Lancaster University Management School

Economics Working Paper Series

2019/001

Higher Order Approximation of IV Estimators with Invalid Instruments

Byunghoon Kang

The Department of Economics
Lancaster University Management School
Lancaster LA1 4YX
UK

Higher Order Approximation of IV Estimators with Invalid Instruments

Byunghoon Kang[†]

[†] *Department of Economics, Lancaster University Management School, Lancaster, LA1 4YX, UK*
E-mail: b.kang1@lancaster.ac.uk

December 2018

Click here for the Online Supplementary Appendix

Abstract This paper considers the instrument selection problem in instrumental variable (IV) regression model when there is a large set of instruments with potential invalidity. I derive higher-order mean square error (MSE) approximation of two-stage least squares (2SLS), limited information maximum likelihood (LIML), Fuller (FULL) and bias-adjusted 2SLS (B2SLS) estimators with allowing for local violation of the instrument-exogeneity conditions. Based on the approximation to the higher-order MSE, I propose instrument selection criteria that are robust to potential invalidity of instruments. Furthermore, I also show the optimality results of instrument selection criteria in Donald and Newey (2001, *Econometrica*) under faster than $N^{-1/2}$ locally invalid instruments specification.

Keywords: *Instrument selection, Invalid instruments, Many instruments, 2SLS, LIML, Fuller estimator, Bias-adjusted 2SLS*

1. INTRODUCTION

The instrumental variable (IV) estimators are widely used in modern economics, and some empirical applications involve a large set of potential instruments and debates about the validity of instruments, which I refer to as an exogeneity condition, i.e., instruments are uncorrelated with the error term in the structural equation. Although researchers have been routinely used the Sargan-Hansen J-test, the validity of instruments is generally uncertain; instruments may have direct effects on the outcome variables, and model misspecification can make instruments invalid.[1] Furthermore, when there are many potential instruments, the finite sample performance of the IV estimator can be sensitive to the choice of instruments. To capture finite sample properties of the IV estimator, higher-order approximation and the instrument selection criteria have been found useful in the literature, e.g., Donald and Newey (2001), Hahn et al. (2004), Kuersteiner and Okui (2010), assuming instruments are valid.

This paper develops Nagar (1959) type higher-order mean square error (MSE) approximation of the k-class estimators (including two-stage least squares (2SLS) estimator, limited information maximum likelihood (LIML), Fuller (FULL), and bias-adjusted version of the 2SLS (B2SLS) estimator) in linear IV model with many instruments allowing locally invalid instruments. Higher-order MSE approximation of k-class estimators

[1] For questionable IVs with potential invalidity in various empirical applications, see Section 2.1 of Guggenberger (2012) and references therein. See also Kolesar et al. (2015) for an interesting empirical application with invalid instruments even when the instruments are assigned randomly.

under the local violation of the instrument-exogeneity includes higher-order bias from many/invalid instruments as well as higher-order variances, and this has not been available in the literature. MSE approximation in this paper not only depends on the number of instruments but also on the degree of instruments invalidity and relative strength of valid and invalid instruments. Our result is also useful to provide which estimator is more robust to local violation of the validity of the instrument among the wide class of the k-class estimators. Furthermore, this paper also generalizes the first-order asymptotic results in Hahn and Hausman (2005) as well as the higher-order expansions in Rothenberg (1984) with invalid instruments, see Remarks 3.3 and 3.4 below.

Higher order approximation in this paper hinges on $N^{-\gamma}(\gamma \geq 1/2)$ local-to-zero specification allowing locally invalid instruments, and the local-misspecification approach has gained considerable new attention recently, e.g., Conley et al. (2012), Andrews et al. (2017), Armstrong and Kolesar (2018), Bonhomme and Weidner (2018).[2] This device allows to develop useful approximation theory of IV estimators with invalid instruments and it requires non-trivial extension of Donald and Newey (2001) because the dominating higher-order terms not only depend on the order of invalid instrument (γ) but also the rate of instruments (K) as well as the estimators considered.

First, I show that dominating terms (that depend on K) of MSE approximation in this paper reduce to those of Donald and Newey (2001) with the knife-edge rate $\gamma = 1/2$ when (i) instruments are all valid or (ii) the direct effect of the instruments on the outcome variable are orthogonal to their effect on the endogenous variable. The second case is closely related to the identification assumption in the literature, e.g., Kolesar et al. (2015). Based on the higher-order MSE approximation, I suggest instrument selection criterion that is robust to locally invalid instruments. We require known to be valid instruments and provide the asymptotic unbiasedness of the criterion function.[3] In the presence of invalid instruments, instruments selection criterion without additional terms in the MSE may lead to a misleading balance of bias and efficiency; including invalid instruments that are a strong predictor of the first-stage can lower MSE although it may slightly increase bias.

Second, I show that instrument selection criteria in Donald and Newey (2001) are robust to a small degree of invalid instruments as the dominating terms in MSE coincide when $\gamma > 1/2$. Interestingly, results in this paper suggest that LIML, FULL, and B2SLS are more robust than 2SLS with locally invalid instrument specifications. To be more specific, we show the asymptotic optimality of Donald and Newey (2001) criterion for LIML, FULL, and B2SLS for all $\gamma > 1/2$, and 2SLS for $\gamma > 1 - \alpha$ under the same rate restriction of $K = N^{\alpha}(0 < \alpha < 1/2)$ imposed in Donald and Newey (2001).

The literature on the higher order approximation of the k-class estimator with valid instruments has a long history such as Nagar (1959), Anderson and Sawa (1973), Morimune (1983), Rothenberg (1984). Based on higher-order approximations, there are many papers about instrument (moment) and/or weight selections in the IV and GMM setup, e.g., Donald and Newey (2001), Donald et al. (2009), Canay (2010), Kuersteiner and Okui (2010), Okui (2011), Carrasco (2012), Kuersteiner (2012), Lee and Zhou (2015), among others. Many papers also developed moment selection procedures to select valid

[2] There are several papers deal with estimation and inference issues with with local violation of exogeneity conditions, Newey (1985), Hahn and Hausman (2005), Berkowitz et al. (2008), Otsu (2011), Berkowitz et al. (2012), Guggenberger (2012), Kraay (2012), Caner (2014), among many others.

[3] The existence of valid instruments is also used for identification and estimation purposes in recent papers, e.g., Nevo and Rosen (2012), Cheng and Liao (2015), DiTraglia (2016).

moments from the set of valid and invalid moment conditions, e.g., Andrews (1999), Andrews and Lu (2001), Hall and Peixe (2003), Hong et al. (2003), Liao (2013). DiTraglia (2016) developed moment selection criterion based on the first-order asymptotic MSE with possible locally invalid moment conditions in GMM setups.

Although we require prior knowledge of the order of the instrument strengths in this paper similar to Donald and Newey (2001), many recent papers developed estimation and moment selection techniques (e.g., Lasso) in high-dimensional setup without requiring order of the instruments such as Belloni et al. (2012), Cheng and Liao (2015), Caner et al. (2018), for examples. Kang et al. (2016) propose Lasso-type methods to identify and select valid instruments, and they do not a require small set of instruments that are known to be valid. Windmeijer et al. (2018) investigate that Lasso procedures may not consistently select the invalid instruments if these are relatively strong, and they propose a median-type estimator that is consistent when more than 50% of the instruments are valid.

The outline of the paper is as follows. Section 2 introduces the basic model setup and notation. Section 3 describes higher-order MSE approximations of the IV estimators. Section 4 provides approximations under the different local sequence of invalid instruments, and Section 5 discusses instrument selection criteria. Section 6 includes simulation results in various Monte Carlo settings and Section 7 concludes. All proofs, additional simulations, and results are provided in the Online Appendix.

2. THE MODEL AND ESTIMATORS

I consider a linear IV model allowing potentially invalid instruments,

$$y_i = W_i'\delta_0 + \varepsilon_i = Y_i'\theta_0 + x_{1i}'\beta_0 + \varepsilon_i \tag{2.1}$$

$$W_i = f(x_i) + u_i = \begin{pmatrix} E[Y_i|x_i] \\ x_{1i} \end{pmatrix} + \begin{pmatrix} \xi_i \\ 0 \end{pmatrix}, \tag{2.2}$$

$$\varepsilon_i = \frac{g(x_i)}{\sqrt{N}} + v_i, \qquad E[v_i|x_i] = 0, \tag{2.3}$$

where y_i is a scalar outcome variable, W_i is a $p \times 1$ vector that includes endogenous variables Y_i and $d \times 1$ vector of exogenous variables x_{1i}. $\delta_0 = (\theta_0', \beta_0')' \in \mathbb{R}^p$ is a parameter of interest and x_{1i} is a subset of the exogenous variables x_i. The number of regressors (p) and the number of exogenous regressors (d) are assumed fixed and they do not depend on the sample size N.

