Alternative Approximations of the Bias and MSE of the IV Estimator Under Weak Identification With an Application to Bias Correction[*]

John Chao
University of Maryland

Norman R. Swanson
Rutgers University

First Draft: January 2000
Current Version: November 2004

Abstract

We provide analytical formulae for the asymptotic bias (ABIAS) and mean squared error (AMSE) of the IV estimator, and obtain approximations thereof based on an asymptotic scheme which essentially requires the expectation of the first stage F-statistic to converge to a finite (possibly small) positive limit as the number of instruments approaches infinity. Our analytical formulae can be viewed as generalizing the bias and MSE results of Richardson and Wu (1971) to the case with non-normal errors and stochastic instruments. Our approximations are shown to compare favorably with approximations due to Morimune (1983) and Donald and Newey (2001), particularly when the instruments are weak. We also construct consistent estimators for the ABIAS and AMSE, and we use these to further construct a number of bias corrected OLS and IV estimators, the properties of which are examined both analytically and via a series of Monte Carlo experiments.

JEL classification: C13, C31.
Keywords: confluent hypergeometric function, Laplace approximation, local-to-zero asymptotics, weak instruments.

1 Introduction

Over the last decade there have been a great number of papers written on the subject of instrumental variables (IV) regression with instruments that are only weakly correlated with the endogenous explanatory variables. A nonexhaustive list of important recent contributions to this growing literature include Nelson and Startz (1990a), Dufour (1997), Hall, Rudebusch and Wilcox (1996), Kleibergen (2002,2004), Shea (1997), Staiger and Stock (1997), Wang and Zivot (1998), Hahn and Inoue (2000), Hall and Peixe (2000), Donald and Newey (2001), Hahn and Kuersteiner (2002), Stock, Wright and Yogo (2002), Stock and Yogo (2002,2003), and the references contained therein.[1] Much of this literature focuses on the impact that the use of weak instruments has on interval estimation and hypothesis testing, with relatively fewer results being obtained on the properties of point estimators under weak identification. This is in spite of the fact that applied researchers who first noted the problem of weak instruments are clearly interested in its consequences for both point estimation and hypothesis testing (see e.g. Nelson and Startz (1990b), Bound, Jaeger, and Baker (1995), and Angrist and Krueger (1995)).

This paper focuses on point estimation properties. In particular, we focus on the IV estimator, and derive explicit analytical formulae for the asymptotic bias (ABIAS) and asymptotic mean-squared error (AMSE) of this estimator under the local-to-zero/weak-instrument framework pioneered by Staiger and Stock (1997). The formulae that we derive correspond to the exact bias and MSE functions of the 2SLS estimator, as derived by Richardson and Wu (1971), under the assumption of a simultaneous equations model with fixed instruments and Gaussian error distribution; and in this sense our results can be viewed as generalizing theirs to the more general setting with possibly non-normal errors and stochastic instruments.

We also derive approximations for the ABIAS and AMSE functions that are based on an expansion that, loosely speaking, allows the number of instruments k_{21} to grow to infinity while keeping the population analog of the F-statistic fixed and possibly small. Given the large k_{21} nature of our approximation, the results we obtain are connected to the many-instrument asymptotic approximation of Morimune (1983) and Bekker (1994). Moreover, since our approximations are extracted by expanding the asymptotic bias and MSE functions, our results can also be interpreted as having been derived from a sequential limit procedure whereby the sample size T is first allowed to grow to infinity, followed by the passage to infinity of the number of instruments, k_{21}. Interestingly, the lead term of our bias expansion (when appropriately standardized by the ABIAS of the OLS estimator) is precisely the relative bias measure given in Staiger and Stock (1997) in the case where there is only one endogenous regressor. In addition, the lead term of the MSE expansion is the square of the lead term of the bias expansion, implying that the variance component of the MSE is of a lower order vis-a-vis the bias component when there are many weak instruments.

Numerical calculations are performed to compare our approximations to bias and MSE estimates based on the asymptotic approximation of Morimune (1983), and to the alternative MSE approximation obtained in Donald and Newey (2001). We find our approximations to perform relatively well compared to those of Morimune (1983) and Donald and Newey (2001), particularly when the instruments are weak and the degree of apparent overidentification (as given by the order condition) is relatively large.[2]

---

[1] Other related work includes papers by Phillips (1989), Choi and Phillips (1992), and Kitamura (1994), which examines the implications for statistical inference when the underlying simultaneous equations model is partially identified, and papers by Forchini and Hillier (1999) and Moreira (2001, 2002) which explore conditional inference under weak identification.

[2] The performance of the MSE approximation of Donald and Newey (2001) is also examined in the Monte Carlo study

A consequence of the sequential limit approach which we adopt here is that consistent estimators for the ABIAS and AMSE can be obtained.[3] This, in turn, enables us to construct bias-corrected OLS and IV estimators, which consistently estimate the structural coefficient of the IV regression, even when the instruments are weak, in the local-to-zero sense. This is in contrast to the standard unadjusted OLS and IV estimators, which are inconsistent under the local-to-zero framework. Additionally, we show that in the conventional setup where the model is fully identified, all but one of our proposed bias-corrected estimators remains consistent. A small Monte Carlo experiment is carried out in order to assess the relative performance of our bias adjusted estimators as compared with OLS, IV, and LIML estimators, and the bias corrected estimators are shown to perform reasonably well in the many weak instrument setting.

The rest of this paper is organized as follows. Section 2 discusses our setup. Section 3 presents formulae for the ABIAS and AMSE and discusses some properties of these formulae. Section 4 outlines our ABIAS and AMSE approximations. Section 5 contains the results of various numerical calculations used to assess the accuracy of our approximations; and Section 6 discusses the consistent estimation of ABIAS and AMSE and suggests a number of bias-corrected OLS and IV estimators. In Section 7, a series of Monte Carlo results are used to illustrate the performance of our bias-corrected estimators. Concluding remarks are given in Section 8. All proofs and technical details are contained in two appendices. Before proceeding, we briefly introduce some notation. In the sequel, the symbols "$\Longrightarrow$" and "$\equiv$" denote convergence and equivalence in distribution, respectively. Also, $P_X = X(X'X)^{-1}X'$ is the matrix which projects orthogonally onto the range space of X and $M_X = I - P_X$.

2 Setup

Consider the simultaneous equations model (SEM):

$$y_1 = y_2\beta + X\gamma + u, \tag{1}$$

$$y_2 = Z\Pi + X\Phi + v, \tag{2}$$

where y_1 and y_2 are $T \times 1$ vectors of observations on the two endogenous variables, X is an $T \times k_1$ matrix of observations on k_1 exogenous variables included in the structural equation (1), Z is a $T \times k_2$ matrix of observations on k_2 exogenous variables excluded from the structural equation, and u and v are $T \times 1$ vectors of random disturbances[4]. Let u_t and v_t denote the tth component of the random vectors u and v, respectively; and let Z'_t and X'_t denote the tth row of the matrices Z and X, respectively. Additionally, let $w_t = (u_t, v_t)'$ (or $w = (u, v)$) and let $\overline{Z}_t = (X'_t, Z'_t)'$ (or $\overline{Z} = (X, Z)$); assume that $E(w_t) = \mathbf{0}$, $E(w_t w'_t) = \Sigma = \begin{pmatrix} \sigma_{uu} & \sigma_{uv} \\ \sigma_{uv} & \sigma_{vv} \end{pmatrix}$, and $E\overline{Z}_t w'_t = \mathbf{0}$ for all t, and assume that $E(w_t w'_s) = \mathbf{0}$ for all $t \neq s$, where $t, s = 1, ..., T$. Following Staiger and Stock (1997), we make the following assumptions.

---

reported in Hahn, Hausman, and Kuersteiner (2002). The results they obtained on the Donald-Newey approximation are in rough agreement with the numerical results reported in Section 5 of this paper.

[3]In a recent paper, Stock and Yogo (2003) give conditions under which sequential limit results are equivalent to results obtained by taking k_{21} and T jointly to infinity. They argue that the sequential asymptotic approach often provides an easier and useful way of calculating results which would also be obtained under a joint asymptotic scheme.

[4]For notational simplicity we only study the case with one endogenous explanatory variable in this paper. However, we conjecture that many of the qualitative conclusions reached here will continue to hold in more general settings.

Assumption 1: $\Pi = \Pi_T = C/\sqrt{T}$, where C is a fixed $k_2 \times 1$ vector.

Assumption 2: *The following limits hold jointly: (i)* $(u'u/T, u'v/T, v'v/T) \xrightarrow{p} (\sigma_{uu}, \sigma_{uv}, \sigma_{vv})$, *(ii)* $\overline{Z}'\overline{Z}/T \xrightarrow{p} Q$, *and (iii)* $(T^{-1/2}u'X, T^{-1/2}u'Z, T^{-1/2}v'X, T^{-1/2}v'Z)' \Longrightarrow (\psi'_{Xu}, \psi'_{Zu}, \psi'_{Xv}, \psi'_{Zv})'$, *where* $Q = E(\overline{Z}_t \overline{Z}'_t)$ *and where* $\psi \equiv (\psi'_{Xu}, \psi'_{Zu}, \psi'_{Xv}, \psi'_{Zv})'$ *is distributed* $N(\mathbf{0}, (\Sigma \otimes Q))$.

We consider IV estimation of β, where the IV estimator may not make use of all available instruments. Define $\widehat{\beta}_{IV} = (y'_2(P_H - P_X)y_2)^{-1}(y'_2(P_H - P_X)y_1)$, where $H = (Z_1, X)$ is an $T \times (k_{21} + k_1)$ matrix of instruments, and Z_1 is an $T \times k_{21}$ submatrix of Z formed by column selection. It will prove convenient to partition Z as $Z = (Z_1, Z_2)$, where Z_2 is an $T \times k_{22}$ matrix of observations of the excluded exogenous variables not used as instruments in estimation. Note that when $Z_1 = Z$ and $H = [Z, X]$ (i.e. when all available instruments are used), the IV estimator defined above is equivalent to the 2SLS estimator. Additionally, partition Π_T, $T^{-\frac{1}{2}}Z'u$, $T^{-\frac{1}{2}}Z'v$, ψ_{Zu}, and ψ_{Zv} conformably with $Z = (Z_1, Z_2)$ by writing $\Pi_T = (\Pi'_{1,T}, \Pi'_{2,T})' = (C'_1/\sqrt{T}, C'_2/\sqrt{T})'$, $T^{-\frac{1}{2}}Z'u = (T^{-\frac{1}{2}}u'Z_1, T^{-\frac{1}{2}}u'Z_2)'$, $T^{-\frac{1}{2}}Z'v = (T^{-\frac{1}{2}}v'Z_1, T^{-\frac{1}{2}}v'Z_2)'$, $\psi_{Zu} = (\psi'_{Z_1 u}, \psi'_{Z_2 u})'$, and $\psi_{Zv} = (\psi'_{Z_1 v}, \psi'_{Z_2 v})'$, where from part (iii) of Assumption 2 we have that $(T^{-\frac{1}{2}}u'Z_1, T^{-\frac{1}{2}}u'Z_2, T^{-\frac{1}{2}}v'Z_1, T^{-\frac{1}{2}}v'Z_2)' \Rightarrow (\psi'_{Z_1 u}, \psi'_{Z_2 u}, \psi'_{Z_1 v}, \psi'_{Z_2 v})'$. Furthermore, partition Q conformably with $\overline{Z} = (X, Z_1, Z_2)$ as

$$Q = \left(\begin{array}{ccc} Q_{XX} & Q_{XZ_1} & Q_{XZ_2} \\ Q_{Z_1 X} & Q_{Z_1 Z_1} & Q_{Z_1 Z_2} \\ Q_{Z_2 X} & Q_{Z_2 Z_1} & Q_{Z_2 Z_2} \end{array} \right). \tag{3}$$

Finally, define

$$\Omega = \left(\begin{array}{cc} \Omega_{11} & \Omega_{12} \\ \Omega'_{12} & \Omega_{22} \end{array} \right) = \left(\begin{array}{cc} Q_{Z_1 Z_1} - Q_{Z_1 X}Q_{XX}^{-1}Q_{XZ_1} & Q_{Z_1 Z_2} - Q_{Z_1 X}Q_{XX}^{-1}Q_{XZ_2} \\ Q_{Z_2 Z_1} - Q_{Z_2 X}Q_{XX}^{-1}Q_{XZ_1} & Q_{Z_2 Z_2} - Q_{Z_2 X}Q_{XX}^{-1}Q_{XZ_2} \end{array} \right) \tag{4}$$

and $\Omega_{1*} = (\Omega_{11}, \Omega_{12})$. To ensure that the ABIAS and AMSE of the IV estimator are well-behaved assume that:

Assumption 3: *There exists a finite positive integer* T_0 *such that* $\sup_{T \geq T_0} E(|U_T|^{2+\delta}) < \infty$, *for some* $\delta > 0$, *where* $U_T = \widehat{\beta}_{IV,T} - \beta_0$, $\widehat{\beta}_{IV,T}$ *denotes the IV estimator of* β *for a sample of size* T, *and where* β_0 *is the true value of* β.

Assumption 3 is sufficient for the uniform integrability of $(\widehat{\beta}_{IV,T} - \beta_0)^2$ (see Billingsley (1968), pp. 32). Under Assumption 3, $\lim_{T \to \infty} E(\widehat{\beta}_{IV,T} - \beta_0) = E(U)$ and $\lim_{T \to \infty} E(\widehat{\beta}_{IV,T} - \beta_0)^2 = E(U^2)$, where U is the limiting random variable of the sequence $\{U_T\}$, whose explicit form is given in Lemma A1 in Appendix A. Hence, the ABIAS and AMSE correspond to the bias and MSE implied by the limiting distribution of $\widehat{\beta}_{IV,T}$. Note also that for the special case where $(u_t, v_t)' \sim i.i.d.$ $N(0, \Sigma)$, $k_{21} \geq 4$ implies Assumption 3, since it is well-known that the IV estimator of β under Gaussianity has finite sample moments which exist up to and including the degree of apparent overidentification, as given by the order condition (see e.g. Sawa (1969)). Throughout this paper, we shall assume that $k_{21} \geq 4$ so as to ensure that our results encompass the Gaussian case. In addition, note that Assumption 3 rules out the limited information maximum likelihood (LIML) estimator in the Gaussian case, since no positive integer moment exists for the finite sample distribution of LIML in this case (see e.g. Mariano and Sawa (1972), Mariano and McDonald (1979), and Phillips (1984, 1985)).

3 Bias and MSE Formulae and Their Approximations

We begin with a proposition which gives explicit analytical formulae for the ABIAS and AMSE of the IV estimator under weak instruments and which also characterizes some of the properties of the (asymptotic) bias and MSE functions.

Proposition 3.1: *Given the SEM described above, and under Assumptions 1, 2, and 3, the following results hold for $k_{21} \geq 4$:*

(Bias)

(a)

$$b_{\widehat{\beta}_{IV}}\left(\mu'\mu, k_{21}\right) = \sigma_{uu}^{1/2}\sigma_{vv}^{-1/2}\rho e^{-\frac{\mu'\mu}{2}}{}_1F_1\left(\frac{k_{21}}{2}-1;\frac{k_{21}}{2};\frac{\mu'\mu}{2}\right), \tag{5}$$

where $b_{\widehat{\beta}_{IV}}\left(\mu'\mu, k_{21}\right) = \lim_{T\to\infty} E(\widehat{\beta}_{IV,T}-\beta_0)$ is the asymptotic bias function of the IV estimator which we write as a function of $\mu'\mu = \sigma_{vv}^{-1}C'\Omega'_{1*}\Omega_{11}^{-1}\Omega_{1*}C$ and k_{21}, and where $\rho = \sigma_{uv}\sigma_{uu}^{-\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}$, $\Gamma(\cdot)$ denotes the gamma function, and ${}_1F_1(\cdot;\cdot;\cdot)$ denotes the confluent hypergeometric function to be described in Remark 3.2(i) below;

(b) For k_{21} fixed, as $\mu'\mu \to \infty$, $b_{\widehat{\beta}_{IV}}\left(\mu'\mu, k_{21}\right) \to 0$;

(c) For $\mu'\mu$ fixed, as $k_{21}\to\infty$, $b_{\widehat{\beta}_{IV}}\left(\mu'\mu, k_{21}\right) \to \sigma_{uv}/\sigma_{vv} = \sigma_{uu}^{1/2}\sigma_{vv}^{-1/2}\rho$;

(d) The absolute value of the asymptotic bias function (i.e. $|b_{\widehat{\beta}_{IV}}\left(\mu'\mu, k_{21}\right)|$) is a monotonically decreasing function of $\mu'\mu$ for k_{21} fixed and $\sigma_{uv} \neq 0$;

(e) The absolute value of the bias function (i.e. $|b_{\widehat{\beta}_{IV}}\left(\mu'\mu, k_{21}\right)|$) is a monotonically increasing function of k_{21} for $\mu'\mu$ fixed and $\sigma_{uv} \neq 0$;

(MSE)

(f)

$$\begin{aligned}m_{\widehat{\beta}_{IV}}\left(\mu'\mu, k_{21}\right) &= \sigma_{uu}\sigma_{vv}^{-1}\rho^2 e^{-\frac{\mu'\mu}{2}}\left[\frac{1}{\rho^2}\left(\frac{1}{k_{21}-2}\right){}_1F_1\left(\frac{k_{21}}{2}-1;\frac{k_{21}}{2};\frac{\mu'\mu}{2}\right)\right.\\&\quad\left.+\left(\frac{k_{21}-3}{k_{21}-2}\right){}_1F_1\left(\frac{k_{21}}{2}-2;\frac{k_{21}}{2};\frac{\mu'\mu}{2}\right)\right],\end{aligned}\tag{6}$$

where $m_{\widehat{\beta}_{IV}}\left(\mu'\mu, k_{21}\right) = \lim_{T\to\infty} E\left(\widehat{\beta}_{IV,T}-\beta_0\right)^2$ is the asymptotic mean squared error function of the IV estimator;

(g) For k_{21} fixed, as $\mu'\mu\to\infty$, $m_{\widehat{\beta}_{IV}}\left(\mu'\mu, k_{21}\right) \to 0$;

(h) For $\mu'\mu$ fixed, as $k_{21}\to\infty$, $m_{\widehat{\beta}_{IV}}\left(\mu'\mu, k_{21}\right) \to \sigma_{uv}^2/\sigma_{vv}^2 = \sigma_{uu}\sigma_{vv}^{-1}\rho^2$;

(i) The asymptotic mean squared error function $m_{\widehat{\beta}_{IV}}\left(\mu'\mu, k_{21}\right)$ is a monotonically decreasing function of $\mu'\mu$ for k_{21} fixed and $\sigma_{uv} \neq 0$.

Remark 3.2: (i) Note that the ABIAS and AMSE formulae, given by expressions (5) and (6) above, involve confluent hypergeometric functions denoted by the symbol ${}_1F_1(\cdot;\cdot;\cdot)$. This is a well-known special function in applied mathematics, which, in the theory of ordinary differential equations, arises as a solution of the confluent hypergeometric differential equation (also called Kummer's differential equation)[5]. In addition, it is well known that confluent hypergeometric functions have infinite series representations (e.g. see Slater

---

[5]See Lebedev (1972) for more detailed discussions of confluent hypergeometric functions.

(1960), pp.2), so that $_1F_1\left(a;\ b;\ x\right) = \sum_{j=0}^{\infty} \frac{(a)_j}{(b)_j} \frac{x^j}{j!}$, where $(a)_j$ denotes the Pochhammer symbol (i.e. $(a)_j = \Gamma(a+j)/\Gamma(a)$ for integer $j \geq 1$, and $(a)_j = 1$ for $j = 0$). It follows that the expressions for the bias and MSE can be written in infinite series form:

$$b_{\widehat{\beta}_{IV}}\left(\mu'\mu, k_{21}\right) = \sigma_{uu}^{1/2}\sigma_{vv}^{-1/2}\rho e^{-\frac{\mu'\mu}{2}}\left[\sum_{j=0}^{\infty} \frac{(\frac{k_{21}}{2}-1)_j}{(\frac{k_{21}}{2})_j}\frac{\left(\frac{\mu'\mu}{2}\right)^j}{j!}\right], \tag{7}$$

$$m_{\widehat{\beta}_{IV}}\left(\mu'\mu, k_{21}\right) = \sigma_{uu}\sigma_{vv}^{-1}\rho^2 e^{-\frac{\mu'\mu}{2}}\left[\frac{1}{\rho^2}\left(\frac{1}{k_{21}-2}\right)\sum_{j=0}^{\infty}\frac{(\frac{k_{21}}{2}-1)_j}{(\frac{k_{21}}{2})_j}\frac{\left(\frac{\mu'\mu}{2}\right)^j}{j!}\right.$$
$$\left.+ \left(\frac{k_{21}-3}{k_{21}-2}\right)\sum_{j=0}^{\infty}\frac{(\frac{k_{21}}{2}-2)_j}{(\frac{k_{21}}{2})_j}\frac{\left(\frac{\mu'\mu}{2}\right)^j}{j!}\right]. \tag{8}$$

The main merit of these infinite series representations is that they provide explicit formulae for the ABIAS and AMSE of the IV estimator under weak identification, which can be used in numerical calculations. We also make extensive use of these representations in deriving the properties of the ABIAS and AMSE reported in the above proposition.

