Econometric Theory, **40**, 2024, 752–789.
doi:10.1017/S0266466622000597

HIGHER-ORDER APPROXIMATION OF IV ESTIMATORS WITH INVALID INSTRUMENTS

BYUNGHOON KANG
Lancaster University

This paper analyzes the higher-order approximation of instrumental variable (IV) estimators in a linear homoskedastic IV regression model when a large set of instruments with potential invalidity is present. We establish theoretical results on the higher-order mean-squared error (MSE) approximation of the two-stage least-squares (2SLS), the limited information maximum likelihood (LIML), the Fuller (FULL), the bias-adjusted 2SLS, and jackknife version of the LIML and FULL estimators by allowing for local violations of the instrument exogeneity conditions. Based on the approximation to the higher-order MSE, we consider the instrument selection criteria that can be used to choose among the set of available instruments. We demonstrate the asymptotic optimality of the instrument selection procedure proposed by Donald and Newey (2001, *Econometrica* 69, 1161–1191) in the presence of locally (faster than $N^{-1/2}$) invalid instruments in the sense that the dominant term in the MSE with the chosen instrument is asymptotically equivalent to the infeasible optimum. Furthermore, we propose instrument selection procedures to choose instruments among the sets of conservative (known) valid instruments and potentially locally ($N^{-1/2}$) invalid instruments based on the higher-order MSE of the IV estimators by considering the bias-variance trade-off.

1. INTRODUCTION

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 the instruments which we refer to as an exogeneity condition, i.e., the instruments are uncorrelated with the error term in the structural equation. Although researchers have routinely used the Sargan–Hansen J-test, the validity of instruments is generally uncertain; instruments may have direct effects on the

We thank the Editor (Peter Phillips), the Co-Editor (Patrik Guggenberger), and three anonymous referees for thoughtful comments that significantly improved the paper. 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, and NEM 2019. This paper is a revised version of the second chapter in the author's 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." Address correspondence to Byunghoon Kang, Department of Economics, Lancaster University, Lancaster, UK; e-mail: b.kang1@lancaster.ac.uk.

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 an IV estimator, higher-order approximation and instrument selection criteria have been found useful in the literature (e.g., Donald and Newey, 2001; Hahn, Hausman, and Kuersteiner, 2004; Kuersteiner and Okui, 2010), assuming that the instruments are valid.

This paper develops Nagar (1959)-type higher-order mean-squared error (MSE) approximations of the k-class estimators (including the two-stage least-squares [2SLS] estimator, the limited information maximum likelihood [LIML] estimator, the Fuller [FULL] estimator, and the bias-adjusted version of the 2SLS [B2SLS] estimator) in a linear homoskedastic IV model with many instruments while allowing for locally invalid instruments. Higher-order MSE approximations of general k-class estimators under local violations of instrument exogeneity include higher-order biases from many/invalid instruments as well as higher-order variances, and such a result is not available in the literature. MSE approximation depends not only on the number of instruments, but also on the degree of instrument invalidity and the relative strengths of the valid and invalid instruments. For the 2SLS estimator, 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).

Although the theoretical results in this paper are based on homoskedasticity assumption and the rate conditions $K/N \to 0$, we also provide higher-order MSE results for the estimators that are robust to heteroskedasticity and many instruments, such as the jackknife IV estimator (JIVE) and jackknife versions of the LIML and FULL (HLIM/HFUL) estimators considered in Hausman et al. (2012).[2] The higher-order results of the jackknife versions of the k-class estimators are new with or without invalid instruments, and they complement the results in the literature, such as Chao et al. (2012) and Hausman et al. (2012) (see Remarks 3.7 and 3.8).

The higher-order approximations in this paper hinge on $N^{-\gamma} (\gamma \geq 1/2)$ local-to-zero specification that allows for locally invalid instruments, and the local-misspecification approach has recently attracted considerable new attention (e.g., Conley, Hansen, and Rossi, 2012; Andrews, Gentzkow, and Shapiro, 2017; Armstrong and Kolesár, 2021; Bonhomme and Weidner, 2022).[3] This device

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

[2] When the number of instruments increases at the same rate as the sample size, i.e., $K/N \to \alpha, 0 < \alpha < 1$ (Bekker, 1994), the LIML, FULL, and B2SLS estimators are inconsistent with heteroskedasticity (Bekker and van der Ploeg, 2005; Hausman et al., 2012).

[3] Several important papers deal with estimation and inference issues involving local violations of the exogeneity conditions (e.g., Newey, 1985; Hahn and Hausman, 2005; Berkowitz, Caner, and Fang, 2008, 2012; Otsu, 2011; Guggenberger, 2012; Guggenberger and Kumar, 2012; Kraay, 2012; Caner, 2014, among many others).

https://doi.org/10.1017/S0266466622000597 Published online by Cambridge University Press

allows for the development of useful approximation theory for IV estimators with invalid instruments by focusing on a common structural parameter of interest, and this approach requires a nontrivial extension of Donald and Newey (2001; hereafter DN) because the dominating higher-order terms depend not only on the order of the invalid instruments (γ), but also on the rate of the number of instruments (K) as well as the estimators considered.

We show that the robustness of the higher-order MSE approximations (and instrument selection criteria) in DN under the presence of locally invalid instruments with $\gamma > \frac{1}{2}$ as the dominating terms that depend on K are the same as those of DN. We establish the asymptotic optimality of the instrument selection criteria in DN even with the presence of a small degree of invalid instruments ($\gamma > \frac{1}{2}$) in the sense that the dominant term in the MSE with the chosen $\widehat{K}$ achieves the infeasible optimum asymptotically (e.g., Li, 1987; see equation (5.5) for a formal definition). While the selection criteria for other k-class estimators are asymptotically optimal for all $\gamma > \frac{1}{2}$, in contrast, the criterion for the 2SLS estimator is optimal if $\gamma > 1-\alpha$ with $K = N^{\alpha}, \alpha < \frac{1}{2}$, i.e., when the degree of invalidity is sufficiently small (γ is sufficiently large).

We also consider invalidity-robust (IR) instrument selection criteria based on a higher-order approximation with $\gamma = \frac{1}{2}$; these criteria include higher-order bias/variance terms due to invalid instruments. In the presence of $N^{-1/2}$ locally invalid instruments, instrument selection criteria without additional terms in the MSE may lead to a misleading balance between bias and efficiency; the inclusion of invalid instruments that are strong first-stage predictors can reduce the MSE, although this may slightly increase the bias. Since the MSE in this case contains terms that cannot be estimated in the absence of valid instruments, the proposed criteria require a known set of valid instruments. Such criteria can be useful when researchers have a "conservative" set of valid instruments and explore all other candidate instruments that are potentially invalid considering the bias–variance trade-off induced by including them.[4] We show that the IR criterion is an asymptotically unbiased estimator of the dominant terms in the higher-order MSE approximation (Proposition 5.2).

In this paper, we also investigate the finite-sample properties of IV estimators (with or without instrument selection) under potentially invalid instruments with various Monte Carlo experiments. Our main findings are in line with those in the literature (Hahn et al., 2004; Guggenberger, 2008; Hausman et al., 2012), who recommend utilizing estimators with finite-sample moments instead of "no-moment" estimators. The interdecile range and root MSE of the LIML/HLIM and JIVE estimators are considerably larger than those of the FULL/HFUL and

[4] Some recent papers in the generalized method of moments (GMM) setup also assume that there is a subset of moment conditions that are known to be valid. When the correctly specified moment conditions identify a parameter of interest, Liao (2013) and Cheng and Liao (2015) consider consistent moment selection methods based on shrinkage estimation. In DiTraglia (2016), valid moment conditions are used to provide an asymptotically unbiased estimator of the first-order asymptotic MSE, which is similar to our paper.

https://doi.org/10.1017/S0266466622000597 Published online by Cambridge University Press

(overidentified) 2SLS estimators with many instruments regardless of instrument invalidity γ, especially in the weakly identified cases.[5]

Although it is highlighted in the literature that the Fuller and HFUL estimators perform well under many weak instrument setups, our simulation evidence further suggests that these estimators combined with instrument selection procedures can perform well even when the instruments are potentially invalid. We find that the FULL/HFUL estimators combined with the DN criterion can lead to a reduction in the MSE compared with that obtained by the estimators when using all instrument sets or only valid instruments even when instruments are slightly invalid and the correct specification case. This is expected from the theory in Proposition 5.1, which shows that the DN criteria are asymptotically optimal when we suspect a small degree of invalid instruments ($\gamma > \frac{1}{2}$). The FULL/HFUL estimators based on IR criteria have similar or slightly larger MSEs than those of the DN criterion across different values of γ. 2SLS also performs well, and the median bias of 2SLS can be lower than those of the other estimators with invalid instruments because the misspecification bias and the many-instrument bias can have opposite signs so that they offset each other, as expected from the theory (Remark 3.4).

It may be true that the MSE approximations derived in this paper may not provide good practical guidance for "no-moment" estimators, especially in the weak instrument scenarios (Hahn et al., 2004). Although this paper makes no theoretical contribution with regard to the "no-moments problem," in some simulation evidence, we find that instrument selection and model averaging can mitigate the moment problem for "no-moment" estimators, such as LIML, in terms of a significant reduction in the trimmed MSE and median absolute deviation (MAD).

The literature on higher-order approximations of k-class estimators with valid instruments has a long history, such as Nagar (1959), Anderson and Sawa (1973), Morimune (1983), and Rothenberg (1984). Phillips (1980) provides the exact distribution theory for the IV estimators in the general simultaneous equations models and a higher-order expansion of that distribution using the Laplace approximation, which can be considerably more accurate than the Edgeworth expansions. The exact distribution theory remains valid even when no moments are finite and also provides more information on the distribution of IVs than the MSEs (see Phillips, 1983, for an excellent review of the exact small sample theory of the IV estimators). Newey and Smith (2004) derive the higher-order asymptotic properties of the GMM and the generalized empirical likelihood (EL) estimators under correct specifications. Schennach (2007) proposes an exponentially tilted EL estimator that is robust under globally misspecified models while achieving the same higher-order properties of the EL estimator under correct specifications.

[5] It is well known that the LIML estimator has no finite moments (Mariano and Sawa, 1972) and that the 2SLS estimator has moments up to the degree of overidentification (Kinal, 1980; Phillips, 1980). The Fuller (1977) modifications of the LIML and HFUL estimators have finite-sample moments (Hausman et al., 2012). For models with general moment conditions, see also Hausman et al. (2011), who provide some modifications for the continuous updating estimator (CUE) to solve the moment problems associated with the CUE.

https://doi.org/10.1017/S0266466622000597 Published online by Cambridge University Press

Based on higher-order approximations, there are many papers that study the instrument (moment) and/or weight selections in the IV and GMM setups (e.g., DN; Donald, Imbens, and Newey, 2009; Canay, 2010; Kuersteiner and Okui, 2010; Okui, 2011; Carrasco, 2012; Kuersteiner, 2012; Lee and Zhou, 2015, among others). Many papers have also developed moment selection procedures to select valid moments from the sets of valid and invalid moment conditions (e.g., Andrews, 1999; Andrews and Lu, 2001; Hall and Peixe, 2003; Hong, Preston, and Shum, 2003; Liao, 2013). DiTraglia (2016) develops a moment selection criterion based on the first-order asymptotic MSE with possible locally invalid moment conditions in GMM setups.

Several important papers have investigated the asymptotic properties of IV and GMM estimators in misspecified moment condition models, such as Maasoumi and Phillips (1982) and Hall and Inoue (2003). Kitamura, Otsu, and Evdokimov (2013) propose an estimator that achieves optimal robust minimax properties under local model perturbations. For locally misspecified moment condition models, Andrews et al. (2017) propose a measure of the sensitivity of parameter estimates for minimum distance estimators, and Armstrong and Kolesár (2021) develop a method for constructing valid confidence intervals that are robust to misspecification. Andrews (2019) characterizes the estimands of the linear GMM under global misspecification. In an IV model with heterogeneous treatment effects, the moment condition is misspecified, and a 2SLS estimand can be characterized as a weighted average of the local average treatment effects (LATEs) (see Imbens and Angrist, 1994). Under treatment effect heterogeneity, Kolesár (2013) shows that LIML estimands may be outside of the convex hulls of the individual treatment effects and Evdokimov and Kolesár (2019) describe the estimands of the 2SLS and JIVE estimators with many instruments/covariates and provide valid asymptotic variance formulas.

Asymptotic optimality results (Proposition 5.1) require prior knowledge regarding the order of the instrument strengths, similar to DN; however, many recent papers, such as Belloni et al. (2012), Cheng and Liao (2015), and Caner, Han, and Lee (2018), have developed estimation and moment selection techniques (e.g., Lasso) in high-dimensional setups without requiring an order of the instruments. Kang et al. (2016) propose Lasso-type methods to identify and select valid instruments, and these approaches do not require set of instruments that are known to be valid. Windmeijer et al. (2018) determine that Lasso procedures may not consistently select the invalid instruments if they are relatively strong, and the authors propose a median-type estimator that is consistent when more than 50% of the instruments are valid.

The outline of this paper is as follows. Section 2 introduces the basic model setup and assumptions. Section 3 provides the theoretical results on the higher-order MSE approximations of IV estimators with $\gamma = \frac{1}{2}$. Section 4 establishes the MSE approximations under different local sequences of invalid instruments ($\gamma > \frac{1}{2}$). Section 5 provides the asymptotic optimality of instrument selection according to the DN criteria and proposes IR instrument selection criteria. Section 6 includes

https://doi.org/10.1017/S0266466622000597 Published online by Cambridge University Press

the simulation results obtained under various Monte Carlo settings, and Section 7 concludes the paper. All proofs, additional simulations, and auxiliary results are provided in the Supplementary Material.

2. THE MODEL AND ESTIMATORS

We consider a linear IV model allowing for potentially invalid instruments:

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

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

$$\varepsilon_i = \frac{g(x_i)}{N^\gamma} + v_i, \qquad \mathbb{E}[v_i|x_i] = 0, \quad \text{for } i = 1, \ldots, N, \tag{2.3}$$

where y_i is a scalar outcome variable and 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 to be fixed, and they do not depend on the sample size N. Note that all statements and conditions involving conditional expectations (conditional on $X = [x_1, \ldots, x_N]'$) hold with probability approaching one (w.p.a.1.). We suppress the subscript N hereafter and omit the "w.p.a.1." to simplify the notation. Although our results are based on the homoskedastic assumption of error terms (v_i, ξ_i'), the presence of invalid instruments leads to heteroskedastic errors in the estimation equation, i.e., $\mathbb{E}[\varepsilon_i^2|x_i]$ depends on x_i for any finite N.

Let $\psi_i^K \equiv \psi^K(x_i) = (\psi_{1K}(x_i), \ldots, \psi_{KK}(x_i))'$ be a $K \times 1 (K \geq p)$ vector of IVs (or basis functions), and we assume that ψ_i^K contains the included exogenous variables x_{1i}. Here, K indicates both the number of instruments and the index of the instrument sets. Different groups of instruments are allowed for different K, and the potential candidate instrument set does not necessarily have to be a nested set. That is, we allow different approximating functions in the series (sieves) terms or qualitatively different instruments to be used for different Ks. We further note that the growth rate of K is restricted ($K/N \to 0$) in this paper.