(2.3) imposes local-to-zero specification allowing invalid instruments. Under this setup, any potential instruments $\psi(x_i)$ are asymptotically valid as $N \to \infty$, but $E[\psi(x_i)\varepsilon_i] = 0$ do not necessarily hold in finite samples. Note that the direct effect of the instruments $g(x)$ can be allowed to be large numbers for any finite N, and we do not restrict the functional form of $g(x)$.

REMARK 2.1. With this knife-edge rate ($N^{-1/2}$), the stochastic order of bias of IV estimators from the invalid instrument is equal to the first-order asymptotic variance. This rate provides the right-balance for a useful theory to understand the finite sample behavior of IV estimators with possibly invalid instruments. We also consider the higher-order MSE approximations under faster rates ($N^{-\gamma}, \gamma > 1/2$) in Section 4 which requires different analysis because of the dominating term changes.

We may also consider $\gamma = 0$ (global misspecification) or $0 < \gamma < 1/2$ where bias from invalid instruments dominates. We can still provide MSE approximations centered with pseudo-true value which is the probability limit of IV estimators (or sequence of pseudo-true value depending on N), e.g., instruments specific LATE (Local Average Treatment Effect) parameters for 2SLS in the presence of heterogeneous treatment effect. However, different choice of instruments and estimators can lead to a different pseudo-true value, and this makes MSE comparisons difficult. Thus, we focus on the δ_0 and compare MSE among different instruments as well as different IV estimators under our setup with $\gamma \geq 1/2$. Identification issues with globally invalid instruments in IV framework can be found in Nevo and Rosen (2012), Kolesar et al. (2015), Kang et al. (2016).

Now, we consider several different k-class estimators which are widely used in the linear IV model. We first consider the 2SLS estimator,

$$\widehat{\delta}_{\mathrm{2SLS}}(K) = (W'P^K W)^{-1}(W'P^K y), \tag{2.4}$$

where $y = (y_1, \cdots, y_N)', W = [Y, X_1], Y = [Y_1, \cdots, Y_N]', X_1 = [x_{11}, \cdots, x_{1N}]'$ and $P^K = \Psi^K(\Psi^{K'}\Psi^K)^-\Psi^{K'}$ is the projection matrix for the instrument vector $\Psi^K = [\psi_1^K, ..., \psi_N^K]'$ with $K \times 1 (K \geq p)$ vector of instrumental variables (or basis functions) $\psi_i^K \equiv \psi^K(x_i) = (\psi_{1K}(x_i), ..., \psi_{KK}(x_i))'$. I assume ψ_i^K includes exogenous variables x_{1i}, and K indicates both the number of instruments and the index of the instrument sets following Donald and Newey (2001).

Next, we consider LIML estimator

$$\widehat{\delta}_{\mathrm{LIML}}(K) = (W'P^K W - \hat{\Lambda}(K)W'W)^{-1}(W'P^K y - \hat{\Lambda}(K)W'y), \tag{2.5}$$

where

$$\hat{\Lambda}(K) = \min_{\delta} \frac{(y - W\delta)'P^K(y - W\delta)}{(y - W\delta)'(y - W\delta)}.$$

We also consider the Fuller (1977) estimator (FULL),

$$\widehat{\delta}_{\mathrm{FULL}}(K) = (W'P^K W - \check{\Lambda}(K)W'W)^{-1}(W'P^K y - \check{\Lambda}(K)W'y), \tag{2.6}$$

where

$$\check{\Lambda}(K) = \frac{\hat{\Lambda}(K) - \frac{C}{N-K}(1 - \hat{\Lambda}(K))}{1 - \frac{C}{N-K}(1 - \hat{\Lambda}(K))}$$

for some constant C. Popular choices are $C = 1$ or $C = 4$ due to their higher-order unbiasedness or minimum MSE property. Finally, we consider bias-adjusted 2SLS estimator (B2SLS) from Donald and Newey (2001) as a modification of the Nagar (1959) estimator with $\bar{\Lambda}(K) = (K - d - 2)/N$,

$$\widehat{\delta}_{\mathrm{B2SLS}}(K) = (W'P^K W - \bar{\Lambda}(K)W'W)^{-1}(W'P^K y - \bar{\Lambda}(K)W'y). \tag{2.7}$$

3. ASSUMPTIONS AND HIGHER-ORDER MSE RESULTS

I derive the Nagar (1959) type higher-order asymptotic MSE for the IV estimators with locally invalid instruments setup. I will find a decomposition for IV estimators $\widehat{\delta}(K)$

above with the following form,

$$N(\widehat{\delta}(K)-\delta_0)(\widehat{\delta}(K)-\delta_0)' = \widehat{Q}(K)+\widehat{r}(K),$$
$$E[\widehat{Q}(K)|X] = \sigma_v^2 H^{-1} + H^{-1}H_g H_g' H^{-1} + G + L(K) + T(K), \quad (3.1)$$
$$[\widehat{r}(K)+T(K)]/tr(G+L(K)) = o_p(1), \quad K\to\infty, N\to\infty,$$

where $H = f'f/N, H_g = f'g/N, f = [f_1, \cdots, f_N]', f_i = f(x_i)$, $g = [g_1, \cdots, g_N]', g_i = g(x_i)$, and $X = [x_1, ..., x_N]'$. I also define $\sigma_{uv} = E[u_i v_i|x_i], \sigma_v^2 = E[v_i^2|x_i], \sigma_\varepsilon^2 = E[\varepsilon_i^2|x_i]$ and $\Sigma_u = E[u_i u_i'|x_i]$.

The dominating terms in the conditional MSE approximation (3.1) are $\sigma_v^2 H^{-1}$ and $H^{-1}H_g H_g' H^{-1}$ that correspond to the first-order asymptotic variance and the square of the asymptotic bias from the locally invalid instrument, respectively. These are $O_p(1)$ terms, but they do not depend on K in our large K approximation. Next leading terms are G and $L(K)$. They include the higher-order bias and variance due to many and invalid instruments and have different forms for each IV estimator. G includes terms that do not depend on K. $\widehat{r}(K)$ and $T(K)$ are the remainder terms converge to 0 faster than G and $L(K)$.

We impose the following assumptions similar to Donald and Newey (2001).

Assumption 3.1. *(a)* $\{y_i, Y_i, x_i\}_{i=1}^N$ *are independent and identically distributed (i.i.d.); (b)* $E[v_i^2|x_i] = \sigma_v^2 > 0$, *and* $E[\|\xi_i\|^4|x_i], E[|v_i|^4|x_i]$ *are bounded.*

Assumption 3.2. *(a)* $\bar{H} = E[f_i f_i']$ *exists and is nonsingular,* $\bar{H}_g = E[f_i g_i]$ *exists; (b) there exists* π_K, π_K^g *such that* $E[\|f(x) - \pi_K \psi^K(x)\|^2] \to 0$ *and* $E[|g(x) - \pi_K^g \psi^K(x)|^2] \to 0$ *as* $K \to \infty$.

Assumption 3.3. *(a)* $E[(v_i, \xi_i')'(v_i, \xi_i')|x_i]$ *is constant; (b)* ${\Psi^K}'\Psi^K$ *is nonsingular with probability approaching one; (c)* $max_{i\le N} P_{ii}^K \xrightarrow{p} 0$. *(iv)* f_i *and* g_i *are bounded.*

Assumption 3.1 imposes boundedness of the fourth conditional moments of the error terms. Assumption 3.2(a) is imposed for a usual identification assumption and for the existence of the first-order bias from invalid instruments. Assumption 3.2(b) requires the mean square approximation error of the $f(x)$ and $g(x)$ by the linear combination of instruments $\psi^K(x)$ goes to 0 as the number of instrument increases. Assumption 3.3 also imposes homoskedasticity and restricts the growth rate of K. For example, $K = O(N)$ is not allowed under 3.3(c), see van Hasselt (2010), Anatolyev and Yaskov (2017), and references therein.

Our first result gives the MSE approximation for the 2SLS estimator. Proposition 3.1 is a generalization of the result in Donald and Newey (2001) allowing possibly invalid instruments.