(ii) Part (g) of Proposition 3.1 states that the MSE function for $\widehat{\beta}_{IV,T}$ approaches zero as $\mu'\mu \to \infty$. Note that the case where $\mu'\mu \to \infty$ corresponds roughly to the case where the available instruments are not weak, but are instead fully relevant. In this case, then, Proposition 3.1 part (g) establishes that $\widehat{\beta}_{IV,T}$ converges in a mean squared sense to the true value, β_0. It follows, then, that in this case $\widehat{\beta}_{IV,T}$ is a (weakly) consistent estimator of β, a result which also follows from conventional asymptotic analysis with a fully identified model. Hence, results associated with the standard textbook case of good instruments are a limiting special case of our results.

(iii) It should be of interest to also derive the asymptotic bias and MSE of the OLS estimator under the weak instrument/local-to-zero framework, so that comparison can be made to the IV bias and MSE given in Proposition 3.1. To proceed, let $\widehat{\beta}_{OLS,T}$ denote the OLS estimator, and define the OLS asymptotic bias and MSE as $\lim_{T\to\infty} E\left(\widehat{\beta}_{OLS,T} - \beta_0\right)$ and $\lim_{T\to\infty} E\left(\widehat{\beta}_{OLS,T} - \beta_0\right)^2$, respectively. Then, under a condition similar to Assumption 3 above, we can easily derive the following result.

Lemma 3.3: *Suppose that Assumptions 1 and 2 hold for the SEM described by equations (1) and (2). Suppose further that there exists a finite positive integer T^* such that $\sup_{T \geq T^*} E[|U_T^*|^{2+\delta}] < \infty$ for some $\delta > 0$, where $U_T^* = \widehat{\beta}_{OLS,T} - \beta_0$, and where β_0 denotes the true value of β. Then, $\lim_{T\to\infty} E\left(\widehat{\beta}_{OLS,T} - \beta_0\right) = \sigma_{uv}/\sigma_{vv}$ and $\lim_{T\to\infty} E\left(\widehat{\beta}_{OLS,T} - \beta_0\right)^2 = \sigma_{uu}\sigma_{vv}^{-1}\rho^2$.*

Note that the assumption that $\sup_{T \geq T^*} E[|U_T^*|^{2+\delta}] < \infty$ holds for any $\delta > 0$, under Gaussian error assumptions and for T sufficiently large, since it is well-known that the finite sample distribution of the OLS estimator in this case has moments which exist up to order $T - 2$ (see Sawa (1969) for a more detailed discussion of the existence of moments of the OLS estimator). Hence, this assumption is not vacuous. Moreover, comparing

5

the OLS asymptotic bias and MSE given in Lemma 3.3 with the IV bias and MSE results obtained in parts (c) and (h) of Proposition 3.1, we see that for fixed $\mu'\mu$, the ABIAS and AMSE of the IV estimator both converge to the ABIAS and AMSE of the OLS estimator, as $k_{21} \to \infty$. Thus, if additional instruments do not increase the value of the concentration parameter $\mu'\mu$ (i.e. they are completely irrelevant), then adding such instruments will eventually make the bias as bad as that of the OLS estimator.

(iv) Write the asymptotic bias function of $\widehat{\beta}_{IV}$ as $b_{\widehat{\beta}_{IV}}(\mu'\mu, k_{21}) = \sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}\rho f(\mu'\mu, k_{21})$, where $f(\mu'\mu, k_{21}) = e^{-\frac{\mu'\mu}{2}} {}_1F_1\left(\frac{k_{21}}{2} - 1; \frac{k_{21}}{2}; \frac{\mu'\mu}{2}\right)$. From the proof of part (d) of Proposition 3.1, note that $0 < f(\mu'\mu, k_{21}) < 1$, for $\mu'\mu \in (0, \infty)$ and for positive integer k_{21} which is large enough so that the bias function exists. Since $\sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}\rho = \sigma_{uv}/\sigma_{vv}$ is simply the asymptotic bias of the OLS estimator, it follows that the bias of the IV estimator given in equation (5) has the same sign as the OLS bias. Additionally, note that $|b_{\widehat{\beta}_{IV}}(\mu'\mu, k_{21})| = |\sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}\rho|f(\mu'\mu, k_{21}) < |\sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}\rho|$. Hence, even when the instruments are weak in the sense of Staiger and Stock (1997), the ABIAS of the IV estimator is less in absolute magnitude than that of the OLS estimator, as long as $\mu'\mu \neq 0$ and as long as k_{21} is large enough so that the IV bias function exists. IV ABIAS only tends to the OLS bias as $k_{21} \to \infty$. On the other hand, when $\mu'\mu = 0$, then the asymptotic biases of the two estimators are exactly equal for all values of k_{21} which are large enough so that the IV bias function exists. Our results, therefore, formalize the intuitive discussions given in Bound, Jaeger, and Baker (1995) and Angrist and Krueger (1995) which suggest that with weak instruments, the IV estimator is biased in the direction of the OLS estimator, and the magnitude of the bias approaches that of the OLS estimator as the R^2 between the instruments and the endogenous explanatory variable approaches zero (i.e. as $\mu'\mu \to 0$). Our results also generalize characterizations of the IV bias given in Nelson and Startz (1990a,b) for a simple Gaussian model with a single fixed instrument and a single endogenous regressor to the more general case of an SEM with an arbitrary number of possibly stochastic instruments and with possible non-normal errors.

(v) Unlike the ABIAS of $\widehat{\beta}_{IV}$, the AMSE of $\widehat{\beta}_{IV}$ with weak instruments may, depending on the size of the concentration parameter $(\mu'\mu)$ and the number of instruments used (k_{21}), be either greater or less than the AMSE of $\widehat{\beta}_{OLS}$. To see this, consider the example where $\mu'\mu = 0$, and note that, in this case, the infinite series form of $m_{\widehat{\beta}_{IV}}(\mu'\mu, k_{21})$ reduces to:

$$m_{\widehat{\beta}_{IV}}(\mu'\mu = 0, k_{21}) = \sigma_{uu}\sigma_{vv}^{-1}\rho^2 \left[1 + \left(\frac{1-\rho^2}{\rho^2}\right)\left(\frac{1}{k_{21}-2}\right)\right], \tag{9}$$

which is greater than the AMSE of the OLS estimator for $\rho^2 < 1$ and for values of k_{21} large enough such that the AMSE of the IV estimator exists. On the other hand, we have already shown that as $\mu'\mu \to \infty$, the AMSE of the IV estimator approaches zero for k_{21} fixed; so that, as $\mu'\mu$ grows, the AMSE of the IV estimator will eventually become smaller than that of the OLS estimator.

(vi) Our results can also be compared with those obtained in the extensive literature on the finite sample properties of IV estimators and, in particular, with the results of Richardson and Wu (1971), who obtained the exact bias and MSE of the $2SLS$ estimator for a fixed instrument/Gaussian model[6]. To proceed with such a comparison, note first that the SEM given by expressions (1) and (2) can alternatively be written in

---

[6] Other papers which have studied the bias and/or MSE of the IV estimator, but for a fully identified model, include Richardson (1968), Hillier, Kinal, and Srivastava (1984), and Buse (1992).

reduced form. Namely:

$$y_1 = Z\Gamma_1 + X\Gamma_2 + \varepsilon_1, \tag{10}$$

$$y_2 = Z\Pi + X\Phi + \varepsilon_2, \tag{11}$$

where $\Gamma_1 = \Pi\beta$, $\Gamma_2 = \Phi\beta + \gamma$, $\varepsilon_2 = v$, and $\varepsilon_1 = u + v\beta = u + \varepsilon_2\beta$. In the finite sample literature on IV estimators, a Gaussian assumption is often made on the disturbances of this reduced form model; that is, it is often assumed that $(\varepsilon_{1t}, \varepsilon_{2t})' \equiv i.i.d.N(\mathbf{0}, G)$, where ε_{1t} and ε_{2t} denote the tth component of the $T \times 1$ random vectors ε_1 and ε_2, respectively, and where G can be partitioned conformably with $(\varepsilon_{1t}, \varepsilon_{2t})'$ as $G = \begin{pmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{pmatrix}$. Now, consider the case where all available instruments are used (i.e. the case where the IV estimator is simply the $2SLS$ estimator). Then, it follows that $\mu'\mu = \sigma_{vv}^{-1}C'\Omega C = \sigma_{vv}^{-1}C'(Q_{ZZ} - Q_{ZX}Q_{XX}^{-1}Q_{XZ})C$. In addition, note that in terms of the elements of the reduced form error covariance matrix, G, the elements of the structural error covariance matrix Σ given earlier in Section 2 can be written as: $\sigma_{uu} = g_{11} - 2g_{12}\beta + g_{22}\beta^2$, $\sigma_{uv} = g_{12} - g_{22}\beta$, and $\sigma_{vv} = g_{22}$. Substituting these expressions into (5) and (6), we see that:

$$b_{\widehat{\beta}_{IV}}(\mu'\mu, k_{21}) = -\frac{g_{22}\beta - g_{12}}{g_{22}}e^{-\frac{\mu'\mu}{2}} {}_1F_1\left(\frac{k_{21}}{2} - 1; \frac{k_{21}}{2}; \frac{\mu'\mu}{2}\right), \tag{12}$$

$$m_{\widehat{\beta}_{IV}}(\mu'\mu, k_{21}) = \frac{g_{11}g_{22} - g_{12}^2}{g_{22}}\left(\frac{1}{k_{21} - 2}\right)(1 + \overline{\beta}^2)e^{-\frac{\mu'\mu}{2}} {}_1F_1\left(\frac{k_{21}}{2} - 1; \frac{k_{21}}{2}; \frac{\mu'\mu}{2}\right)$$
$$+ \left(\frac{k_{21} - 3}{k_{21} - 2}\right)\overline{\beta}^2 e^{-\frac{\mu'\mu}{2}} {}_1F_1\left(\frac{k_{21}}{2} - 2; \frac{k_{21}}{2}; \frac{\mu'\mu}{2}\right), \tag{13}$$

where $\overline{\beta} = (g_{22}\beta - g_{12})(g_{11}g_{22} - g_{12}^2)^{-\frac{1}{2}}$. Comparing expressions (12) and (13) with equations (3.1) and (4.1) of Richardson and Wu (1971), we see that in this case the formulae for the ABIAS and AMSE are virtually identical to the exact bias and MSE derived under the assumption of a fixed instrument/Gaussian model - the only minor difference being that the (population) concentration parameter $\mu'\mu$ enters into the asymptotic formulae given in expressions (12) and (13) above, whereas the expression $\sigma_{vv}^{-1}\Pi'Z'M_X Z\Pi$ (with $M_X = I_T - X(X'X)^{-1}X'$) appears in the exact formulae reported in Richardson and Wu (1971). Hence, our bias and MSE results are consistent with the point made by Staiger and Stock (1997) that the limiting distribution of the $2SLS$ estimator under the local-to-zero assumption is the same as the exact distribution of the estimator under the more restrictive assumptions of fixed instruments and Gaussian errors.

4 Approximation Results for the Bias and MSE

In this section, we construct approximations for the bias and MSE that greatly simplify the more complicated expressions given in (7) and (8). To proceed, assume that:

Assumption 4: $\frac{\mu'\mu}{k_{21}} = \tau^2 + O\left(k_{21}^{-2}\right)$ *for some constant* $\tau^2 \in (0, \infty)$, *as* $\mu'\mu$, $k_{21} \to \infty$.

The next result gives a formal statement of our approximations based on Assumption 4.

7

Theorem 4.1 (Approximations): *Suppose that Assumption 4 holds. Write $\mu'\mu = \tau^2 k_{21} + O\left(k_{21}^{-1}\right) = \mu'\mu(\tau^2, k_{21})$, say, and reparameterize the bias and MSE functions given in equations (5) and (6) in terms of τ^2 and k_{21} so that:*

$$b_{\widehat{\beta}_{IV}}\left(\tau^2, k_{21}\right) = \sigma_{uu}^{1/2}\sigma_{vv}^{-1/2}\rho e^{-\frac{\mu'\mu(\tau^2,k_{21})}{2}} \; {}_1F_1\left(\frac{k_{21}}{2}-1;\frac{k_{21}}{2};\frac{\mu'\mu(\tau^2,k_{21})}{2}\right), \tag{14}$$

$$m_{\widehat{\beta}_{IV}}\left(\tau^2, k_{21}\right) = \sigma_{uu}\sigma_{vv}^{-1}\rho^2 e^{-\frac{\mu'\mu(\tau^2,k_{21})}{2}}\left[\frac{1}{\rho^2}\left(\frac{1}{k_{21}-2}\right) \; {}_1F_1\left(\frac{k_{21}}{2}-1;\frac{k_{21}}{2};\frac{\mu'\mu(\tau^2,k_{21})}{2}\right)\right.$$
$$\left. + \left(\frac{k_{21}-3}{k_{21}-2}\right) \; {}_1F_1\left(\frac{k_{21}}{2}-2;\frac{k_{21}}{2};\frac{\mu'\mu(\tau^2,k_{21})}{2}\right)\right]. \tag{15}$$

Then, as $k_{21} \to \infty$, the following results hold:
(i)

$$b_{\widehat{\beta}_{IV}}\left(\tau^2, k_{21}\right) = \sigma_{uu}^{1/2}\sigma_{vv}^{-1/2}\rho\left\{\left(\frac{1}{1+\tau^2}\right) - \frac{2}{k_{21}}\left(\frac{1}{1+\tau^2}\right)\left(\frac{\tau^2}{1+\tau^2}\right)^2\right\} + O(k_{21}^{-2}) \tag{16}$$

(ii)

$$m_{\widehat{\beta}_{IV}}\left(\tau^2, k_{21}\right) = \sigma_{uu}\sigma_{vv}^{-1}\rho^2\left\{\left(\frac{1}{1+\tau^2}\right)^2 + \left(\frac{1-\rho^2}{\rho^2}\right)\left(\frac{1}{k_{21}}\right)\left(\frac{1}{1+\tau^2}\right) + \left(\frac{1}{k_{21}}\right)\left(\frac{1}{1+\tau^2}\right)\right.$$
$$\left.\left[1 - 7\left(\frac{1}{1+\tau^2}\right) + 12\left(\frac{1}{1+\tau^2}\right)^2 - 6\left(\frac{1}{1+\tau^2}\right)^3\right]\right\} + O(k_{21}^{-2}) \tag{17}$$

Remark 4.2: (i) Observe that Assumption 4 imposes a condition on the ratio $\frac{\mu'\mu}{k_{21}}$ which is referred to as the "population analog" of the first-stage F-statistic for testing the relevance of the instruments used (i.e. see Bound, Jaeger, and Baker (1995)). The magnitude of this ratio, or its numerator, $\mu'\mu$, has long been recognized to be a natural measure of the strength of identification, or, alternatively, the strength of the instruments (cf. Phillips (1983), Rothenberg (1983), Staiger and Stock (1997), and Stock and Yogo (2003)).[7] Indeed, when the model is strongly identified, we expect the first-stage F statistic to diverge asymptotically, so that the null hypothesis of identification failure is rejected with probability approaching one, asymptotically. On the other hand, Assumption 4 requires that $\frac{\mu'\mu}{k_{21}}$ does not diverge, but instead approaches a (non-zero) constant. Thus, roughly speaking, Assumption 4 corresponds to the case where the instruments are weaker than that typically assumed under conventional strong-identification asymptotics.

In addition, note that Assumption 4 bears some resemblance to the situation assumed in Case (ii) of Morimune (1983). However, a key difference between Assumption 4 and Case (ii) of Morimune (1983) is that Morimune's Case (ii) requires both the concentration parameter, $\mu'\mu$, and the number of instruments, k_{21}, to be of the same order of magnitude as the sample size, T. Thus, his assumption corresponds more closely to the case of strong identification, since conventional asymptotics also assumes that $\mu'\mu$ is of the same order as T. On the other hand, Assumption 4 only assumes that $\mu'\mu$ and k_{21} are of the same order of magnitude; and thus we do not preclude situations where $\mu'\mu$ and k_{21} may be of a lower order, relative to

---

[7] Under Gaussian error assumptions, properties of the exact finite sample distribution of the $2SLS$ estimator have also been shown to depend importantly on the magnitude of $\mu'\mu$, as has been shown in the extensive numerical calculations reported in Anderson and Sawa (1979).

T. It follows that our asymptotic setup might be expected to be more appropriate for situations where the instruments are weak, in the sense that $\mu'\mu$ is small relative to the sample size[8].

(ii) Set

$$\widehat{b}_{\widehat{\beta}_{IV}}\left(\tau^2, k_{21}\right) = \sigma_{uu}^{1/2}\sigma_{vv}^{-1/2}\rho\left[\left(\frac{1}{1+\tau^2}\right) - \frac{2}{k_{21}}\left(\frac{1}{1+\tau^2}\right)\left(\frac{\tau^2}{1+\tau^2}\right)^2\right]. \tag{18}$$

Recall from Remark 3.2(iii) that the ABIAS of the OLS estimator is given by $b_{\widehat{\beta}_{OLS}} = \sigma_{uu}^{1/2}\sigma_{vv}^{-1/2}\rho$. It follows that, by taking the ratio of the two, we obtain the relative bias measure:

$$\frac{\widehat{b}_{\widehat{\beta}_{IV}}\left(\tau^2, k_{21}\right)}{b_{\widehat{\beta}_{OLS}}} = \left(\frac{1}{1+\tau^2}\right) - \frac{2}{k_{21}}\left(\frac{1}{1+\tau^2}\right)\left(\frac{\tau^2}{1+\tau^2}\right)^2. \tag{19}$$

Observe that the lead term of expression (19) is $(1+\tau^2)^{-1} = (1 + \mu'\mu/k_{21})^{-1}$. Note that when all available instruments are used so that $IV = 2SLS$, $(1 + \mu'\mu/k_{21})^{-1}$ is the relative bias measure given in Staiger and Stock (1997), for the case where there is only a single endogenous explanatory variable (see Staiger and Stock (1997), pp. 566 and 575). Staiger and Stock point out that this measure of relative bias is given by an approximation which holds for large k_{21} and/or large $\mu'\mu/k_{21}$. Our analysis shows that their relative bias measure can also be obtained from an approximation that requires $\mu'\mu/k_{21}$ to approach a finite limit as $\mu'\mu$, $k_{21} \to \infty$.[9]

(iii) Although Assumption 4 requires that $\tau^2 > 0$, it is easy to see, by following the proof of Theorem 4.1, that the bias and MSE expansion given by expressions (16) and (17) are valid even for $\tau^2 = 0$. However, the condition $\tau^2 > 0$ is assumed because, as explained in Remark 6.2.3(i) of Section 6, τ^2 must not be zero if our objective is the construction of consistent estimators of the lead term of the bias and MSE expansions, and the subsequent construction of bias-adjusted estimators.

5   Numerical Results

In order to assess the potential usefulness of our approximations, we carried out some numerical calculations using a canonical SEM, where the reduced form error covariance matrix is taken to be the identity matrix (i.e. $G = I$, using the notation introduced in Remark 3.2(vi)). We report two sets of numerical results. The first set is based on a simple regression analysis which we used in order to evaluate the accuracy of our bias and MSE approximations, and in order to compare our approximation (referred to as the *CS Approximations*) with an alternative MSE approximation first derived in Donald and Newey (2001) and further examined in Hahn, Hausman, and Kuersteiner (2002) (referred to as the *DN Approximations*). The regressions that we ran are of the form:

---

[8]Note further that even within a weak instrument setup, it is reasonable to think that $\mu'\mu$ might increase as one uses more instruments, so long as the added instruments are not completely uncorrelated with the endogenous regressor (i.e. so long as we are dealing with weak, but not completely irrelevant, instruments).

[9]The approximate bias formula presented here has also been discussed in Hahn and Hausman (2002), although, in that paper, the bias approximation is not justified using the Laplace approximation method, as is done here.

Regression for the CS Approximation:

$$b_{\widehat{\beta}_{IV}}\left(\mu'\mu, k_{21}\right) \quad = \quad \phi_0 + \phi_1 \left[-\beta\left(1 + \mu'\mu/k_{21}\right)^{-1}\right] + error, \tag{20}$$

$$m_{\widehat{\beta}_{IV}}\left(\tau^2, k_{21}\right) \quad = \quad \pi_0 + \pi_1 \left[\beta^2\left(1 + \mu'\mu/k_{21}\right)^{-2}\right] + error, \tag{21}$$

Regression for the DN Approximation:

$$m_{\widehat{\beta}_{IV}}\left(\tau^2, k_{21}\right) = \pi_0^* + \pi_1^* \left[\left(1 + \beta^2\right)/\mu'\mu + k_{21}^2\beta^2/(\mu'\mu)^2\right] + error. \tag{22}$$

Note that the explanatory variables (i.e. the terms in square brackets) in regressions (20) and (21) are the first order terms, respectively, of our bias and MSE approximations, both specialized to the canonical case using the fact that $\sigma_{vv} = 1$, $\sigma_{uu} = 1 + \beta^2$, and $\rho = -\beta/\sqrt{1 + \beta^2}$ in this case. The independent variable in (22) is the *DN Approximation* for the MSE, also specialized to the case where the underlying model is a canonical SEM. The dependent variables in regressions (20)-(22), on the other hand, are calculated using the analytical formulae for the bias and MSE given by expressions (7) and (8), again specialized to the canonical case. Values for both the dependent and the independent variables were calculated for $\beta = \{-0.5, -1.0, -1.5, ..., -10\}$, $\mu'\mu = \{2, 4, 6, 8, ..., 100\}$, and $k_{21} = \{3, 5, 7, 9, 11, ..., 101\}$, so that 50,000 observations were generated by taking all possible combinations.[10] A total of 50 regressions were run for each approximation, with each regression including 1000 observations for a given value of $\mu'\mu$. Of note is that these regressions thus all include observations for low values of k_{21}.