Although the theoretical results regarding the asymptotic MSE below do not need such modifications, we restrict the rate of the total number of candidate sets for the asymptotic optimality of the instrument selection criteria, similar to DN. In particular, we do not allow for the consideration of all possible combinations of instrument sets. Having prior knowledge of the order of instruments may reduce the number of candidate sets and greatly reduce the computational cost in practice. In general, researchers may not have prior beliefs about which instrument groups have large impacts on the reduced-form equation. However, in certain setups, it may be natural to have this ordering, e.g., including lower orders first for a power series approximation or including main instruments first among vectors of main instruments and various interaction terms.

https://doi.org/10.1017/S0266466622000597 Published online by Cambridge University Press

We allow for possibly invalid instruments in (2.3) and consider a local-to-zero specification by defining

$$\mathbb{E}[\varepsilon_i|x_i] = g_N(x_i) = \frac{g(x_i)}{N^\gamma}, \quad \text{for } \gamma \geq 1/2. \tag{2.4}$$

Under this setup, any potential instruments $\psi(x_i)$ are asymptotically valid as $N \to \infty$, but $\mathbb{E}[\psi(x_i)\varepsilon_i] = 0$ does not necessarily hold in finite samples. Note that the direct effects of the instruments $g(x_i)$ can be large numbers for any finite N, and we do not restrict the functional form of $g(x_i)$.

Although our framework deals with mixed drifting sequences, it is important to deal with the cases where $\gamma = \frac{1}{2}$ and $\gamma > \frac{1}{2}$ separately in a Nagar (1959)-type approximation to distinguish $O_p(1)$, the leading higher-order terms, and the remaining terms that are of smaller orders. With the knife-edge rate $\gamma = \frac{1}{2}$ (Section 3), the stochastic order of the biases of the IV estimators from an invalid instrument is equal to the standard deviation. This rate provides the right balance for a useful theory to understand the finite-sample behaviors of IV estimators with possibly invalid instruments. When $\gamma > \frac{1}{2}$ (Section 4), a different analysis is required because the dominating term in the higher-order MSE approximation changes.

Remark 2.1. We may consider $\gamma = 0$ (global misspecification or globally invalid instruments) or $0 < \gamma < \frac{1}{2}$, where the bias from invalid instruments dominates. We can still provide MSE approximations centered with a pseudo-true value which is the probability limit of an IV estimator (or a sequence of pseudo-true values depending on N), e.g., instrument-specific LATE parameters for 2SLS in the presence of a heterogeneous treatment effect. However, different choices of instruments and estimators can lead to different pseudo-true values regardless of misspecification; even under a correct specification (without invalid instruments), IV estimators can have different probability limits in the presence of many instruments. Under (global or local) misspecification, the problem can be more severe, and pseudo-true values are generally difficult to characterize in misspecified models, which makes MSE comparisons difficult across different estimators. Here, we focus on the common parameter of interest δ_0 and compare the MSEs among different instruments K as well as different IV estimators with $\gamma \geq \frac{1}{2}$. With globally invalid instruments, Nevo and Rosen (2012), Kolesár et al. (2015), and Kang et al. (2016) provide (set or point) identification results with respect to δ_0 in an IV framework.

Now, we consider several k-class estimators that are widely used in linear IV models. We first consider the 2SLS estimator:

$$\widehat{\delta}_{2\text{SLS}}(K) = (W'P^K W)^{-1}(W'P^K y), \tag{2.5}$$

where $y = (y_1,\ldots,y_N)', W = [Y, X_1], Y = [Y_1,\ldots,Y_N]', X_1 = [x_{11},\ldots,x_{1N}]'$, and $P^K = \Psi^K(\Psi^{K'}\Psi^K)^-\Psi^{K'}$ is the projection matrix for the instrument vector $\Psi^K = [\psi_1^K,\ldots,\psi_N^K]'$.

https://doi.org/10.1017/S0266466622000597 Published online by Cambridge University Press

Next, we consider the LIML estimator:

$$\widehat{\delta}_{\text{LIML}}(K) = (W'P^K W - \hat{\Lambda}(K)W'W)^{-1}(W'P^K y - \hat{\Lambda}(K)W'y), \tag{2.6}$$

where

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

LIML is known to be median unbiased to the second-order and is also known not to have finite-sample moments. We next consider the Fuller (1977) estimator (FULL), which solves the nonexistence of moments in LIML:

$$\widehat{\delta}_{\text{FULL}}(K) = (W'P^K W - \check{\Lambda}(K)W'W)^{-1}(W'P^K y - \check{\Lambda}(K)W'y), \tag{2.7}$$

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 the minimum MSE property. Next, we consider the bias-adjusted 2SLS estimator (B2SLS) from DN as a modification of the Nagar (1959) estimator with $\bar{\Lambda}(K) = (K-d-2)/N$:

$$\widehat{\delta}_{\text{B2SLS}}(K) = (W'P^K W - \bar{\Lambda}(K)W'W)^{-1}(W'P^K y - \bar{\Lambda}(K)W'y). \tag{2.8}$$

Finally, we consider the following jackknife-version k-class estimators of the form:

$$\widehat{\delta}(K) = \left(W'P^K W - \sum_{i=1}^{n} P_{ii}^K W_i W_i' - \bar{\lambda}(K)W'W\right)^{-1}\left(W'P^K y - \sum_{i=1}^{n} P_{ii}^K W_i y_i - \bar{\lambda}(K)W'y\right), \tag{2.9}$$

where P_{ii}^K denotes the diagonal elements of P^K. $\bar{\lambda}(K) = 0$ corresponds to the JIVE2 estimator with $\widehat{\delta}_{\text{JIVE2}}(K)$ considered in Angrist, Imbens, and Krueger (1999), Newey and Windmeijer (2009), and Chao et al. (2012), among others. When $\bar{\lambda}(K) = \hat{\lambda}(K)$,

$$\hat{\lambda}(K) = \min_{\delta} \frac{(y-W\delta)'(P^K - D^K)(y-W\delta)}{(y-W\delta)'(y-W\delta)},$$

where $D^K = diag(P_{ii}^K)$ is a diagonal matrix, $\widehat{\delta}(K) = \widehat{\delta}_{\text{HLIM}}(K)$ is a jackknife version of the LIML estimator that is robust to heteroskedasticity (HLIM; considered in Hausman et al., 2012).[6] When $\bar{\lambda}(K) = \frac{\hat{\lambda}(K)-C/N(1-\hat{\lambda}(K))}{1-C/N(1-\hat{\lambda}(K))}$, $\widehat{\delta}(K) = \widehat{\delta}_{\text{HFUL}}(K)$ is a heteroskedasticity robust version of the Fuller (1977) (HFUL) estimator, which is also considered in Hausman et al. (2012). Hausman et al. (2012) recommend $C = 1$ for the HFUL estimator.

[6] HLIM can be equivalently defined with $\hat{\lambda}(K)$ as the smallest eigenvalue of $\left(\overline{W}'\overline{W}\right)^{-1}\left(\overline{W}'P\overline{W} - \sum_{i=1}^{n} P_{ii}^K \overline{W}_i \overline{W}_i'\right)$ with $\overline{W} = [y, W]$.

https://doi.org/10.1017/S0266466622000597 Published online by Cambridge University Press

2.1. Assumptions and Higher-Order MSEs

We derive the Nagar (1959)-type higher-order asymptotic MSEs for the IV estimators with locally invalid instrument setups. As in Nagar (1959) and Rothenberg (1984), a higher-order MSE can be calculated from the first few terms of the stochastic expansion of an estimator. Even when some IV estimators do not possess finite moments, the moments of an approximating distribution can be defined, and the calculated moments can be interpreted as "pseudomoments" (see Pfanzagl and Wefelmeyer, 1978; Sargan, 1982; Phillips, 2003, for detailed discussions).

We consider conditional (with respect to the exogenous variables $X = [x_1,\dots,x_N]'$) MSEs for the IV estimators $\widehat{\delta}(K)$ and find a decomposition with the following form:

$$\begin{aligned}
&N(\widehat{\delta}(K)-\delta_0)(\widehat{\delta}(K)-\delta_0)' = \widehat{Q}(K)+\widehat{r}(K),\\
&\mathbb{E}[\widehat{Q}(K)|X] = \Phi + G + L(K) + T(K),\\
&[\widehat{r}(K)+T(K)]/tr(G+L(K)) = o_p(1), \quad K\to\infty, N\to\infty, \qquad \textbf{(2.10)}
\end{aligned}$$

where the main terms Φ, G, and $L(K)$ in (2.10) will be defined later in Sections 3 and 4 as they have different forms for each IV estimator and γ.

The dominant term in (2.10) is Φ, which is $O_p(1)$, and it does not depend on K in our large-K approximation. Furthermore, note that different γ values lead to changes in the order of the stochastic terms, and thus this requires a separate analysis. For example, when $\gamma = \frac{1}{2}$, Φ includes the first-order asymptotic variance and the square of the asymptotic bias from the locally invalid instrument. However, when $\gamma > \frac{1}{2}$, the bias from invalid instruments becomes the higher-order term, so Φ only includes the first-order variance term. It is important to note that Φ is omitted in the approximate MSE criteria (Section 5) to be used for selecting K, as it is the same for all estimators and does not depend on K.

G and $L(K)$ are the next leading terms, which include the higher-order bias and variance due to the presence of many invalid instruments. G includes terms that do not depend on K, and the instrument selection criteria are based only on $L(K)$, which is the leading higher-order term that depends on K in the conditional MSE approximation (2.10). $L(K)$ contain the important higher-order terms for all the subsequent analyses in the following sections. $\widehat{r}(K)$ and $T(K)$ are the remainder terms that converge to 0 faster than $G+L(K)$.

We impose the following assumptions, similar to DN. Let $f_i = f(x_i)$ and $g_i = g(x_i)$.

Assumption 2.1. (a) Let $\{(v_i, Y_i, x_i) : i = 1,\dots,N; N \geq 1\}$ be i.i.d. random vectors satisfying the models (2.1)–(2.3). (b) $\mathbb{E}[v_i^2|x_i] = \sigma_v^2 > 0$, and $\mathbb{E}[\|\xi_i\|^4|x_i]$, $\mathbb{E}[|v_i|^4|x_i]$ are finite.

Assumption 2.2. (a) $\bar{H} = \mathbb{E}[f_i f_i']$ exists and is nonsingular, and $\bar{H}_g = \mathbb{E}[f_i g_i]$ exists. (b) There exist π_K, π_K^g such that $\mathbb{E}[\|f(x)-\pi_K\psi^K(x)\|^2] \to 0$ and $\mathbb{E}[|g(x)-\pi_K^g\psi^K(x)|^2] \to 0$ as $K\to\infty$.

https://doi.org/10.1017/S0266466622000597 Published online by Cambridge University Press

Assumption 2.3. (a) $\mathbb{E}[(v_i,\xi_i')'(v_i,\xi_i')|x_i]$ is constant. (b) $\Psi^{K'}\Psi^{K}$ is nonsingular w.p.a.1. (c) $max_{i\leq N}P^K_{ii}\xrightarrow{p}0$. (d) f_i and g_i are bounded.

Assumption 2.1(b) imposes homoskedasticity and boundedness of the fourth conditional moments of the error terms. Assumption 2.2(a) is imposed for a usual identification assumption and the existence of first-order bias from invalid instruments. Assumption 2.2(b) requires the mean-squared approximation error of the unknown $f(x)$ and $g(x)$ by the linear combination of instruments $\psi^K(x)$ to go to 0 as the number of instruments increases. Assumption 2.3 imposes homoskedasticity and restricts the growth rate of K. For example, $K=O(N)$, as in Bekker (1994), is not allowed under Assumption 2.3(c) (see van Hasselt, 2010; Anatolyev and Yaskov, 2017, and the references therein). For the data-generating process (DGP), we consider in the Monte Carlo simulation (Section 6, equation (6.11)), we can easily verify Assumptions 2.2 and 2.3. For example, Assumption 2.2(b) is automatically satisfied with linear specifications of $f(\cdot)$ and $g(\cdot)$, and Assumption 2.3(b) is satisfied using the standard arguments in the nonparametric series regression literature (e.g., Newey, 1997).

3. HIGHER-ORDER MSE RESULTS WITH $\gamma=\frac{1}{2}$

This section provides higher-order MSE approximations of the k-class estimators, including 2SLS, LIML, FULL, and B2SLS, by allowing locally invalid instruments when $\gamma=\frac{1}{2}$. Some of the results below extend the results presented in DN, and the results on the higher-order MSEs of the jackknife versions of k-class estimators (JIVE2, HLIM, and HFUL) in the homoskedastic linear IV setup are new to the literature with (or without) locally invalid instruments.

Our first result gives the MSE approximation for the 2SLS estimator. Proposition 3.1 is a generalization of the result in DN that allows possibly (locally) invalid instruments. Let $H=f'f/N, H_g=f'g/N, f=[f_1,\dots,f_N]', g=[g_1,\dots,g_N]'$, $\sigma_{uv}=\mathbb{E}[u_iv_i|x_i]$, and $\sigma_v^2=\mathbb{E}[v_i^2|x_i]$.

Proposition 3.1. *If Assumptions 2.1–2.3 are satisfied with* $\gamma=\frac{1}{2}$, $\sigma_{uv}\neq 0$, $H_g\neq 0$, *and* $K^2/N\to 0$, *then the approximate MSE for the 2SLS estimator satisfies the decomposition (2.10) with* $\Phi=\sigma_v^2H^{-1}+H^{-1}H_gH_g'H^{-1}, G=0$, *and the following terms:*

$$L(K)=H^{-1}\Big[\frac{K}{N^{1/2}}(H_g\sigma_{uv}'+\sigma_{uv}H_g')+\sigma_{uv}\sigma_{uv}'\frac{K^2}{N}+\sigma_v^2\frac{f'(I-P^K)f}{N}+L_g(K)\Big]H^{-1}, \tag{3.1}$$

where

$$L_g(K)=H_gH_g'H^{-1}\frac{f'(I-P^K)f}{N}+\frac{f'(I-P^K)f}{N}H^{-1}H_gH_g' -\frac{f'(I-P^K)g}{N}H_g'-H_g\frac{g'(I-P^K)f}{N}.$$

https://doi.org/10.1017/S0266466622000597 Published online by Cambridge University Press

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

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

With invalid instruments ($H_g \neq 0$), it is important to point out that the dominating term of $L(K)$ in Proposition 3.1 is proportional to $K/\sqrt{N}$, and it comes from the cross-product of the bias from many instruments and invalid instruments. Differently from DN, this higher-order bias term dominates the squared bias from many instruments, which is proportional to K^2/N since $K^2/N = o(K/\sqrt{N})$ as $K \to \infty$.