Proposition 3.1. *If Assumptions 3.1, 3.2, 3.3 are satisfied,* $\sigma_{uv} \neq 0, H_g\sigma_{uv}' \neq 0, H_g \neq 0$, *and* $K^2/N \to 0$, *then the approximate MSE for the 2SLS estimator satisfies decompo-*

sition (3.1) with $G = 0$ *and the following terms*

$$L(K) = H^{-1}\big[\frac{K}{N^{1/2}}(H_g\sigma_{uv}' + \sigma_{uv}H_g') + \sigma_{uv}\sigma_{uv}'\frac{K^2}{N} + \sigma_v^2\frac{f'(I-P^K)f}{N} + H_gH_g'H^{-1}\frac{f'(I-P^K)f}{N} + \frac{f'(I-P^K)f}{N}H^{-1}H_gH_g' - \frac{f'(I-P^K)g}{N}H_g' - H_g\frac{g'(I-P^K)f}{N}\big]H^{-1}. \quad (3.2)$$

Moreover, ignoring terms of order $O_p(K^2/N) = o_p(K/\sqrt{N})$,

$$L(K) = H^{-1}\big[\frac{K}{N^{1/2}}(H_g\sigma_{uv}' + \sigma_{uv}H_g') + \sigma_v^2\frac{f'(I-P^K)f}{N} + H_gH_g'H^{-1}\frac{f'(I-P^K)f}{N} + \frac{f'(I-P^K)f}{N}H^{-1}H_gH_g' - \frac{f'(I-P^K)g}{N}H_g' - H_g\frac{g'(I-P^K)f}{N}\big]H^{-1}. \quad (3.3)$$

Remark 3.1. When $H_g = 0$, $L(K)$ in Proposition 3.1 reduces to Proposition 1 in Donald and Newey (2001),

$$H^{-1}[\sigma_{uv}\sigma_{uv}'\frac{K^2}{N} + \sigma_v^2\frac{f'(I-P^K)f}{N}]H^{-1}.$$

MSE approximation results in Proposition 3.1 includes higher-order terms from many instruments as well as additional terms due to invalid instruments.

Interestingly, $H_g = 0$ not only holds when $g(x) = 0$ (i.e., exclusion restriction holds) but also holds when the direct effect of the instruments to the outcome variable (g) are orthogonal to the effect of the instruments on the endogenous variable (f) allowing $g(x) \neq 0$. Suppose $f = \Psi\pi, g = \Psi\tau$, then $H_g = 0$ holds when $\pi'\Psi'\Psi\tau/N = 0$ and this is closely related to the identification assumption (Assumption 5) in Kolesar et al. (2015) for the consistency of k-class estimator under many invalid instruments setup. This observation suggests that the robustness of MSE approximation results (thus, instrument selection criteria) of Donald and Newey (2001) to the possibly invalid instruments provided that the direct effects of instruments are uncorrelated with the effects on the first-stage. See also discussions in Kolesar et al. (2015) with empirical examples such as Chetty et al. (2011) where $H_g = 0$ may be a reasonable assumption.

Remark 3.2. The first three terms of $L(K)$ in (3.2) and $H^{-1}H_gH_g'H^{-1}$ correspond to the square and cross-product of the two sources of bias we consider: bias from many instruments and invalid instruments. With invalid instruments, the dominating terms in MSE approximation are order $O_p(K/\sqrt{N})$ as in (3.3). This dominates the bias from many instruments $O_p(K^2/N)$ in Donald and Newey (2001). The remaining terms in $L(K)$ regarding $f'(I-P^K)f/N$ represent higher order variance term and it decreases as K increases. Note that locally invalid instrument specifications change not only the order of the bias from many instruments but also weights on the higher order variance $f'(I-P^K)f/N$. If the chosen instruments K are independent with the direct effect of invalid instruments g, then $P^Kg/N = 0$ and the last two terms in $L(K)$ reduce to $-2H_gH_g'$. Therefore excluding invalid instruments can help to reduce MSE. However, including (locally) invalid instruments that are a strong predictor of the first-stage also can lower MSE. Proposition 3.1 provides robust bias and variance trade-off in the presence of invalid instruments

REMARK 3.3. In an online appendix (Section S2.1), I provide MSE approximation of 2SLS under $K = O(\sqrt{N})$ and this generalizes the first-order asymptotic MSE results of the Hahn and Hausman (2005).[4] For a linear specification of $f = \Psi\pi, g = \Psi\tau$ with a scalar endogenous variable Y_i and no included exogenous variables, dominating bias term in MSE approximation becomes

$$H^{-1}H_gH_g'H^{-1} + H^{-1}[\frac{K}{\sqrt{N}}(H_g\sigma_{uv}' + \sigma_{uv}H_g') + \sigma_{uv}\sigma_{uv}'\frac{K^2}{N}]H^{-1} = (\frac{H_g + \alpha\sigma_{uv}}{H})^2$$

where $H_g = \pi'\Psi'\Psi\tau/N, \alpha = K/\sqrt{N}$, and this corresponds to Theorem 3 of Hahn and Hausman (2005). Our results imply that the normal distribution of the error terms and linearity assumption of f and g are not essential for results in Hahn and Hausman (2005).

REMARK 3.4. In an online appendix (Section S2.2), I also provide bias and variance approximation of 2SLS with invalid instruments similar to approaches in Rothenberg (1984). Consider a model $y = W\delta_0 + \Psi\tau/\mu + v, W = \Psi\pi + u$, where $\mu^2 = \pi'\Psi'\Psi\pi/\sigma_u^2$ is a concentration parameter and δ_0 is a scalar. Under conventional asymptotics (μ is large, K is small), the bias of the 2SLS estimator can be approximated by

$$E(\widehat{\delta}_{2\mathrm{SLS}}(K) - \delta_0) \approx \frac{\sigma_{uv}}{\sigma_u^2}(\frac{K-2}{\mu^2}) + \frac{\sigma_v}{\sigma_u}\frac{\tilde{\mu}^2}{\mu^3}$$

where $\tilde{\mu}^2 = \pi'\Psi'\Psi\tau/(\sigma_u\sigma_v)$. The 2SLS bias depends on the strength (μ^2) and the number of instruments (K) as well as the invalidity of instruments ($\tilde{\mu}^2$). Although our theory does not rely on the weak instrument or many-weak instrument asymptotics, approximation above is useful to understand relative magnitude of bias due to many instruments and invalid instruments. For example, when $K/\mu \gg \tilde{\mu}^2/\mu^2$, then the first term dominates, and vice-versa.

Next, I give the MSE approximation for the LIML and FULL estimator. Unlike 2SLS, the order of dominating terms that depend on K for the LIML and FULL ($O_p(K/N)$) remain same as in Donald and Newey (2001). Note that the LIML and FULL estimator has the same approximate MSE to the order we consider here.

Let $\eta_i = u_i - v_i\sigma_{uv}/\sigma_v^2$ and $\Sigma_\eta = E[\eta_i\eta_i']$.

PROPOSITION 3.2. *If Assumptions 3.1, 3.2, 3.3 are satisfied,* $E[v_i^2\eta_i|x_i] = 0$*,* $K/N \to 0$*,* $\Sigma_\eta \neq 0, H_g \neq 0$ *and* $E[\|\xi_i\|^5|x_i], E[|v_i|^5|x_i]$ *are bounded, then the approximate MSE for the LIML and FULL estimator satisfies decomposition (3.1) with*

$$L(K) = H^{-1}[\sigma_v^2\Sigma_\eta\frac{K}{N} + \sigma_v^2\frac{f'(I-P^K)f}{N} + H_gH_g'H^{-1}\frac{f'(I-P^K)f}{N} + \frac{f'(I-P^K)f}{N}H^{-1}H_gH_g' - \frac{f'(I-P^K)g}{N}H_g' - H_g\frac{g'(I-P^K)f}{N}]H^{-1} \quad (3.4)$$

[4]A version of similar result can be also found in Lee and Okui (2012) where they derive the first-order asymptotic bias and variance of 2SLS under $K = O(N)$ with locally invalid IV in the proof of Theorem 4.

and

$$G = H^{-1}\big[(H_g\sigma'_{uv} + \sigma_{uv}H'_g)(\frac{1}{\sqrt{N}} - \frac{\sum_i f'_i H^{-1} f_i}{2N^{3/2}}) - \frac{1}{\sqrt{N}}(HH'_g H^{-1}\sigma_{uv} + \sigma'_{uv}H^{-1}H_g H)$$
$$- \sum_i [f_i H'_g H^{-1} f_i \sigma'_{uv} + \sigma_{uv} f'_i H^{-1} H_g f'_i + f_i \sigma'_{uv} H^{-1} f_i H'_g + H_g f'_i H^{-1}\sigma_{uv} f'_i]/N^{3/2}\big]H^{-1}.$$

REMARK 3.5. It is important to note that $G = O_p(1/\sqrt{N})$ do not depend on K, thus it will not be used for instrument selection criterion although it can be easily estimated. With possibly invalid instruments, the dominating term in MSE approximation (that depends on K) is order $O_p(K/N)$ which is same as those of Donald and Newey (2001). For the LIML or FULL, $L(K)$ does not include higher order bias from many instruments estimator, and the terms in $L(K)$ show higher-order variance trade-off with many invalid instruments. The third-moment condition $E[v_i^2\eta_i|x_i] = 0$ holds when $(v_i, \eta'_i)'$ is normally distributed, and this is imposed for the simplification similar to Donald and Newey (2001). Without this condition, $L(K)$ will have additional terms which are provided in proof of Proposition 3.2.

Finally, I provide a result for B2SLS estimator.