The results of our regression analysis are reported in Table 1. Observe that both our bias and MSE approximation fare very well, with R^2 values very near to unity regardless of the strength of the instruments, as measured by the magnitude of the concentration parameter, $\mu'\mu$. In fact, the R^2 value for our bias approximation never drops below 0.9977 while the R^2 value for our MSE approximation never drops below 0.9980. On the other hand, the regression based on the *DN Approximation* has a relatively low R^2 value of 0.4665 when $\mu'\mu = 2$, and R^2 values remain low when instruments are weak, although they rise steadily as the instruments become stronger; ultimately resulting in an R^2 value of 0.961 for the boundary case where $\mu'\mu = 100$. In addition to the R^2 values, another indicator of the accuracy of our approximations is the fact that for both the bias regression (20) and the MSE regression (21), the estimated coefficients $\widehat{\phi}_1$ and $\widehat{\pi}_1$ are very close to unity in all cases, as we would expect them to be if the approximations are good. In contrast, the value of the estimated coefficient $\widehat{\pi}_1^*$ in regression (22) never exceeds 0.2753, which is its value when $\mu'\mu = 100$.

The second set of numerical calculations compares our bias and MSE approximations with the bias and MSE estimates arising from the asymptotic approximation of Morimune (1983). Eight data generating processes (DGPs) are considered for this comparison. The first four of these DGPs correspond to models A-D in Section 7 of Morimune (1983). Following Morimune (1983), we set $k_1 = 4$, $\alpha = \left(g_{22}\beta - g_{12}\right)/\sqrt{|G|} = 1$, and $\mu'\mu = 25$ for all four of these DGPs.[11] In addition, we set $k_{21} = k_2 = 6$, 11, 16, and 21 for these four DGPs, conforming to the number of instruments specified for the four simulation models in Morimune

---

[10] Note that, for our regressions analysis, we have chosen only negative values of β. This choice does not bias our numerical results, as the IV bias function is perfectly symmetrical with respect to positive and negative values of β, and the IV MSE function only depends on β^2.

[11] Note that our notation differs from Morimune's original notation. For example, he uses Ω, σ^2, Z_1, and Z_2 to denote,

(1983).[12] In the remaining four DGPs, we take $\mu'\mu = 10$, and set $k_2 = 30, 40, 50$, and 100. These DGPs were not considered in Morimune's study, and we consider them here because they involve cases where the instruments are weaker (as measured by the magnitude of $\mu'\mu$) and where number of instruments is greater than in those cases considered in the first four DGPs.

Table 2 summarizes the bias and MSE calculations for these eight DGPs. To facilitate comparison with Morimune's results, we report in Table 2 the bias and MSE for the standardized estimator $\sqrt{\sigma_{vv}\sigma_{uu}^{-1}\mu'\mu}\left(\widehat{\beta} - \beta_0\right)$.[13] The first two columns of Table 2 give the exact bias and MSE values calculated using the analytical expressions (7) and (8), while the next two columns contain bias and MSE values based on the Monte Carlo simulation reported in Morimune (1983).[14] Comparing the first four columns of Table 2 for the first four DGPs, we see that the simulated bias and MSE values correspond very closely with the exact bias and MSE based on our analytical formulae, as expected, given that the Morimune simulation values are calculated using a large number of draws (i.e. $20,000$ draws)[15]. Note also that simulated bias and MSE values for the last four DGPs are not available from Morimune (1983), since these DGPs were not considered in his paper. Next, turning our attention to the remaining columns of Table 2, note that columns 5-8 report values of our bias and MSE approximations, where CS Bias1 and CS Bias2 denote bias approximations based on our first- and second-order approximations, respectively, and where CS MSE1 and CS MSE2 are defined similarly. Columns 9 and 10 report bias and MSE approximations based on the asymptotic expansion of Morimune (1983) under what he refers to as Case (ii), while the last column gives values of the MSE approximation of Donald and Newey (2001). Comparing our approximations to those of Morimune (1983), we see that when the degree of overidentification is relatively modest, such as the cases where $k_2 = 6, 11$, and 16, Morimune's approximation for both the bias and the MSE outperform ours. However, as the degree of overidentification (as given by the order condition) increases and as the available instruments become weaker (as measured

respectively, the reduced form error covariance matrix, the error variance of the structural equation, the included exogenous variables, and the excluded exogenous variables, whereas our notation which corresponds to these varialbes uses G, σ_{uu}, X, and Z, respectively. Moreover, he defines the standardized concentration parameter, $\mu^2 = \frac{\sigma_{uu}}{|G|}\Pi'Z'M_XZ\Pi$ (written here in our notation), which differs slightly from our $\mu'\mu$ defined in Section 3. In particular, μ^2 is related to our $\mu'\mu$ by the approximate formula $\mu^2 \approx \frac{\sigma_{uu}\sigma_{vv}}{|G|}\mu'\mu$. Thus, assuming $\alpha = 1$ and a canonical reduced form error covariance structure, we have the approximate relationship that $\mu^2 \approx 2\mu'\mu$.

[12] Note that for the purpose of our numerical evaluation, we use all available instruments in constructing the IV estimator so that $k_{21} = k_2$.

[13] Note that the standardized estimator, $\sqrt{\sigma_{vv}\sigma_{uu}^{-1}\mu'\mu}\left(\widehat{\beta} - \beta_0\right)$, in our notation corresponds to the standardized estimator, $\frac{\delta\sqrt{\omega_{22}}}{\sigma}\left(\widehat{\beta} - \beta\right)$, using the notation of Morimune (1983). Reporting bias and MSE values for this standardized estimator ensures that our results are directly comparable with the numerical results given in Section 7 of Morimune (1983).

[14] Morimune (1983) actually reports variance instead of MSE. We convert values which he reports for the variance into values for MSE in order to facilitate comparison of his results with ours.

[15] Note that, for the purpose of comparing our numerical results with those of Morimune (1983), it is sensible to think of the analytical expressions (7) and (8) as giving the exact bias and MSE of the IV estimator, even though earlier we have referred to these expressions as being the *asymptotic* bias and MSE under the local-to-zero/weak instrument framework. This is because the numerical calculations reported in Morimune (1983) are carried out under the assumptions of fixed instruments and Gaussian errors; and, as we have discussed in Remark 3.2(vi) above, the asymptotic bias and MSE under the local-to-zero framework are in fact the exact bias and MSE under these assumptions.

by a smaller value of $\mu'\mu$), our approximations become just as good and, in some cases, better than the approximations of Morimune (1983). In particular, note that for the last four DGPs, where $\mu'\mu$ is smaller and the degree of overidentification is larger than that for the first four DGPs, our CS MSE2 provides a more accurate approximation for the exact MSE than the Morimune approximation. Thus, and as expected, our approximations appear to be most useful in many weak instrument cases.

6   Estimation of Bias and MSE and Bias Correction

6.1   Consistent Estimation of the Bias and MSE

In this subsection, we obtain consistent estimators for the lead terms of the bias and MSE expansions given in Theorem 4.1. To proceed, first define M_i $(i = 1, 2)$, such that $M_1 = M_{\overline{Z}} = I_T - \overline{Z}\left(\overline{Z}'\overline{Z}\right)^{-1}\overline{Z}'$ and $M_2 = M_X = I_T - X\left(X'X\right)^{-1}X'$, where $\overline{Z} = (Z, X)$. Further, let $s_{uv,i} = \frac{(y_1 - y_2\widehat{\beta}_{IV})'M_i y_2}{T}$ (for $i = 1, 2$) and $s_{uu} = \frac{(y_1 - y_2\widehat{\beta}_{IV})'M_2(y_1 - y_2\widehat{\beta}_{IV})}{T}$. Next, define the following statistics:

$$\widehat{\sigma}_{vv,i} = \frac{y_2'M_i y_2}{T}, \text{ for } i = 1, 2; \tag{23}$$

$$W_{k_{21},T} = \left[\frac{y_2'(P_H - P_X)y_2}{\widehat{\sigma}_{vv,1}}\right]k_{21}^{-1} = \frac{W_{k_{21},T}^*}{k_{21}}; \tag{24}$$

$$\widehat{\sigma}_{uv,i} = s_{uv,i}\left(\frac{W_{k_{21},T}}{W_{k_{21},T} - 1}\right) = s_{uv,i}\left(\frac{1}{1 - \frac{1}{W_{k_{21},T}}}\right), \text{ for } i = 1, 2; \tag{25}$$

$$\widehat{\sigma}_{uu,i} = s_{uu} + 2\frac{\widehat{\sigma}_{uv,i}^2}{\widehat{\sigma}_{vv,i}}\left(\frac{1}{W_{k_{21},T}}\right) - \frac{\widehat{\sigma}_{uv,i}^2}{\widehat{\sigma}_{vv,i}}\left(\frac{1}{W_{k_{21},T}}\right)^2, \text{ for } i = 1, 2. \tag{26}$$

The following lemma shows that we can obtain consistent estimates of the quantities σ_{vv}, σ_{uv}, σ_{uu}, and $\left(1 + \tau^2\right)$ under a sequential limit approach.

Lemma 6.1.1: *Suppose that Assumptions 1 and 2 hold. Let $T \to \infty$, and then let $k_{21}, \mu'\mu \to \infty$ such that Assumption 4 holds. Then: (a) $\widehat{\sigma}_{vv,i} \xrightarrow{p} \sigma_{vv}$, for $i = 1, 2$; (b) $W_{k_{21},T} \xrightarrow{p} 1 + \tau^2$; (c) $\widehat{\sigma}_{uv,i} \xrightarrow{p} \sigma_{uv}$, for $i = 1, 2$; (d) $\widehat{\sigma}_{uu,i} \xrightarrow{p} \sigma_{uu}$, for $i = 1, 2$.*

Based on these estimators, we propose four different estimators for the ABIAS and six different estimators for the AMSE, as follows:

$$\widehat{BIAS}_i = \frac{\widehat{\sigma}_{uv,i}}{\widehat{\sigma}_{vv,i}}\left(\frac{1}{W_{k_{21},T}}\right), \text{ for } i = 1, 2; \tag{27}$$

$$\widetilde{BIAS}_i = \frac{\widehat{\sigma}_{uv,i}}{\widehat{\sigma}_{vv,i}}\left[\left(\frac{1}{W_{k_{21},T}}\right) - \frac{1}{k_{21}}\left(\frac{1}{W_{k_{21},T}}\right)\left\{2 - 4\left(\frac{1}{W_{k_{21},T}}\right) + 2\left(\frac{1}{W_{k_{21},T}}\right)^2\right\}\right]$$

$$= \frac{\widehat{\sigma}_{uv,i}}{\widehat{\sigma}_{vv,i}}\left[\left(\frac{1}{W_{k_{21},T}}\right) - \frac{2}{k_{21}}\left(\frac{1}{W_{k_{21},T}}\right)\left(\frac{W_{k_{21},T} - 1}{W_{k_{21},T}}\right)^2\right], \text{ for } i = 1, 2; \tag{28}$$

$$\widehat{MSE}_i = \frac{\widehat{\sigma}_{uv,i}^2}{\widehat{\sigma}_{vv,i}^2}\left(\frac{1}{W_{k_{21},T}}\right)^2, \text{ for } i = 1, 2; \tag{29}$$

12

$$\widetilde{MSE}_i = \frac{\widehat{\sigma}_{uv,i}^2}{\widehat{\sigma}_{vv,i}^2} \left[\left(\frac{1}{W_{k_{21},T}} \right)^2 + \frac{1}{k_{21}} \left(\frac{\widehat{\sigma}_{uu,i}\widehat{\sigma}_{vv,i} - \widehat{\sigma}_{uv,i}^2}{\widehat{\sigma}_{uv,i}^2} \right) \left(\frac{1}{W_{k_{21},T}} \right) \right.$$
$$\left. + \frac{1}{k_{21}} \left(\frac{1}{W_{k_{21},T}} \right) \left(1 - \frac{7}{W_{k_{21},T}} + \frac{12}{W_{k_{21},T}^2} - \frac{6}{W_{k_{21},T}^3} \right) \right], \text{ for } i = 1, 2; \qquad (30)$$

$$\overline{MSE}_i = \frac{\widehat{\sigma}_{uv,i}^2}{\widehat{\sigma}_{vv,i}^2} \left[\left(\frac{1}{W_{k_{21},T}} \right)^2 + \frac{1}{k_{21}} \left(\frac{\widehat{g}_{11}\widehat{g}_{22} - \widehat{g}_{12}^2}{\widehat{\sigma}_{uv,i}^2} \right) \left(\frac{1}{W_{k_{21},T}} \right) \right.$$
$$\left. + \frac{1}{k_{21}} \left(\frac{1}{W_{k_{21},T}} \right) \left(1 - \frac{7}{W_{k_{21},T}} + \frac{12}{W_{k_{21},T}^2} - \frac{6}{W_{k_{21},T}^3} \right) \right], \text{ for } i = 1, 2, \qquad (31)$$

where $\widehat{g}_{ij} = \frac{y_i' M_1 y_j}{T}$, for $i, j = 1, 2$. Note that the difference between the "hat" estimators and the "tilde" estimators is that the "hat" estimators are constructed based only on the lead term of the expansions given in Theorem 4.1 while the "tilde" estimators make use of both the lead term and the second order term. In addition, the difference between $\widetilde{MSE}_i$ and $\overline{MSE}_i$ lies in the fact that, given the equivalence $\sigma_{uu}\sigma_{vv} - \sigma_{uv}^2 = g_{11}g_{22} - g_{12}^2$, there are two ways we can estimate the quantity $\sigma_{uu}\sigma_{vv} - \sigma_{uv}^2$ (i.e. we can either estimate $\sigma_{uu}\sigma_{vv} - \sigma_{uv}^2$, directly leading to the pair of estimators $\widetilde{MSE}_i$ $(i = 1, 2)$, or estimate it indirectly as $g_{11}g_{22} - g_{12}^2$, leading to the alternative pair of estimators $\overline{MSE}_i$ $(i = 1, 2)$). The next theorem derives the probability limits of the estimators given by expressions (27)-(31).

Theorem 6.1.2: *Suppose that Assumptions 1 and 2 hold. Let $T \to \infty$, and then let $k_{21}, \mu'\mu \to \infty$, such that Assumption 4 holds. Then: (a) $\widehat{BIAS}_i \overset{p}{\to} \sigma_{uu}^{1/2}\sigma_{vv}^{-1/2}\rho \left(\frac{1}{1+\tau^2} \right)$, for $i = 1, 2$; (b) $\widetilde{BIAS}_i \overset{p}{\to}$*
$\sigma_{uu}^{1/2}\sigma_{vv}^{-1/2}\rho \left(\frac{1}{1+\tau^2} \right)$, for $i = 1, 2$; (c) $\widehat{MSE}_i \overset{p}{\to} \sigma_{uu}\sigma_{vv}^{-1}\rho^2 \left(\frac{1}{1+\tau^2} \right)^2$, for $i = 1, 2$; (d) $\widetilde{MSE}_i \overset{p}{\to} \sigma_{uu}\sigma_{vv}^{-1}\rho^2 \left(\frac{1}{1+\tau^2} \right)^2$,
for $i = 1, 2$; (e) $\overline{MSE}_i \overset{p}{\to} \sigma_{uu}\sigma_{vv}^{-1}\rho^2 \left(\frac{1}{1+\tau^2} \right)^2$, for $i = 1, 2$.

Remark 6.1.3: (i) The estimators defined in equations (27)-(31) are all weakly consistent, in the sense that each bias estimator converges in probability to the lead term of the bias expansion given in (16), while each MSE estimator converges in probability to the lead term of the MSE expansion given in (17). These results suggest that there is information which can be exploited when a large number of weakly correlated instruments are available, so that consistent estimation may be achieved when the number of instruments is allowed to approach infinity.

(ii) It is also of interest to analyze the asymptotic properties of our bias and MSE estimators in the conventional framework, where the instruments are not assumed to be weak in the local-to-zero sense but rather the usual identification condition is assumed to hold, even as the sample size approaches infinity. Hence, in place of the local-to-zero condition of Assumption 1, we make the alternative identification assumption:

Assumption 1*: Let Π be a fixed $k_2 \times 1$ vector such that $\Pi \neq 0$.

In order to obtain the probability limits of our estimators under Assumption 1*, we first give a lemma which contains limiting results for our estimators of the parameters $\sigma_{vv}, \sigma_{uv}, \sigma_{uu}$, and for the Wald statistic for testing instrument relevance, $W_{k_{21},T}$.

Lemma 6.1.4: *Suppose that Assumptions 1* and 2 hold. Then, as $T \to \infty$, the following limit results hold: (a) $\widehat{\sigma}_{vv,1} \overset{p}{\to} \sigma_{vv}$; (b) $\widehat{\sigma}_{vv,2} \overset{p}{\to} \Pi'\Omega\Pi + \sigma_{vv}$; (c) $W_{k_{21},T} = Op(T)$; (d) $\widehat{\sigma}_{uv,i} \overset{p}{\to} \sigma_{uv}$, for $i = 1, 2$; (e) $\widehat{\sigma}_{uu,i} \overset{p}{\to} \sigma_{uu}$, for $i = 1, 2$.*

13

In view of expressions (27)-(31), the next theorem follows as an immediate consequence of Lemma 6.1.4 and the Slutsky Theorem:

Theorem 6.1.5: *Suppose that Assumptions 1* and 2 hold. Then, as $T \to \infty$, the following limit results hold: (a)* $\widehat{BIAS}_i \overset{p}{\to} 0$, *for* $i = 1, 2$; *(b)* $\widetilde{BIAS}_i \overset{p}{\to} 0$, *for* $i = 1, 2$; *(c)* $\widehat{MSE}_i \overset{p}{\to} 0$, *for* $i = 1, 2$; *(d)* $\widetilde{MSE}_i \overset{p}{\to} 0$, *for* $i = 1, 2$; *(e)* $\overline{MSE}_i \overset{p}{\to} 0$, *for* $i = 1, 2$.

Note that, all of these bias and MSE estimators approaches zero, as $T \to \infty$ (as they should in the case of full identification, since the IV estimator is weakly consistent in this case). These results suggest that our estimators behave in a reasonable manner even in the conventional case where instruments are fully relevant (i.e. when Assumption 1* holds).

6.2 Bias Correction

The results of the last subsection can be used to construct bias-adjusted OLS and IV estimators. In particular, we propose five alternative bias-corrected estimators:

$$\widetilde{\beta}_{OLS,i} = \widehat{\beta}_{OLS} - \frac{\widehat{\sigma}_{uv,i}}{\widehat{\sigma}_{vv,i}}, \text{ for } i = 1, 2; \tag{32}$$

$$\widetilde{\beta}_{IV} = \widehat{\beta}_{IV} - \widehat{BIAS}_1; \text{ and} \tag{33}$$

$$\widetilde{\widetilde{\beta}}_{IV,i} = \widehat{\beta}_{IV} - \widetilde{BIAS}_i, \text{ for } i = 1, 2. \tag{34}$$

The next two theorems give the probability limits of these bias-corrected estimators both for the case where the local-to-zero/weak instrument condition (Assumption 1) is assumed and for the case where the conventional strong-identification condition (Assumption 1^*) is assumed.

Theorem 6.2.1: *Suppose that Assumptions 1 and 2 hold. Let $T \to \infty$, and then let $k_{21}, \mu'\mu \to \infty$, such that Assumption 4 holds. Then: (a)* $\widetilde{\beta}_{OLS,i} \overset{p}{\to} \beta_0$ *for* $i = 1, 2$; *(b)* $\widetilde{\beta}_{IV} \overset{p}{\to} \beta_0$; *(c)* $\widetilde{\widetilde{\beta}}_{IV,i} \overset{p}{\to} \beta_0$ *for* $i = 1, 2$.

Theorem 6.2.2: *Suppose that Assumptions 1* and 2 hold. Then, as $T \to \infty$, the following limit results hold: (a)* $\widetilde{\beta}_{OLS,1} \overset{p}{\to} \beta_0 - \frac{\sigma_{uv}}{\sigma_{vv}} \left(\frac{\Pi'\Omega\Pi}{\Pi'\Omega\Pi + \sigma_{vv}} \right)$; *(b)* $\widetilde{\beta}_{OLS,2} \overset{p}{\to} \beta_0$; *(c)* $\widetilde{\beta}_{IV} \overset{p}{\to} \beta_0$; *(d)* $\widetilde{\widetilde{\beta}}_{IV,i} \overset{p}{\to} \beta_0$ *for* $i = 1, 2$.