Remark 3.1. The MSE approximation results in Proposition 3.1 include higher-order terms from many instruments as well as additional terms due to invalid instruments. When $H_g = 0$ w.p.a.1., it can be shown that Proposition 3.1 reduces to Proposition 1 in DN:

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

Note that the condition $H_g \xrightarrow{p} 0$ is closely related to the identification assumption in Kolesár et al. (2015, Assumps. 4 and 5) for the consistency of k-class estimators under many (global) invalid instrument setups. Interestingly, $H_g \xrightarrow{p} 0$ not only holds when $g(x) = 0$ (i.e., exclusion restriction holds), but also holds when the direct effects of the instruments on the outcome variable are orthogonal to their effects on the endogenous variable with a random effect structure, as in Kolesár et al. (2015). Consider a linear model $f = \Psi\pi, g = \Psi\tau$, and suppose that we treat the coefficients π, τ as random vectors (conditional on x_i). As in Kolesár et al. (2015), if we orthogonalize the coefficient $(\tilde{\tau}, \tilde{\pi}) = (\alpha\Psi'\Psi)^{1/2}(\tau, \pi), \alpha = K/N$ and assume that the pairs $(\tilde{\tau}_k, \tilde{\pi}_k)$ for $k = 1, \ldots, K$ are i.i.d. random vectors with mean (μ_τ, μ_π) and variance–covariance Ξ, then $H_g = \pi'\Psi'\Psi\tau/N = \sum_{k=1}^{K}\tilde{\pi}_k\tilde{\tau}_k/K \xrightarrow{p} \mu_\tau\mu_\pi + \Xi_{12}$, and the identification condition $H_g \xrightarrow{p} 0$ holds when $\mu_\tau = 0$ and $\Xi_{12} = 0$ (covariance of $\tilde{\pi}_k$ and $\tilde{\tau}_k$). Kolesár et al. (2015) discuss an empirical example such as that in Chetty et al. (2011), where $H_g \xrightarrow{p} 0$ may be a reasonable assumption.

The leading higher-order terms $L(K)$ can be reduced to those of DN by allowing $g(x) \neq 0$, and this observation suggests that the instrument selection criteria in DN can be shown to be asymptotically optimal even with possibly invalid instruments when $\gamma = \frac{1}{2}$, i.e., the dominant terms in the MSE with the selected $\widehat{K}$ are asymptotically equivalent to the infeasible optimum in the presence of locally invalid instruments when $H_g \xrightarrow{p} 0$. See Proposition 5.1 for formal asymptotic optimality results and similar implications when $\gamma > \frac{1}{2}$ without imposing $H_g \xrightarrow{p} 0$.

Remark 3.2. The first three terms of $L(K)$ in (3.1) and $H^{-1}H_gH_g'H^{-1}$ in Φ correspond to the square and cross-product of the two bias sources that we

https://doi.org/10.1017/S0266466622000597 Published online by Cambridge University Press

consider: the bias from many instruments and that from invalid instruments. The remaining terms in $L(K)$ represent higher-order variance terms, and the term $f'(I-P^K)f/N$ decreases as K increases. Note that locally invalid instrument specifications change not only the order of the biases from many instruments, but also the weights of the higher-order variances $f'(I-P^K)f/N$. If $P^Kg/N \xrightarrow{p} 0$ (which holds when the chosen instruments K are independent of the invalid instruments that have direct effects), then the last two terms in $L_g(K)$ reduce to $-2H_gH_g'$. Therefore, the exclusion of invalid instruments can help to reduce the MSE. However, including (locally) invalid instruments that are strong predictors of the first-stage can also lower the MSE. Proposition 3.1 provides a bias and variance trade-off in the presence of invalid instruments.

Remark 3.3. In Section S2.1 of the Supplementary Material, we provide an MSE approximation of 2SLS under $K=O(\sqrt{N})$, which generalizes the first-order asymptotic MSE results of Hahn and Hausman (2005).[7] For a linear specification of $f=\Psi\pi, g=\Psi\tau$ with a scalar endogenous variable Y_i and no included exogenous variables, the dominating bias term in the MSE approximation becomes

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

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 regarding f and g are not essential for the results in Hahn and Hausman (2005).

Remark 3.4. In Section S2.2 of the Supplementary Material, we also provide bias and variance approximations for 2SLS with invalid instruments by using a similar method to the 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 and K is small), the bias of the 2SLS estimator can be approximated by

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

where $\tilde{\mu}=\pi'\Psi'\Psi\tau/(\sigma_u\sigma_v)$. The 2SLS bias depends on the strength (μ^2) and the number of instruments (K), as well as the invalidity of the instruments ($\tilde{\mu}$). When $K/\mu^2 \gg \tilde{\mu}/\mu^3$, the first term dominates, and vice versa. Although our theory does not rely on weak instrument asymptotics, the above approximation can be useful for understanding the relative magnitude of bias due to many instruments and invalid instruments. For 2SLS, the misspecification bias and the many-instrument

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

https://doi.org/10.1017/S0266466622000597 Published online by Cambridge University Press

bias can have opposite signs and cancel out each other so that 2SLS can have a smaller finite-sample bias than those of other IV estimators.[8]

Next, we give the MSE approximations for the LIML and FULL estimators. Unlike that of 2SLS, the order of the dominating terms that depend on K for the LIML and FULL estimators remain the same ($O_p(K/N)$), as in DN. Note that the LIML and FULL estimators have the same approximate MSEs for the order we consider here.

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

PROPOSITION 3.2. *If Assumptions 2.1–2.3 are satisfied with* $\gamma = \frac{1}{2}$, $\mathbb{E}[v_i^2\eta_i|x_i] = 0$, $\Sigma_\eta \neq 0, H_g \neq 0$, $K/N \to 0$, *and* $\mathbb{E}[\|\xi_i\|^5|x_i], \mathbb{E}[|v_i|^5|x_i]$ *are finite, then the approximate MSEs for the LIML and FULL estimators satisfy the decomposition (2.10) with* $\Phi = \sigma_v^2 H^{-1} + H^{-1}H_g H_g' H^{-1}$, $G = G_{\text{LIML}}$ *defined in the Supplementary Material, and the following terms:*

$$L(K) = H^{-1}\big[\sigma_v^2\Sigma_\eta\frac{K}{N} + \sigma_v^2\frac{f'(I-P^K)f}{N} + L_g(K)\big]H^{-1}, \tag{3.3}$$

where $L_g(K)$ *is as defined in Proposition 3.1.*

Remark 3.5. It is important to note that $G = G_{\text{LIML}} = O_p(1/\sqrt{N})$ does not depend on K. Although G can be estimated with sample analogs, it is not used for the instrument selection criterion. With possibly invalid instruments, the dominating term in the MSE approximation (which depends on K) is proportional to K/N, which is the same as that in DN. For LIML or FULL, $L(K)$ does not include a higher-order bias from the presence of many instruments, and the terms in $L(K)$ show higher-order variance trade-offs with many invalid instruments. The third-moment condition $\mathbb{E}[v_i^2\eta_i|x_i] = 0$ holds when $(v_i, \eta_i')'$ is normally distributed, and this is imposed for a simplification similar to DN. Without this condition, $L(K)$ have additional terms that can be estimated, and they are provided in the proof of Proposition 3.2.

Next, we provide a result for the B2SLS estimator.

PROPOSITION 3.3. *If Assumptions 2.1–2.3 are satisfied with* $\gamma = \frac{1}{2}$, $\sigma_{uv} \neq 0$, $H_g \neq 0, \mathbb{E}[v_i^2 u_i|x_i] = 0$, *and* $K/N \to 0$, *then the approximate MSE for the B2SLS estimator satisfies the decomposition (2.10) with* $\Phi = \sigma_v^2 H^{-1} + H^{-1}H_g H_g' H^{-1}$, $G = G_{\text{B2SLS}}$ *defined in the Supplementary Material, and the following terms:*

$$L(K) = H^{-1}\big[(\sigma_v^2\Sigma_\eta + 2\sigma_{uv}\sigma_{uv}')\frac{K}{N} + \sigma_v^2\frac{f'(I-P^K)f}{N} + L_g(K)\big]H^{-1}, \tag{3.4}$$

where $L_g(K)$ *is as defined in Proposition 3.1.*

[8] We thank the referee for pointing this out.

https://doi.org/10.1017/S0266466622000597 Published online by Cambridge University Press

Remark 3.6. B2SLS is shown to be less efficient than LIML in the absence of invalid instruments ($H_g = 0$). Although $L(K)$ in the MSE approximation for B2SLS is larger than those of LIML/FULL, it seems difficult to show the higher-order efficiency of the LIML and FULL estimators with locally invalid instruments because of the different terms, i.e., G_{LIML} in Proposition 3.2 and G_{B2SLS} in Proposition 3.3.

We next consider the JIVE2, HLIM, and HFUL estimators. Under the rate condition in this paper ($K/N \to 0$), all estimators considered are consistent. However, JIVE2, HLIM, and HFUL are consistent under many-instruments sequences ($K/N \to \alpha > 0$) and heteroskedasticity, whereas 2SLS, LIML, and B2SLS are not consistent. Although our results are based on homoskedastic error terms, the higher-order theory of the jackknife versions of the k-class estimators complements the literature, and some of the theoretical results are consistent with the (first-order) asymptotic results in Chao et al. (2012) and Hausman et al. (2012).

Furthermore, the higher-order MSEs of the JIVE2 and HLIM/HFUL estimators do not contain the terms derived from the third moments, unlike those of B2SLS and LIML/FULL. For example, Chao et al. (2012) and Hausman et al. (2012, p. 223) also show that the first-order asymptotic variance of JIVE2/HLIM does not depend on the third and fourth moment terms of the error disturbances under homoskedasticity.

We next provide a result for JIVE2 estimator. Let $D^K = diag(P_{ii}^K)$ be a diagonal matrix.

Proposition 3.4. *If Assumptions 2.1–2.3 are satisfied with* $\gamma = \frac{1}{2}$, $\sigma_{uv} \neq 0$, $H_g \neq 0$, *and* $K/N \to 0$, *then the approximate MSE for the JIVE2 estimator satisfies the decomposition (2.10) with* $\Phi = \sigma_v^2 H^{-1} + H^{-1}H_gH_g'H^{-1}$, $G = G_{\text{JIVE2}}$ *defined in the Supplementary Material, and the following terms:*

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

where

$$\begin{aligned} L_{g,D}(K) = {} & H_gH_g'H^{-1}\frac{f'(I-(P^K-D^K))f}{N} + \frac{f'(I-(P^K-D^K))f}{N}H^{-1}H_gH_g' \\ & -\frac{f'(I-(P^K-D^K))g}{N}H_g' - H_g\frac{g'(I-(P^K-D^K))f}{N}. \end{aligned}$$

Remark 3.7. When $H_g = 0$ w.p.a.1., $L(K)$ in Proposition 3.4 reduces to that of B2SLS in Proposition 3.3 without assuming zero third-moment conditions. Without invalid instruments, Hahn et al. (2004) provide the same higher-order MSE results for the jackknife 2SLS estimator with a scalar endogenous variable, no included exogenous variables, jointly normal residuals $(v_i, u_i')'$, and an additional assumption $\max_i P_{ii}^K = O_p(N^{-1})$, which is slightly stronger than that which we impose here (Assumption 2.3(c)). Donald and Newey (1999) derive similar results for the JIVE1 estimator considered in Phillips and Hale (1977) and Angrist et al. (1999). Higher-order MSE derivations for the JIVE2 estimator have not been available in the literature, and we also contribute to the literature by extending the existing results to a setup with locally invalid instruments. However, the presence of invalid instruments complicates higher-order MSE comparisons among B2SLS, JIVE2, and LIML.

Finally, we give results for the HLIM and HFUL estimators.

PROPOSITION 3.5. *If Assumptions 2.1–2.3 are satisfied with $\gamma = \frac{1}{2}$, $\Sigma_\eta \neq 0$, $H_g \neq 0$, and $K/N \to 0$, then the approximate MSEs for the HLIM and HFUL estimators satisfy the decomposition (2.10) with $\Phi = \sigma_v^2 H^{-1} + H^{-1} H_g H_g' H^{-1}$, $G = G_{\text{HLIM}}$ defined in the Supplementary Material, and the following terms:*

$$L(K) = H^{-1}\left[\sigma_v^2 \Sigma_\eta \frac{K}{N} + \sigma_v^2 \frac{f'(I-P^K)f}{N} + L_{g,D}(K)\right] H^{-1}, \tag{3.6}$$

where $L_{g,D}(K)$ is as defined in Proposition 3.4.

Remark 3.8. When $H_g = 0$, dominant terms of the MSE in Proposition 3.5 (without imposing symmetry-type conditions on the error terms) are identical to those of LIML/FULL in Proposition 3.2. Hausman et al. (2012, p. 224) show that the (first-order) asymptotic variances of HLIM and LIML are the same in the presence of many weak instruments, $K/N \to 0$, $\max_i P_{ii}^K = o_p(1)$, and homoskedasticity. Although $G_{\text{LIML}} = G_{\text{HLIM}}$ (see equation (S1.1) in the Supplementary Material), however, $L(K)$ is different in the presence of invalid instruments, and neither of the estimators dominates each other in terms of their higher-order MSEs.

Proposition 3.5 also provides the higher-order efficiency of HLIM/HFUL relative to that of JIVE2 in the absence of invalid instruments, as the dominating terms $L(K)$ in the higher-order MSE are smaller than those for JIVE2 in Proposition 3.4. This complements the results in Chao et al. (2012) and Hausman et al. (2012), where the authors show that HLIM is asymptotically more efficient than JIVE2 under many-weak instruments and homoskedasticity.

4. HIGHER-ORDER MSE RESULTS WITH $\gamma > \frac{1}{2}$

In this section, we consider faster rates of local-to-zero specification (i.e., a smaller degree of invalidity) than the $N^{-1/2}$ rates considered in Section 3. All estimators are consistent, and there are no first-order asymptotic biases due to invalid instruments under the conditions imposed here, but higher-order theory is still useful for capturing changes in the orders of the bias and variance that the first-order asymptotic theory does not capture.

Although we expect the stochastic orders of the higher-order bias and variance from invalid instruments to become smaller than the terms due to the many

https://doi.org/10.1017/S0266466622000597 Published online by Cambridge University Press

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 4.1–4.5) is that the changes in the degree of invalid instruments (γ) affect the finite-sample behaviors of IV estimators differently.

The following proposition provides a higher-order MSE approximation result for the 2SLS estimator with $\gamma > \frac{1}{2}$.

PROPOSITION 4.1. *Suppose that Assumptions 2.1–2.3 are satisfied with $\gamma > \frac{1}{2}$. If $\sigma_{uv} \neq 0$, $H_g \neq 0$, and $K^2/N \to 0$, then the approximate MSE for the 2SLS estimator satisfies the decomposition (2.10) with $\Phi = \sigma_v^2 H^{-1}, G = \frac{1}{N^{2\gamma-1}} H^{-1} H_g H_g' H^{-1}$, and*

$$L(K) = H^{-1}\Big[\frac{K}{N^\gamma}(H_g \sigma_{uv}' + \sigma_{uv} H_g') + \sigma_{uv}\sigma_{uv}' \frac{K^2}{N} + \sigma_v^2 \frac{f'(I-P^K)f}{N}\Big]H^{-1}. \tag{4.1}$$

If we further assume that $\frac{K}{N^{1-\gamma}} \to \infty$, then (2.10) holds with $\Phi = \sigma_v^2 H^{-1}, G = 0$, and

$$L(K) = H^{-1}\Big[\sigma_{uv}\sigma_{uv}' \frac{K^2}{N} + \sigma_v^2 \frac{f'(I-P^K)f}{N}\Big]H^{-1}. \tag{4.2}$$

Remark 4.1. In (4.1), G and $L(K)$ contain higher-order bias terms due to the presence of many invalid instruments, as well as higher-order variance terms because of the many instruments. $L(K)$ includes the (squared) bias from many instruments (proportional to K^2/N) and the interactions from many invalid instruments (proportional to K/N^γ). Note that the order of the interaction biases increases with K and decreases with γ and cannot be ignored in general. When $\gamma = \frac{1}{2}$, this interaction bias dominates the bias from many instruments, as in Proposition 3.1. However, if we restrict the rate of K, the second result of Proposition 4.1 shows that the higher-order bias from many instruments dominates the bias terms due to the invalid instruments, and $L(K)$ in (4.2) has the same form in DN. The assumption of the second result holds when $\gamma > 1-\alpha$ ($K = O(N^\alpha)$), and this always holds for $\gamma \geq 1$.