PROPOSITION 3.3. *If Assumptions 3.1, 3.2, 3.3 are satisfied,* $\sigma_{uv} \neq 0, H_g \neq 0, E[v_i^2 u_i|x_i] = 0$, $K/N \to 0$, *then the approximate MSE for the B2SLS estimator satisfies decomposition (3.1) with*

$$L(K) = H^{-1}\big[(\sigma_v^2\Sigma_\eta + 2\sigma_{uv}\sigma'_{uv})\frac{K}{N} + \sigma_v^2\frac{f'(I-P^K)f}{N} + H_g H'_g H^{-1}\frac{f'(I-P^K)f}{N}$$
$$+ \frac{f'(I-P^K)f}{N}H^{-1}H_g H'_g - \frac{f'(I-P^K)g}{N}H'_g - H_g\frac{g'(I-P^K)f}{N}\big]H^{-1}, \quad (3.5)$$

and

$$G = H^{-1}\big[(H_g\sigma'_{uv} + \sigma_{uv}H'_g)(\frac{d+3}{\sqrt{N}} - \frac{\sum_i f'_i H^{-1} f_i}{N^{3/2}}) - \frac{1}{\sqrt{N}}(HH'_g H^{-1}\sigma_{uv} + \sigma'_{uv}H^{-1}H_g H)$$
$$- \sum_i [f_i H'_g H^{-1} f_i \sigma'_{uv} + \sigma_{uv} f'_i H^{-1} H_g f'_i + f_i \sigma'_{uv} H^{-1} f_i H'_g + H_g f'_i H^{-1}\sigma_{uv} f'_i]/N^{3/2}\big]H^{-1}.$$

REMARK 3.6. Although $L(K)$ in MSE approximations for B2SLS is larger than those of LIML and FULL, it seems difficult to show the higher order efficiency of LIML or FULL estimator with locally invalid instruments because of different dominating term G in Propositions 3.2 and 3.3.

4. HIGHER-ORDER MSE RESULTS UNDER DIFFERENT RATES OF LOCALLY INVALID INSTRUMENTS

In this section, we consider faster rates of local-to-zero specification (i.e., a smaller degree of invalidity) than $N^{-1/2}$ rates considered in Sections 2 and 3. Under the rate we consider in this section, 2SLS, LIML, FULL, and B2SLS estimators are all consistent, but higher-order theory is still useful to capture changes in the order of the bias and variance that the first-order asymptotics cannot capture.

Although, we expect the stochastic order of higher-order bias and variance from invalid instruments become smaller than the terms due to the many instruments, the results generally depend not only on the drifting sequences but also on the specific rate of K. The key insight from the main results in this section (Propositions 4.1-4.3) is that the changes in the order of invalid instruments affect the finite sample behavior of IV estimators differently.

Specifically, I consider the following model with $\gamma > 1/2$

$$y_i = W_i'\delta_0 + \frac{g(x_i)}{N^\gamma} + v_i, \qquad E[v_i|x_i] = 0, \tag{4.1}$$

$$W_i = f(x_i) + u_i, \tag{4.2}$$

where all variables are defined same as in Section 2. Under the model (4.1)-(4.2) with $\gamma > 1/2$, I will find the following decomposition for IV estimator $\hat{\delta}(K)$,

$$\begin{aligned} &N(\widehat{\delta}(K) - \delta_0)(\widehat{\delta}(K) - \delta_0)' = \widehat{Q}(K) + \widehat{r}(K), \\ &E[\widehat{Q}(K)|X] = \sigma_v^2 H^{-1} + G + L(K) + T(K), \\ &[\widehat{r}(K) + T(K)]/tr(G + L(K)) = o_p(1), \quad K \to \infty, N \to \infty. \end{aligned} \tag{4.3}$$

We note that the first-order variance ($\sigma_v^2 H^{-1}$) is the only $O_p(1)$ term in conditional MSE approximations, and the bias from invalid instruments (i.e., $H^{-1}H_g H_g' H^{-1}$ in (3.1)) becomes the higher-order term. Moreover, the leading higher-order terms, G and $L(K)$, are different than in previous sections, and this requires separate analysis with modifications of the results in Propositions 3.1-3.3.

Following proposition provides a higher-order MSE approximation result for 2SLS estimators.

PROPOSITION 4.1. *Suppose Assumptions 3.1, 3.2 and 3.3 are satisfied with the model* (4.1)-(4.2). *If $K^2/N \to 0$, $\sigma_{uv} \neq 0, H_g\sigma_{uv}' \neq 0$, and $H_g \neq 0$, then the approximate MSE for the 2SLS estimator satisfies decomposition* (4.3) *with $G = \frac{1}{N^{2\gamma-1}} H^{-1}H_g H_g' H^{-1}$ and*

$$L(K) = H^{-1}\big[\frac{K}{N^\gamma}(H_g\sigma_{uv}' + \sigma_{uv}H_g') + \sigma_{uv}\sigma_{uv}'\frac{K^2}{N} + \sigma_v^2\frac{f'(I - P^K)f}{N}\big]H^{-1}. \tag{4.4}$$

If we further assume $\frac{K}{N^{1-\gamma}} \to \infty$, then (4.3) *holds with $G = 0$ and*

$$L(K) = H^{-1}\big[\sigma_{uv}\sigma_{uv}'\frac{K^2}{N} + \sigma_v^2\frac{f'(I - P^K)f}{N}\big]H^{-1}. \tag{4.5}$$

REMARK 4.1. In (4.4), G and $L(K)$ contain all four higher-order bias terms due to many and invalid instruments as well as a higher-order variance from many instruments. If we restrict the rate of K, the second result of Proposition 4.1 shows that the higher-order bias from many instruments dominates all other bias terms due to the invalid instruments, and $L(K)$ in (4.5) has the same form in Donald and Newey (2001). The assumption of the second result holds when $\gamma > 1 - \alpha$ ($K = N^\alpha$). This always holds when $\gamma \geq 1$, thus MSE approximations in Donald and Newey (2001) are still valid under the same rate conditions ($K^2/N \to 0$).

Proposition 4.1 shows that Donald and Newey (2001)'s MSE approximation for 2SLS estimator is robust to a small degree of invalid instruments. Moreover, without estimating H_g and $g(\cdot)$, their instrument selection criterion based on (4.5) also satisfy optimality results (Proposition 4 in Donald and Newey (2001)) under $\gamma \geq 1$. Nevertheless of these intuitive results, I quantify the robustness of the MSE approximation of 2SLS estimator in Donald and Newey (2001) and this is non-trivial as the dominating terms in $L(K)$ not only depend on the order of invalid instrument γ, but also the rates of K.

Next two results are for LIML/FULL and B2SLS estimators.

PROPOSITION 4.2. *Suppose Assumptions 3.1, 3.2 and 3.3 are satisfied with the model* (4.1)-(4.2). *Assume* $K/N \to 0, \Sigma_\eta \neq 0, H_g \neq 0, E[v_i^2 \eta_i | x_i] = 0$, *and* $E[\|\xi_i\|^5 | x_i], E[|v_i|^5 | x_i]$ *are bounded. Then the approximate MSE for the LIML or FULL estimator satisfies decomposition* (4.3) *with* $G = \frac{1}{N^{2\gamma-1}} H^{-1} H_g H_g' H^{-1}$, *and the following terms*

$$L(K) = H^{-1}\big[\sigma_v^2 \Sigma_\eta \frac{K}{N} + \sigma_v^2 \frac{f'(I-P^K)f}{N}\big] H^{-1}. \tag{4.6}$$

PROPOSITION 4.3. *Suppose Assumptions 3.1, 3.2 and 3.3 are satisfied with the model* (4.1)-(4.2). *Assume* $K/N \to 0, \sigma_{uv} \neq 0, H_g \neq 0$, *and* $E[v_i^2 u_i | x_i] = 0$. *Then the approximate MSE for the B2SLS estimator satisfies* (4.3) *with* $G = \frac{1}{N^{2\gamma-1}} H^{-1} H_g H_g' H^{-1}$,

$$L(K) = H^{-1}\big[(\sigma_v^2 \Sigma_\eta + 2\sigma_{uv}\sigma_{uv}') \frac{K}{N} + \sigma_v^2 \frac{f'(I-P^K)f}{N}\big] H^{-1}. \tag{4.7}$$

REMARK 4.2. Propositions 4.2 and 4.3 show that the leading term $L(K)$ in MSE approximation for LIML, FULL and B2SLS estimator is same as the leading term in Donald and Newey (2001) for all $\gamma > 1/2$. Note that MSE approximation still includes higher-order bias from locally invalid instruments G which do not depend on K.

Propositions 4.2 and 4.3 show that robustness of the MSE approximation (and instrument selection criterion) of LIML, FULL, B2SLS estimator in Donald and Newey (2001) under the presence of locally invalid instruments ($\gamma > 1/2$). With smaller higher-order bias from many instruments, MSE approximation in Donald and Newey (2001) for LIML, FULL, and B2SLS are robust to the wider range of γ than the 2SLS estimator.