Remark 6.2.3: (i) Note that, under the local-to-zero framework with many instruments, the bias-corrected estimators are consistent. This is in contrast to the uncorrected OLS and IV estimators, which are not consistent in this case. It should also be noted that if we fix k_{21} and only allow T to approach infinity, then none of the estimators are consistent. In fact, in this case, both the uncorrected and the bias-corrected IV estimators converge weakly to random variables, although the form of the random limit clearly varies depending on the estimator. See Staiger and Stock (1997) for a precise characterization of the limiting distribution of the (uncorrected) IV estimator in the case where k_{21} is held fixed.

Moreover, if $\tau^2 = 0$, then the bias-adjusted estimators introduced above would not consistently estimate the structural coefficient, β. This is due to the fact that the bias-adjusted estimators make use of covariance estimators of the form: $s_{uv,i} = \frac{(y_1 - y_2\widehat{\beta}_{IV})' M_i y_2}{T}$, for $i = 1, 2$; and $s_{uv,i} \overset{p}{\to} \sigma_{uv} \left(\frac{\tau^2}{1 + \tau^2} \right)$, for $i = 1, 2$ (as shown in the proof of part (c) of Lemma 6.1.1). Consequently, if $\tau^2 = 0$, $s_{uv,i} \overset{p}{\to} 0$ for $i = 1, 2$; Hence, asymptotically, neither $s_{uv,1}$ nor $s_{uv,2}$ carries any information about the value of σ_{uv} and, thus, the traditional estimators cannot be adjusted in order to obtain consistent estimators. Indeed, since $\tau^2 = 0$ arises either because the model is unidentified in the traditional sense (i.e. $C = 0$) or, more generally, because all but a finite number of instruments are completely uncorrelated with the endogenous explanatory variable as the number of

14

instruments approaches infinity, we would not expect consistent estimation of β to be possible when $\tau^2 = 0$. Our results suggest that if one is faced with a situation where only a great many weak instruments are available; then, it may still be worthwhile to make use of these poor quality instruments in constructing bias-corrected estimators of β, so long as the instruments are not completely uncorrelated with the endogenous explanatory variable (i.e. so long as the data are not well-modeled by assuming that $\tau^2 = 0$).

(ii) Theorem 6.2.1 establishes the consistency of the bias-corrected estimators on the basis of a sequential asymptotic scheme. Under stronger but more primitive conditions than those stipulated in this paper, the bias-corrected estimators proposed here can be shown to be consistent under a pathwise asymptotic scheme, whereby the number of instruments approaches infinity as a function of the sample size. (see e.g. Stock and Yogo (2003), who give general conditions under which sequential limit results coincide with results obtained by letting k_{21} and T approach infinity jointly).

(iii) Theorem 6.2.2 shows that in the conventional case where the instruments are fully relevant, all but one of the bias-corrected estimators are still consistent. Indeed, only $\widetilde{\beta}_{OLS,1}$ is inconsistent under Assumption 1*. The reason for this inconsistency is that in this case it can easily be shown that $\widehat{\beta}_{OLS} \xrightarrow{p} \beta_0 + \frac{\sigma_{uv}}{\Pi'\Omega\Pi + \sigma_{vv}}$, whereas the bias-correction factor ,$-(\widehat{\sigma}_{uv,1}/\widehat{\sigma}_{vv,1})$, is a consistent estimator which converges in probability to $-\sigma_{uv}/\sigma_{vv}$. In summary, we see that with the exception of $\widetilde{\beta}_{OLS,1}$, the bias-corrected estimators do not impose a penalty on the user, as far as consistency is concerned, even if the available instruments turn out to be good.[16]

7   Monte Carlo Results

In this section, we report the results of a Monte Carlo study of the sampling behavior of the bias adjusted estimators introduced in Section 6.2. Our experimental setup is based on a special case of the SEM given by equations (1) and (2), with $\gamma = 0$ and $\Phi = 0$. We can write this model in terms of the tth observation as

$$y_{1t} = y_{2t}\beta + u_t, \tag{35}$$

$$y_{2t} = Z_t'\Pi + v_t, \tag{36}$$

where $t = 1, ..., T$, and where the definitions of y_{1t} (1×1), y_{2t} (1×1), Z_t $(k_2 \times 1)$, u_t (1×1), and v_t (1×1) follow directly, given the discussion in Section 2. Note that equation (36) is already written in its reduced form, since Z_t is presumed to be exogenous, and, from Remark 3.2(vi), note that we can also write equation (35) in terms of its reduced form representation as:

$$y_{1t} = Z_t'\Gamma_1 + \varepsilon_{1t}, \quad t = 1, ..., T. \tag{37}$$

Data for our Monte Carlo experiments are generated using the canonical version of the model described in expressions (35)-(37) above. In particular, set $\varepsilon_{2t} = v_t$ for all t, and assume that $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})' \equiv$ i.i.d. $N(0, I_2)$. Also, given the vector of structural disturbances, $w_t = (u_t, v_t)'$, with $E(w_t w_t') = \Sigma = \begin{pmatrix} \sigma_{uu} & \sigma_{vu} \\ \sigma_{vu} & \sigma_{vv} \end{pmatrix}$, we note that our canonical model specification implies that $\sigma_{uu} = 1 + \beta^2$, $\sigma_{uv} = -\beta$,

---

[16]Note also that since the correction terms in the bias-corrected IV estimators are all of order $O_p(T^{-1})$ under Assumption 1* and 2, these estimators are also asymptotically normal in the usual sense under these assumptions.

and $\sigma_{vv} = 1$. Thus, the degree of endogeneity in our data generating processes is determined by the value of the parameter β, or, alternatively, by the value of the correlation coefficient, $\rho_{uv} = \frac{-\beta}{\sqrt{1+\beta^2}}$. We take $\rho_{uv} = \{0.3, 0.5\}$ in generating the data for our experiments. In addition, we examine three (T, k_2) configurations: (i) $T = 2000$, $k_2 = 20$; (ii) $T = 500$, $k_2 = 20$; and (iii) $T = 500$, $k_2 = 100$. We also assume that $Z_t \equiv i.i.d.\ N(0, I_{k_2})$, and we use all k_2 instruments in constructing all estimators. Moreover, we set $\Pi = (\pi_1, \pi_2, ..., \pi_{50})' = (\overline{\pi}, \overline{\pi}, ..., \overline{\pi})'$, so that the degree of relevance of each instrument is assumed to be the same. Given this setup, there is a simple relationship linking $\overline{\pi}$ and the theoretical R^2 of the first stage regression (as defined on page 11 of Hahn, Hausman, and Kuersteiner (2002)). Namely, $R^2_{relev} = \frac{E(\Pi' Z_t Z_t' \Pi)}{E(\Pi' Z_t Z_t' \Pi) + \sigma_{vv}} = \frac{k_{21} \overline{\pi}^2}{k_{21} \overline{\pi}^2 + 1}$, or $\overline{\pi} = \sqrt{\frac{R^2_{relev}}{k_{21}\left(1 - R^2_{relev}\right)}}$. We control for the degree of instrument weakness in our experiments by varying the value of R^2_{relev} over the set $\{0.01, 0.05, 0.10, 0.20\}$. Alternatively, the degree of instrument weakness can also be expressed in terms of $\tau^2 = \frac{\mu' \mu}{k_{21}}$, since it follows that $\tau^2 = \frac{T R^2_{relev}}{k_{21}\left(1 - R^2_{relev}\right)}$ for the canonical case consider here. In the tables below, we report the value of τ^2 for each experiment so as to give an idea of how our results vary with τ^2.

We present two different sets of Monte Carlo results. The first set of results, reported in Tables 3-5, gives median bias (i.e. the median of $\widehat{\beta} - \beta_0$) for the various alternative estimators, while the second set of results, reported in Tables 6-8, gives probabilities of concentration, defined as $P\left[\left|\left(\frac{\mu' \mu}{\sigma_{uu}}\right)\left(\widehat{\beta} - \beta_0\right)\right| \leq \xi\right]$, where $\widehat{\mu' \mu} = \Pi' Z' Z \Pi$.[17] Both set of results are computed on the basis of 5000 simulation draws per experiment.[18]

Turning first to our results on median bias, note that when $T = 500$ (see Tables 3 and 4), $\widehat{\beta}_{LIML}$ tends to be the estimator with the smallest median bias, although this top ranking is not attained uniformly across all experiments in Table 4, where $k_{21} = 20$. Following close behind is one of our bias-adjusted IV estimators, $\widetilde{\beta}_{IV}$, which ranks second or third in most experiments reported in Tables 3 and 4 and even has a top ranking in one experiment. On the other hand, the bias-adjusted OLS estimator, $\widetilde{\beta}_{OLS,1}$ performs well when $k_{21} = 100$ (i.e. when the degree of overidentification is relatively large), but this estimator does not perform nearly as well when the degree of overidentification is more modest (e.g. when $k_{21} = 20$). When the sample size is increased (i.e. $T = 2000$), $\widehat{\beta}_{LIML}$ is no longer the top performer in terms of median bias (see Table 5). In this case, $\widehat{\beta}_{LIML}$ often ranks behind $\widetilde{\beta}_{IV}$, $\widetilde{\widetilde{\beta}}_{IV,1}$, and $\widetilde{\beta}_{OLS,2}$, with $\widetilde{\widetilde{\beta}}_{IV,1}$ being the overall top performer. Note further that the unadjusted OLS estimator, $\widehat{\beta}_{OLS}$, and the unadjusted IV estimator, $\widehat{\beta}_{IV}$, are the two worst performers in terms of median bias. In particular, $\widehat{\beta}_{OLS}$ finishes last in every experiment while $\widehat{\beta}_{IV}$ places either sixth or seventh. These results suggest that our bias correction is effective in reducing the median bias of the (unadjusted) OLS and IV estimators. Finally, observe that, as expected, the median bias of all estimators tends to decrease as the value of τ^2 increases, since a larger value of τ^2 is associated with stronger instruments. Note, however, that the relative ranking of the estimators does not seem to vary in a systematic way with variation in the value of τ^2.

Next, we turn our attention to Tables 6-8, which report probabilities of concentration for the various estimators. Overall, although not uniformly across all experiments, we find $\widehat{\beta}_{LIML}$ to be the estimator

---

[17] We follow Morimune (1983) in our definition of probabilities of concentration (see Section 7.2 of his paper for details). We thank an anonymous referee for suggesting to us the idea of reporting probabilities of concentration.

[18] Of note is that we report median bias and probabilities of concentration in lieu of bias and MSE because these measures do not require existence of moments (as mentioned in Section 2, it is well-known that the bias and MSE of the LIML estimator do not exist under the Gaussian error assumption).

whose sampling distribution is most concentrated around β_0, the true value of β. However, $\widehat{\beta}_{LIML}$ seems to do less well relative to some of our bias-corrected estimators in those cases where the instruments are weak but the degree of overidentification (as given by the order condition) is relative large, as can be seen by looking at entries associated with $R^2_{relev} = 0.01$, 0.05 in Table 7, where $k_{21} = 100$. Indeed, in these "many-weak- instrument" cases, we actually find the bias-adjusted OLS estimator, $\widetilde{\beta}_{OLS,1}$, to perform quite well, often attaining the highest concentration probability. Moreover, $\widetilde{\beta}_{IV}$ and $\widetilde{\widetilde{\beta}}_{IV,1}$ are also seen in Table 7 to have higher concentration probabilities than $\widehat{\beta}_{LIML}$ in a number of cases, when R^2_{relev} is low.

Another setting where $\widehat{\beta}_{LIML}$ seems to do less well, in terms of relative ranking, is when we consider concentration probabilities in a relatively large neighborhood of β_0, such as the case with $\xi = 2.5$, as opposed to the smaller neighborhoods where $\xi = 0.5$ and $\xi = 1.0$. In particular, looking across the three tables, we see that for the cases with $\xi = 0.5$ and $\xi = 1.0$, $\widehat{\beta}_{LIML}$ has the highest concentration probability in 19 out of 24 total experiments and in 20 out of 24 experiments, respectively. On the other hand, for the case where $\xi = 2.5$, $\widehat{\beta}_{LIML}$ has the highest concentration probability in only 7 out of 24 experiments. In fact, focusing specifically on those cases with $R^2_{relev} = 0.01$ in Table 7, we see that the concentration probability of $\widehat{\beta}_{LIML}$ in the larger ($\xi = 2.5$) neighborhood is 0.664 when $\rho_{uv} = 0.3$ and is 0.654 when $\rho_{uv} = 0.5$. These probabilities are substantially lower than the corresponding concentration probabilities of any of our bias-corrected estimators, which are in the range $0.856 - 0.890$. The relatively low concentration probability of $\widehat{\beta}_{LIML}$ in these cases may be attributable, at least in part, to the well-known fact that the finite sample distribution of $\widehat{\beta}_{LIML}$ has thick tails, which become more apparent when there are many weak instruments.

At the other end of the spectrum, the unadjusted OLS estimator, $\widehat{\beta}_{OLS}$, is without question the worst performer in terms of concentration probability. As a result of its large bias, $\widehat{\beta}_{OLS}$ in many instances has a concentration probability of zero, even in the large ($\xi = 2.5$) neighborhood. The unadjusted IV estimator, $\widehat{\beta}_{IV}$, also has a relatively large bias, so it too ranks near the bottom in terms of concentration probabilities in the smaller neighborhoods (i.e. $\xi = 0.5$ and $\xi = 1.0$). Interestingly, however, because the finite sample distribution of $\widehat{\beta}_{IV}$ has relatively thin tails, in many instances, $\widehat{\beta}_{IV}$ actually has a higher concentration probability in the larger ($\xi = 2.5$) neighborhood than most if not all of the other estimators.

8 Concluding Remarks

In this paper, we construct approximations for the ABIAS and AMSE of the IV estimator when the available instruments are assumed to be weakly correlated with the endogenous explanatory variable, using the local-to-zero framework of Staiger and Stock (1997). These approximations are shown, via a series of numerical computations, to be quite accurate. Our approximations, thus, offer useful alternatives to other approximations which have been proposed in the literature (see e.g. Morimune (1983), Donald and Newey (2001), and Hahn, Hausman, and Kuersteiner (2002)), particularly when the instruments are weak. Additionally, we are able to construct a variety of bias-corrected OLS and IV estimators which are consistent under a sequential asymptotic scheme, even when the instruments are weak in the local-to-zero sense. Interestingly, except for one of our bias-corrected OLS estimators, bias correction does not come at a cost, so that consistency is attained under the usual strong-identification assumption as well.

Appendix A

This appendix collects a number of lemmas which we will use to establish the main results of our paper. Before presenting the lemmas, however, we first introduce some notation which will appear in the statements and proofs of some of our lemmas and theorems. To begin, define: $Z_{u,1} = \Omega_{11}^{-\frac{1}{2}'}(\psi_{Z_1 u} - Q_{Z_1 X}Q_{XX}^{-1}\psi_{Xu})\sigma_{uu}^{-\frac{1}{2}}$ and $Z_{v,1} = \Omega_{11}^{-\frac{1}{2}'}(\psi_{Z_1 v} - Q_{Z_1 X}Q_{XX}^{-1}\psi_{Xv})\sigma_{vv}^{-\frac{1}{2}}$, and note that

$$\begin{pmatrix} Z_{u,1} \\ Z_{v,1} \end{pmatrix} \sim N\left(0, \left(\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \otimes I_{k_{21}}\right)\right). \tag{38}$$

In addition, define

$$v_1(\mu'\mu, k_{21}) = (\mu + Z_{v,1})'(\mu + Z_{v,1}) = \sum_{i=1}^{k_{21}} \left(\mu_i + Z_{v,1}^i\right)^2, \tag{39}$$

$$v_2(\mu'\mu, k_{21}) = (\mu + Z_{v,1})'Z_{u,1} = \sum_{i=1}^{k_{21}} \left(\mu_i + Z_{v,1}^i\right) Z_{u,1}^i, \tag{40}$$

where μ_i, $Z_{u,1}^i$, and $Z_{v,1}^i$ are the ith component of μ, $Z_{u,1}$, and $Z_{v,1}$, respectively. Note that we have written $v_1(\cdot, \cdot)$ as a function of $\mu'\mu$ and not μ because v_1 is a noncentral χ^2 random variable which depends on μ only through the noncentrality parameter $\mu'\mu$. In addition, since $\mu'Z_{u,1} \equiv N(0, \mu'\mu)$, $v_2(\mu'\mu, k_{21}) = \mu'Z_{u,1} + Z_{v,1}'Z_{u,1}$ also depends on μ only through $\mu'\mu$. To simplify notations, we will often write v_1 and v_2 instead of $v_1(\mu'\mu, k_{21})$ and $v_2(\mu'\mu, k_{21})$ in places where no confusion is caused by not making explicit the dependence of v_1 and v_2 on $\mu'\mu$ and k_{21}.

The following lemmas will be used in Appendix B to establish the main results of our paper:

Lemma A1: Let $\widehat{\beta}_{IV,T}$ be the IV estimator defined in Section 2 and suppose that (1), (2) and Assumptions 1 and 2 hold. Then, as $T \to \infty$

$$\widehat{\beta}_{IV,T} - \beta_0 \Longrightarrow \sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}} v_1^{-1} v_2. \tag{41}$$

Proof: The proof follows from slight modification of the proof of Theorem 1, part (a) of Staiger and Stock (1997) and is, thus, omitted.

Lemma A2: If $x > 0$ and $a, c > 0$, then as $x \to \infty$,

$$_1F_1(a; c; x) = \frac{\Gamma(c)}{\Gamma(a)} e^x x^{-(c-a)} \left[\sum_{j=0}^{p-1} \frac{(c-a)_j(1-a)_j}{j!} x^{-j} + O(|x|^{-p})\right]. \tag{42}$$

Proof: See Lebedev (1972), pp. 268-271.

Lemma A3: Suppose x is bounded and suppose $a, c \to \infty$ such that $\lim_{a,c \to \infty} \frac{(c-a)x}{c} = 0$. Then,

$$_1F_1(a; c; x) = e^x \left[\sum_{j=0}^{p-1} \frac{(c-a)_j(-x)^j}{(c)_j \, j!} + O(|c|^{-p})\right]. \tag{43}$$

Proof: The proof follows from Kummer's transform. See, for example, Slater (1960), pp.12, 65-66.

Lemma A4: Let $\chi_q^2(\mu'\mu)$ denote a non-central chi-square random variable with noncentrality parameter $\mu'\mu$ and q degrees of freedom. Also let r denote a positive integer such that $q > 2p$. Then,

$$\begin{aligned} E\left[\left(\chi_q^2(\mu'\mu)\right)^{-p}\right] &= 2^{-p} e^{-\frac{\mu'\mu}{2}} \sum_{j=0}^{\infty} \frac{(\frac{\mu'\mu}{2})^j}{j!} \frac{\Gamma\left(\frac{1}{2}(q+2j) - p\right)}{\Gamma\left(\frac{1}{2}(q+2j)\right)} \\ &= 2^{-p} e^{-\frac{\mu'\mu}{2}} \frac{\Gamma\left(\frac{q}{2} - p\right)}{\Gamma(\frac{q}{2})} {}_1F_1\left(\frac{1}{2}q - p; \frac{1}{2}q; \frac{\mu'\mu}{2}\right). \end{aligned} \tag{44}$$

18

Proof: See Ullah (1974), pp. 145-148.

Lemma A5: If the $(J \times 1)$ vector w is distributed normally with mean vector θ and covariance matrix I_J and suppose $\phi(\cdot)$ is a Borel measurable function. Then,

$$E\left[\phi(w'w)w\right] = \theta\, E\left[\phi\left(\chi^2_{J+2}(\theta'\theta)\right)\right]. \tag{45}$$

Proof: See Judge and Bock (1978), Theorem 1 of Appendix B.2, pp.321-322.

Lemma A6: Let the $(J \times 1)$ random vector, w, be normally distributed with mean vector θ and covariance matrix I_J; and suppose that $\phi(\cdot)$ is a Borel measurable function. Then,

$$E\left[\phi(w'w)ww'\right] = E\left[\phi\left(\chi^2_{J+2}(\theta'\theta)\right)\right] I_J + E\left[\phi\left(\chi^2_{J+4}(\theta'\theta)\right)\right]\theta\theta'. \tag{46}$$

Proof: See Judge and Bock (1978), Theorem 3 of Appendix B.2, pp. 323.