In (4.2), the dominant terms $L(K)$ remain the same as those in DN, which shows that the MSE approximation in DN for the 2SLS estimator is robust to a small degree of invalid instruments. Nevertheless of these intuitive results, we quantify the robustness of the MSE approximation of the 2SLS estimator in DN, and this is nontrivial as the dominating terms in $L(K)$ depend not only on the order of invalid instruments γ, but also on the rate of K. Moreover, we establish the asymptotic optimality of the instrument selection criterion for 2SLS, which does not require the estimation of H_g and $g(\cdot)$ (see Proposition 5.1).

The next two results are for the LIML/FULL and B2SLS estimators.

PROPOSITION 4.2. *Suppose that Assumptions 2.1–2.3 are satisfied with $\gamma > \frac{1}{2}$. Assume that $\Sigma_\eta \neq 0, H_g \neq 0, \mathbb{E}[v_i^2 \eta_i | x_i] = 0$, $K/N \to 0$, and $\mathbb{E}[\|\xi_i\|^5 | x_i], \mathbb{E}[|v_i|^5 | x_i]$ are finite. Then, the approximate MSEs for the LIML and FULL estimators satisfy* *the decomposition (2.10) with* $\Phi = \sigma_v^2 H^{-1}, G = \frac{1}{N^{2\gamma-1}} H^{-1} H_g H_g' H^{-1}$*, and the following terms:*

$$L(K) = H^{-1}\big[\sigma_v^2 \Sigma_\eta \frac{K}{N} + \sigma_v^2 \frac{f'(I-P^K)f}{N}\big] H^{-1}. \tag{4.3}$$

PROPOSITION 4.3. *Suppose that Assumptions 2.1–2.3 are satisfied with* $\gamma > \frac{1}{2}$*. Assume that* $\sigma_{uv} \neq 0, H_g \neq 0$*,* $\mathbb{E}[v_i^2 u_i | x_i] = 0$*, and* $K/N \to 0$*. Then, the approximate MSE for the B2SLS estimator satisfies the decomposition (2.10) with* $\Phi = \sigma_v^2 H^{-1}, G = \frac{1}{N^{2\gamma-1}} H^{-1} H_g H_g' H^{-1}$*, and*

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

Remark 4.2. Propositions 4.2 and 4.3 show that the leading term $L(K)$ in the MSE approximations for the LIML, FULL, and B2SLS estimators is the same as the leading term in DN for all $\gamma > \frac{1}{2}$ under the same rate conditions $K/N \to 0$. Note that the MSE approximations can still capture the higher-order biases from locally invalid instruments G, which do not depend on K. The above results show the robustness of the MSE approximations (and instrument selection criteria) of the LIML, FULL, and B2SLS estimators in DN under the presence of locally invalid instruments in the sense that the dominant terms (that depend on K) in the higher-order MSEs remain the same with possibly invalid instruments ($\gamma > \frac{1}{2}$). With a smaller higher-order bias from many instruments, the dominating terms in the MSE approximation coincide with those of DN for LIML/FULL and B2SLS only if $\gamma > \frac{1}{2}$; in contrast, 2SLS requires a more stringent condition.

The next two results are valid for the JIVE2 and HLIM/HFUL estimators. Their implications remain the same as in Remark 4.2.

PROPOSITION 4.4. *Suppose that Assumptions 2.1–2.3 are satisfied with* $\gamma > \frac{1}{2}$*. Assume that* $\sigma_{uv} \neq 0, H_g \neq 0$*, and* $K/N \to 0$*. Then, the approximate MSE for the JIVE2 estimator satisfies the decomposition (2.10) with* $\Phi = \sigma_v^2 H^{-1}$*,* $G = \frac{1}{N^{2\gamma-1}} H^{-1} H_g H_g' H^{-1}$*, and* $L(K)$ *defined the same as in Proposition 4.3.*

PROPOSITION 4.5. *Suppose that Assumptions 2.1–2.3 are satisfied with* $\gamma > \frac{1}{2}$*. Assume* $\Sigma_\eta \neq 0, H_g \neq 0$*, and* $K/N \to 0$*. Then, the approximate MSEs for the HLIM and HFUL estimators satisfy the decomposition (2.10) with* $\Phi = \sigma_v^2 H^{-1}$*,* $G = \frac{1}{N^{2\gamma-1}} H^{-1} H_g H_g' H^{-1}$*, and* $L(K)$ *defined the same as in Proposition 4.2.*

5. INSTRUMENT SELECTION CRITERIA

In this section, we consider instrument selection criteria based on the higher-order MSE approximations provided in Sections 3 and 4.

We first establish the asymptotic optimality property of DN's criteria under a specification with $N^{-\gamma}(\gamma > \frac{1}{2})$ locally invalid instrument based on

Propositions 4.1–4.5. To simplify the results, we consider a case with a scalar endogenous regressor (i.e., Y_i is scalar) where the covariates have already been partialled out.[9] For the details of the case with general vector endogenous variables Y_i, see Section S3 of the Supplementary Material.

5.1. Optimality of DN's Criteria under $\gamma > \frac{1}{2}$

We choose K to minimize an estimate of the dominant term $L(K)$ based on the higher-order MSE approximation:

$$\widehat{K} = \arg\min_{K \in \mathcal{K}} \widehat{L}(K), \tag{5.1}$$

where $\mathcal{K} = \mathcal{K}_N = \{K_m : 1 \le m \le M_N\}$ is a set of instruments, and the total number of instrument sets M_N depends on the sample size.

We require preliminary estimates of the model and the goodness of fit criterion for the first-stage reduced-form equation. Let $\widetilde{\delta}$ be some preliminary estimator, e.g., an IV estimator where the instruments $\widetilde{K}$ are chosen to minimize the cross validation (CV) or Mallows (1973) criteria for the reduced-form equation. Let $\widetilde{\varepsilon} = y - W\widetilde{\delta}$ be residuals, and let $\hat{H} = W'P^{\widetilde{K}}W/N - \hat{\sigma}_u^2\widetilde{K}/N$ be a preliminary estimator of $H = f'f/N$. Additionally, let $\widetilde{u} = (I - P^{\widetilde{K}})W$ be a residual vector of the first-stage reduced-form regression; it is important that all preliminary estimates remain fixed and do not depend on K, while the criterion is calculated for different instrument sets.

The criteria based on Propositions 4.1–4.5 are defined below and are equivalent to those of DN:

$$\text{2SLS}: \widehat{L}_{DN}(K) = \hat{\sigma}_{uv}^2 \frac{K^2}{N} + \hat{\sigma}_v^2 \left(\hat{R}(K) - \hat{\sigma}_u^2 \frac{K}{N} \right), \tag{5.2}$$

$$\text{LIML, FULL, HLIM, HFUL}: \widehat{L}_{DN}(K) = \hat{\sigma}_v^2 \left(\hat{R}(K) - \frac{\hat{\sigma}_{uv}^2}{\hat{\sigma}_v^2} \frac{K}{N} \right), \tag{5.3}$$

$$\text{B2SLS, JIVE2}: \widehat{L}_{DN}(K) = \hat{\sigma}_v^2 \left(\hat{R}(K) + \frac{\hat{\sigma}_{uv}^2}{\hat{\sigma}_v^2} \frac{K}{N} \right), \tag{5.4}$$

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

[9] 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 the exogenous covariates x_{1i}.

https://doi.org/10.1017/S0266466622000597 Published online by Cambridge University Press

We show that the instrument selection criteria in DN are asymptotically optimal in the sense of Li (1987) under locally invalid instruments, i.e., $\widehat{K}$ in (5.1) with $\widehat{L}_{DN}(K)$ satisfies equation (5.5).

Assumption 5.1. W_i is a scalar, $\hat{\sigma}_v^2 - \sigma_v^2 = o_p(1), \hat{\sigma}_u^2 - \sigma_u^2 = o_p(1), \hat{\sigma}_{uv} - \sigma_{uv} = o_p(1), \hat{H} - \bar{H} = o_p(1), \bar{H}^{-1}\sigma_{uv} \neq 0$, and $var(\bar{H}^{-1}\eta_i) > 0$.

Assumption 5.2. Assume that $\mathbb{E}[u_i^8|x_i] < \infty$ and $\sum_{K\in\mathcal{K}_N}(NR(K))^{-1} \to 0$, where $R(K) = \sigma_u^2(K/N) + \bar{H}^{-1}[f'(I-P^K)f/N]\bar{H}^{-1}$. Furthermore, assume that $\sup_K \sup_i P_{ii}^K \xrightarrow{p} 0$ when $\hat{R}(K)$ is the CV criterion.

Assumption 5.1 imposes consistency of the preliminary estimators. Assumption 5.2 is a standard assumption in the literature (e.g., Li, 1987) regarding the asymptotic optimality of the model selection criteria. As $\sum_{K\in\mathcal{K}_N}(NR(K))^{-1} \leq M_N(\inf_{K\in\mathcal{K}_N} NR(K))^{-1}$, Assumption 5.2 can be replaced by $\inf_{K\in\mathcal{K}_N} NR(K) \to \infty$ with restrictions on the set of possible instrument sets M_N, and this excludes the case where f has a finite-order representation with a number of instruments and rules out $\mathcal{K}_N$ including all possible combinations of instruments.

Proposition 5.1. *Suppose that Assumptions 5.1 and 5.2 hold. Under the same assumptions as in Proposition 4.1, the following holds for 2SLS with* $\widehat{K} = \arg\min_{K\in\mathcal{K}} \widehat{L}_{DN}(K)$ *if* $\frac{K}{N^{1-\gamma}} \to \infty$:

$$\frac{L(\widehat{K})}{\inf_K L(K)} \xrightarrow{p} 1, \tag{5.5}$$

where $L(K)$ *is as defined in (4.2).*

For the LIML (FULL), B2SLS, JIVE2, and HLIM (HFUL) criteria, equation (5.5) holds for all $\gamma > \frac{1}{2}$ *under the same assumptions as in Propositions 4.2–4.5, respectively.*

Remark 5.1. Proposition 5.1 provides the asymptotic optimality of the instrument selection criteria in Donald and Newey, 2001 and shows that their criteria for choosing K are robust to locally invalid instruments ($\gamma > \frac{1}{2}$). The optimality result, equation (5.5), implies that $L(\widehat{K})$ with $\widehat{K}$ obtained by the selection procedure is asymptotically equivalent to minimizing the unknown $L(K)$ directly in the presence of locally invalid instruments. Asymptotic optimality results for the JIVE2, HLIM, and HFUL estimators are new with or without invalid instruments.

While the selection criteria for LIML/FULL, B2SLS, JIVE2, and HLIM/HFUL are asymptotically optimal for all $\gamma > \frac{1}{2}$, the optimality of the 2SLS criterion requires $\gamma > 1-\alpha$, where $K = O(N^\alpha), 0 < \alpha < \frac{1}{2}$. The criterion for 2SLS is optimal when the degree of invalidity is sufficiently small (γ is sufficiently large).

5.2. Invalidity-Robust Instrument Selection Criteria

We next consider instrument selection criteria based on Propositions 3.1–3.5. Although Proposition 5.1 shows that the instrument selection criteria in DN are

https://doi.org/10.1017/S0266466622000597 Published online by Cambridge University Press

robust to a very small degree of invalid instruments ($\gamma > \frac{1}{2}$), the higher-order bias/variance terms due to invalid instruments may not be negligible when the degree of invalidity increases (i.e., $\gamma = \frac{1}{2}$), as shown in Section 3.

We propose *IR* instrument selection criteria that are robust to a larger degree of invalid instruments ($\gamma = \frac{1}{2}$) based on Propositions 3.1–3.5. Because the estimation of the MSE requires some preliminary estimates of $g(x)$, we require a known set of valid instruments z_i: $\mathbb{E}[z_i\varepsilon_i] = 0, z_i \in \mathbb{R}^q (q \geq p)$. We consider the case in which researchers have a "conservative" set of valid instruments and explore all other candidate instruments that are potentially invalid considering the bias–variance trade-off. Although this is a critical assumption, our derivations of the MSE approximations in earlier sections do not require known valid instruments. Recent papers that have addressed similar questions regarding the general GMM setup (such as Liao, 2013; Cheng and Liao, 2015; DiTraglia, 2016; Cheng, Liao, and Shi, 2019) also assume that there is a subset of moments that is known to be valid for identification and estimation.