5. INSTRUMENT SELECTION CRITERIA

In this section, I consider instrument selection criteria based on the MSE approximation in Propositions 4.1-4.3 which coincide to Donald and Newey (2001)'s criteria and show optimality property under $N^{-\gamma}$ ($\gamma > 1/2$) locally invalid instrument specification. Then, I propose *invalidity-robust* instrument selection criteria based on Propositions 3.1-3.3 under $\gamma = 1/2$.

To simplify the results, we consider a simple case with scalar endogenous regressor (i.e., Y_i is scalar) where covariates have already been partialled out.[5] For the general vector endogenous variables Y_i case, see the Online Appendix (Section S3).

We choose K to minimize $\widehat{L}(K)$ which is an estimate of $L(K)$ provided in MSE approximations, and this requires preliminary estimates of the model and goodness of fit criterion for the first-stage reduced form equation. Let $\widetilde{\delta}$ be some preliminary estimator, e.g., IV estimator where the instruments $\widetilde{K}$ are chosen to minimize cross-validation

[5]Specifically, from the original data, $(\widetilde{y}, \widetilde{Y}, \widetilde{X})$, let $y = M_{X_1}\widetilde{y}, Y = M_{X_1}\widetilde{Y}, X = M_{X_1}\widetilde{X}$ where $M_{X_1} = I - X_1(X_1'X_1)^{-}X_1'$ is the orthogonal projection matrix of exogenous covariates x_{1i}.

(CV) or Mallows (1973) criteria for the reduced form equation. Let $\widetilde{\varepsilon} = y - W\widetilde{\delta}$ as residuals, and let $\hat{H} = W'P^{\widetilde{K}}W/N$ as a preliminary estimator of $H = f'f/N$. Also, let $\widetilde{u} = (I - P^{\widetilde{K}})W$ as a residual vector of the first-stage reduced-form regression, and all preliminary estimates remain fixed while the criterion is calculated for different K.

We first consider following criteria based on $L(K)$ provided in Propositions 4.1-4.3 which were considered in Donald and Newey (2001),

$$\text{2SLS}: \ \widehat{L}_{DN}(K) = \hat{\sigma}_{uv}^2 \frac{K^2}{N} + \hat{\sigma}_v^2(\hat{R}(K) - \hat{\sigma}_u^2 \frac{K}{N}), \tag{5.1}$$

$$\text{LIML, FULL}: \ \widehat{L}_{DN}(K) = \hat{\sigma}_v^2(\hat{R}(K) - \frac{\hat{\sigma}_{uv}^2}{\hat{\sigma}_v^2}\frac{K}{N}), \tag{5.2}$$

$$\text{B2SLS}: \ \widehat{L}_{DN}(K) = \hat{\sigma}_v^2(\hat{R}(K) + \frac{\hat{\sigma}_{uv}^2}{\hat{\sigma}_v^2}\frac{K}{N}), \tag{5.3}$$

where $\hat{\sigma}_{uv} = \widetilde{u}'\widetilde{\varepsilon}/N, \hat{\sigma}_v^2 = \widetilde{\varepsilon}'\widetilde{\varepsilon}/N, \hat{\sigma}_u^2 = \widetilde{u}'\widetilde{u}/N$, and the Mallows' or CV criterion

$$\hat{R}(K) = \frac{\hat{u}^{K'}\hat{u}^K}{N} + 2\hat{\sigma}_u^2\frac{K}{N}, \quad \hat{R}(K) = \frac{1}{N}\sum_{i=1}^N \frac{(\hat{u}_i^K)^2}{(1 - P_{ii}^K)^2}$$

with residual vectors $\hat{u}^K = (I - P^K)W$.

To provide optimality of above criteria under locally invalid instruments, I impose following assumptions as in Donald and Newey (2001).

Assumption 5.1. *W_i is scalar, $\hat{\sigma}_v^2 - \sigma_v^2 = o_p(1), \hat{\sigma}_u^2 - \sigma_u^2 = o_p(1), \hat{\sigma}_{uv} - \sigma_{uv} = o_p(1)$, $\hat{H} - H = o_p(1)$, $\bar{H}^{-1}\sigma_{uv} \neq 0$, and $var(H^{-1}\eta_i) > 0$. Also assume,* $\sup_K \sup_i P_{ii}^K \xrightarrow{p} 0$, $E[u_i^8|x_i] < \infty$, $\inf_K NR(K) \to \infty$ *where* $R(K) = \sigma_u^2(K/N) + H^{-1}[f'(I - P^K)f/N]H^{-1}$.

Proposition 5.1. *Suppose Assumption 5.1 holds and the same assumptions as in Proposition 4.1 hold. For 2SLS estimator with $\widehat{K} = \arg\min_{K\in\mathcal{K}} \widehat{L}_{DN}(K)$, following holds for all $\gamma \geq 1$,*

$$\frac{L(\widehat{K})}{\inf_K L(K)} \xrightarrow{p} 1 \tag{5.4}$$

where $L(K)$ defined in (4.5).

For LIML (FULL) and B2SLS estimators, (5.4) *holds for all $\gamma > 1/2$ under the same assumptions as in Proposition 4.2 and Proposition 4.3, respectively.*

Proposition 5.1 provides optimality of instrument selection criteria in Donald and Newey (2001) under $\gamma > 1/2$ allowing same rates of K. While LIML (FULL) and B2SLS criteria are robust to all $\gamma > 1/2$, 2SLS criterion is only robust to $\gamma \geq 1$ and it requires restrictive rates of K to be robust to $1/2 < \gamma < 1$ by Proposition 4.1. This result suggests that instrument selection criteria in Donald and Newey (2001) are robust to a small degree of invalid instruments, but LIML (FULL) and B2SLS are more robust than 2SLS.

Next, we provide instrument selection criteria robust to invalid instruments based on Propositions 3.1-3.3 and show asymptotic unbiasedness of the criteria. Estimation of $L(K)$ requires some preliminary estimates of $g(x)$. I assume that we have some known to be valid instrument sets z_i. Note that our derivation of the MSE approximation in earlier sections does not require a valid instrument. The assumption of having a small number of valid instruments is also used in recent papers which address similar questions, e.g., Cheng and Liao (2015), DiTraglia (2016).

Instrument selection criteria robust to the validity of instruments are as follows;

$$
\begin{aligned}
\text{2SLS}: \widehat{L}_{IR}(K) &= 2\hat{H}_g\hat{\sigma}_{uv}\frac{K}{\sqrt{N}} + \hat{\sigma}_{uv}^2\frac{K^2}{N} \\
&\quad + (\hat{\sigma}_v^2 + \frac{2\hat{H}_g^2}{\hat{H}})(\hat{R}(K) - \hat{\sigma}_u^2\frac{K}{N}) - 2\hat{H}_g\hat{G}(K) \qquad (5.5) \\
\text{LIML}: \widehat{L}_{IR}(K) &= (\hat{\sigma}_v^2\hat{\sigma}_u^2 - \hat{\sigma}_{uv}^2)\frac{K}{N} + (\hat{\sigma}_v^2 + \frac{2\hat{H}_g^2}{\hat{H}})(\hat{R}(K) - \hat{\sigma}_u^2\frac{K}{N}) - 2\hat{H}_g\hat{G}(K) \quad (5.6) \\
\text{B2SLS}: \widehat{L}_{IR}(K) &= (\hat{\sigma}_v^2\hat{\sigma}_u^2 + \hat{\sigma}_{uv}^2)\frac{K}{N} + (\hat{\sigma}_v^2 + \frac{2\hat{H}_g^2}{\hat{H}})(\hat{R}(K) - \hat{\sigma}_u^2\frac{K}{N}) - 2\hat{H}_g\hat{G}(K) \quad (5.7)
\end{aligned}
$$

where $\hat{H}_g = W'P^{\widetilde{K}}\widehat{\varepsilon}/\sqrt{N}$, $\hat{G}(K) = W'(I - P^K)\widehat{\varepsilon}/\sqrt{N}$, $\widehat{\varepsilon} = y - W\widehat{\delta}$, and $\widehat{\delta}$ is preliminary estimator using valid instruments z_i. Note that $\widehat{L}_{IR}(K)$ reduce to $\widehat{L}_{DN}(K)$ in (5.1)-(5.3) when $\hat{H}_g = 0$.

PROPOSITION 5.2. *Suppose Assumptions 3.1, 3.2, 3.3, 5.1 are satisfied, $\sigma_{uv} \neq 0$, and assume that we have vector of instruments $z_i \in \mathbb{R}^q (q \geq p)$ such that $E[z_i\varepsilon_i] = 0$. Then $\widehat{L}_{IR}(K)$ given in* (5.5)-(5.7) *for the 2SLS, LIML (FULL), and B2SLS estimator satisfies following:*

$$E[\widehat{L}_{IR}(K)|X] = L(K) + \Sigma + o_p(1).$$

where $L(K)$ defined in Proposition 3.1-3.3, and Σ do not depend on K.