Lemma A7: Suppose that Assumption 4 holds. Write $\mu'\mu = \tau^2 k_{21} + R^*(k_{21}) = \mu'\mu(\tau^2, k_{21})$ (say), where $R^*(k_{21}) = O\left(k_{21}^{-1}\right)$. Then, for a given value of τ^2, as $k_{21} \to \infty$, the following results hold

(a)

$$
\begin{aligned}
& {}_1F_1\left(k_{21}/2 - 1;\ k_{21}/2;\ \mu'\mu/2\right)\exp\left\{-\left(\mu'\mu/2\right)\right\} \\
= &\ {}_1F_1\left(k_{21}/2 - 1;\ k_{21}/2;\ \mu'\mu(\tau^2, k_{21})/2\right)\exp\left\{-\left(\mu'\mu(\tau^2, k_{21})/2\right)\right\} \\
= &\ \left(1+\tau^2\right)^{-1} - k_{21}^{-1}\left(1+\tau^2\right)^{-1}\left[2 - 4\left(1+\tau^2\right)^{-1} + 2\left(1+\tau^2\right)^{-2}\right] \\
& - k_{21}^{-2}\left(1+\tau^2\right)^{-2}\left[8 - 28\left(1+\tau^2\right)^{-1} + 32\left(1+\tau^2\right)^{-2} - 12\left(1+\tau^2\right)^{-3}\right] \\
& - R^*(k_{21})k_{21}^{-1}\left(1+\tau^2\right)^{-2} + O\left(k_{21}^{-3}\right),
\end{aligned} \tag{47}
$$

(b)

$$
\begin{aligned}
& {}_1F_1\left(k_{21}/2 - 2;\ k_{21}/2 - 1;\ \mu'\mu/2\right)\exp\left\{-\left(\mu'\mu/2\right)\right\} \\
= &\ {}_1F_1\left(k_{21}/2 - 2;\ k_{21}/2 - 1;\ \mu'\mu(\tau^2, k_{21})/2\right)\exp\left\{-\left(\mu'\mu(\tau^2, k_{21})/2\right)\right\} \\
= &\ \left(1+\tau^2\right)^{-1} - k_{21}^{-1}\left(1+\tau^2\right)^{-1}\left[4 - 6\left(1+\tau^2\right)^{-1} + 2\left(1+\tau^2\right)^{-2}\right] \\
& - k_{21}^{-2}\left(1+\tau^2\right)^{-2}\left[24 - 56\left(1+\tau^2\right)^{-1} + 44\left(1+\tau^2\right)^{-2} - 12\left(1+\tau^2\right)^{-3}\right] \\
& - R^*(k_{21})k_{21}^{-1}\left(1+\tau^2\right)^{-2} + O\left(k_{21}^{-3}\right).
\end{aligned} \tag{48}
$$

Proof: We shall only prove part (a) since the proof for part (b) follows in an analogous manner. To show (a), we make use of a well-known integral representation of the confluent hypergeometric function (see Lebedev (1972) pp. 266) to write

$$
\begin{aligned}
& {}_1F_1\left(k_{21}/2 - 1;\ k_{21}/2;\ \mu'\mu(\tau^2, k_{21})/2\right)\exp\left\{-\left(\mu'\mu(\tau^2, k_{21})/2\right)\right\} \\
= &\ \left[0.5\left(k_{21} - 2\right)\right]\int_0^1 \exp\left\{k_{21}h_1(t)\right\}\exp\left\{0.5R^*(k_{21})(t-1)\right\}dt,
\end{aligned} \tag{49}
$$

where $h_1(t) = 0.5\left[\tau^2(t-1) + \log t\right] - (2/k_{21})\log t$. Given the integral representation (49), we can obtain the expansion given by the right-hand side of expression (47) by applying a Laplace approximation to this integral representation. We note that the maximum of the integrand of (49) in the interval $[0, 1]$ occurs at the

boundary point $t = 1$, and as $k_{21} \to \infty$ the mass of the integral becomes increasingly concentrated in some neighborhood of $t = 1$. Hence, we can obtain an accurate approximation for this integral by approximating the integrand with its Taylor expansion in some shrinking neighborhood of $t = 1$ and by showing that integration over the domain outside of this shrinking neighborhood becomes negligible as k_{21} becomes large. To proceed, notice that the RHS of equation (49) can be written as:

$$[0.5\,(k_{21} - 2)] \int_{1-1/\sqrt{k_{21}}}^{1} \exp\left\{k_{21}h_1(t)\right\} \exp\left\{0.5R^*(k_{21})(t-1)\right\} dt$$

$$+ [0.5\,(k_{21} - 2)] \int_{0}^{1-1/\sqrt{k_{21}}} \exp\left\{k_{21}h_1(t)\right\} \exp\left\{0.5R^*(k_{21})(t-1)\right\} dt = I_1 + I_2 \quad (say), \tag{50}$$

Now, note that:

$$I_2 \;\leq\; [0.5\,(k_{21} - 2)] \exp\left\{-\left(0.5\tau^2\sqrt{k_{21}}\right)\right\} \left(1 - k_{21}^{-\frac{1}{2}}\right)^{(k_{21}-2)/2} \exp\left\{-0.5 k_{21}^{-\frac{1}{2}} R^*(k_{21})\right\}$$

$$=\; O\left(k_{21} \exp\left\{-\left(0.5\tau^2\sqrt{k_{21}}\right)\right\} \left(1 - k_{21}^{-\frac{1}{2}}\right)^{(k_{21}-2)/2}\right), \tag{51}$$

where the inequality holds for $k_{21} \geq 4$. Now, turning our attention to I_1, we first make the change of variable $r = t - 1$ and rewrite $I_1 = [0.5\,(k_{21} - 2)] \int_{-1/\sqrt{k_{21}}}^{0} \exp\left\{k_{21}h_2(r)\right\} \exp\left\{0.5R^*(k_{21})r\right\} dr$ where $h_2(r) = 0.5\left[\tau^2 r + \log(1+r)\right] - (2/k_{21})\log(1+r)$. With this change of variable, we note that the maximum of the integrand of I_1 in the interval $[-1/\sqrt{k_{21}}, 0]$ now occurs at the boundary point $r = 0$. To apply the Laplace approximation to I_1, note first that the derivatives of $h_2(r)$ evaluated at $r = 0$ have the explicit forms: $h_2'(0) = 0.5\left(1 + \tau^2\right) - 2k_{21}^{-1}$ and $h_2^{(i)}(0) = (-1)^{i-1}(i-1)!\left[0.5 - 2k_{21}^{-1}\right]$ for integer $i \geq 2$. By Taylor's formula, we can expand $h_2(r)$ about the point $r = 0$ as follows

$$h_2(r) = h_2(0) + h_2'(0)r + \left(h_2^{(2)}(0)/2!\right) r^2 + \left(h_2^{(3)}(0)/3!\right) r^3 + \left(h_2^{(4)}(r^*)/4!\right) r^4, \tag{52}$$

where r^* lies on the line segment between r and 0 and $h_2(0) = 0$. Moreover, for $-1/\sqrt{k_{21}} \leq r \leq 0$, $|h_2^{(4)}(r^*)| = |3 - 12k_{21}^{-1}|(1+r)^{-4} \leq |3 - 12k_{21}^{-1}|k_{21}^2\left(\sqrt{k_{21}} - 1\right)^{-4} = M(k_{21})$ (say), and note that $M(k_{21}) \to 3$ as $k_{21} \to \infty$. Hence, for $-1/\sqrt{k_{21}} \leq r \leq 0$,

$$\left|h_2(r) - \sum_{i=1}^{3} \frac{h_2^{(i)}(0)}{i!} r^i\right| = \left|\frac{h_2^{(4)}(r^*)}{4!} r^4\right| \leq \left[M(k_{21})r^4\right]/4!. \tag{53}$$

It follows that $\sum_{i=1}^{3} \frac{h_2^{(i)}(0)}{i!} r^i - \frac{M(k_{21})r^4}{4!} \leq h_2(r) \leq \sum_{i=1}^{3} \frac{h_2^{(i)}(0)}{i!} r^i + \frac{M(k_{21})r^4}{4!}$, so that

$$\left(\frac{k_{21}-2}{2}\right) \int_{-\frac{1}{\sqrt{k_{21}}}}^{0} \exp\left\{k_{21}\left(\sum_{i=1}^{3} \frac{h_2^{(i)}(0)}{i!} r^i - \frac{M(k_{21})r^4}{4!}\right)\right\} \exp\left\{\frac{r}{2}R^*(k_{21})\right\} dr$$

$$\leq\; \left(\frac{k_{21}-2}{2}\right) \int_{-1/\sqrt{k_{21}}}^{0} \exp\left\{k_{21}h_2(r)\right\} \exp\left\{0.5rR^*(k_{21})\right\} dr$$

$$\leq\; \left(\frac{k_{21}-2}{2}\right) \int_{-\frac{1}{\sqrt{k_{21}}}}^{0} \exp\left\{k_{21}\left(\sum_{i=1}^{3} \frac{h_2^{(i)}(0)}{i!} r^i + \frac{M(k_{21})r^4}{4!}\right)\right\} \exp\left\{\frac{r}{2}R^*(k_{21})\right\} dr. \tag{54}$$

Let I_3 denote the upper bound integral in expression (54). To evaluate I_3, we rewrite it as

$$I_3 = \left(\frac{k_{21}-2}{2}\right) \int_{-\frac{1}{\sqrt{k_{21}}}}^{0} \exp\left\{k_{21}h_2'(0)r\right\} \exp\left\{k_{21}\left(\sum_{i=2}^{3} \frac{h_2^{(i)}(0)}{i!} r^i + \frac{M(k_{21})r^4}{4!}\right)\right\} \exp\left\{\frac{r}{2}R^*(k_{21})\right\} dr.$$

$$\tag{55}$$

Expanding the latter two exponentials in the integrand above in power series and integrating term-by-term while noting the absolute and uniform convergence of the series involved in the interval $r \in [-1/\sqrt{k_{21}}, 0]$ for $k_{21} \geq 4$; we obtain, after some tedious but straightforward calculations,

$$
\begin{aligned}
I_3 &= (0.5\,(k_{21}-2)) \left[\int_{-1/\sqrt{k_{21}}}^{0} \exp\{k_{21}h_2'(0)r\} \left(1 + \left[k_{21}h_2^{(2)}(0)/2!\right]r^2 \right. \right. \\
&\quad \left. \left. + \left[k_{21}h_2^{(3)}(0)/3!\right]r^3 + \left[k_{21}\left(h_2^{(2)}(0)\right)^2 (2!)^{-3}\right]r^4 + 0.5R^*(k_{21})r\right)dr + O\left(k_{21}^{-4}\right)\right] \\
&= \left(1+\tau^2\right)^{-1} - k_{21}^{-1}\left(1+\tau^2\right)^{-1}\left[2 - 4\left(1+\tau^2\right)^{-1} + 2\left(1+\tau^2\right)^{-2}\right] - k_{21}^{-2}\left(1+\tau^2\right)^{-2}[8- \\
&\quad 28\left(1+\tau^2\right)^{-1} + 32\left(1+\tau^2\right)^{-2} - 12\left(1+\tau^2\right)^{-3}] - R^*(k_{21})k_{21}^{-1}\left(1+\tau^2\right)^{-3} + O\left(k_{21}^{-3}\right). \quad (56)
\end{aligned}
$$

By a similar argument, it can be shown that the lower bound integral in expression (54) can also be approximated by the right-hand side of expression (56). It, thus, follows that

$$
\begin{aligned}
&(0.5\,(k_{21}-2)) \int_{-1/\sqrt{k_{21}}}^{0} \exp\{k_{21}h_2(r)\} \exp\{0.5R^*(k_{21})r\}\,dr \\
&= \left(1+\tau^2\right)^{-1} - k_{21}^{-1}\left(1+\tau^2\right)^{-1}\left[2 - 4\left(1+\tau^2\right)^{-1} + 2\left(1+\tau^2\right)^{-2}\right] - k_{21}^{-2}\left(1+\tau^2\right)^{-2}[8- \\
&\quad 28\left(1+\tau^2\right)^{-1} + 32\left(1+\tau^2\right)^{-2} - 12\left(1+\tau^2\right)^{-3}] - R^*(k_{21})k_{21}^{-1}\left(1+\tau^2\right)^{-3} + O\left(k_{21}^{-3}\right) \quad (57)
\end{aligned}
$$

Finally, the result given in part (a) follows immediately from expressions (51) and (57). □

Lemma A8: Suppose that (1), (2) and Assumptions 1 and 2 hold. Then, the following convergence results hold jointly as $T \to \infty$: Then,

(a) $(u'M_X u/T,\ y_2'M_X u/T,\ y_2'M_X y_2/T) \xrightarrow{p} (\sigma_{uu}, \sigma_{uv}, \sigma_{vv})$;

(b) $Z_1'M_X Z_1/T \xrightarrow{p} \Omega_{11}$, where $\Omega_{11} = Q_{Z_1 Z_1} - Q_{Z_1 X}Q_{XX}^{-1}Q_{XZ_1}$;

(c) $\left\{(Z_1'M_X Z_1)^{-\frac{1}{2}}Z_1'M_X u, (Z_1'M_X Z_1)^{-\frac{1}{2}}Z_1'M_X v\right\} \Longrightarrow \left\{Z_{u,1}\sigma_{uu}^{\frac{1}{2}}, Z_{v,1}\sigma_{vv}^{\frac{1}{2}}\right\}$, where $(Z_{u,1}', Z_{v,1}')'$ has joint normal distribution given by (38);

(d) $(Z_1'M_X Z_1/T)^{-\frac{1}{2}}\left(Z_1'M_X y_2/\sqrt{T}\right) \Longrightarrow (\mu + Z_{v,1})\sigma_{vv}^{\frac{1}{2}}$;

(e) $(y_2'M_X Z_1(Z_1'M_X Z_1)^{-1}Z_1'M_X u, y_2'M_X Z_1(Z_1'M_X Z_1)^{-1}Z_1'M_X y_2, u'M_X Z_1(Z_1'M_X Z_1)^{-1}Z_1'M_X u) \Longrightarrow$
$\left(\sigma_{vv}^{\frac{1}{2}}v_2\sigma_{uu}^{\frac{1}{2}},\ \sigma_{vv}v_1,\ \sigma_{uu}Z_{u,1}'Z_{u,1}\right)$;

(f) $(u'M_{(Z,\,X)}u/T,\ y_2'M_{(Z,\,X)}u/T,\ y_2'M_{(Z,\,X)}y_2/T) \xrightarrow{p} (\sigma_{uu}, \sigma_{uv}, \sigma_{vv})$;

(g) $(y_1'M_{(Z,\,X)}y_1/T,\ y_1'M_{(Z,\,X)}y_2/T) \xrightarrow{p} (g_{11}, g_{12})$, where g_{11} and g_{12} are elements of the reduced form error covariance matrix G;

Proof: Part (a) is identical to part (a) of Lemma A1 of Staiger and Stock (1997), and is proved there. Parts (b)-(e) are similar to parts (b)-(e) of Lemma A1 of Staiger and Stock (1997); the only difference being that Lemma A1 of Staiger and Stock (1997) gives convergence results for sample moments involving the entire instrument matrix, Z, whereas our lemma involves Z_1, the submatrix of Z obtained via column selection. Hence, parts (b)-(e) can be proved by minor modifications of the proof of parts (b)-(e) of Lemma A1 of Staiger and Stock (1997), noting, in particular, that joint convergence holds as a result of Assumption 2 and the continuous mapping theorem.

To show part (f), define $u^\perp = M_X u$, $Z^\perp = M_X Z$, and $y_2^\perp = M_X y_2$ and write $\frac{u'M_{(Z,\,X)}u}{T} = \frac{1}{T}u'M_X u - \frac{1}{T}u^{\perp\prime}P_{Z^\perp}u^\perp$, $\frac{y_2'M_{(Z,\,X)}u}{T} = \frac{1}{T}y_2'M_X u - \frac{1}{T}y_2^{\perp\prime}P_{Z^\perp}u^\perp$, and $\frac{y_2'M_{(Z,\,X)}y_2}{T} = \frac{1}{T}y_2'M_X y_2 - \frac{1}{T}y_2^{\perp\prime}P_{Z^\perp}y_2^\perp$, where

21

$P_{Z\perp} = M_X Z \left(Z'M_X Z\right)^{-1} Z'M_X$. Now, part (e) of Lemma A1 of Staiger and Stock (1997) implies that $\frac{1}{T}u^{\perp\prime}P_{Z\perp}u^{\perp} = Op\left(\frac{1}{T}\right)$, $\frac{1}{T}y_2^{\perp\prime}P_{Z\perp}u^{\perp} = Op\left(\frac{1}{T}\right)$, and $\frac{1}{T}y_2^{\perp\prime}P_{Z\perp}y_2^{\perp} = Op\left(\frac{1}{T}\right)$. The results of part (f) then follow immediately from part (a) of this lemma and the Slutsky Theorem.

To show part (g), first note that g_{11} and g_{12} are related to elements of the structural error covariance matrix Σ by the relations: $g_{11} = \sigma_{uu} + 2\sigma_{uv}\beta + \sigma_{vv}\beta^2$ and $g_{11} = \sigma_{uu} + \sigma_{vv}\beta$, where $\sigma_{vv} = g_{22}$. Next, observe that $\frac{y_1'M_{(Z,\ X)}y_1}{T} = \frac{u'M_{(Z,\ X)}u}{T} + 2\frac{y_2'M_{(Z,\ X)}u}{T}\beta + \frac{y_2'M_{(Z,\ X)}y_2}{T}\beta^2$ and $\frac{y_1'M_{(Z,\ X)}y_2}{T} = \frac{y_2'M_{(Z,\ X)}u}{T} + \frac{y_2'M_{(Z,\ X)}y_2}{T}\beta$. Thus, it follows immediately from part (f) of this lemma and the Slutsky Theorem that $\frac{y_1'M_{(Z,\ X)}y_1}{T} \xrightarrow{p} \sigma_{uu} + 2\sigma_{uv}\beta + \sigma_{vv}\beta^2 = g_{11}$ and $\frac{y_1'M_{(Z,\ X)}y_2}{T} \xrightarrow{p} \sigma_{uu} + \sigma_{vv}\beta = g_{12}$.

Lemma A9: Let Assumption 4 hold, so that $\mu'\mu/k_{21} = \tau^2 + O(k_{21}^{-2})$, for a fixed constant, $\tau^2 \in (0,\infty)$, and write $\mu'\mu = \tau^2 k_{21} + O(k_{21}^{-1}) = \mu'\mu(\tau^2, k_{21})$. Then, as $\mu'\mu$, $k_{21} \to \infty$, (a) $\frac{v_1(\mu'\mu(\tau^2,\ k_{21}),k_{21})}{k_{21}} \xrightarrow{p} (1+\tau^2)$; and (b) $\frac{v_2(\mu'\mu(\tau^2,\ k_{21}),k_{21})}{k_{21}} \xrightarrow{p} \rho$.

Proof: To prove (a), write $\frac{v_1(\mu'\mu(\tau^2,\ k_{21}),k_{21})}{k_{21}} = \frac{Z'_{v,1}Z_{v,1}}{k_{21}} + 2\frac{\mu'Z_{v,1}}{k_{21}} + \frac{\mu'\mu}{k_{21}}$. Next, note that $\frac{\mu'Z_{v,1}}{k_{21}} \equiv N\left(0, \frac{\mu'\mu}{k_{21}^2}\right)$ so that $E\left(\frac{2\mu'Z_{v,1}}{k_{21}}\right)^2 = \frac{4\mu'\mu}{k_{21}^2} = \frac{4\tau^2}{k_{21}} + O(k_{21}^{-3}) = O(k_{21}^{-1})$, and, thus, $2\frac{\mu'Z_{v,1}}{k_{21}} \xrightarrow{p} 0$ as $k_{21} \to \infty$ and $\mu'\mu \to \infty$ under Assumption 4. Moreover, note that, as $k_{21} \to \infty$ and $\mu'\mu \to \infty$ under Assumption 4, $E\left(\frac{Z'_{v,1}Z_{v,1}}{k_{21}} - 1\right)^2 = \frac{2}{k_{21}} \to 0$, so that $\frac{Z'_{v,1}Z_{v,1}}{k_{21}} \xrightarrow{p} 1$, and note also that $\frac{\mu'\mu}{k_{21}} \to \tau^2$. It follows by the Slutsky Theorem that $\frac{Z'_{v,1}Z_{v,1}}{k_{21}} + 2\frac{\mu'Z_{v,1}}{k_{21}} + \frac{\mu'\mu}{k_{21}} \xrightarrow{p} 1 + \tau^2$ under Assumption 4.

To show (b), write $\frac{v_2(\mu'\mu(\tau^2,\ k_{21}),k_{21})}{k_{21}} = \frac{\mu'Z_{u,1}}{k_{21}} + \frac{Z'_{v,1}Z_{u,1}}{k_{21}}$. First, from expression (38), we see that $Z_{u,1} \equiv N(0, I_{k_{21}})$, $Z_{v,1} \equiv N(0, I_{k_{21}})$, and $E(Z_{u,1}Z'_{v,1}) = \rho I_{k_{21}}$. It follows from Khinchine's weak law of large numbers that, as $k_{21} \to \infty$ and $\mu'\mu \to \infty$ under Assumption 4, $\frac{Z'_{v,1}Z_{u,1}}{k_{21}} = (1/k_{21})\sum_{i=1}^{k_{21}} Z^i_{v,1}Z^i_{u,1} \xrightarrow{p} \rho$, where $Z^i_{v,1}$ and $Z^i_{u,1}$ denote the ith component of $Z_{v,1}$ and $Z_{u,1}$, respectively. In addition, note that $\frac{\mu'Z_{u,1}}{k_{21}} \equiv N\left(0, \frac{\mu'\mu}{k_{21}^2}\right)$ so that $E\left(\frac{\mu'Z_{u,1}}{k_{21}}\right)^2 = \frac{\mu'\mu}{k_{21}^2} = \frac{\tau^2}{k_{21}} + O(k_{21}^{-3}) = O(k_{21}^{-1})$, and, thus, $\frac{\mu'Z_{u,1}}{k_{21}} \xrightarrow{p} 0$ under Assumption 4. The desired result follows by the Slutsky Theorem.