The proposed instrument selection criteria $\widehat{L}_{IR}(K)$ are as follows:

$$\text{2SLS} : \widehat{L}_{IR}(K) = \widehat{L}_{DN}(K) + 2\hat{H}_g\hat{\sigma}_{uv}\frac{K}{\sqrt{N}} + \widehat{L}_g(K), \tag{5.6}$$

$$\text{LIML/FULL/B2SLS} : \widehat{L}_{IR}(K) = \widehat{L}_{DN}(K) + \widehat{L}_g(K), \tag{5.7}$$

$$\text{JIVE2/HLIM/HFUL} : \widehat{L}_{IR}(K) = \widehat{L}_{DN}(K) + \widehat{L}_{g,D}(K), \tag{5.8}$$

where $\widehat{L}_{DN}(K)$ is defined as in (5.2)–(5.4),

$$\widehat{L}_g(K) = \frac{2\hat{H}_g^2}{\hat{H}}(\hat{R}(K) - \hat{\sigma}_u^2\frac{K}{N}) - 2\hat{H}_g\hat{G}(K) + \hat{R}_g(K),$$

$$\widehat{L}_{g,D}(K) = \frac{2\hat{H}_g^2}{\hat{H}}\hat{R}_D(K) - 2\hat{H}_g\hat{G}_D(K) + \hat{R}_{g,D}(K),$$

$\hat{H}_g = W'P^{\widetilde{K}}\widehat{\varepsilon}/\sqrt{N} - \widetilde{K}/\sqrt{N}\hat{\sigma}_{uv}$, $\hat{G}(K) = W'(I - P^K)\widehat{\varepsilon}/\sqrt{N} + K/\sqrt{N}\hat{\sigma}_{uv}$, $\hat{R}_g(K) = 2K/N\hat{H}\hat{H}_z^{-1}\hat{\sigma}_{v,z}^2\hat{\sigma}_u^2(I - \hat{H}\hat{H}_z^{-1}) - 2\hat{H}\hat{H}_z^{-1}\hat{\sigma}_v^2\hat{R}_z(K) + 2\hat{H}_{z,D}\hat{H}_z^{-1}\hat{\sigma}_v^2\hat{R}(K), \hat{H}_z = W'P_zW/N, \hat{R}_z(K) = W'P_z(I - P^K)W/N + 2\hat{\sigma}_u^2K/N, \hat{H}_{z,D} = W'D_zW/N$, where P_z is the projection matrix using valid instrument $z = [z_1, \ldots, z_N]'$, and $D_z = diag(P_{ii,z})$ is a diagonal matrix. Let also $\hat{R}_D(K) = \frac{W'(I-(P^K-D^K))W}{N}, \hat{R}_{g,D}(K) = W'D^KW/N(\hat{\sigma}_v^2I - \hat{H}\hat{H}_z^{-1}\hat{\sigma}_{v,z}^2) - 2\hat{H}\hat{H}_z^{-1}\hat{\sigma}_v^2\hat{R}_z(K) + 2\hat{H}_{z,D}\hat{H}_z^{-1}\hat{\sigma}_v^2\hat{R}_D(K)$, $\hat{G}_D(K) = W'(I - (P^K - D^K))\widehat{\varepsilon}/\sqrt{N}$, $\hat{\sigma}_{v,z}^2 = \widehat{\varepsilon}'P_z\widehat{\varepsilon}/N$, $\widehat{\varepsilon} = y - W\widehat{\delta}$, and $\widehat{\delta}$ is a preliminary estimator obtained using valid instruments $z = [z_1, \ldots, z_N]'$ with the projection matrix P_z.[10]

[10] Note that we considered a simplified version of the criteria $\widehat{L}_{IR}(K)$ without $\hat{R}_g(K), \hat{R}_{g,D}(K)$ for the simulation. If we use estimates $\hat{H}_z = \hat{H} = \hat{H}_{z,D}, \hat{\sigma}_{v,z}^2 = \hat{\sigma}_v^2, \hat{R}_z(K) = \hat{R}(K) = \hat{R}_D(K)$, then $\hat{R}_g(K) = \hat{R}_{g,D}(K) = 0$, and there are no notable differences in the simulation evidence with or without these additional terms. For the simplified version, $\widehat{L}_{IR}(K)$ in (5.6)–(5.8) reduces to $\widehat{L}_{DN}(K)$ in (5.2)–(5.4) when $\hat{H}_g = 0$.

https://doi.org/10.1017/S0266466622000597 Published online by Cambridge University Press

We show below that $\widehat{L}_{IR}(K)$ is an asymptotically unbiased estimator of the leading terms in higher-order MSE approximation that depend on K defined in Sections 3 and 4. Details on the proof and sufficient conditions for the higher-order decomposition are provided in Section S1.3 of the Supplementary Material. Let $\bar{\rho}_{K,N} = tr(L(K))$ and $\sigma^2_{v,z} = \mathbb{E}[v_z' v_z/N|X]$, $v_z = P_z v$.[11]

We impose the following assumption on the consistency of the preliminary estimators. Assumption 5.3 also imposes high-level assumption for a preliminary estimator with the instruments $\widetilde{K}$. For example, the higher-order decomposition $\hat{H}_g - (H_g + \Lambda_g) = T_g + o_p(\bar{\rho}_{K,N}), \hat{H} - H = T_H + o_p(\bar{\rho}_{K,N})$ holds when $\bar{\rho}_{\widetilde{K},N} = o(\bar{\rho}_{K,N})$ as $\widetilde{K}, K \to \infty$.

Assumption 5.3. W_i is a scalar, $\hat{\sigma}^2_v - \sigma^2_v = o_p(1), \hat{\sigma}^2_u - \sigma^2_u = o_p(1), \hat{\sigma}_{uv} - \sigma_{uv} = o_p(1), \hat{\sigma}^2_{v,z} - \sigma^2_{v,z} = o_p(1)$, $\hat{H}_g - (H_g + \Lambda_g) = T_g + o_p(\bar{\rho}_{K,N})$, and $\hat{H} - H = T_H + o_p(\bar{\rho}_{K,N})$, $\hat{G}(K) = T_G + o_p(\bar{\rho}_{K,N})$ for $K \in \mathcal{K}_N$, where $\Lambda_g = O_p(1), T_H = o_p(1)$, $T_g = o_p(1), T_G = o_p(1), \bar{\rho}_{K,N} = o_p(1), ||T_g||||T_G|| = o_p(\bar{\rho}_{K,N})$.

Proposition 5.2. *Suppose that we have a vector of valid instruments $z_i \in \mathbb{R}^q (q \geq p)$ such that $\mathbb{E}[z_i \varepsilon_i] = 0$ for finite q. In addition, suppose that Assumption 5.3 holds and $\mathbb{E}[v_i^2 u_i|x_i] = 0$. Under the same assumptions in Propositions 3.1–3.5 with $\gamma = \frac{1}{2}$, $\widehat{L}_{IR}(K)$ given in (5.6)–(5.8) for the 2SLS, LIML (FULL), B2SLS, JIVE2, and HLIM (HFUL) estimators satisfy the following decomposition:*

$$\widehat{L}_{IR}(K) = \widehat{Q}_L(K) + \widehat{r}_L(K),$$

$$\mathbb{E}[\widehat{Q}_L(K)|X] = L(K) + \bar{r}_L(K),$$

$$[\widehat{r}_L(K) + \bar{r}_L(K)]/tr(L(K)) = o_p(1), \quad K \to \infty,\ N \to \infty, \tag{5.9}$$

with $L(K)$ defined in Propositions 3.1–3.5, respectively.

Alternatively, under the same assumptions in Propositions 4.1–4.5 with $\gamma > \frac{1}{2}$, the decomposition (5.9) holds with $L(K)$ defined in (4.2)–(4.4) for the 2SLS, LIML (FULL), B2SLS, JIVE2, and HLIM (HFUL) estimators, respectively.

Assuming that we have a known set of valid instruments, Proposition 5.2 implies that $\widehat{L}_{IR}(K)$ is an asymptotically unbiased estimator of the leading terms (that depend on K) $L(K)$ in the higher-order MSE approximation provided in Propositions 3.1–3.5 and the remainder terms $(\bar{r}_L(K), \widehat{r}_L(K))$ go to zero faster than the $L(K)$. This result also holds with $\gamma > \frac{1}{2}$ and the $L(K)$ defined in Section 4.[12]

[11] Because the choice of K is unaffected by subtracting constants from $\widehat{L}_{IR}(K)$, we can assume without loss of generality that $\widehat{L}_{IR}(K)$ can be constructed using $\tilde{R}(K) = \hat{R}(K) - u'u/N, \tilde{G}(K) = \hat{G}(K) - \bar{G}$ where $\bar{G} = O_p(1)$ that do not depend on K.

[12] The third moment assumption $\mathbb{E}[v_i^2 u_i|x_i] = 0$ is imposed for the simplification of the criteria similar to Propositions 3.2 and 3.3. Without this condition, $\widehat{L}_{IR}(K)$ have additional terms that can be estimated. We have not included these terms in the instrument selection criteria for simplicity, and they provide similar results in our simulations.

It is important to note that while $\widehat{L}_{IR}(K)$ satisfies (5.9) for a nonrandom sequence of $K \to \infty$, this does not necessarily imply that $\widehat{L}_{IR}(\widehat{K})$ with $\widehat{K}$ selected by the IR criterion approximates the infeasible $\inf_K L(K)$ well in practice. It would be desirable to justify $\widehat{L}_{IR}(K)$ in terms of optimality as in Proposition 5.1, but this is a difficult problem, as it requires dealing with the estimation of $g(\cdot)$, which is not $\sqrt{N}$-estimable. Although (5.9) shows that $\widehat{L}_{IR}(K)$ is an unbiased estimate of $L(K)$, the sampling variability of $\widehat{L}_{IR}(K)$ can be large which makes uniform consistency fail to hold, but we do not investigate this direction in the paper. We further note that the preliminary estimator $\widehat{\delta}$ may affect the finite-sample behavior of the IR criterion; however, simulation evidence suggests that an IR instrument selection criterion combined with an IV estimator that has a small bias property under many instruments without moment problems, such as the Fuller/HFUL estimators, works well.

6. MONTE CARLO SIMULATION

We investigate the finite-sample performance of the IV estimators based on the instrument selection criteria considered in this paper. We use the same simulation design of DN and Kuersteiner and Okui (2010), allowing for potentially invalid instruments (with an intercept in the model). The model to be estimated is

$$\begin{aligned} y_i &= \beta_0 + x_i\beta_1 + \varepsilon_i, \\ \mathbb{E}(z_i\varepsilon_i) &= 0, \end{aligned} \tag{6.10}$$

where x_i and β_0, β_1 are scalars, and z_i is a $(K+1) \times 1$ vector of IVs. We assume that z_i always contains a constant and K denotes the number of excluded exogenous variables.

We estimate β_0, β_1 by various IV estimators with $\widehat{K}$ chosen via instrument selection criteria. Our DGP is

$$\begin{aligned} y_i &= \beta_0 + x_i\beta_1 + \frac{\tau' Z_i}{N^\gamma} + v_i, \\ x_i &= \pi' Z_i + u_i, \\ Z_i &\sim N\left(0, I_{\bar{K}}\right), \\ \begin{pmatrix} v_i \\ u_i \end{pmatrix} &\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & \sigma_{uv} - \pi'\tau/N^\gamma \\ \sigma_{uv} - \pi'\tau/N^\gamma & 1 \end{bmatrix} \right). \end{aligned} \tag{6.11}$$

We set $\beta_0 = 1, \beta_1 = 0.1$ and vary $(N, \bar{K}, R^2, \pi, \tau)$ in the simulation experiments. We set the maximum number of instruments $\bar{K} = 20$ when the number of observations $N = 100$ and $\bar{K} = 30$ for $N = 1{,}000$. We set the first-stage R^2 as $\{0.1, 0.01\}$. We also set the endogeneity of x_i as $Cov(x_i, \varepsilon_i) = \sigma_{uv} = 0.5$ and perform 10,000 simulation replications.

As in DN, we set the first-stage coefficient $\pi = (\pi_1, \ldots, \pi_{\bar{K}})'$ as

$$\pi_k = c(\bar{K})(1 - k/(\bar{K}+1))^4,$$

https://doi.org/10.1017/S0266466622000597 Published online by Cambridge University Press

where $c(\bar{K})$ is chosen to make R^2 either 0.1 or 0.01. We consider the case in which there is some prior information about which instruments are strong.

When we fix $\tau \neq 0$, the key parameter is γ ("degree of invalidity"), and we vary $\gamma = \infty, 1, \frac{1}{2}$, and $\frac{1}{3}$. When $\gamma = \infty$, any instruments z_i from the full set of instruments Z_i are valid. For $0 < \gamma < \infty$, we find that $\mathbb{E}(z_i \varepsilon_i) = \tau / N^\gamma \neq 0$, so the moment condition (6.10) fails to hold in any finite sample. The DGP can also be written as a globally misspecified model such that $\gamma = 0$ and τ is not too large, but we use the locally misspecified setup to be consistent with the results in Sections 3 and 4.

We consider the following specification for $\tau = (\tau_1, \ldots, \tau_{\bar{K}})'$:

$$\tau_k = 0, \text{ for } k = 1, \quad \tau_k = 0.5, \text{ for } k = 2, \ldots, \bar{K}/2, \text{ and } \quad \tau_k = 0, \text{ for } k > \bar{K}/2.$$

We assume that only the first instrument is known to be valid. Valid instruments are included in all candidate instruments and used for preliminary estimates in the IR instrument selection criteria. We consider that the next "strong" IVs are potentially invalid. This is empirically relevant, as IVs that are strongly correlated with the endogenous regressor are more likely to be correlated with the dependent variable, and there exists a bias–variance trade-off when using invalid but relevant instruments. Moreover, the last half of the "weak" IVs are valid, so there also exists a trade-off when using valid but weak instruments.[13]

We focus on β_1, and Tables 1–4 report the median bias (Bias), MAD, interdecile range (IDR), root mean squared error (MSE), and root trimmed mean-squared error (TMSE) of 2SLS, LIML, FULL (with $C = 1$), JIVE2, HLIM, and HFUL (with $C = 1$). We use these robust measures of central tendency and dispersion due to concerns on the existence of moments for some estimators.[14] For comparison purposes, we include the OLS and GMM averaging estimator (GMM-AVE) by Cheng et al. (2019), which combines a conservative GMM estimator with valid moment conditions and a GMM estimator based on all (possibly invalid) instruments, where the weights are given in Cheng et al. (2019, eqn. (4.7)).[15] For all IV estimators, we consider four different cases: using all available instruments (all), using the valid instrument only (val), utilizing the instruments chosen by DN's criterion (DN), and the instruments selected by the IR criterion in this

[13] In the additional simulation results reported in the Supplementary Material, we also investigate the same simulation design of DN without an intercept in the model. For example, we consider different specifications such as $\sigma_{uv} = 0.2, 0.8$, different π and τ, and a heteroskedastic setup. In particular, we consider (1) $\pi_k = \sqrt{R^2/\bar{K}(1-R^2)}$, the case where the instruments have equal strengths as in DN (Model 2), (2) $\pi_k = c(\bar{K})(1 - (\bar{K}+1-k)/(\bar{K}+1))^4$, the case where the order of the IV strengths is wrong, (3) $\tau \propto (0,1,1,1,0,\ldots,0)$, (4) $\tau \propto (0,1,0,\ldots,0)$, and (5) $\tau \propto (0,0,0,0,1,\ldots,1)$. Furthermore, we also explore the simulation design in Hausman et al. (2012) with an intercept, and we note that theoretical results of HLIM/HFUL in Hausman et al. (2012, Assump. 1) include an intercept in the model. See the Supplementary Material for further discussion. In an earlier version of the paper, we provide the results assuming the first two instruments are known to be valid.

[14] To be specific, we compute a root trimmed mean square, $\sqrt{\mathbb{E}[\min\{(\hat{\beta}_1 - \beta_1)^2, 100\}]}$ similar to Okui (2011), but with a larger trimming parameter.

[15] We do not report the results for the Fuller/HFUL estimator with a constant $C = 4$, as they are mostly similar to those in the $C = 1$ case. We also omit the results for B2SLS, as it is dominated by the other estimators in most cases.

