Time Series Econometrics 1(1), Article 4.

Staiger, D., & Stock, J. (1997). Instrumental variables regression with weak instruments. *Econometrica*, *65*(3), 557–586.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of Royal Statististical Society, Series B*, *58*, 267–288.