Lemma A10: Suppose that (1), (2) and Assumptions 1* and 2 hold. Then, the following convergence results hold as $T \to \infty$: (a) $(u'M_X u/T,\ y_2'M_X u/T,\ y_2'M_X y_2/T) \xrightarrow{p} (\sigma_{uu},\ \sigma_{uv},\ \Pi'\Omega\Pi + \sigma_{vv})$, (b) $(u'M_{(Z,\ X)}u/T,\ y_2'M_{(Z,\ X)}u/T,\ y_2'M_{(Z,\ X)}y_2/T) \xrightarrow{p} (\sigma_{uu},\ \sigma_{uv},\ \sigma_{vv})$, and (c) $(Z_1'M_X Z_1/T,\ Z_1'M_X y_2/T) \xrightarrow{p} (\Omega_{11},\Omega_{1*}\Pi)$.

Proof: Each part of this lemma follows directly from Assumptions 1* and 2 and the Slutsky Theorem. The arguments are standard and well-known, so we omit the details.

<center>

Appendix B

</center>

Proof of Proposition 3.1: To show part (a), we note that by Lemma A1, $U_T = \widehat{\beta}_{IV,T} - \beta_0 \Rightarrow \sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}v_1^{-1}v_2 \equiv U$ (say). Moreover, given Assumption 3, we have by Theorem 5.4 of Billingsley (1968) that $\lim_{T\to\infty} E(U_T) = \lim_{T\to\infty} E\left[\widehat{\beta}_{IV,T} - \beta_0\right] = E\left[\sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}v_1^{-1}v_2\right] = E(U)$. It follows that to derive the asymptotic bias of $\widehat{\beta}_{IV}$, we need merely to give an explicit form for $E\left[\sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}v_1^{-1}v_2\right]$. To proceed, note that, given (38), we can write $Z_{u,1} = Z_{v,1}\rho + Z_{u1.v1}$, where $Z_{u1.v1} \sim N(0,(1-\rho^2)I_{k_{21}})$ represents the projection error and is, thus,

independent of $Z_{v,1}$. Next, we rewrite the limiting random variable U as:

$$
\begin{aligned}
U &= \sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}v_1^{-1}v_2 = \sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}\left[(\mu+Z_{v,1})'(\mu+Z_{v,1})\right]^{-1}(\mu+Z_{v,1})'Z_{u,1} \\
&= \sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}\left[(\mu+Z_{v,1})'(\mu+Z_{v,1})\right]^{-1}(\mu+Z_{v,1})'(Z_{v,1}\rho+Z_{u1.v1}),
\end{aligned}
\tag{58}
$$

so that making use of the law of iterated expectations, we have

$$
\begin{aligned}
E(U) &= E_{Z_{v,1}}\left[E_{Z_{u,1}|Z_{v,1}}\left(\sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}\left[(\mu+Z_{v,1})'(\mu+Z_{v,1})\right]^{-1}(\mu+Z_{v,1})'(Z_{v,1}\rho+Z_{u1.v1})\right)\right] \\
&= \sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}E_{Z_{v,1}}\left[\left[(\mu+Z_{v,1})'(\mu+Z_{v,1})\right]^{-1}(\mu+Z_{v,1})'Z_{v,1}\rho\right],
\end{aligned}
\tag{59}
$$

where $E_{Z_{v,1}}(\cdot)$ and $E_{Z_{u,1}|Z_{v,1}}(\cdot)$ denote, respectively, the expectation taken with respect to the marginal density of $Z_{v,1}$ and the expectation taken with respect to the conditional density of $Z_{u,1}$ given $Z_{v,1}$. Now, to evaluate the right-hand side of (59), we note that:

$$
\begin{aligned}
&\sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}E_{Z_{v,1}}\left[\left[(\mu+Z_{v,1})'(\mu+Z_{v,1})\right]^{-1}(\mu+Z_{v,1})'Z_{v,1}\rho\right] \\
&= \sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}E_{Z_{v,1}}\left[\left[(\mu+Z_{v,1})'(\mu+Z_{v,1})\right]^{-1}(\mu+Z_{v,1})'(\mu+Z_{v,1}-\mu)\rho\right] \\
&= \sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}\rho - \sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}E_{Z_{v,1}}\left[\left[(\mu+Z_{v,1})'(\mu+Z_{v,1})\right]^{-1}(\mu+Z_{v,1})'\mu\rho\right] \\
&= \sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}\rho - \sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}\rho\mu'\mu E\left(\left[\chi_{k_{21}+2}^2(\mu'\mu)\right]^{-1}\right).
\end{aligned}
\tag{60}
$$

where the last line of expression (60) follows from Lemma A5, by noting that $(\mu+Z_{v,1})\sim N(\mu,I_{k_{21}})$, and so $(\mu+Z_{v,1})'(\mu+Z_{v,1})\sim\chi_{k_{21}}^2(\mu'\mu)$. Finally, applying Lemma A4 to (60), we obtain:

$$
\begin{aligned}
\lim_{T\to\infty}E\left[\widehat{\beta}_{IV,T}-\beta_0\right] &= E(U) \\
&= \sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}\rho\left[1-\left(\frac{\mu'\mu}{2}\right)e^{-\frac{\mu'\mu}{2}}\frac{\Gamma(k_{21}/2)}{\Gamma(k_{21}/2+1)}{}_1F_1\left(\frac{k_{21}}{2};\frac{k_{21}}{2}+1;\frac{\mu'\mu}{2}\right)\right] \\
&= \sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}\rho\left[1-e^{-\frac{\mu'\mu}{2}}\left\{\Gamma(k_{21}/2)/\Gamma(k_{21}/2+1)\right\}(k_{21}/2)\right. \\
&\quad\times\left.\left\{{}_1F_1\left(k_{21}/2;\ k_{21}/2;\ \mu'\mu/2\right)-{}_1F_1\left(k_{21}/2-1;\ k_{21}/2;\ \mu'\mu/2\right)\right\}\right] \\
&= \sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}\rho\left[1-e^{-\frac{\mu'\mu}{2}}{}_1F_1\left(k_{21}/2;\ k_{21}/2;\ \mu'\mu/2\right)\right. \\
&\quad\left.+e^{-\frac{\mu'\mu}{2}}{}_1F_1\left(k_{21}/2-1;\ k_{21}/2;\ \mu'\mu/2\right)\right] \\
&= \sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}\rho e^{-\frac{\mu'\mu}{2}}{}_1F_1\left(k_{21}/2-1;\ k_{21}/2;\ \mu'\mu/2\right),
\end{aligned}
\tag{61}
$$

where the third equality above follows from the following recurrence relation:
$z\,{}_1F_1\left(\alpha+1;\gamma+1;z\right)=\gamma\times\left[{}_1F_1\left(\alpha+1;\gamma;z\right)-{}_1F_1\left(\alpha;\gamma;z\right)\right]$; and where the fifth equality above follows from the fact that ${}_1F_1\left(\alpha;\alpha;z\right)=e^z$.

To show part (b), note that $\frac{\mu'\mu}{2}>0,\frac{k_{21}}{2}>0$, and $\frac{k_{21}}{2}+1>0$. Hence, direct application of Lemma A2 yields
$\sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}\rho e^{-\frac{\mu'\mu}{2}}{}_1F_1\left(k_{21}/2-1;\ k_{21}/2;\mu'\mu/2\right)=\sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}\rho e^{-\frac{\mu'\mu}{2}}\left(\Gamma\left(k_{21}/2\right)/\Gamma\left(k_{21}/2-1\right)\right)e^{\frac{\mu'\mu}{2}}\left(\mu'\mu/2\right)^{-1}\left[1+O((\mu'\mu)^{-1})\right]$
$\sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}\rho(k_{21}-2)\left(\mu'\mu\right)^{-1}\left[1+O((\mu'\mu)^{-1})\right]=O((\mu'\mu)^{-1}).$

To show part (c), note that $\lim_{k_{21}\to\infty}\frac{(\frac{1}{2}k_{21}-\frac{1}{2}k_{21}+1)\left(\frac{\mu'\mu}{2}\right)}{(\frac{1}{2}k_{21})}=\lim_{k_2\to\infty}\left(\frac{\mu'\mu}{k_{21}}\right)=0$. Hence, direct application of Lemma A3 gives $\sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}\rho e^{-\frac{\mu'\mu}{2}}{}_1F_1\left(k_{21}/2-1;\ k_{21}/2;\ \mu'\mu/2\right)=\sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}\rho e^{-\frac{\mu'\mu}{2}}e^{\frac{\mu'\mu}{2}}\left[1+O(k_{21}^{-1})\right]=$
$\sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}\rho\left[1+O(k_{21}^{-1})\right]\to\sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}\rho$ as $k_{21}\to\infty$.

To show (d), note that, based on (7), we can write the bias formula in its infinite series form as follows:

$$
\begin{aligned}
b_{\widehat{\beta}_{IV}}\left(\mu'\mu, k_{21}\right) &= \sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}\rho e^{-\frac{\mu'\mu}{2}}\left[\sum_{j=0}^{\infty}\frac{\left(k_{21}/2 - 1\right)_{j}}{\left(k_{21}/2\right)j}\frac{\left(\mu'\mu/2\right)^{j}}{j!}\right] \\
&= \sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}\rho e^{-\frac{\mu'\mu}{2}}\left[\sum_{j=0}^{\infty}\left(\frac{k_{21} - 2}{k_{21} + 2j - 2}\right)\frac{\left(\mu'\mu/2\right)^{j}}{j!}\right]
\end{aligned}
\tag{62}
$$

.

Let $f\left(\mu'\mu, k_{21}\right) = e^{-\frac{\mu'\mu}{2}}\left[\sum_{j=0}^{\infty}\left(\frac{k_{21}-2}{k_{21}+2j-2}\right)\frac{\left(\mu'\mu/2\right)^{j}}{j!}\right]$, and observe that $f\left(\mu'\mu = 0, k_{21}\right) = 1$. Also, from the proof of part (b) above, we know that $\lim_{\mu'\mu\to\infty}f\left(\mu'\mu, k_{21}\right) = 0$. Moreover, note that, under the assumption that $k_{21} \geq 4$,

$$
\begin{aligned}
\frac{\partial f\left(\mu'\mu, k_{21}\right)}{\partial\left(\mu'\mu\right)} &= \left(\frac{1}{2}\right)e^{-\frac{\mu'\mu}{2}}\sum_{j=0}^{\infty}\left(\frac{k_{21}-2}{k_{21}+2j-2}\right)\frac{j\left(\mu'\mu/2\right)^{j-1}}{j!} - \sum_{j=0}^{\infty}\left(\frac{k_{21}-2}{k_{21}+2j-2}\right)\frac{\left(\mu'\mu/2\right)^{j}}{j!} \\
&= \left(\frac{1}{2}\right)e^{-\frac{\mu'\mu}{2}}\sum_{j=1}^{\infty}\left(\frac{k_{21}-2}{k_{21}+2j-2}\right)\frac{\left(\frac{\mu'\mu}{2}\right)^{j-1}}{(j-1)!} - \sum_{j=0}^{\infty}\left(\frac{k_{21}-2}{k_{21}+2j-2}\right)\frac{\left(\mu'\mu/2\right)^{j}}{j!} \\
&= \left(\frac{1}{2}\right)e^{-\frac{\mu'\mu}{2}}\sum_{j=0}^{\infty}\frac{\left(\mu'\mu/2\right)^{j}}{j!}\frac{(k_{21}-2)\left[(k_{21}+2j-2)-(k_{21}+2j)\right]}{(k_{21}+2j-2)(k_{21}+2j)} \\
&= -e^{-\frac{\mu'\mu}{2}}\left[\sum_{j=0}^{\infty}\left(\frac{k_{21}-2}{(k_{21}+2j-2)(k_{21}+2j)}\right)\frac{\left(\mu'\mu/2\right)^{j}}{j!}\right] < 0,
\end{aligned}
\tag{63}
$$

where term-by-term differentiation is justified by the absolute and uniform convergence of the infinite series representation of $f\left(\mu'\mu, k_{21}\right)$ and of the infinite series (63). It follows that $0 \leq f\left(\mu'\mu, k_{21}\right) \leq 1$, and is a monotonically decreasing function of $\left(\mu'\mu\right)$ for $\left(\mu'\mu\right) \in [0, \infty)$. Moreover, from expression (62) and the definition of $f\left(\mu'\mu, k_{21}\right)$, we see that $\left|b_{\widehat{\beta}_{IV}}\left(\mu'\mu, k_{21}\right)\right| = |\rho|\sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}f\left(\mu'\mu, k_{21}\right)$, so that $\left|b_{\widehat{\beta}_{IV}}\left(\mu'\mu, k_{21}\right)\right|$ depends on $\mu'\mu$ only through the factor $f\left(\mu'\mu, k_{21}\right)$. Hence, $\left|b_{\widehat{\beta}_{IV}}\left(\mu'\mu, k_{21}\right)\right|$ is a monotonically decreasing function of $\mu'\mu$ for $\mu'\mu \in [0, \infty)$ and $\sigma_{uv} \neq 0$.

To show (e), we differentiate the infinite series representation of $f\left(\mu'\mu, k_{21}\right)$ term-by-term to obtain:

$$
\frac{\partial f\left(\mu'\mu, k_{21}\right)}{\partial k_{21}} = e^{-\frac{\mu'\mu}{2}}\sum_{j=1}^{\infty}\left[\frac{2j}{(k_{21}+2j-2)^{2}}\right]\frac{\left(\mu'\mu/2\right)^{j}}{j!} > 0,
\tag{64}
$$

noting that interchanging the operations of differentiation and summation is justified by the absolute and uniform convergence of the infinite series involved for $k_{21} \geq 4$. It follows that $f\left(\mu'\mu, k_{21}\right)$ and, thus, $\left|b_{\widehat{\beta}_{IV}}\left(\mu'\mu, k_{21}\right)\right|$ are monotonically increasing functions of k_{21} for $\mu'\mu$ fixed and $\sigma_{uv} \neq 0$.

To show part (f), note that by Theorem 5.4 of Billingsley (1968) and Lemma A1, we have that $\lim_{T\to\infty}E\left[\widehat{\beta}_{IV,T} - \beta_{0}\right]^{2} = \lim_{T\to\infty}E(U_{T}^{2}) = E(U^{2}) = E[\sigma_{uu}\sigma_{vv}^{-1}v_{1}^{-1}v_{2}^{2}v_{1}^{-1}]$. Hence, similar to the proof of part (a) above, derivation of the AMSE only entails the derivation of an explicit form for $E[\sigma_{uu}\sigma_{vv}^{-1}v_{1}^{-1}v_{2}^{2}v_{1}^{-1}]$. To proceed, note that, using expression (38) and the decomposition $Z_{u,1} = Z_{v,1}\rho + Z_{u1.v1}$, we can write $U^{2} = \sigma_{uu}\sigma_{vv}^{-1}v_{1}^{-1}v_{2}^{2}v_{1}^{-1} = \sigma_{uu}\sigma_{vv}^{-1}\left[(\mu + Z_{v,1})'(\mu + Z_{v,1})\right]^{-1}(\mu+Z_{v,1})'(Z_{v,1}\rho + Z_{u1.v1})(Z_{v,1}\rho + Z_{u1.v1})'(\mu+Z_{v,1})\left[(\mu + Z_{v,1})'(\mu + Z_{v,1})\right]^{-1}$,

so that making use of the law of iterated expectations, we have:

$$
\begin{aligned}
E(U^2) &= \sigma_{uu}\sigma_{vv}^{-1}E_{Z_{v,1}}\left[E_{Z_{u,1}|Z_{v,1}}\left([(\mu+Z_{v,1})'(\mu+Z_{v,1})]^{-1}(\mu+Z_{v,1})'(Z_{v,1}\rho+Z_{u1.v1})\right.\right.\\
&\qquad\left.\left.(Z_{v,1}\rho+Z_{u1.v1})'(\mu+Z_{v,1})[(\mu+Z_{v,1})'(\mu+Z_{v,1})]^{-1}\right)\right]\\
&= \sigma_{uu}\sigma_{vv}^{-1}E_{Z_{v,1}}\left([(\mu+Z_{v,1})'(\mu+Z_{v,1})]^{-1}(\mu+Z_{v,1})'\left(Z_{v,1}Z_{v,1}'\rho^2+(1-\rho^2)I_{k_{21}}\right)\right.\\
&\qquad\left.(\mu+Z_{v,1})[(\mu+Z_{v,1})'(\mu+Z_{v,1})]^{-1}\right)\\
&= \sigma_{uu}\sigma_{vv}^{-1}\rho^2E_{Z_{v,1}}\left([(\mu+Z_{v,1})'(\mu+Z_{v,1})]^{-1}(\mu+Z_{v,1})'Z_{v,1}Z_{v,1}'(\mu+Z_{v,1})\right.\\
&\qquad\left.[(\mu+Z_{v,1})'(\mu+Z_{v,1})]^{-1}\right)+\sigma_{uu}\sigma_{vv}^{-1}(1-\rho^2)E_{Z_{v,1}}\left([(\mu+Z_{v,1})'(\mu+Z_{v,1})]^{-1}\right)\\
&= \sigma_{uu}\sigma_{vv}^{-1}\rho^2E_{Z_{v,1}}\left([(\mu+Z_{v,1})'(\mu+Z_{v,1})]^{-1}(\mu+Z_{v,1})'(\mu+Z_{v,1}-\mu)(\mu+Z_{v,1}-\mu)'\right.\\
&\qquad\left.(\mu+Z_{v,1})[(\mu+Z_{v,1})'(\mu+Z_{v,1})]^{-1}\right)\\
&\qquad+\sigma_{uu}\sigma_{vv}^{-1}(1-\rho^2)E_{Z_{v,1}}\left([(\mu+Z_{v,1})'(\mu+Z_{v,1})]^{-1}\right)\\
&= \sigma_{uu}\sigma_{vv}^{-1}\rho^2\left\{1-2E_{Z_{v,1}}\left([(\mu+Z_{v,1})'(\mu+Z_{v,1})]^{-1}(\mu+Z_{v,1})'\mu\right)\right.\\
&\qquad+E_{Z_{v,1}}\left([(\mu+Z_{v,1})'(\mu+Z_{v,1})]^{-1}\mu'(\mu+Z_{v,1})(\mu+Z_{v,1})'\mu\right.\\
&\qquad\left.[(\mu+Z_{v,1})'(\mu+Z_{v,1})]^{-1}\right)+\rho^{-2}\left(1-\rho^2\right)E_{Z_{v,1}}\left([(\mu+Z_{v,1})'(\mu+Z_{v,1})]^{-1}\right)\Bigg\},
\end{aligned}
\tag{65}
$$

where $E_{Z_{v,1}}(\cdot)$ and $E_{Z_{u,1}|Z_{v,1}}(\cdot)$ are expectation operators as defined in the proof of part (a) above. Now, to evaluate the expression to the right of the last equality sign above, we note that since $(\mu+Z_{v,1})\sim N(\mu,I_{k_{21}})$ and $(\mu+Z_{v,1})'(\mu+Z_{v,1})\sim\chi^2_{k_{21}}(\mu'\mu)$, we can apply Lemmas A5 and A6 to (65) above to obtain:

$$
\begin{aligned}
E(U^2) &= \sigma_{uu}\sigma_{vv}^{-1}\rho^2\left\{1-2E\left(\left(\chi^2_{k_{21}+2}(\mu'\mu)\right)^{-1}\right)\mu'\mu+\mu'\left[E\left(\left(\chi^2_{k_{21}+2}(\mu'\mu)\right)^{-2}\right)I_{k_{21}}\right.\right.\\
&\qquad\left.\left.+\left(E\left(\chi^2_{k_{21}+4}(\mu'\mu)\right)^{-2}\right)\mu\mu'\right]\mu+\rho^{-2}\left(1-\rho^2\right)E\left(\left(\chi^2_{k_{21}}(\mu'\mu)\right)^{-1}\right)\right\}.
\end{aligned}
\tag{66}
$$