Table 1. Monte Carlo results: $R^2 = 0.1, N = 100$

	$\gamma = \infty$					$\gamma = 1$					$\gamma = \frac{1}{2}$					$\gamma = \frac{1}{3}$				
$N = 100$	Bias	MAD	IDR	MSE	TMSE	Bias	MAD	IDR	MSE	TMSE	Bias	MAD	IDR	MSE	TMSE	Bias	MAD	IDR	MSE	TMSE
OLS	0.45	0.45	0.22	0.46	0.46	0.45	0.45	0.22	0.46	0.46	0.45	0.45	0.22	0.46	0.46	0.45	0.45	0.23	0.46	0.46
2SLS-all	0.32	0.32	0.42	0.36	0.36	0.33	0.33	0.42	0.37	0.37	0.41	0.41	0.43	0.44	0.44	0.51	0.51	0.47	0.54	0.54
2SLS-val	0.02	0.35	1.66	12.63	1.56	0.04	0.34	1.64	18.12	1.64	0.02	0.35	1.70	14.74	1.65	0.02	0.36	1.74	31.06	1.63
2SLS-DN	0.19	0.25	0.73	1.95	0.56	0.20	0.26	0.72	8.50	0.53	0.34	0.36	0.71	6.63	0.67	0.50	0.51	0.74	28.96	0.76
2SLS-IR	0.08	0.25	0.98	4.85	0.71	0.11	0.26	0.98	2.92	0.72	0.17	0.31	1.06	1.84	0.68	0.24	0.40	1.23	3.75	0.80
GMM-AVE	0.17	0.25	0.79	0.37	0.37	0.19	0.26	0.79	0.37	0.37	0.21	0.28	0.85	0.40	0.40	0.23	0.31	0.94	0.44	0.44
LIML-all	0.04	0.33	1.61	7.67	1.69	0.07	0.34	1.66	17.63	1.73	0.30	0.47	1.92	22.51	1.94	0.73	0.96	3.76	363.52	2.82
LIML-val	0.02	0.35	1.66	12.63	1.56	0.04	0.34	1.64	18.12	1.64	0.02	0.35	1.70	14.74	1.65	0.02	0.36	1.74	31.06	1.63
LIML-DN	0.10	0.24	0.88	2.42	0.78	0.13	0.25	0.90	8.35	0.82	0.26	0.33	0.96	14.61	0.97	0.46	0.51	1.23	33.07	1.15
LIML-IR	0.02	0.26	1.15	3.40	0.96	0.05	0.26	1.14	9.12	1.05	0.12	0.31	1.23	15.85	1.01	0.21	0.42	1.55	39.42	1.32
FULL-all	0.08	0.29	1.21	0.53	0.53	0.10	0.30	1.22	0.54	0.54	0.31	0.42	1.40	0.66	0.66	0.69	0.78	2.10	1.04	1.04
FULL-val	0.15	0.26	0.88	0.39	0.39	0.16	0.26	0.88	0.39	0.39	0.15	0.26	0.90	0.39	0.39	0.15	0.27	0.95	0.40	0.40
FULL-DN	0.14	0.22	0.73	0.34	0.34	0.16	0.23	0.73	0.34	0.34	0.29	0.32	0.78	0.43	0.43	0.45	0.47	0.99	0.61	0.61
FULL-IR	0.11	0.23	0.84	0.37	0.37	0.13	0.23	0.84	0.38	0.38	0.20	0.28	0.93	0.43	0.43	0.27	0.36	1.19	0.60	0.60
JIVE-all	0.10	0.46	2.74	28.11	2.35	0.12	0.45	2.66	65.70	2.32	0.33	0.53	2.29	47.49	2.11	0.60	0.73	2.49	94.61	2.25
JIVE-val	0.17	0.60	3.46	410.25	2.63	0.16	0.59	3.34	76.33	2.62	0.16	0.61	3.55	119.05	2.68	0.15	0.62	3.55	136.79	2.69

(Continues)

Table 1. Continued

	$\gamma=\infty$					$\gamma=1$					$\gamma=\frac{1}{2}$					$\gamma=\frac{1}{3}$				
$N=100$	Bias	MAD	IDR	MSE	TMSE	Bias	MAD	IDR	MSE	TMSE	Bias	MAD	IDR	MSE	TMSE	Bias	MAD	IDR	MSE	TMSE
JIVE-DN	0.07	0.30	1.34	407.81	1.48	0.10	0.30	1.33	13.20	1.46	0.26	0.38	1.28	51.21	1.43	0.47	0.54	1.38	135.54	1.49
JIVE-IR	0.03	0.42	2.05	45.67	1.92	0.04	0.41	2.01	71.11	1.93	0.16	0.42	1.76	120.89	1.67	0.32	0.50	1.70	10.57	1.56
HLIM-all	0.04	0.33	1.64	30.80	1.71	0.07	0.34	1.65	24.55	1.74	0.30	0.47	1.96	68.76	1.97	0.74	0.96	3.77	678.44	2.83
HLIM-val	0.02	0.35	1.67	60.36	1.56	0.04	0.34	1.64	11.57	1.67	0.02	0.35	1.71	54.62	1.63	0.02	0.36	1.75	36.67	1.66
HLIM-DN	0.10	0.24	0.89	3.53	0.79	0.13	0.25	0.90	4.92	0.83	0.27	0.33	0.96	4.47	0.98	0.45	0.50	1.22	3.65	1.16
HLIM-IR	0.02	0.26	1.13	6.80	0.49	0.04	0.26	1.14	5.29	0.49	0.16	0.32	1.16	14.49	0.52	0.32	0.46	1.47	5.18	0.68
HFUL-all	0.09	0.28	1.15	0.50	0.50	0.11	0.29	1.18	0.51	0.51	0.32	0.41	1.33	0.63	0.63	0.69	0.75	1.94	0.99	0.99
HFUL-val	0.15	0.26	0.88	0.38	0.38	0.16	0.26	0.88	0.38	0.38	0.15	0.26	0.90	0.39	0.39	0.15	0.27	0.94	0.40	0.40
HFUL-DN	0.14	0.22	0.73	0.33	0.33	0.16	0.23	0.73	0.34	0.34	0.29	0.32	0.78	0.42	0.42	0.45	0.46	0.98	0.60	0.60
HFUL-IR	0.10	0.22	0.83	0.36	0.36	0.13	0.23	0.84	0.36	0.36	0.20	0.28	0.90	0.42	0.42	0.32	0.38	1.11	0.56	0.56

Notes: (i) all—IV estimators using all instruments; (ii) val—using known valid instruments; (iii) DN—using instruments based on DN's criterion $\widehat{L}_{DN}(K)$; (iv) IR—based on the IR criterion $\widehat{L}_{IR}(K)$; and (v) GMM-AVE—averaging GMM estimator in Cheng et al. (2019), which combines a GMM with valid instruments and a GMM using all instruments.

Table 2. Monte Carlo results: $R^2 = 0.01, N = 100$

	$\gamma = \infty$					$\gamma = 1$					$\gamma = \frac{1}{2}$					$\gamma = \frac{1}{3}$				
$N = 100$	Bias	MAD	IDR	MSE	TMSE	Bias	MAD	IDR	MSE	TMSE	Bias	MAD	IDR	MSE	TMSE	Bias	MAD	IDR	MSE	TMSE
OLS	0.50	0.50	0.22	0.50	0.50	0.50	0.50	0.22	0.50	0.50	0.49	0.49	0.23	0.50	0.50	0.49	0.49	0.24	0.50	0.50
2SLS-all	0.48	0.48	0.50	0.52	0.52	0.48	0.48	0.51	0.52	0.52	0.51	0.51	0.54	0.55	0.55	0.56	0.56	0.63	0.61	0.61
2SLS-val	0.37	0.83	4.78	795.38	3.06	0.37	0.86	4.67	93.12	3.06	0.36	0.85	4.80	65.09	3.09	0.35	0.88	4.87	46.08	3.13
2SLS-DN	0.44	0.52	1.51	15.88	1.86	0.44	0.52	1.54	87.60	1.85	0.51	0.59	1.57	46.74	1.83	0.60	0.69	1.91	15.17	2.05
2SLS-IR	0.43	0.50	1.38	795.09	1.41	0.44	0.51	1.42	11.99	1.44	0.50	0.58	1.53	18.14	1.49	0.58	0.67	1.77	14.70	1.59
GMM-AVE	0.44	0.46	0.92	0.58	0.58	0.45	0.46	0.92	0.58	0.58	0.47	0.48	0.97	0.60	0.60	0.49	0.51	1.05	0.64	0.64
LIML-all	0.37	0.88	4.90	43.92	3.07	0.40	0.87	4.80	57.82	3.11	0.61	1.15	5.97	389.16	3.40	1.28	2.36	12.47	135.58	4.78
LIML-val	0.37	0.83	4.78	795.38	3.06	0.37	0.86	4.67	93.12	3.06	0.36	0.85	4.80	65.09	3.09	0.35	0.88	4.87	46.08	3.13
LIML-DN	0.40	0.62	2.80	18.58	2.40	0.39	0.64	2.88	88.53	2.42	0.50	0.71	2.88	51.43	2.43	0.63	0.90	3.39	34.76	2.65
LIML-IR	0.38	0.65	3.08	1437.86	2.50	0.39	0.67	3.19	95.35	2.62	0.50	0.76	3.50	59.06	2.68	0.68	1.04	5.38	55.58	3.41
FULL-all	0.39	0.59	2.08	0.89	0.89	0.42	0.60	2.03	0.89	0.89	0.58	0.75	2.31	1.03	1.03	1.01	1.23	3.27	1.46	1.46
FULL-val	0.46	0.47	0.92	0.58	0.58	0.46	0.46	0.92	0.58	0.58	0.45	0.46	0.95	0.58	0.58	0.45	0.47	0.98	0.59	0.59
FULL-DN	0.45	0.46	0.93	0.58	0.58	0.45	0.46	0.93	0.58	0.58	0.49	0.50	0.98	0.63	0.63	0.54	0.56	1.25	0.76	0.76
FULL-IR	0.45	0.47	1.01	0.61	0.61	0.44	0.47	1.01	0.62	0.62	0.50	0.51	1.09	0.68	0.68	0.55	0.59	1.49	0.87	0.87
JIVE-all	0.48	0.77	3.66	48.10	2.66	0.49	0.78	3.63	280.29	2.66	0.48	0.82	3.90	1159.16	2.84	0.51	0.91	4.75	104.24	3.08
JIVE-val	0.51	0.60	2.22	313.13	2.17	0.50	0.59	2.27	26.86	2.19	0.50	0.60	2.40	22.85	2.25	0.50	0.61	2.51	62.93	2.28

(Continues)

Table 2. Continued

	$\gamma = \infty$					$\gamma = 1$					$\gamma = \frac{1}{2}$					$\gamma = \frac{1}{3}$				
$N = 100$	Bias	MAD	IDR	MSE	TMSE	Bias	MAD	IDR	MSE	TMSE	Bias	MAD	IDR	MSE	TMSE	Bias	MAD	IDR	MSE	TMSE
JIVE-DN	0.48	0.59	2.16	312.95	2.09	0.48	0.58	2.09	15.17	2.07	0.51	0.61	2.19	22.89	2.14	0.55	0.67	2.51	63.06	2.26
JIVE-IR	0.49	0.58	1.93	357.42	1.95	0.49	0.57	1.89	18.40	1.95	0.51	0.58	2.00	16.12	2.02	0.53	0.61	2.24	100.66	2.14
HLIM-all	0.38	0.88	4.93	170.64	3.07	0.39	0.87	4.84	70.73	3.09	0.61	1.13	5.98	351.77	3.41	1.27	2.34	12.18	310.14	4.76
HLIM-val	0.36	0.84	4.70	105.46	3.05	0.36	0.86	4.56	104.97	3.03	0.36	0.86	4.78	34.48	3.09	0.36	0.88	4.82	370.69	3.14
HLIM-DN	0.39	0.62	2.75	103.63	2.40	0.39	0.64	2.87	102.95	2.39	0.49	0.71	2.86	18.37	2.44	0.64	0.90	3.38	65.92	2.64
HLIM-IR	0.39	0.62	2.63	34.06	0.85	0.40	0.63	2.66	64.96	0.86	0.50	0.72	2.80	43.60	0.91	0.66	1.00	4.02	299.53	1.04
HFUL-all	0.41	0.56	1.90	0.84	0.84	0.42	0.56	1.87	0.83	0.83	0.57	0.70	2.09	0.97	0.97	0.96	1.12	2.96	1.34	1.34
HFUL-val	0.46	0.46	0.91	0.57	0.57	0.45	0.46	0.91	0.57	0.57	0.45	0.46	0.94	0.58	0.58	0.45	0.46	0.97	0.59	0.59
HFUL-DN	0.45	0.46	0.92	0.57	0.57	0.44	0.46	0.92	0.57	0.57	0.49	0.49	0.96	0.62	0.62	0.54	0.55	1.20	0.73	0.73
HFUL-IR	0.45	0.46	0.93	0.58	0.58	0.45	0.46	0.93	0.58	0.58	0.48	0.49	0.97	0.62	0.62	0.52	0.54	1.23	0.73	0.73

Notes: (i) all—IV estimators using all instruments; (ii) val—using known valid instruments; (iii) DN—using instruments based on DN's criterion $\widehat{L}_{DN}(K)$; (iv) IR—based on the IR criterion $\widehat{L}_{IR}(K)$; and (v) GMM-AVE—averaging GMM estimator in Cheng et al. (2019), which combines a GMM with valid instruments and a GMM using all instruments.

Table 3. Monte Carlo results: $R^2 = 0.1, N = 1{,}000$

	$\gamma = \infty$					$\gamma = 1$					$\gamma = \frac{1}{2}$					$\gamma = \frac{1}{3}$				
$N = 1{,}000$	Bias	MAD	IDR	MSE	TMSE	Bias	MAD	IDR	MSE	TMSE	Bias	MAD	IDR	MSE	TMSE	Bias	MAD	IDR	MSE	TMSE
OLS	0.45	0.45	0.07	0.45	0.45	0.45	0.45	0.07	0.45	0.45	0.45	0.45	0.07	0.45	0.45	0.45	0.45	0.07	0.45	0.45
2SLS-all	0.10	0.11	0.20	0.13	0.13	0.11	0.11	0.20	0.13	0.13	0.20	0.20	0.20	0.22	0.22	0.41	0.41	0.21	0.42	0.42
2SLS-val	−0.00	0.13	0.50	0.21	0.21	−0.00	0.13	0.49	0.21	0.21	−0.00	0.13	0.51	0.21	0.21	−0.00	0.13	0.50	0.21	0.21
2SLS-DN	0.04	0.07	0.24	0.10	0.10	0.05	0.07	0.24	0.10	0.10	0.15	0.15	0.24	0.17	0.17	0.40	0.40	0.24	0.41	0.41
2SLS-IR	−0.01	0.08	0.32	0.14	0.14	−0.00	0.08	0.32	0.13	0.13	0.05	0.10	0.35	0.15	0.15	0.13	0.15	0.39	0.20	0.20
GMM-AVE	0.05	0.10	0.37	0.16	0.16	0.05	0.10	0.36	0.16	0.16	0.07	0.13	0.42	0.18	0.18	0.06	0.13	0.49	0.20	0.20
LIML-all	0.00	0.07	0.27	0.11	0.11	0.00	0.07	0.27	0.11	0.11	0.12	0.13	0.27	0.16	0.16	0.39	0.39	0.34	0.90	0.43
LIML-val	−0.00	0.13	0.50	0.21	0.21	−0.00	0.13	0.49	0.21	0.21	−0.00	0.13	0.51	0.21	0.21	−0.00	0.13	0.50	0.21	0.21
LIML-DN	0.01	0.07	0.25	0.10	0.10	0.02	0.07	0.25	0.10	0.10	0.13	0.13	0.25	0.16	0.16	0.37	0.37	0.29	0.39	0.39
LIML-IR	−0.02	0.08	0.31	0.14	0.14	−0.02	0.08	0.31	0.14	0.14	0.04	0.10	0.36	0.15	0.15	0.12	0.15	0.39	0.19	0.19
FULL-all	0.01	0.07	0.27	0.10	0.10	0.01	0.07	0.27	0.11	0.11	0.12	0.13	0.26	0.16	0.16	0.39	0.39	0.34	0.41	0.41
FULL-val	0.02	0.12	0.47	0.19	0.19	0.02	0.12	0.46	0.19	0.19	0.01	0.12	0.47	0.19	0.19	0.01	0.12	0.47	0.19	0.19
FULL-DN	0.02	0.07	0.25	0.10	0.10	0.02	0.07	0.25	0.10	0.10	0.13	0.13	0.24	0.16	0.16	0.37	0.37	0.28	0.39	0.39
FULL-IR	−0.01	0.07	0.30	0.13	0.13	−0.01	0.08	0.30	0.13	0.13	0.05	0.10	0.35	0.15	0.15	0.13	0.15	0.37	0.19	0.19
JIVE-all	−0.01	0.08	0.29	0.12	0.12	−0.01	0.08	0.29	0.12	0.12	0.12	0.13	0.28	0.16	0.16	0.40	0.40	0.28	0.42	0.42
JIVE-val	−0.04	0.14	0.58	0.30	0.30	−0.04	0.14	0.57	0.33	0.30	−0.04	0.14	0.59	0.34	0.32	−0.04	0.14	0.58	0.30	0.29