Proposition 5.2 implies that invalidity-robust instrument selection criterion is an asymptotically unbiased estimators of higher order approximation $L(K)$ in Propositions 3.1-3.3 up to a constant which does not depend on K, and $o_p(1)$ terms. Preliminary estimator $\widehat{\delta}$ using valid instruments affects the finite sample behavior of the invalidity-robust criterion as well as the choice of instruments. It would be desirable to justify $\widehat{L}_{IR}(K)$ in terms of optimality in Proposition 5.1, but this is a difficult problem as it requires to deal with estimation of $g(\cdot)$ which is not $\sqrt{n}$-estimable. However, simulation evidence suggests that invalidity-robust instrument selection criterion combined with IV estimator that has a small bias property under many instruments, such as LIML or Fuller estimator with constant $C = 1$, works well.

6. MONTE CARLO SIMULATION

We investigate the finite sample performance of the 2SLS, LIML, FULL, and B2SLS estimators based on instrument selection criteria in Donald and Newey (2001) and invalidity-robust criteria proposed in this paper.

We use a simple linear IV regression with potentially invalid instruments. The model to be estimated is

$$
\begin{aligned}
y_i &= x_i\beta_0 + \varepsilon_i \\
E(z_i\varepsilon_i) &= 0 \qquad (6.8)
\end{aligned}
$$

where x_i and β_0 are scalar and z_i is a $K \times 1$ vector of instrumental variables. We estimate

β_0 by 2SLS, LIML, Fuller (FULL) and bias-adjusted 2SLS (B2SLS) estimator with $\widehat{K}$ chosen by instrument selection criteria considered in this paper.

Our data-generating process (DGP) is

$$\begin{aligned} y_i &= x_i\beta_0 + \frac{\tau' Z_i}{N^\gamma} + v_i, \qquad (6.9)\\ x_i &= \pi' Z_i + u_i,\\ Z_i &\sim N\left(0, I_{\bar{K}}\right),\\ \begin{pmatrix} v_i \\ u_i \end{pmatrix} &\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & \sigma_{uv} - \pi'\tau/N^\gamma \\ \sigma_{uv} - \pi'\tau/N^\gamma & 1 \end{bmatrix}\right). \end{aligned}$$

We set $\beta_0 = 0.1$ and set the maximum number of instruments $\bar{K}$ as 20 when the number of observations $N = 100$ and 30 for $N = 500$. As in Donald and Newey (2001), we set the first-stage coefficient $\pi = (\pi_1, \cdots, \pi_{\bar{K}})', \pi_k = c(\bar{K})(1 - k/(\bar{K}+1))^4$ where $c(\bar{K})$ is chosen so that the first-stage $R^2 = 0.2$ or 0.02. We also set the endogeneity of x_i as $Cov(x_i, \varepsilon_i) = \sigma_{uv} = 0.5$. The number of Monte Carlo simulation is 10,000. In additional simulation results reported in the Supplementary Appendix, we also investigate the different specifications such as $\sigma_{uv} = 0.2, 0.8$, different π, τ, and a heteroskedastic setup.

If we fix $\tau \neq 0$, the key parameter is γ, and we vary $\gamma = \infty, 1, 2/3$ and $1/2$. When $\gamma = \infty$, any instruments z_i from the full set of instruments Z_i are valid. For $0 < \gamma < \infty$, we find $E(z_i\varepsilon_i) = \tau/N^\gamma \neq 0$, so the moment condition (6.8) fails to hold in any finite samples. The DGP can also be written as a globally misspecified model such that $\gamma = 0$ and τ is not that large, but we use the locally misspecified setup to be consistent with results in Sections 3 and 4.

We consider a following specification for $\tau = (\tau_1, \cdots, \tau_{\bar{K}})'$,

$$\tau_k = 0 \text{ for } k = 1, 2, \quad \tau_k = 1 \text{ for } k = 3, \cdots, \bar{K}/2, \quad \tau_k = 0 \text{ for } k > \bar{K}/2 + 1.$$

So we assume the first two instruments are known to be valid and it will be used for preliminary estimates in the invalidity-robust instrument selection criteria. We also consider the next "strong" IVs are invalid. This is empirically relevant as IV that is strongly correlated with the endogenous regressor is more likely to be correlated with the dependent variable, and there exists bias-variance trade-off of using invalid but relevant instruments. Moreover, the last half of "weak" IVs are valid, so there also exists trade-off of using valid, but weak instruments.

Median bias (Bias), interquartile range (IQR), and mean square error (MSE) of the 2SLS, LIML, FULL (with $C = 1$), B2SLS are reported in Tables 1 and 2. For all IV estimators, we consider four different instrument choices; using all available instruments (all), using the first two valid instruments (val), using instruments chosen by Donald and Newey (2001)'s criterion (DN), and using instruments selected by invalidity-robust criterion in this paper (IR). In addition, Table 3 reports a median of the selected number of instruments $\widehat{K}$ and optimal K^* that minimizes the true MSE of estimators.[6]

Table 1 shows that IV estimators based on Donald and Newey (2001) (DN) criterion have smaller IQR and MSE than the estimators using only valid instrument sets even

[6]For MSE, we compute trimmed mean square, $E[\min\{(\hat{\beta}-\beta)^2, 2\}]$ as in Okui (2011) due to the concerns on large outliers across the simulations. Results for MSE without trimming is available upon request, but results are qualitatively similar, and MSE for Fuller estimator is same as trimmed MSE. Results for Fuller estimator with constant $C = 4$ are similar with $C = 1$, thus omitted here.

when instruments are slightly invalid as well as when instruments are all valid ($\gamma = \infty$). This is consistent with our theory that DN criterion is robust to 'slightly invalid' instruments, and DN includes few more invalid but strong instruments. However, estimators with DN criterion may have large bias and this bias increases as γ decreases (the degree of invalidity increases) and does not disappear with larger sample sizes $N = 500$. Table 1 also shows estimators based on invalidity-robust criterion (IR) have lower median bias than estimators with DN criterion, and it achieves similar or smaller MSE when the degree of invalid instruments is large $\gamma = 1/2$. When $\gamma = 1/2$, DN criterion tends to choose more invalid instruments than the optimal K^*. The results are qualitatively similar for sample sizes of $N = 500$.

There is no unique ranking of estimators which clearly dominates the others in terms of bias, IQR, and MSE. However, Fuller estimator performs well in terms of lowest IQR and MSE. For FULL, the selected $\widehat{K}$ based on DN or IR is also close to the optimal K^*. LIML has the smallest median bias in many cases as LIML is known to be higher-order median unbiased. However, LIML has larger IQR as it is also known not to have any finite moments. Similar performances for LIML under valid instruments can be also found in Hahn et al. (2004). B2SLS performs poorly with IR criteria when $\gamma = 1/2$.

Results for the weak instrument case are reported in Table 2. Bias and MSE approximations, as well as instrument selection criteria may perform poorly in many weak instrument situations. However, Fuller estimator combined with DN and IR criteria performs surprisingly well; it has smaller MSE, IQR and median of selected $\widehat{K}$ is close to optimal K^* in many cases. Although it is highlighted in the literature that the Fuller estimator performs well under many weak instruments setup, our simulation suggests that Fuller estimator combined with instrument selection criterion can still be useful when instruments are potentially invalid.

7. CONCLUSIONS

This paper develops higher-order MSE approximation of the k-class estimators in the linear IV model with many and possibly invalid instruments. Based on the higher-order approximation, I propose instrument selection criteria that are robust to the locally invalid instruments. I also show that the optimality of instrument selection criteria in Donald and Newey (2001) robust to the certain type of locally invalid instrument specifications.

ACKNOWLEDGEMENTS

The author is grateful to Bruce Hansen for his thoughtful discussions. The author also acknowledges helpful conversations with Jack Porter, Xiaoxia Shi, Francis DiTraglia, Joachim Freyberger, Seojeong Lee, Adam McCloskey, and seminar participants at UW-Madison, MEG 2013. This paper is a revised version of the second chapter in my Ph.D. thesis at UW-Madison and the previous version was also circulated under the title "Choosing a Set of Instruments among Many and Possibly Invalid Instruments".

REFERENCES

Anatolyev, Stanislav and P. Yaskov (2017). Asymptotics of diagonal elements of projection matrices under many instruments/regressors. *Econometric Theory 33*, 717–738.

Anderson, T. W. and T. Sawa (1973). Distributions of Estimates of Coefficients of a Single

Equation in a Simultaneous System and Their Asymptotic Expansions. *Econometrica 41*, 683–714.

Andrews, D. W. K. (1999). Consistent Moment Selection Procedures for Generalized Method of Moments Estimation. *Econometrica 67*, 543–563.

Andrews, D. W. K. and B. Lu. (2001). Consistent Model and Moment Selection Procedures for GMM Estimation with Application to Dynamic Panel Data Models. *Journal of Econometrics 101*, 123–164.

Andrews I, Gentzkow M, Shapiro J.M. (2017). Measuring the Sensitivity of Parameter Estimates to Estimation Moments. *Quarterly Journal of Economics 132*, 1553–1592.