Finally, applying Lemma A4 to expression (66) above, we obtain:

$$
\begin{aligned}
\lim_{T\to\infty}E\left[\widehat{\beta}_{IV,T}-\beta_0\right]^2 &= E(U^2)\\
&= \sigma_{uu}\sigma_{vv}^{-1}\rho^2\\
&\qquad\left\{1-(\mu'\mu)\,e^{-\frac{\mu'\mu}{2}}\left[\Gamma(k_{21}/2)/\Gamma(k_{21}/2+1)\right]\,{}_1F_1\left(k_{21}/2;\ k_{21}/2+1;\ \mu'\mu/2\right)\right.\\
&\qquad+(\mu'\mu/4)\,e^{-\frac{\mu'\mu}{2}}\left[\Gamma(k_{21}/2-1)/\Gamma(k_{21}/2+1)\right]\,{}_1F_1\left(k_{21}/2-1;k_{21}/2+1;\mu'\mu/2\right)\\
&\qquad+(\mu'\mu/2)^2\,e^{-\frac{\mu'\mu}{2}}\left[\Gamma(k_{21}/2)/\Gamma(k_{21}/2+2)\right]\,{}_1F_1\left(k_{21}/2;\ k_{21}/2+2;\ \mu'\mu/2\right)\\
&\qquad+\left(2\rho^2\right)^{-1}\left(1-\rho^2\right)e^{-\frac{\mu'\mu}{2}}\left[\Gamma(k_{21}/2-1)/\Gamma(k_{21}/2)\right]\times\\
&\qquad\left.{}_1F_1\left(k_{21}/2-1;\ k_{21}/2;\ \mu'\mu/2\right)\right\}\\
&= \sigma_{uu}\sigma_{vv}^{-1}\rho^2A\ (say).
\end{aligned}
\tag{67}
$$

Now, rewrite A as:

$$\begin{aligned}
A &= 1 - \left(\mu'\mu\right) e^{-\left(\mu'\mu/2\right)} 2k_{21}^{-1} \, {}_1F_1\left(k_{21}/2; \ k_{21}/2 + 1; \ \mu'\mu/2\right) \\
&\quad + \left(\mu'\mu/4\right) e^{-\left(\mu'\mu/2\right)} 4 \left(k_{21}(k_{21} - 2)\right)^{-1} \, {}_1F_1\left(k_{21}/2 - 1; \ k_{21}/2 + 1; \ \mu'\mu/2\right) \\
&\quad + \left(\mu'\mu/2\right)^2 e^{-\left(\mu'\mu/2\right)} 4 \left(k_{21}(k_{21} + 2)\right)^{-1} \, {}_1F_1\left(k_{21}/2; \ k_{21}/2 + 2; \ \mu'\mu/2\right) \\
&\quad + \left(2\rho^2\right)^{-1} \left(1 - \rho^2\right) e^{-\left(\mu'\mu/2\right)} 2 \left(k_{21} - 2\right)^{-1} \, {}_1F_1\left(k_{21}/2 - 1; \ k_{21}/2; \ \mu'\mu/2\right). \quad (68)
\end{aligned}$$

Next, note that successive application of the recurrence relation $z \, {}_1F_1\left(\alpha + 1; \gamma + 1; z\right) = \gamma\left[{}_1F_1\left(\alpha + 1; \gamma; z\right) - {}_1F_1\left(\alpha; \gamma; z\right)\right]$ yields:

$$\begin{aligned}
A &= 1 - \left[\left(\mu'\mu\right) e^{-\left(\mu'\mu/2\right)} 2k_{21}^{-1} \, {}_1F_1\left(k_{21}/2; k_{21}/2 + 1; \mu'\mu/2\right)\right] + \left[\left(\mu'\mu/4\right) e^{-\left(\mu'\mu/2\right)} 4 \times\right. \\
&\quad \left. \left(k_{21}(k_{21} - 2)\right)^{-1} \, {}_1F_1\left(k_{21}/2 - 1; k_{21}/2 + 1; \mu'\mu/2\right)\right] + \left[\left(\mu'\mu/2\right) e^{-\left(\mu'\mu/2\right)} 4 \left(k_{21}(k_{21} + 2)\right)^{-1} \times\right. \\
&\quad \left. \left((k_{21} + 2)/2\right) \left\{{}_1F_1\left(k_{21}/2; \ k_{21}/2 + 1; \ \mu'\mu/2\right) - {}_1F_1\left(k_{21}/2 - 1; \ k_{21}/2 + 1; \ \mu'\mu/2\right)\right\}\right] \\
&\quad + \left[\left(2\rho^2\right)^{-1}\left(1 - \rho^2\right) e^{-\left(\mu'\mu/2\right)} 2 \left(k_{21} - 2\right)^{-1} \, {}_1F_1\left(k_{21}/2 - 1; k_{21}/2; \mu'\mu/2\right)\right] \\
&= 1 - \left[\left(\mu'\mu/2\right) e^{-\left(\mu'\mu/2\right)} 2k_{21}^{-1} \, {}_1F_1\left(k_{21}/2; \ k_{21}/2 + 1; \ \mu'\mu/2\right)\right] + \left[e^{-\left(\mu'\mu/2\right)} 2 \left(k_{21}(k_{21} - 2)\right)^{-1} \times\right. \\
&\quad \left. \left(k_{21}/2\right)\left\{{}_1F_1\left(k_{21}/2 - 1; \ k_{21}/2; \ \mu'\mu/2\right) - {}_1F_1\left(k_{21}/2 - 2; \ k_{21}/2; \ \mu'\mu/2\right)\right\}\right] - \left[\left(\mu'\mu/2\right) \times\right. \\
&\quad \left. e^{-\left(\mu'\mu/2\right)} \left(2/k_{21}\right) \, {}_1F_1\left(k_{21}/2 - 1; k_{21}/2 + 1; \mu'\mu/2\right)\right] + \left[\rho^{-2}\left(1 - \rho^2\right) e^{-\left(\mu'\mu/2\right)} \left(k_{21} - 2\right)^{-1} \times\right. \\
&\quad \left. {}_1F_1\left(k_{21}/2 - 1; k_{21}/2; \mu'\mu/2\right)\right] \\
&= 1 - \left[e^{-\left(\mu'\mu/2\right)}\left\{{}_1F_1\left(k_{21}/2; k_{21}/2; \mu'\mu/2\right) - {}_1F_1\left(k_{21}/2 - 1; k_{21}/2; \mu'\mu/2\right)\right\}\right] - \left[e^{-\left(\mu'\mu/2\right)} \times\right. \\
&\quad \left. \left(k_{21} - 2\right)^{-1} \, {}_1F_1\left(k_{21}/2 - 2; k_{21}/2; \mu'\mu/2\right)\right] - \left[e^{-\left(\mu'\mu/2\right)} \left\{{}_1F_1\left(k_{21}/2 - 1; k_{21}/2; \mu'\mu/2\right)\right.\right. \\
&\quad \left.\left. - {}_1F_1\left(k_{21}/2 - 2; k_{21}/2; \mu'\mu/2\right)\right\}\right] + \left[\rho^{-2} e^{-\left(\mu'\mu/2\right)} \left(k_{21} - 2\right)^{-1} \, {}_1F_1\left(k_{21}/2 - 1; k_{21}/2; \mu'\mu/2\right)\right] (69)
\end{aligned}$$

Finally, noting that ${}_1F_1\left(\alpha; \alpha; z\right) = e^z$, we can simplify the expression above by writing:

$$\begin{aligned}
A &= -e^{-\left(\mu'\mu/2\right)}\left(k_{21} - 2\right)^{-1} \, {}_1F_1\left(k_{21}/2 - 2; k_{21}/2; \mu'\mu/2\right) + e^{-\left(\mu'\mu/2\right)} \, {}_1F_1\left(k_{21}/2 - 2; k_{21}/2; \mu'\mu/2\right) \\
&\quad + \rho^{-2} e^{-\left(\mu'\mu/2\right)}\left(k_{21} - 2\right)^{-1} \, {}_1F_1\left(k_{21}/2 - 1; k_{21}/2; \mu'\mu/2\right) \\
&= e^{-\left(\mu'\mu/2\right)}\left[\left(k_{21} - 3\right)/\left(k_{21} - 2\right)\right] \, {}_1F_1\left(k_{21}/2 - 2; k_{21}/2; \mu'\mu/2\right) + \\
&\quad \rho^{-2} e^{-\left(\mu'\mu/2\right)}\left(k_{21} - 2\right)^{-1} \, {}_1F_1\left(k_{21}/2 - 1; k_{21}/2; \mu'\mu/2\right). \quad (70)
\end{aligned}$$

To show (g), first assume that $k_{21} > 4$, so that $\frac{\mu'\mu}{2} > 0$, $\frac{k_{21}}{2} - 1 > 0$, $\frac{k_{21}}{2} > 0$, and $\frac{k_{21}}{2} - 2 > 0$. It follows that we can apply Lemma A2 to each of the confluent hypergeometric functions ${}_1F_1(\cdot; \cdot; \cdot)$ which appear in (6) to obtain:

$$\begin{aligned}
m_{\widehat{\beta}_{IV}}\left(\mu'\mu, k_{21}\right) &= \sigma_{uu}\sigma_{vv}^{-1}\rho^2 e^{-\frac{\mu'\mu}{2}}\left[\left(\frac{k_{21} - 3}{k_{21} - 2}\right)\frac{\Gamma(k_{21}/2)}{\Gamma(k_{21}/2 - 2)} e^{\frac{\mu'\mu}{2}}\left(\frac{\mu'\mu}{2}\right)^{-2}\left(1 + O((\mu'\mu)^{-1})\right)\right. \\
&\quad \left. + \rho^{-2}\left(k_{21} - 2\right)^{-1}\left[\Gamma(k_{21}/2)/\Gamma(k_{21}/2 - 1)\right] e^{\left(\mu'\mu/2\right)}\left(\mu'\mu/2\right)^{-1}\left(1 + O((\mu'\mu)^{-1})\right)\right] \\
&= \sigma_{uu}\sigma_{vv}^{-1}\rho^2\left[\left(\left(k_{21} - 3\right)/\left(k_{21} - 2\right)\right)\left(\left(k_{21} - 2\right)/2\right)\left(\left(k_{21} - 4\right)/2\right)\left(\mu'\mu/2\right)^{-2}\right. \\
&\quad \left. \times \left(1 + O((\mu'\mu)^{-1})\right) + \rho^{-2} 2^{-1}\left(\mu'\mu/2\right)^{-1}\left(1 + O((\mu'\mu)^{-1})\right)\right] \\
&= O((\mu'\mu)^{-1}). \quad (71)
\end{aligned}$$

Next, assume that $k_{21} = 4$, and observe that, in this case, $e^{-\frac{\mu'\mu}{2}} [(k_{21} - 3)/(k_{21} - 2)] \, _1F_1(k_{21}/2 - 2; k_{21}/2; \mu'\mu/2) = e^{-\frac{\mu'\mu}{2}} (1/2) \, _1F_1(0; 2; \mu'\mu/2) = e^{-\frac{\mu'\mu}{2}} (1/2) = O\left(e^{-\frac{\mu'\mu}{2}}\right)$. It follows that

$$m_{\widehat{\beta}_{IV}}(\mu'\mu, 4) = \sigma_{uu}\sigma_{vv}^{-1}\rho^2 \left[e^{-\frac{\mu'\mu}{2}} (1/2) {}_1F_1(0; 2; \mu'\mu/2) + \rho^{-2} e^{-\frac{\mu'\mu}{2}} (1/2) {}_1F_1(1; 2; \mu'\mu/2) \right] = O\left(e^{-\frac{\mu'\mu}{2}}\right) + O\left((\mu'\mu)^{-1}\right)$$

$$= O\left((\mu'\mu)^{-1}\right).$$

To show (h), note that $\lim\limits_{k_{21} \to \infty} \frac{\left(\frac{1}{2}k_{21} - \left(\frac{1}{2}k_{21} - 1\right)\right)\left(\frac{\mu'\mu}{2}\right)}{\left(\frac{1}{2}k_{21}\right)} = \lim\limits_{k_2 \to \infty} \left(\frac{\mu'\mu}{k_{21}}\right) = 0$ and $\lim\limits_{k_{21} \to \infty} \frac{\left(\frac{1}{2}k_{21} - \left(\frac{1}{2}k_{21} - 2\right)\right)\left(\frac{\mu'\mu}{2}\right)}{\left(\frac{1}{2}k_{21}\right)}$

$= \lim\limits_{k_2 \to \infty} \left(\frac{2\mu'\mu}{k_{21}}\right) = 0$. Hence, each of the $_1F_1(\cdot; \cdot; \cdot)$ functions appearing in expression (6) satisfies the

conditions of Lemma A3 so we may apply this lemma to obtain $m_{\widehat{\beta}_{IV}}(\mu'\mu, k_{21}) = \sigma_{uu}\sigma_{vv}^{-1}\rho^2 e^{-\frac{\mu'\mu}{2}} e^{\frac{\mu'\mu}{2}}$

$\left[\left((k_{21} - 3)/(k_{21} - 2)\right)\left(1 + O(k_{21}^{-1})\right) + \rho^{-2}(k_{21} - 2)^{-1}\left(1 + O(k_{21}^{-1})\right) \right] = \sigma_{uu}\sigma_{vv}^{-1}\rho^2 \times \left[1 + O(k_{21}^{-1})\right] \to \sigma_{uu}\sigma_{vv}^{-1}\rho^2$

as $k_{21} \to \infty$.

To show part (i), it suffices to show that $\frac{\partial m_{\widehat{\beta}_{2SLS}}(\mu'\mu, k_{21})}{\partial(\mu'\mu)} < 0$, for all fixed integer $k_{21} \geq 4$. To proceed, write the MSE formula in its infinite series representation as given by (8):

$$
\begin{aligned}
m_{\widehat{\beta}_{IV}}(\mu'\mu, k_{21}) = {}& \sigma_{uu}\sigma_{vv}^{-1}\rho^2 e^{-\frac{\mu'\mu}{2}} \left[\left(\frac{1}{\rho^2}\right)\left(\frac{1}{k_{21} - 2}\right) \sum_{j=0}^{\infty} \frac{(k_{21}/2 - 1)_j}{(k_{21}/2)_j} \frac{(\mu'\mu/2)^j}{j!} \right. \\
& \left. + \left(\frac{k_{21} - 3}{k_{21} - 2}\right) \sum_{j=0}^{\infty} \frac{(k_{21}/2 - 2)_j}{(k_{21}/2)_j} \frac{(\mu'\mu/2)^j}{j!} \right] \\
= {}& \sigma_{uu}\sigma_{vv}^{-1}\rho^2 e^{-\frac{\mu'\mu}{2}} \left[\left(\frac{1}{\rho^2}\right)\left(\frac{1}{k_{21} - 2}\right) \left(1 + \sum_{j=1}^{\infty} \frac{(k_{21}/2 - 1)}{(k_{21}/2 + j - 1)} \frac{(\mu'\mu/2)^j}{j!}\right) + \right. \\
& \left. \left(\frac{k_{21} - 3}{k_{21} - 2}\right) \left(1 + \sum_{j=1}^{\infty} \frac{(k_{21}/2 - 1)(k_{21}/2 - 2)}{(k_{21}/2 + j - 1)(k_{21}/2 + j - 2)} \frac{(\mu'\mu/2)^j}{j!}\right) \right]. \quad (72)
\end{aligned}
$$

Now, differentiating (72) term-by-term, we obtain:

$$
\begin{aligned}
\frac{\partial m_{\widehat{\beta}_{IV}}(\mu'\mu, k_{21})}{\partial(\mu'\mu)} = {}& \left(\frac{1}{2}\right)\sigma_{uu}\sigma_{vv}^{-1}\rho^2 e^{-\frac{\mu'\mu}{2}} \left\{ \left[\frac{1}{\rho^2}\left(\frac{1}{k_{21} - 2}\right) \sum_{j=1}^{\infty} \frac{(k_{21}/2 - 1)}{(k_{21}/2 + j - 1)} \frac{j(\mu'\mu/2)^{j-1}}{j!} + \right. \right. \\
& \left. \left(\frac{k_{21} - 3}{k_{21} - 2}\right) \left(\sum_{j=1}^{\infty} \frac{(k_{21}/2 - 1)(k_{21}/2 - 2)}{(k_{21}/2 + j - 1)(k_{21}/2 + j - 2)} \frac{j(\mu'\mu/2)^{j-1}}{j!}\right) \right] \\
& - \left[\left(\frac{1}{\rho^2}\right)\left(\frac{1}{k_{21} - 2}\right) \left(1 + \sum_{j=1}^{\infty} \frac{(k_{21}/2 - 1)}{(k_{21}/2 + j - 1)} \frac{(\mu'\mu/2)^j}{j!}\right) \right. \\
& \left. \left. + \left(\frac{k_{21} - 3}{k_{21} - 2}\right) \left(1 + \sum_{j=1}^{\infty} \frac{(k_{21}/2 - 1)(k_{21}/2 - 2)}{(k_{21}/2 + j - 1)(k_{21}/2 + j - 2)} \frac{(\mu'\mu/2)^j}{j!}\right) \right] \right\}
\end{aligned}
$$

27

$$
\begin{aligned}
= \quad & -0.5\sigma_{uu}\sigma_{vv}^{-1}e^{-\left(\mu'\mu/2\right)}\left(k_{21}-2\right)^{-1} \times \\
& \left[\frac{2}{k_{21}} + \sum_{j=1}^{\infty}\frac{\left(k_{21}/2-1\right)}{\left(k_{21}/2+j\right)\left(k_{21}/2+j-1\right)}\frac{\left(\mu'\mu/2\right)^j}{j!}\right] - \sigma_{uu}\sigma_{vv}^{-1}\rho^2 e^{-\frac{\mu'\mu}{2}}\left(\frac{k_{21}-3}{k_{21}-2}\right) \\
& \times \left[\frac{2}{k_{21}} + \sum_{j=1}^{\infty}\frac{\left(k_{21}/2-1\right)\left(k_{21}/2-2\right)}{\left(k_{21}/2+j\right)\left(k_{21}/2+j-1\right)\left(k_{21}/2+j-2\right)}\frac{\left(\mu'\mu/2\right)^j}{j!}\right] \\
< \quad & 0 \text{ for } k_{21} \geq 4,
\end{aligned}
\tag{73}
$$

where interchanging the operations of differentiations and summation is justified by the absolute and uniform convergence of the infinite series (72) and (73).