(Continues)

TABLE 3. Continued

	$\gamma=\infty$					$\gamma=1$					$\gamma=\frac{1}{2}$					$\gamma=\frac{1}{3}$				
$N=1{,}000$	Bias	MAD	IDR	MSE	TMSE	Bias	MAD	IDR	MSE	TMSE	Bias	MAD	IDR	MSE	TMSE	Bias	MAD	IDR	MSE	TMSE
JIVE-DN	0.00	0.07	0.26	0.10	0.10	0.01	0.07	0.26	0.10	0.10	0.12	0.13	0.26	0.15	0.15	0.38	0.38	0.26	0.39	0.39
JIVE-IR	−0.07	0.11	0.48	0.29	0.28	−0.06	0.11	0.48	0.32	0.28	−0.04	0.13	0.55	0.28	0.28	0.05	0.15	0.52	0.22	0.22
HLIM-all	0.00	0.07	0.27	0.11	0.11	0.00	0.07	0.27	0.11	0.11	0.12	0.13	0.27	0.16	0.16	0.39	0.39	0.34	0.78	0.43
HLIM-val	−0.00	0.13	0.50	0.21	0.21	−0.00	0.13	0.49	0.21	0.21	−0.00	0.13	0.50	0.21	0.21	−0.00	0.13	0.50	0.21	0.21
HLIM-DN	0.01	0.07	0.25	0.10	0.10	0.02	0.07	0.25	0.10	0.10	0.13	0.13	0.25	0.16	0.16	0.37	0.37	0.29	0.39	0.39
HLIM-IR	−0.04	0.09	0.40	0.18	0.18	−0.04	0.09	0.40	0.18	0.18	−0.01	0.12	0.46	0.20	0.20	0.08	0.14	0.46	0.20	0.20
HFUL-all	0.01	0.07	0.27	0.10	0.10	0.01	0.07	0.27	0.11	0.11	0.12	0.13	0.26	0.16	0.16	0.39	0.39	0.34	0.41	0.41
HFUL-val	0.02	0.12	0.47	0.19	0.19	0.02	0.12	0.46	0.19	0.19	0.01	0.12	0.47	0.19	0.19	0.01	0.12	0.47	0.19	0.19
HFUL-DN	0.02	0.07	0.25	0.10	0.10	0.02	0.07	0.25	0.10	0.10	0.13	0.13	0.24	0.16	0.16	0.37	0.37	0.28	0.39	0.39
HFUL-IR	−0.02	0.09	0.37	0.16	0.16	−0.02	0.09	0.36	0.16	0.16	0.01	0.11	0.43	0.18	0.18	0.08	0.14	0.44	0.19	0.19

Notes: (i) all—IV estimators using all instruments; (ii) val—using known valid instruments; (iii) DN—using instruments based on DN's criterion $\widehat{L}_{DN}(K)$; (iv) IR—based on the IR criterion $\widehat{L}_{IR}(K)$; and (v) GMM-AVE—averaging GMM estimator in Cheng et al. (2019), which combines a GMM with valid instruments and a GMM using all instruments.

Table 4. Monte Carlo results: $R^2 = 0.01, N = 1,000$

	$\gamma = \infty$					$\gamma = 1$					$\gamma = \frac{1}{2}$					$\gamma = \frac{1}{3}$				
$N = 1,000$	Bias	MAD	IDR	MSE	TMSE	Bias	MAD	IDR	MSE	TMSE	Bias	MAD	IDR	MSE	TMSE	Bias	MAD	IDR	MSE	TMSE
OLS	0.49	0.49	0.07	0.50	0.50	0.50	0.50	0.07	0.50	0.50	0.49	0.49	0.07	0.50	0.50	0.49	0.49	0.07	0.50	0.50
2SLS-all	0.37	0.37	0.37	0.40	0.40	0.38	0.38	0.37	0.40	0.40	0.48	0.48	0.36	0.50	0.50	0.70	0.70	0.47	0.73	0.73
2SLS-val	0.05	0.43	2.18	15.12	2.01	0.06	0.43	2.17	70.62	2.03	0.05	0.43	2.18	28.35	2.05	0.07	0.44	2.16	13.06	2.03
2SLS-DN	0.28	0.30	0.65	0.97	0.61	0.28	0.31	0.65	1.01	0.51	0.45	0.46	0.58	2.58	0.63	0.77	0.78	0.71	5.98	1.03
2SLS-IR	0.18	0.24	0.78	1.74	0.64	0.19	0.24	0.78	12.10	0.62	0.29	0.34	0.86	0.78	0.64	0.50	0.60	1.38	1.97	0.90
GMM-AVE	0.25	0.30	0.79	0.40	0.40	0.25	0.30	0.78	0.40	0.40	0.30	0.34	0.85	0.43	0.43	0.35	0.40	1.03	0.53	0.53
LIML-all	0.05	0.38	1.93	13.51	1.90	0.07	0.37	1.92	25.43	1.88	0.43	0.60	2.44	20.09	2.27	2.29	2.62	8.39	169.41	4.61
LIML-val	0.05	0.43	2.18	15.12	2.01	0.06	0.43	2.17	70.62	2.03	0.05	0.43	2.18	28.35	2.05	0.07	0.44	2.16	13.06	2.03
LIML-DN	0.13	0.26	0.98	3.23	1.05	0.14	0.26	0.96	14.49	0.92	0.33	0.40	1.06	4.19	1.08	0.88	0.92	1.78	11.37	1.75
LIML-IR	0.07	0.26	1.15	10.50	1.07	0.08	0.26	1.14	10.55	1.01	0.22	0.32	1.20	9.35	1.12	0.62	0.75	2.49	16.30	2.00
FULL-all	0.08	0.33	1.39	0.60	0.60	0.11	0.33	1.39	0.61	0.61	0.44	0.53	1.64	0.79	0.79	1.91	1.93	3.05	2.09	2.09
FULL-val	0.22	0.30	0.94	0.43	0.43	0.22	0.30	0.93	0.43	0.43	0.22	0.30	0.94	0.43	0.43	0.22	0.31	0.96	0.44	0.44
FULL-DN	0.17	0.24	0.77	0.36	0.36	0.18	0.24	0.76	0.36	0.36	0.35	0.37	0.82	0.48	0.48	0.80	0.80	1.37	0.99	0.99
FULL-IR	0.16	0.24	0.87	0.41	0.41	0.17	0.24	0.88	0.42	0.42	0.28	0.32	0.94	0.51	0.51	0.53	0.60	1.88	1.03	1.03
JIVE-all	0.14	0.51	2.99	40.03	2.48	0.15	0.52	3.05	437.76	2.47	0.46	0.63	2.50	40.66	2.30	1.11	1.32	4.71	465.87	3.22
JIVE-val	0.36	0.68	3.71	195.36	2.75	0.35	0.67	3.64	207.72	2.69	0.35	0.67	3.78	20.79	2.73	0.36	0.67	3.72	43.14	2.74

(Continues)

https://doi.org/10.1017/S0266466622000597 Published online by Cambridge University Press

Table 4. Continued

	$\gamma = \infty$					$\gamma = 1$					$\gamma = \frac{1}{2}$					$\gamma = \frac{1}{3}$				
$N = 1{,}000$	Bias	MAD	IDR	MSE	TMSE	Bias	MAD	IDR	MSE	TMSE	Bias	MAD	IDR	MSE	TMSE	Bias	MAD	IDR	MSE	TMSE
JIVE-DN	0.13	0.33	1.46	62.06	1.64	0.14	0.34	1.48	204.27	1.65	0.36	0.46	1.38	17.28	1.56	0.81	0.86	1.70	27.22	1.80
JIVE-IR	0.17	0.41	1.66	60.57	1.67	0.17	0.41	1.67	29.13	1.68	0.33	0.44	1.36	13.19	1.40	0.57	0.63	1.63	41.60	1.60
HLIM-all	0.04	0.38	1.94	37.21	1.91	0.07	0.37	1.92	1103.02	1.90	0.43	0.61	2.44	70.81	2.27	2.29	2.62	8.29	103.01	4.61
HLIM-val	0.05	0.43	2.19	21.35	2.00	0.06	0.43	2.16	81.26	2.04	0.05	0.43	2.18	29.00	2.02	0.06	0.44	2.19	25.14	2.03
HLIM-DN	0.13	0.26	0.98	14.72	1.05	0.14	0.26	0.96	77.75	0.94	0.33	0.40	1.06	26.29	1.08	0.88	0.92	1.79	11.23	1.77
HLIM-IR	0.10	0.26	1.12	15.19	0.49	0.11	0.26	1.10	8.05	0.49	0.29	0.36	1.12	7.22	0.57	0.74	0.89	2.74	24.48	0.99
HFUL-all	0.08	0.33	1.38	0.59	0.59	0.11	0.33	1.38	0.60	0.60	0.44	0.53	1.62	0.78	0.78	1.90	1.92	3.00	2.07	2.07
HFUL-val	0.22	0.30	0.94	0.43	0.43	0.22	0.30	0.93	0.43	0.43	0.22	0.30	0.94	0.43	0.43	0.22	0.31	0.96	0.44	0.44
HFUL-DN	0.17	0.24	0.76	0.36	0.36	0.19	0.24	0.75	0.36	0.36	0.35	0.37	0.82	0.48	0.48	0.78	0.79	1.35	0.94	0.94
HFUL-IR	0.17	0.25	0.84	0.39	0.39	0.18	0.25	0.84	0.39	0.39	0.30	0.33	0.87	0.47	0.47	0.53	0.59	1.67	0.93	0.93

Notes: (i) all—IV estimators using all instruments; (ii) val—using known valid instruments; (iii) DN—using instruments based on DN's criterion $\widehat{L}_{DN}(K)$; (iv) IR—based on the IR criterion $\widehat{L}_{IR}(K)$; and (v) GMM-AVE—averaging GMM estimator in Cheng et al. (2019), which combines a GMM with valid instruments and a GMM using all instruments.

Table 5. Monte Carlo results: Median of $\widehat{K}$

			$\gamma = \infty$			$\gamma = 1$			$\gamma = \frac{1}{2}$			$\gamma = \frac{1}{3}$		
N	R^2		K^{MSE}	DN	IR	K^{MSE}	DN	IR	K^{MSE}	DN	IR	K^{MSE}	DN	IR
100	0.1	2SLS	7	4	3	7	4	3	4	5	2	20	5	2
		LIML	4	3	3	5	3	3	4	3	2	3	3	2
		FULL	3	4	2	4	4	2	1	3	2	1	3	2
		JIVE	5	3	2	6	3	2	6	3	2	4	3	2
		HLIM	4	3	2	5	3	2	4	3	2	3	3	3
		HFUL	3	3	2	4	3	2	1	3	2	1	3	2
100	0.01	2SLS	20	3	3	20	3	3	20	3	3	20	3	3
		LIML	2	1	2	4	1	2	3	1	2	1	1	2
		FULL	1	1	1	1	1	1	1	1	1	1	1	1
		JIVE	1	1	1	1	1	1	1	1	1	1	1	1
		HLIM	2	1	2	3	1	2	3	1	2	1	1	2
		HFUL	1	1	1	1	1	1	1	1	1	1	1	1
1,000	0.1	2SLS	9	8	7	9	8	7	4	9	5	2	13	3
		LIML	12	11	9	12	11	9	4	10	6	2	10	3
		FULL	12	11	8	12	11	8	4	10	5	1	10	3
		JIVE	11	10	1	10	10	1	5	10	1	2	10	1
		HLIM	12	11	1	12	11	1	4	10	1	2	10	1
		HFUL	12	11	1	12	11	1	4	10	1	1	10	1
1,000	0.01	2SLS	8	7	5	8	7	5	5	9	4	30	8	3
		LIML	6	4	4	8	4	4	5	4	3	1	4	2
		FULL	4	4	4	4	4	4	1	4	4	1	4	2
		JIVE	8	3	3	9	3	3	7	3	4	3	3	2
		HLIM	6	4	5	8	4	5	5	4	5	1	4	3
		HFUL	4	4	3	4	4	3	1	3	4	1	4	2

paper (IR). Additionally, Table 5 reports the median value of $\widehat{K}$ and K^{MSE} that minimizes the trimmed MSE of the estimators.[16]

6.1. MSE, TMSE, and IDR

In terms of the MSEs, our findings are in line with the literature (Hahn et al., 2004; Guggenberger, 2008; Hausman et al., 2012), which recommends utilizing estimators with finite-sample moments over the "no-moment" estimators. The MSE/TMSEs of LIML/HLIM and JIVE2 are considerably larger than those of

[16] Note that this is a different measure with K^* reported in DN (Tables V and VI), which is a median of $\arg\min_{K \in \mathcal{K}} L(K)$.

https://doi.org/10.1017/S0266466622000597 Published online by Cambridge University Press

FULL/HFUL with many instruments regardless of the instrument invalidity γ, especially in the weakly identified cases (Tables 2 and 4). We observe that MSE can be significantly larger than the trimmed MSE for "no-moment" estimators, but MSE for the Fuller/HFUL estimator is the same as trimmed MSE in all cases.

It is true that the MSE approximations derived in this paper may perform poorly for "no-moment" estimators, especially in the weak instrument scenarios (Hahn et al., 2004). However, we find that the FULL/HFUL estimators combined with the DN criterion can lead to a reduction in the IDR and MSE/TMSE compared to those of the estimators using all instrument sets or only valid instruments, even when the instruments are slightly invalid and the correct specification case ($\gamma = \infty$). This is consistent with our theory that the DN criterion is asymptotically optimal under "slightly invalid" ($\gamma > \frac{1}{2}$) instruments. The FULL/HFUL estimators based on the IR criterion have similar MSE to those of DN across different γ, and DN/IR may include a few more invalid but strong instruments for MSE reductions. We find that selecting estimators based on these criteria can mitigate the moment problem even for LIML/HLIM/JIVE2, with significant TMSE reductions, although this is not always the case.