Armstrong, Tim and Kolesár, M (2018). Sensitivity Analysis using Approximate Moment Condition Models. Working Paper

Belloni, A., D. Chen, V. Chernozhukov and C. Hansen (2012). Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain. *Econometrica 80*, 2369–2430.

Berkowitz, D., M. Caner and Y. Fang (2008). Are Nearly Exogenous Instruments Reliable?. *Economics Letters 101*, 20–23.

Berkowitz, D., M. Caner and Y. Fang (2012). The Validity of Instruments Revisited. *Journal of Econometrics 166*, 255–266.

Bonhomme, S., M. Weidner (2018). Minimizing Sensitivity to Model Misspecification. Working Paper

Canay, I. (2010). Simultaneous Selection and Weighting of Moments in GMM Using a Trapezoidal Kernel. *Journal of Econometrics 156*, 284–303.

Caner, M. (2014). Near Exogeneity and Weak Identification in Generalized Empirical Likelihood Estimators: Many Moment Asymptotics. *Journal of Econometrics 182*, 247–268.

Caner, M., X. Han and Y. Lee (2018). Adaptive Elastic Net GMM Estimation With Many Invalid Moment Conditions: Simultaneous Model and Moment Selection. *Journal of Business and Economic Statistics 36*, 24–46.

Carrasco, M. (2012). A Regularization Approach to the Many Instruments Problem. *Journal of Econometrics 170*, 383–398.

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

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D.W. Schanzenbach, and D. Yagan (2011). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star. *Quarterly Journal of Economics 126*, 1593–1660.

Conley T. G., C. B. Hansen, and P. E. Rossi (2012). Plausibly Exogenous. *Review of Economics and Statistics 94*, 260–272.

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

Donald, S. G., G. W. Imbens, and W. K. Newey (2009). Choosing Instrumental Variables in Conditional Moment Restriction Models. *Journal of Econometrics 152*, 28–36.

Donald, S. G. and W. K. Newey (2001). Choosing the number of instruments. *Econometrica*, *69*, 1161–1191.

Fuller, W.A. (1977). Some Properties of a Modification of the Limited Information Estimator. *Econometrica*, *45*, 939–953.

Guggenberger, P. (2012). On the Asymptotic Size Distortion of Tests When Instruments Locally Violate the Exogeneity Assumption. *Econometric Theory*, *28*, 387–421.

Hahn, J. and J. Hausman (2005). Estimation with Valid and Invalid Instruments. *Annales d'Économie et de Statistique*, *79/80*, 25–57.

Hahn, J., J. Hausman, G. Kuersteiner (2004). Estimation with weak instruments: Accuracy of higher-order bias and MSE approximations. *Econometrics Journal*, *7*, 272–306.

Hall, A. R. and F. P. M. Peixe (2003). A Consistent Method for the Selection of Relevant Instruments *Econometric Reviews*, *22*, 269–287.

Hong, H., B. Preston, and M. Shum (2003). Generalized Empirical Likelihood-based Model Selection Criteria for Moment Condition Models. *Econometric Theory*, *19*, 923–943.

Kang, H., A. Zhang, T.T. Cai and D.S. Small, (2016). Instrumental Variables Estimation with some Invalid Instruments and its Application to Mendelian Randomization. *Journal of the American Statistical Association*, *111*, 132–144.

Kolesár, M., R. Chetty, J. Friedman, E. Glaeser, and G. W. Imbens (2015). Identification and Inference With Many Invalid Instruments. *Journal of Business and Economic Statistics*, *33*, 474–484.

Kraay, A. (2012). Instrumental Variables Regressions with Uncertain Exclusion Restrictions: A Bayesian Approach. *Journal of Applied Econometrics*, *27*, 108–128.

Kuersteiner, G. (2012). Kernel Weighted GMM Estimators for Linear Time Series Models. *Journal of Econometrics*, *170*, 399–421.

Kuersteiner, G. and R. Okui (2010). Constructing Optimal Instruments by First-Stage Prediction Averaging. *Econometrica*, *78*, 697–718.

Lee, Y. and R. Okui (2012). Hahn-Hausman test as a specification test. *Journal of Econometrics*, *167*, 133–139.

Lee, Y. and Y. Zhou (2015). Averaged Instrumental Variables Estimators. Working Paper

Liao, Z. (2013). Adaptive GMM Shrinkage Estimation with Consistent Moment Selection. *Econometric Theory*, *29*, 857–904.

Mallows, C. L. (1973). Some comments on C_p. *Technometrics*, *15*, 661–675.

Morimune, K. (1983). Approximate Distributions of k-Class Estimators when the Degree of Overidentifiability is Large Compared with the Sample Size. *Econometrica*, *51*, 821–841.

Nagar, A. L.(1959). The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations. *Econometrica*, *27*, 575–595.

Nevo, A. and A. Rosen (2012). Identification with Imperfect Instruments. *Review of Economics and Statistics*, *94*, 659–671.

Newey, W.K. (1985). Generalized method of moments specification testing. *Journal of Econometrics*, *29*, 229–256.

Okui, R. (2011). Instrumental Variable Estimation in the Presence of Many Moment Conditions. *Journal of Econometrics*, *165*, 70–86.

Otsu, T. (2011). Moderate deviations of generalized method of moments and empirical likelihood estimators. *Journal of Multivariate Analysis*, *102*, 1203–1216.

Rothenberg, Thomas J. (1984). Approximating the distributions of econometric estimators and test statistics. In Griliches, Z. and Intriligator, M. D. (Eds.), *Handbook of Econometrics, Volume 2*, 881–935. Elsevier.

van Hasselt, M. (2010). Many instruments asymptotic approximations under nonnormal error distributions. *Econometric Theory*, *26*, 633–645.

Windmeijer, F., H. Farbmacher, N. Davies, and G. D. Smith. (2018) On the Use of the Lasso for Instrumental Variables Estimation with Some Invalid Instruments. *Journal of the American Statistical Association*, *forthcoming.*

Table 1. Monte Carlo Results: $R^2 = 0.2$.

	$\gamma=\infty$			$\gamma=1$			$\gamma=2/3$			$\gamma=1/2$		
$N=100$	Bias	IQR	MSE	Bias	IQR	MSE	Bias	IQR	MSE	Bias	IQR	MSE
2SLS-all	0.217	0.356	0.066	0.233	0.356	0.074	0.280	0.366	0.099	0.355	0.390	0.150
2SLS-val	0.032	0.667	0.080	0.036	0.671	0.083	0.033	0.677	0.086	0.035	0.694	0.089
2SLS-DN	0.102	0.504	0.052	0.127	0.501	0.056	0.190	0.521	0.076	0.293	0.552	0.128
2SLS-IR	0.058	0.661	0.083	0.074	0.682	0.089	0.081	0.724	0.099	0.093	0.808	0.120
LIML-all	-0.002	0.757	0.134	0.033	0.743	0.126	0.123	0.825	0.171	0.262	1.278	0.377
LIML-val	0.000	0.741	0.109	0.002	0.755	0.115	-0.002	0.766	0.118	0.004	0.775	0.120
LIML-DN	0.051	0.554	0.058	0.075	0.549	0.061	0.136	0.579	0.078	0.231	0.704	0.147
LIML-IR	0.001	0.715	0.101	0.013	0.733	0.103	0.037	0.774	0.115	0.052	0.895	0.156
FULL-all	0.018	0.687	0.092	0.051	0.679	0.092	0.137	0.747	0.130	0.270	1.102	0.292
FULL-val	0.033	0.639	0.069	0.035	0.653	0.072	0.033	0.653	0.073	0.037	0.669	0.077
FULL-DN	0.116	0.439	0.042	0.133	0.439	0.047	0.184	0.459	0.067	0.263	0.554	0.119
FULL-IR	0.105	0.479	0.046	0.116	0.486	0.050	0.138	0.521	0.062	0.158	0.631	0.094
B2SLS-all	0.025	0.803	0.144	0.063	0.765	0.133	0.159	0.745	0.145	0.306	0.753	0.215
B2SLS-val	0.032	0.667	0.080	0.036	0.671	0.083	0.033	0.677	0.086	0.035	0.694	0.089
B2SLS-DN	0.079	0.551	0.069	0.098	0.547	0.069	0.145	0.557	0.077	0.228	0.623	0.113
B2SLS-IR	0.022	0.629	0.076	0.044	0.631	0.079	0.122	0.639	0.084	0.238	0.669	0.129

	$\gamma=\infty$			$\gamma=1$			$\gamma=2/3$			$\gamma=1/2$		
$N=500$	Bias	IQR	MSE	Bias	IQR	MSE	Bias	IQR	MSE	Bias	IQR	MSE
2SLS-all	0.094	0.196	0.014	0.102	0.197	0.016	0.147	0.193	0.027	0.249	0.195	0.067
2SLS-val	0.009	0.335	0.018	0.007	0.338	0.018	0.007	0.337	0.018	0.007	0.345	0.019
2SLS-DN	0.040	0.228	0.009	0.045	0.230	0.010	0.092	0.230	0.016	0.207	0.241	0.050
2SLS-IR	0.005	0.327	0.017	0.008	0.333	0.018	0.019	0.352	0.019	0.015	0.394	0.024
LIML-all	-0.000	0.252	0.010	0.009	0.253	0.010	0.066	0.246	0.014	0.182	0.266	0.043
LIML-val	0.001	0.344	0.019	-0.002	0.347	0.019	-0.001	0.348	0.019	-0.001	0.353	0.020
LIML-DN	0.012	0.236	0.009	0.019	0.238	0.009	0.071	0.233	0.013	0.180	0.245	0.041
LIML-IR	-0.021	0.316	0.018	-0.018	0.324	0.018	0.003	0.336	0.018	0.009	0.383	0.024
FULL-all	0.004	0.248	0.010	0.013	0.249	0.010	0.070	0.244	0.014	0.184	0.263	0.043
FULL-val	0.010	0.333	0.018	0.007	0.337	0.018	0.007	0.338	0.018	0.009	0.342	0.019
FULL-DN	0.027	0.225	0.008	0.033	0.227	0.009	0.084	0.223	0.014	0.189	0.235	0.043
FULL-IR	0.007	0.286	0.013	0.010	0.290	0.013	0.031	0.300	0.015	0.040	0.342	0.020
B2SLS-all	0.004	0.260	0.011	0.014	0.264	0.011	0.073	0.252	0.015	0.204	0.246	0.050
B2SLS-val	0.009	0.335	0.018	0.007	0.338	0.018	0.007	0.337	0.018	0.007	0.345	0.019
B2SLS-DN	0.019	0.236	0.009	0.023	0.237	0.009	0.068	0.240	0.013	0.183	0.250	0.041
B2SLS-IR	0.004	0.252	0.010	0.012	0.255	0.010	0.071	0.247	0.014	0.201	0.240	0.048

Note: (i) all - IV estimators using all instruments; (ii) val - using known valid instruments; (iii) DN - using instruments based on Donald and Newey (2001)'s criterion $\widehat{L}_{DN}(K)$; (iv) IR - based on the invalidity-robust criterion $\widehat{L}_{IR}(K)$.

Table 2. Monte Carlo Results: $R^2 = 0.02$.

	$\gamma = \infty$			$\gamma = 1$			$\gamma = 2/3$			$\gamma = 1/2$		
$N = 100$	Bias	IQR	MSE	Bias	IQR	MSE	Bias	IQR	MSE	Bias	IQR	MSE
2SLS-all	0.453	0.490	0.245	0.459	0.496	0.250	0.485	0.515	0.279	0.529	0.583	0.335
2SLS-val	0.289	1.788	0.522	0.293	1.865	0.541	0.290	1.843	0.536	0.289	1.918	0.551
2SLS-DN	0.381	1.278	0.441	0.399	1.303	0.451	0.442	1.299	0.481	0.510	1.467	0.573
2SLS-IR	0.308	2.240	0.615	0.309	2.243	0.625	0.339	2.230	0.634	0.356	2.359	0.679
LIML-all	0.278	4.413	0.917	0.332	4.320	0.912	0.483	5.041	1.049	0.922	9.498	1.465
LIML-val	0.207	3.684	0.803	0.217	3.752	0.816	0.207	3.787	0.820	0.199	3.797	0.829
LIML-DN	0.311	2.278	0.620	0.313	2.283	0.631	0.374	2.349	0.670	0.447	2.680	0.777
LIML-IR	0.247	2.885	0.720	0.258	2.821	0.734	0.291	3.018	0.755	0.334	3.327	0.829
FULL-all	0.321	2.008	0.590	0.361	1.964	0.592	0.481	2.155	0.722	0.787	2.950	1.094
FULL-val	0.335	1.140	0.311	0.337	1.161	0.323	0.334	1.162	0.319	0.328	1.191	0.325
FULL-DN	0.436	0.551	0.229	0.436	0.555	0.232	0.452	0.563	0.252	0.475	0.681	0.308
FULL-IR	0.430	0.577	0.234	0.435	0.579	0.238	0.449	0.607	0.260	0.475	0.738	0.323
B2SLS-all	0.408	3.275	0.821	0.422	3.195	0.822	0.484	3.519	0.882	0.573	3.789	0.965
B2SLS-val	0.289	1.788	0.522	0.293	1.865	0.541	0.290	1.843	0.536	0.289	1.918	0.551
B2SLS-DN	0.400	1.076	0.364	0.406	1.070	0.365	0.429	1.132	0.400	0.461	1.224	0.456
B2SLS-IR	0.307	1.885	0.545	0.334	1.924	0.557	0.372	2.034	0.603	0.447	2.136	0.666

	$\gamma = \infty$			$\gamma = 1$			$\gamma = 2/3$			$\gamma = 1/2$		
$N = 500$	Bias	IQR	MSE	Bias	IQR	MSE	Bias	IQR	MSE	Bias	IQR	MSE
2SLS-all	0.370	0.371	0.158	0.382	0.365	0.167	0.429	0.368	0.204	0.542	0.404	0.317
2SLS-val	0.107	1.131	0.248	0.111	1.148	0.249	0.102	1.138	0.246	0.103	1.169	0.257
2SLS-DN	0.262	0.668	0.150	0.280	0.663	0.156	0.358	0.637	0.190	0.530	0.661	0.344
2SLS-IR	0.108	1.143	0.268	0.118	1.162	0.272	0.133	1.178	0.273	0.160	1.341	0.328
LIML-all	0.050	1.910	0.484	0.080	1.926	0.494	0.255	2.062	0.560	0.872	5.606	1.243
LIML-val	0.028	1.624	0.402	0.027	1.664	0.411	0.018	1.630	0.403	0.024	1.623	0.408
LIML-DN	0.128	0.966	0.217	0.145	0.978	0.227	0.219	0.991	0.251	0.398	1.346	0.453
LIML-IR	0.029	1.404	0.347	0.036	1.439	0.353	0.060	1.432	0.352	0.116	1.818	0.477
FULL-all	0.088	1.386	0.320	0.115	1.413	0.332	0.277	1.513	0.410	0.815	2.610	1.014
FULL-val	0.126	0.948	0.163	0.124	0.959	0.163	0.120	0.955	0.163	0.123	0.967	0.173
FULL-DN	0.255	0.554	0.113	0.265	0.554	0.118	0.315	0.551	0.148	0.432	0.713	0.277
FULL-IR	0.236	0.597	0.114	0.244	0.600	0.119	0.266	0.605	0.136	0.318	0.845	0.247
B2SLS-all	0.146	2.106	0.533	0.180	2.090	0.537	0.316	1.878	0.542	0.638	2.110	0.805
B2SLS-val	0.107	1.131	0.248	0.111	1.148	0.249	0.102	1.138	0.246	0.103	1.169	0.257
B2SLS-DN	0.172	1.040	0.248	0.192	1.027	0.246	0.252	0.919	0.229	0.385	1.009	0.314
B2SLS-IR	0.069	1.225	0.276	0.083	1.218	0.274	0.155	1.194	0.281	0.329	1.286	0.377

Note: (i) all - IV estimators using all instruments; (ii) val - using known valid instruments; (iii) DN - using instruments based on Donald and Newey (2001)'s criterion $\widehat{L}_{DN}(K)$; (iv) IR - based on the invalidity-robust criterion $\widehat{L}_{IR}(K)$.

Table 3. Monte Carlo Results: Median of $\widehat{K}$.

			$\gamma=\infty$			$\gamma=1$			$\gamma=2/3$			$\gamma=1/2$		
N	R^2		K^*	DN	IR	K^*	DN	IR	K^*	DN	IR	K^*	DN	IR
100	0.2	2SLS	7	5	3	6	5	3	4	5	3	3	6	2
		LIML	5	5	3	9	5	3	3	5	3	2	5	2
		FULL	5	5	4	6	5	4	4	5	3	2	5	3
		B2SLS	2	3	6	2	3	6	3	4	5	2	4	5
500	0.2	2SLS	10	8	6	9	8	6	5	9	5	3	10	3
		LIML	11	11	8	12	11	8	6	11	6	3	11	3
		FULL	11	11	7	12	11	7	6	11	6	3	11	3
		B2SLS	11	7	20	11	7	20	6	7	20	3	9	19
100	0.02	2SLS	20	3	1	20	3	1	20	3	1	20	3	1
		LIML	19	1	1	15	1	1	5	1	1	4	2	1
		FULL	1	1	2	1	1	2	1	1	2	1	1	2
		B2SLS	1	1	3	1	1	3	1	1	3	1	1	3
500	0.02	2SLS	8	7	2	7	7	2	5	8	2	4	9	2
		LIML	6	4	2	11	4	2	11	4	2	4	4	2
		FULL	3	4	3	4	4	3	2	4	3	2	4	2
		B2SLS	1	3	4	1	3	4	1	3	4	1	3	4