For FULL/HFUL, the MSE/TMSE typically increases when γ or R^2 decreases. The IDR/TMSE of the IV estimators with strong invalid instruments (Table 1 with $R^2 = 0.1, \gamma = \frac{1}{3}$) and valid weak instruments (Table 2 with $R^2 = 0.01, \gamma = \infty$) can be similar. We also find that the model averaging estimator can help to mitigate the moment problem by effectively choosing weights between a conservative and an aggressive GMM estimator, and the MSE of GMM-AVE works quite well across simulations. Because 2SLS has finite moments for $K \geq 2$, the MSE of 2SLS when using only valid instruments is considerably larger than those of DN/IR as well as using all instruments.

With larger sample sizes $N = 1,000$, the MSE/TMSEs (IDR) of LIML/HLIM based on the DN/IR criteria are very similar to those of FULL/HFUL across γ in the relatively strongly identified case (Table 3), but not in the weakly identified case (Table 4).

6.2. Median Bias and Median Absolute Deviation

We first compare the IV estimators using all instruments ($K = \bar{K}$). We find that the median bias and MAD are very similar, with the exception that the bias of 2SLS can be substantially larger than those of the other estimators in the strongly identified cases when the instruments are valid or slightly invalid. Similar performances for 2SLS/LIML/FULL under valid instruments can also be found in Hahn et al. (2004). However, the bias/MAD of 2SLS can be lower than those of the other estimators, especially when the degree of invalidity is large ($\gamma = \frac{1}{3}$) because the misspecification bias and the many-instrument bias can have opposite signs and offset each other, as theoretically expected (Remark 3.4). The median bias and MAD of LIML/HLIM under many instruments can be significantly larger than those of FULL/HFUL when the instruments are invalid ($\gamma = \frac{1}{3}$), especially in the weakly identified cases (Table 2).

https://doi.org/10.1017/S0266466622000597 Published online by Cambridge University Press

The DN criterion tends to choose more invalid instruments when the degree of invalidity increases, which leads to a large estimator bias that increases as γ decreases, although the bias decreases when the sample size increases to $N = 1,000$. In general, we find that an estimator based on IR has a lower median bias than an estimator with DN, and it has similar or slightly larger MAD than that obtained using only valid instruments across γ.

In sum, the Fuller and HFUL estimators combined with instrument selection perform very well, with the lowest IDR and MSE/TMSE. For FULL/HFUL, the median of the selected $\widehat{K}$ based on DN or IR is close to K^{MSE} in many cases. Although it is highlighted in the literature that the Fuller or HFUL estimator performs well under many weak instrument setups, our simulation suggests that those estimators combined with instrument selection procedures can also perform well when the instruments are potentially invalid in finite samples.

Remark 6.1. Potentially interesting future research involves the investigation of diagnostic tools that provide empirical researchers with the ability to evaluate the degree of invalidity of a set of available instruments. In practice, researchers can compare estimates with instruments chosen by DN and IR criteria. In our simulation experiments, we find that the differences in the median biases of the 2SLS estimators and the rejection rates of the Sargan–Hansen J-test based on DN and IR tend to increase when the degree of invalidity increases.[17] However, this is not a formal way of distinguishing between the cases with $\gamma = \frac{1}{2}$, $\gamma > \frac{1}{2}$, and $\gamma < \frac{1}{2}$. We leave this topic for future research.

7. CONCLUSIONS

This paper develops higher-order MSE approximations for IV estimators in a linear homoskedastic IV model with many and possibly invalid instruments. We consider various k-class estimators, including 2SLS, LIML, FULL, B2SLS, JIVE2, HLIM, and HFUL. Based on the higher-order MSE approximations, we consider the instrument selection criteria that can be used to choose among the set of available instruments. We demonstrate the asymptotic optimality of the instrument selection criteria in DN under locally $(N^{-\gamma}, \gamma > \frac{1}{2})$ invalid instrument setups. We also propose instrument selection criteria that are applicable when researchers have a "conservative" set of valid instruments by considering additional higher-order terms due to $N^{-1/2}$ locally invalid instruments.

There are some limitations in the present work and scope for future extensions. First, it would be interesting to extend the results for a case involving heteroskedastic error terms. The assumption of conditional homoskedasticity greatly simplifies calculations and helps to investigate higher-order comparisons between estimators. It would be of interest to analyze locally invalid instrument specifications combined with the existing work of Donald et al. (2009) in a GMM setup.

[17] We further note that the conventional overidentification tests may perform poorly when the number of instruments is large (Lee and Okui, 2012; Chao et al., 2014).

https://doi.org/10.1017/S0266466622000597 Published online by Cambridge University Press

Second, our higher-order results do not rely on the weak- or many-weak-instrument asymptotics of Staiger and Stock (1997) and Chao and Swanson (2005), and the MSE approximations in this paper are not valid under such sequences. Here, we consider the scenario where the number of instruments is small relative to the sample size and the instruments are strong, and we believe that this case is important in many empirical applications (see Hansen, Hausman, and Newey, 2008, for a survey of applied microeconomic applications). Our simulation evidence suggests that the proposed instrument selection criteria can be useful for estimators that perform well under many weak instrument setups, such as FULL/HFUL.

Finally, our instrument selection criteria can suffer from post-model-selection problems. Simulation evidence (in Section 6) shows that a selection estimator can outperform an estimator using only valid instruments for some parts of the parameter spaces (mostly where $\gamma > \frac{1}{2}$); the worst-case MSE of the instrument selection procedure increases when γ decreases and can be much worse than that obtained using only valid instruments when the degree of invalidity is large (small γ). Although we justify the choice of $\widehat{K}$ by asymptotic optimality or unbiasedness over large parameter spaces for γ, the randomness in $\widehat{K}$ may not be fully accounted for in the high-order MSE approximation of $\widehat{\delta}(\widehat{K})$. The worst-case (scaled) MSE of the estimator based on the DN procedures can diverge to infinity as the sample size grows when $\gamma < \frac{1}{2}$, as the criteria concern only the bias and variance from many instruments while ignoring the bias from invalid instruments. The MSE of an estimator based on IR procedures can also increase as γ decreases because poor estimation of $g(\cdot)$ can worsen performance. Similar undesirable (related to nonuniformity) properties of the post model selection estimators can be found in Leeb and Pötscher (2005, 2008). In a related study on the pretesting issue, Guggenberger and Kumar (2012) investigate the negative impact of an overidentification pretest on the subsequent inference and show that the asymptotic size of the second-stage test can be equal to one. A recent paper Cheng et al. (2019) has shown that the averaging estimator can have uniformly lower asymptotic risk than an estimator using only the valid moment conditions in the GMM setup.

SUPPLEMENTARY MATERIAL

Kang, B. (2022) Supplement to "Higher-Order Approximation of IV Estimators with Invalid Instruments," Econometric Theory Supplementary Material. To view, please visit: https://doi.org/10.1017/S0266466622000597.

REFERENCES

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

Anderson, T.W. & 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.

https://doi.org/10.1017/S0266466622000597 Published online by Cambridge University Press

Andrews, D.W.K. & 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. (2019) On the structure of IV Estimands. *Journal of Econometrics* 211, 294–307.

Andrews, I., M. Gentzkow, & J.M. Shapiro (2017) Measuring the sensitivity of parameter estimates to estimation moments. *Quarterly Journal of Economics* 132, 1553–1592.

Angrist, J.D., G.W. Imbens, & A.B. Krueger (1999) Jackknife instrumental variables estimation. *Journal of Applied Econometrics* 14, 57–67.

Armstrong, T. & M. Kolesár (2021) Sensitivity analysis using approximate moment condition models. *Quantitative Economics* 12, 77–108.

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

Bekker, P.A. & J. van der Ploeg (2005) Instrumental variable estimation based on grouped data. *Statistica Neerlandica* 59, 239–267.

Belloni, A., D. Chen, V. Chernozhukov, & C. Hansen (2012) Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica* 80, 2369–2430.

Berkowitz, D., M. Caner, & Y. Fang (2008) Are nearly exogenous instruments reliable? *Economics Letters* 101, 20–23.

Berkowitz, D., M. Caner, & Y. Fang (2012) The validity of instruments revisited. *Journal of Econometrics* 166, 255–266.

Bonhomme, S. & M. Weidner (2022) Minimizing sensitivity to model misspecification. *Quantitative Economics* 13, 907–954.

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, & 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.

Chao, J.C., J. Hausman, W.K. Newey, N.R. Swanson, & T. Woutersen (2014) Testing overidentifying restrictions with many instruments and heteroskedasticity. *Journal of Econometrics* 178, 15–21.

Chao, J.C. & N.R. Swanson (2005) Consistent estimation with a large number of weak instruments. *Econometrica* 73, 1673–1692.

Chao, J.C., N.R. Swanson, J. Hausman, W.K. Newey, & T. Woutersen (2012) Asymptotic distribution of JIVE in a heteroskedastic IV regression with many instruments. *Econometric Theory* 28, 42–86.

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

Cheng, X., Z. Liao, & R. Shi (2019) On uniform asymptotic risk of averaging GMM estimator. *Quantitative Economics* 10, 931–979.

Chetty, R., J.N. Friedman, N. Hilger, E. Saez, D.W. Schanzenbach, & 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, & 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, & W.K. Newey (2009) Choosing instrumental variables in conditional moment restriction models. *Journal of Econometrics* 152, 28–36.

Donald, S.G. & W.K. Newey (1999) Choosing the Number of Instruments. Working paper 99-05, Department of Economics, Massachusetts Institute of Technology. https://dspace.mit.edu/bitstream/handle/1721.1/63410/choosingnumberof00dona.pdf?sequence=1.

https://doi.org/10.1017/S0266466622000597 Published online by Cambridge University Press

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

Evdokimov, K. & M. Kolesár (2019) Inference in Instrumental Variable Regression Analysis with Heterogeneous Treatment Effects. Working paper.

Fuller, W.A. (1977) Some properties of a modification of the limited information estimator. *Econometrica* 45, 939–953.

Guggenberger, P. (2008) Finite sample evidence suggesting a heavy tail problem of the generalized empirical likelihood estimator. *Econometric Reviews* 27, 526–541.

Guggenberger, P. (2012) On the asymptotic size distortion of tests when instruments locally violate the exogeneity assumption. *Econometric Theory* 28, 387–421.

Guggenberger, P. & G. Kumar (2012) On the size distortion of tests after an overidentifying restrictions pretest. *Journal of Applied Econometrics* 27, 1138–1160.

Hahn, J. & 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. *The Econometrics Journal* 7, 272–306.

Hall, A.R. & A. Inoue (2003) The large sample behavior of the generalized method of moments estimator in misspecified models. *Journal of Econometrics* 114(2), 361–394.

Hall, A.R. & F.P.M. Peixe (2003) A consistent method for the selection of relevant instruments. *Econometric Reviews* 22, 269–287.

Hansen, C., J. Hausman, & W.K. Newey (2008) Estimation with many instrumental variables. *Journal of Business and Economic Statistics* 26, 398–422.

Hausman, J., R. Lewis, K. Menzel, & W.K. Newey (2011) Properties of the CUE estimator and a modification with moments. *Journal of Econometrics* 165, 45–57.

Hausman, J., W.K. Newey, T. Woutersen, J.C. Chao, & N.R. Swanson (2012) Instrumental variable estimation with heteroskedasticity and many instruments. *Quantitative Economics* 3, 211–255.

Hong, H., B. Preston, & M. Shum (2003) Generalized empirical likelihood-based model selection criteria for moment condition models. *Econometric Theory* 19, 923–943.

Imbens, G.W. & J.D. Angrist (1994) Identification and estimation of local average treatment effects. *Econometrica* 62, 467–475.

Kang, H., A. Zhang, T.T. Cai, & 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.

Kinal, T. (1980) The existence of moments of *k*-class estimators. *Econometrica* 48, 241–249.

Kitamura, Y., T. Otsu, & K. Evdokimov (2013) Robustness, infinitesimal neighborhoods, and moment restrictions. *Econometrica* 81, 1185–1201.

Kolesár, M (2013) Estimation in an Instrumental Variables Model with Treatment Effect Heterogeneity. Working paper. https://www.princeton.edu/~mkolesar/papers/late_estimation.pdf.

Kolesár, M., R. Chetty, J. Friedman, E. Glaeser, & 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. & R. Okui (2010) Constructing optimal instruments by first-stage prediction averaging. *Econometrica* 78, 697–718.

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

Lee, Y. & Y. Zhou (2015) Averaged Instrumental Variables Estimators. Working paper. https://ylee41.expressions.syr.edu/wp-content/uploads/W_AvgIV_LeeZhou.pdf.

Leeb, H. & B.M. Pötscher (2005) Model selection and inference: Facts and fiction. *Econometric Theory* 21, 21–59.

https://doi.org/10.1017/S0266466622000597 Published online by Cambridge University Press

Leeb, H. & B.M. Pötscher (2008) Sparse estimators and the oracle property, or the return of Hodges' estimator. *Journal of Econometrics* 142, 201–211.

Li, K.C. (1987) Asymptotic optimality for C_p, C_L, cross-validation and generalized cross-validation: Discrete index set. *Annals of Statistics* 15, 958–975.

Liao, Z. (2013) Adaptive GMM shrinkage estimation with consistent moment selection. *Econometric Theory* 29, 857–904.

Maasoumi, E. & P.C.B. Phillips (1982) On the behavior of inconsistent instrumental variable estimators. *Journal of Econometrics* 19, 183–201.

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

Mariano, R.S. & T. Sawa (1972) The exact finite-sample distribution of the limited-information maximum likelihood estimator in the case of two included endogenous variables. *Journal of the American Statistical Association* 67, 159–163.

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. & 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.

Newey, W.K. (1997) Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics* 79, 147–168.

Newey, W.K. & R.J. Smith (2004) Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* 72, 219–255.

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

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.

Pfanzagl, J. & W. Wefelmeyer (1978) A third-order optimum property of the maximum likelihood estimator. *Journal of Multivariate Analysis* 8, 1–29.

Phillips, G. & C. Hale (1977) The bias of instrumental variable estimators of simultaneous equation systems. *International Economic Review* 18, 219–228.

Phillips, P.C.B. (1980) The exact distribution of instrumental variable estimators in an equation containing $n+1$ endogenous variables. *Econometrica* 48(4), 861–878.

Phillips, P.C.B. (1983) Exact small sample theory in the simultaneous equations model. In Z. Griliches & M.D. Intriligator (eds.), *Handbook of Econometrics*, vol. 1, pp. 449–516. North-Holland.

Phillips, P.C.B. (2003) Vision and influence in econometrics: John Denis Sargan. *Econometric Theory* 19, 495–511.

Rothenberg, T.J. (1984) Approximating the distributions of econometric estimators and test statistics. In Z. Griliches & M.D. Intriligator (eds.), *Handbook of Econometrics*, vol. 2, pp. 881–935. Elsevier.

Sargan, J.D. (1982) On Monte Carlo estimates of moments that are infinite. In R.I. Basmann & G.F. Rhodes (eds.), *Advances in Econometrics: A Research Annual*, pp. 267–299. JAI Press.

Schennach, S.M. (2007) Point estimation with exponentially tilted empirical likelihood. *Annals of Statistics* 35, 634–672.

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

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

Windmeijer, F., H. Farbmacher, N. Davies, & G.D. Smith (2018) On the use of the Lasso for instrumental variables estimation with some invalid instruments. *Journal of the American Statistical Association* 114, 1339–1350.

https://doi.org/10.1017/S0266466622000597 Published online by Cambridge University Press