Proof of Lemma 3.3: Note first that Assumption 2 imply that $\widehat{\beta}_{OLS,T} - \beta_0 \overset{p}{\to} \sigma_{uv}/\sigma_{vv}$, as was shown in Staiger and Stock (1997). Moreover, given the assumption $\sup_{T\geq T^*} E[|U_T^*|^{2+\delta}] < \infty$ for some $\delta > 0$ and for some positive integer T^*, the conditions of Theorem 5.4 of Billingsley (1968) are satisfied. It then follows directly from Theorem 5.4 of Billingsley (1968) that $\lim_{T\to\infty} E\left(\widehat{\beta}_{OLS,T} - \beta_0\right) = \sigma_{uv}/\sigma_{vv}$ and $\lim_{T\to\infty} E\left(\widehat{\beta}_{OLS,T} - \beta_0\right)^2 = E(\sigma_{uv}/\sigma_{vv})^2 = \sigma_{uu}\sigma_{vv}^{-1}\rho^2.\square$

Proof of Theorem 4.1: To show part (a), note that direct application of part (a) of Lemma A7 to the bias expression (5) yields:

$$
\begin{aligned}
b_{\widehat{\beta}_{IV}}\left(\tau^2,k_{21}\right) = \quad & \sigma_{uu}^{1/2}\sigma_{vv}^{-1/2}\rho\left\{\left(1+\tau^2\right)^{-1} - k_{21}^{-1}\left(1+\tau^2\right)^{-1}\left[2 - 4\left(1+\tau^2\right)^{-1} + 2\left(1+\tau^2\right)^{-2}\right]\right. \\
& \left. -k_{21}^{-2}\left(1+\tau^2\right)^{-2}\left[8 - 28\left(1+\tau^2\right)^{-1} + 32\left(1+\tau^2\right)^{-2} - 12\left(1+\tau^2\right)^{-3}\right]\right. \\
& \left. -R^*(k_{21})k_{21}^{-1}\left(1+\tau^2\right)^{-2} + O\left(k_{21}^{-3}\right)\right\} \\
= \quad & \sigma_{uu}^{1/2}\sigma_{vv}^{-1/2}\rho\left\{\left(1+\tau^2\right)^{-1} - k_{21}^{-1}\left(1+\tau^2\right)^{-1}\left[2 - 4\left(1+\tau^2\right)^{-1} + 2\left(1+\tau^2\right)^{-2}\right]\right\} \\
& +O\left(k_{21}^{-2}\right) \\
= \quad & \sigma_{uu}^{1/2}\sigma_{vv}^{-1/2}\rho\left\{\left(1+\tau^2\right)^{-1} - 2k_{21}^{-1}\left(1+\tau^2\right)^{-1}\left(\tau^2/\left(1+\tau^2\right)\right)^2\right\} + O\left(k_{21}^{-2}\right).
\end{aligned}
\tag{74}
$$

To show part (b), we first rewrite expression (6) as follows:

$$
\begin{aligned}
m_{\widehat{\beta}_{IV}}\left(\tau^2,k_{21}\right) = \quad & \sigma_{uu}\sigma_{vv}^{-1}\rho^2\left[\rho^{-2}\left(k_{21}-2\right)^{-1}{}_1F_1\left(k_{21}/2-1;\ k_{21}/2;\ \mu'\mu(\tau^2,k_{21})/2\right)e^{-\mu'\mu(\tau^2,k_{21})/2}\right. \\
& + \left(\frac{k_{21}-3}{k_{21}-2}\right)\left(\frac{k_{21}-2}{2}\right){}_1F_1\left(\frac{k_{21}}{2}-2;\frac{k_{21}}{2}-1;\frac{\mu'\mu(\tau^2,k_{21})}{2}\right)e^{-\frac{\mu'\mu(\tau^2,k_{21})}{2}} \\
& \left. - \left(\frac{k_{21}-3}{k_{21}-2}\right)\left(\frac{k_{21}-4}{2}\right){}_1F_1\left(\frac{k_{21}}{2}-1;\frac{k_{21}}{2};\frac{\mu'\mu(\tau^2,k_{21})}{2}\right)e^{-\frac{\mu'\mu(\tau^2,k_{21})}{2}}\right],
\end{aligned}
\tag{75}
$$

where we have made use of the identity $(\gamma-\alpha-1)\,{}_1F_1\left(\alpha;\gamma;z\right) = (\gamma-1)\,{}_1F_1\left(\alpha;\gamma-1;z\right) - \alpha\,{}_1F_1\left(\alpha+1;\gamma;z\right)$ in rewriting expression (6). (See Lebedev (1972), pp. 262, for more details on this and identities involving confluent hypergeometric functions.) Applying the results of Lemma A7 to the confluent hypergeometric

functions in expression (6) above, we obtain:

$$
\begin{aligned}
m_{\widehat{\beta}_{IV}}\left(\tau^2, k_{21}\right) &= \sigma_{uu}\sigma_{vv}^{-1}\rho^2\left[\rho^{-2}k_{21}^{-1}\left(1-2/k_{21}\right)^{-1}\left\{\left(1+\tau^2\right)^{-1}-k_{21}^{-1}\left(1+\tau^2\right)^{-1}\left(2-4\left(1+\tau^2\right)^{-1}\right.\right.\right.\\
&\quad \left.\left.+2\left(1+\tau^2\right)^{-2}\right)-k_{21}^{-2}\left(1+\tau^2\right)^{-2}\left(8-28\left(1+\tau^2\right)^{-1}+32\left(1+\tau^2\right)^{-2}\right.\right.\\
&\quad \left.\left.-12\left(1+\tau^2\right)^{-3}\right)-R^*(k_{21})k_{21}^{-1}\left(1+\tau^2\right)^{-2}+O\left(k_{21}^{-3}\right)\right\}+\left(1-2/k_{21}\right)^{-1}\times\\
&\quad \left(1-3/k_{21}\right)\left(k_{21}/2-1\right)\left\{\left(1+\tau^2\right)^{-1}-k_{21}^{-1}\left(1+\tau^2\right)^{-1}\left(4-6\left(1+\tau^2\right)^{-1}+\right.\right.\\
&\quad \left.\left. 2\left(1+\tau^2\right)^{-2}\right)-k_{21}^{-2}\left(1+\tau^2\right)^{-2}\left(24-56\left(1+\tau^2\right)^{-1}+44\left(1+\tau^2\right)^{-2}\right.\right.\\
&\quad \left.\left.-12\left(1+\tau^2\right)^{-3}\right)-R^*(k_{21})k_{21}^{-1}\left(1+\tau^2\right)^{-2}+O\left(k_{21}^{-3}\right)\right\}-\\
&\quad \left(1-3/k_{21}\right)\left(1-2/k_{21}\right)^{-1}\left(k_{21}/2-2\right)\left\{\left(1+\tau^2\right)^{-1}-k_{21}^{-1}\left(1+\tau^2\right)^{-1}\left(2\right.\right.\\
&\quad \left.\left.-4\left(1+\tau^2\right)^{-1}+2\left(1+\tau^2\right)^{-2}\right)-k_{21}^{-2}\left(1+\tau^2\right)^{-2}\left(8-28\left(1+\tau^2\right)^{-1}+\right.\right.\\
&\quad \left.\left.\left. 32\left(1+\tau^2\right)^{-2}-12\left(1+\tau^2\right)^{-3}\right)-R^*(k_{21})k_{21}^{-1}\left(1+\tau^2\right)^{-2}+O\left(k_{21}^{-3}\right)\right\}\right.
\end{aligned}
\tag{76}
$$

Expanding $\left(1-2/k_{21}\right)^{-1}$ in the binomial series series $\left(1-2/k_{21}\right)^{-1}=1+2/k_{21}+4/k_{21}^2+O\left(k_{21}^{-3}\right)$; and, after some tedious but straightforward calculations, it can be shown that:

$$
\begin{aligned}
m_{\widehat{\beta}_{IV}}\left(\tau^2, k_{21}\right) &= \sigma_{uu}\sigma_{vv}^{-1}\rho^2\left\{\left(1+\tau^2\right)^{-2}+\rho^{-2}k_{21}^{-1}\left(1+\tau^2\right)^{-1}-k_{21}^{-1}\left(1+\tau^2\right)^{-2}\left[7\right.\right.\\
&\quad \left.\left.-12\left(1+\tau^2\right)^{-1}+6\left(1+\tau^2\right)^{-2}\right]+O(k_{21}^{-2})\right\}\\
&= \sigma_{uu}\sigma_{vv}^{-1}\rho^2\left\{\left(1+\tau^2\right)^{-2}+\left(\left(1-\rho^2\right)/\rho^2\right)k_{21}^{-1}\left(1+\tau^2\right)^{-1}+k_{21}^{-1}\left(1+\tau^2\right)^{-1}\times\right.\\
&\quad \left.\left[1-7\left(1+\tau^2\right)^{-1}+12\left(1+\tau^2\right)^{-2}-6\left(1+\tau^2\right)^{-3}\right]\right\}+O(k_{21}^{-2}).
\end{aligned}
\tag{77}
$$

Proof of Lemma 6.1.1: To show part (a), note that since $\widehat{\sigma}_{vv,1}=\frac{y_2'M_{(Z,X)}y_2}{T}$ and $\widehat{\sigma}_{vv,2}=\frac{y_2'M_X y_2}{T}$, it follows directly from part (a) and (f) of Lemma A8 that, as $T\to\infty$, $\widehat{\sigma}_{vv,1}\xrightarrow{p}\sigma_{vv}$ and $\widehat{\sigma}_{vv,2}\xrightarrow{p}\sigma_{vv}$. Note, of course, that these limits do not depend on either k_{21} or $\mu'\mu$. Part (a) follows immediately.

To show part (b), note that it follows from part (d) of Lemma A8, part (a) of this lemma, and the continuous mapping theorem, that as $T\to\infty$, $W_{k_{21},T}=\frac{\left(\frac{y_2'M_X Z_1}{\sqrt{T}}\right)\left(\frac{Z_1'M_X Z_1}{T}\right)^{-1}\left(\frac{Z_1'M_X y_2}{\sqrt{T}}\right)/k_{21}}{\widehat{\sigma}_{vv,1}}\Longrightarrow$ $\frac{(\mu+Z_{v,1})'(\mu+Z_{v,1})}{k_{21}}=\frac{v_1(\mu'\mu,k_{21})}{k_{21}}$. It then follows directly from part (a) of Lemma A9 that, as $k_{21}\to\infty$ and $\mu'\mu\to\infty$ under Assumption 4, $\frac{v_1(\mu'\mu,k_{21})}{k_{21}}\xrightarrow{p}1+\tau^2$, as desired.

To prove part (c), write $\widehat{\sigma}_{uv,1}=\frac{\left(y_1-y_2\widehat{\beta}_{IV}\right)'M_{(Z,X)}y_2}{T}\left(\frac{W_{k_{21},T}}{W_{k_{21},T}-1}\right)=\left[\frac{u'M_{(Z,X)}y_2}{T}-\left(\widehat{\beta}_{IV}-\beta_0\right)\right.$ $\left.\frac{y_2'M_{(Z,X)}y_2}{T}\right]\left(\frac{W_{k_{21},T}}{W_{k_{21},T}-1}\right)$ and $\widehat{\sigma}_{uv,2}=\left[\frac{\left(y_1-y_2\widehat{\beta}_{IV}\right)'M_X y_2}{T}-\left(\widehat{\beta}_{IV}-\beta_0\right)\frac{y_2'M_X y_2}{T}\right]\left(\frac{W_{k_{21},T}}{W_{k_{21},T}-1}\right)$. Applying Lemma A1, parts (a) and (f) of Lemma A8, and the continuous mapping theorem, we see immediately that $\widehat{\sigma}_{uv,1}\Longrightarrow\left[\sigma_{uv}-\sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{\frac{1}{2}}\left(\frac{v_1(\mu'\mu,k_{21})}{k_{21}}\right)^{-1}\left(\frac{v_2(\mu'\mu,k_{21})}{k_{21}}\right)\right]\left(\frac{v_1(\mu'\mu,k_{21})}{k_{21}}-1\right)^{-1}\left(\frac{v_1(\mu'\mu,k_{21})}{k_{21}}\right)=\mathcal{A}_{k_{21},\mu'\mu}$ say, and also that $\widehat{\sigma}_{uv,2}\Longrightarrow\mathcal{A}_{k_{21},\mu'\mu}$ as $T\to\infty$, so that both estimators approach the same random limit as the sample size approaches infinity.[19]   Moreover, applying Lemma A9 and the Slutsky Theorem, we deduce

---

[19] Note that the continuous mapping theorem is applicable here because Assumption 2 implies the joint convergence of the components of $\widehat{\sigma}_{uv,1}$ and $\widehat{\sigma}_{uv,2}$.

that, as $k_{21} \to \infty$ and $\mu'\mu \to \infty$ under Assumption 4, $\mathcal{A}_{k_{21},\mu'\mu} \xrightarrow{p} \left[\sigma_{uv} - \sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{\frac{1}{2}}\rho\left(1+\tau^2\right)^{-1}\right]\left(\frac{1+\tau^2}{\tau^2}\right)$
$= \sigma_{uv}\left[1 - \frac{1}{1+\tau^2}\right]\left(\frac{1+\tau^2}{\tau^2}\right) = \sigma_{uv}$, thus establishing the desired results.

To show part (d), we write $\widehat{\sigma}_{uu,1} = s_{uu} + 2\frac{\widehat{\sigma}_{uv,1}^2}{\widehat{\sigma}_{vv,1}}\left(\frac{1}{W_{k_{21},T}}\right) - \frac{\widehat{\sigma}_{uv,1}^2}{\widehat{\sigma}_{vv,1}}\left(\frac{1}{W_{k_{21},T}}\right)^2$ and $\widehat{\sigma}_{uu,2} = s_{uu} + 2\frac{\widehat{\sigma}_{uv,2}^2}{\widehat{\sigma}_{vv,2}}\left(\frac{1}{W_{k_{21},T}}\right) - \frac{\widehat{\sigma}_{uv,2}^2}{\widehat{\sigma}_{vv,2}}\left(\frac{1}{W_{k_{21},T}}\right)^2$. Note first that $s_{uu} = \frac{(y_1-y_2\widehat{\beta}_{IV})'M_X(y_1-y_2\widehat{\beta}_{IV})}{T} = \frac{u'M_X u}{T} - 2\left(\widehat{\beta}_{IV} - \beta_0\right)\frac{y_2'M_X u}{T} + \left(\widehat{\beta}_{IV} - \beta_0\right)^2\frac{y_2'M_X y_2}{T}$. Moreover, the proofs of parts (a)-(c) above show that $\widehat{\sigma}_{vv,1} \xrightarrow{p} \sigma_{vv}$, $W_{k_{21},T} \Longrightarrow \frac{v_1(\mu'\mu,k_{21})}{k_{21}}$, and $\widehat{\sigma}_{uv,1} \Longrightarrow$ $\mathcal{A}_{k_{21},\mu'\mu}$ as $T \to \infty$. It then follows from Lemma A1, part (a) of Lemma A8, and the continuous mapping theorem, that as $T \to \infty$, $\widehat{\sigma}_{uu,1} \Longrightarrow \sigma_{uu} - 2\sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}\sigma_{uv}\left(\frac{v_1(\mu'\mu,k_{21})}{k_{21}}\right)^{-1}\left(\frac{v_2(\mu'\mu,k_{21})}{k_{21}}\right) + \sigma_{uu}\left(\frac{v_1(\mu'\mu,k_{21})}{k_{21}}\right)^{-2}\left(\frac{v_2(\mu'\mu,k_{21})}{k_{21}}\right)^2 +$ $2\frac{\mathcal{A}_{k_{21},\mu'\mu}^2}{\sigma_{vv}}\left(\frac{v_1(\mu'\mu,k_{21})}{k_{21}}\right)^{-1} - \frac{\mathcal{A}_{k_{21},\mu'\mu}^2}{\sigma_{vv}}\left(\frac{v_1(\mu'\mu,k_{21})}{k_{21}}\right)^{-2} = \mathcal{B}_{k_{21},\mu'\mu}$ (say). Similarly, $\widehat{\sigma}_{uu,2} \Longrightarrow \mathcal{B}_{k_{21},\mu'\mu}$ as $T \to \infty$. Finally, since $\mathcal{A}_{k_{21},\mu'\mu} \xrightarrow{p} \sigma_{uv}$ under Assumption 4, as was shown in the proof of part (c) above; applying Lemma A9 and the Slutsky Theorem, we deduce that $\mathcal{B}_{k_{21},\mu'\mu} \xrightarrow{p} \sigma_{uu} - 2\sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}\sigma_{uv}\left(\frac{\rho}{1+\tau^2}\right)$ $+\sigma_{uu}\left(\frac{\rho}{1+\tau^2}\right)^2 + 2\frac{\sigma_{uv}^2}{\sigma_{vv}}\left(\frac{1}{1+\tau^2}\right) - \frac{\sigma_{uv}^2}{\sigma_{vv}}\left(\frac{1}{1+\tau^2}\right)^2 = \sigma_{uu}$, as $k_{21} \to \infty$ and $\mu'\mu \to \infty$ under Assumption 4, thus, establishing the desired results.

Proof of Theorem 6.1.2: For each part of this theorem, we will only prove the convergence result for the estimator with subscript $i = 1$ since the proofs for the estimators with subscript $i = 2$ follow in a like manner. First, to show (a) for the case $i = 1$, write $\widehat{BIAS}_1 = \frac{\widehat{\sigma}_{uv,1}}{\widehat{\sigma}_{vv,1}}\left(\frac{1}{W_{k_{21},T}}\right)$. Next, note that the proofs of parts (a), (b), and (c) of Lemma 6.1.1 above show that $\widehat{\sigma}_{vv,1} \xrightarrow{p} \sigma_{vv}$, $W_{k_{21},T} \Longrightarrow \frac{v_1(\mu'\mu,k_{21})}{k_{21}}$, and $\widehat{\sigma}_{uv,1} \Longrightarrow \mathcal{A}_{k_{21},\mu'\mu}$ as $T \to \infty$. It follows from the continuous mapping theorem that, as $T \to \infty$, $\widehat{BIAS}_1 \Longrightarrow$ $\frac{\mathcal{A}_{k_{21},\mu'\mu}}{\sigma_{vv}}\left(\frac{v_1(\mu'\mu,k_{21})}{k_{21}}\right)^{-1} = \mathcal{C}_{k_{21},\mu'\mu}$, say. Since $\mathcal{A}_{k_{21},\mu'\mu} \xrightarrow{p} \sigma_{uv}$ under Assumption 4, as was shown in the proof of Lemma 6.1.1(c) above; applying Lemma A9 and the Slutsky Theorem, we deduce that, as $k_{21} \to \infty$ and $\mu'\mu \to \infty$ under Assumption 4, $\mathcal{C}_{k_{21},\mu'\mu} \xrightarrow{p} \frac{1}{\sigma_{vv}}\left(\sigma_{uv} - \sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{\frac{1}{2}}\left(\frac{\rho}{1+\tau^2}\right)\right)\left(\frac{1+\tau^2}{\tau^2}\right)\left(\frac{1}{1+\tau^2}\right) = \sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}\rho\left(\frac{1}{1+\tau^2}\right)$, as required.

To show part (b) for the case $i = 1$, write $\widetilde{BIAS}_1 = \widehat{BIAS}_1 - \left(\frac{2}{k_{21}}\right)\left[\left(\frac{\widehat{\sigma}_{uv,1}}{\widehat{\sigma}_{vv,1}}\right)\left(\frac{1}{W_{k_{21},T}}\right)\left(\frac{W_{k_{21},T}-1}{W_{k_{21},T}}\right)^2\right]$. Again, note that, as $T \to \infty$, $\widehat{\sigma}_{vv,1} \xrightarrow{p} \sigma_{vv}$, $W_{k_{21},T} \Longrightarrow \frac{v_1(\mu'\mu,k_{21})}{k_{21}}$, $\widehat{\sigma}_{uv,1} \Longrightarrow \mathcal{A}_{k_{21},\mu'\mu}$, and $\widehat{BIAS}_1 \Longrightarrow$ $\mathcal{C}_{k_{21},\mu'\mu}$, as were shown in the proofs of parts (a), (b), and (c) of Lemma 6.1.1, and in the proof of part (a) of this theorem. It then follows from the continuous mapping theorem that as $T \to \infty$, $\widetilde{BIAS}_1 \Longrightarrow \mathcal{C}_{k_{21},\mu'\mu} -$ $\frac{2}{k_{21}}\left[\frac{\mathcal{A}_{k_{21},\mu'\mu}}{\sigma_{vv}}\left(\frac{v_1(\mu'\mu,k_{21})}{k_{21}}\right)^{-1}\left(\frac{v_1(\mu'\mu,k_{21})}{k_{21}}\right)^{-2}\left(\frac{v_1(\mu'\mu,k_{21})}{k_{21}} - 1\right)^2\right] = \mathcal{E}_{k_{21},\mu'\mu}$, say. Moreover, note that using Lemma A9 and the Slutsky Theorem, it is easy to show that $\mathcal{E}_{k_{21},\mu'\mu} = \mathcal{C}_{k_{21},\mu'\mu} + O_p\left(\frac{1}{k_{21}}\right)$; and, from the proof of part (a) above, we have that $\mathcal{C}_{k_{21},\mu'\mu} \xrightarrow{p} \sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}\rho\left(\frac{1}{1+\tau^2}\right)$, under Assumption 4. Thus, we deduce that, as $k_{21} \to \infty$ and $\mu'\mu \to \infty$ under Assumption 4, $\mathcal{E}_{k_{21},\mu'\mu} \xrightarrow{p} \sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}\rho\left(\frac{1}{1+\tau^2}\right)$, as required.

To show part (c) for the case $i = 1$, write $\widehat{MSE}_1 = \left(\widehat{BIAS}_1\right)^2$. It follows immediately from the proof of part (a) above and the continuous mapping theorem, that as $T \to \infty$, $\widehat{MSE}_1 \Longrightarrow \mathcal{C}_{k_{21},\mu'\mu}^2$. Moreover, since $\mathcal{C}_{k_{21},\mu'\mu} \xrightarrow{p} \sigma_{uu}^{\frac{1}{2}}\sigma_{vv}^{-\frac{1}{2}}\rho\left(\frac{1}{1+\tau^2}\right)$, as $k_{21} \to \infty$ and $\mu'\mu \to \infty$ under Assumption 4, it follows immediately by the Slutsky Theorem that $\mathcal{C}_{k_{21},\mu'\mu}^2 \xrightarrow{p} \sigma_{uu}\sigma_{vv}^{-1}\rho^2\left(\frac{1}{1+\tau^2}\right)^2$.

To show part (d) for the case $i = 1$, write

30

$$\widetilde{MSE}_1 = \widehat{MSE}_1 + \frac{1}{k_{21}}\left(\frac{\widehat{\sigma}_{uv,1}^2}{\widehat{\sigma}_{vv,1}^2}\right)\left(\frac{1}{W_{k_{21},T}}\right)\left[\left(\frac{\widehat{\sigma}_{uu,1}\widehat{\sigma}_{vv,1}-\widehat{\sigma}_{uv,1}^2}{\widehat{\sigma}_{uv,1}^2}\right)+\left(1-\frac{7}{W_{k_{21},T}}+\frac{12}{W_{k_{21},T}^2}-\frac{6}{W_{k_{21},T}^3}\right)\right].$$

Next, note that, as $T\to\infty$, $\widehat{\sigma}_{vv,1}\xrightarrow{p}\sigma_{vv}$, $W_{k_{21},T}\Longrightarrow\frac{v_1(\mu'\mu,k_{21})}{k_{21}}$, $\widehat{\sigma}_{uv,1}\Longrightarrow\mathcal{A}_{k_{21},\mu'\mu}$, $\widehat{\sigma}_{uu,1}\Longrightarrow\mathcal{B}_{k_{21},\mu'\mu}$, and $\widehat{MSE}_1\Longrightarrow\mathcal{C}_{k_{21},\mu'\mu}^2$, as were shown in the proof of Lemma 6.1.1 and the proof of part (c) above. It then follows by the continuous mapping theorem that as $T\to\infty$, $\widetilde{MSE}_1\Longrightarrow\mathcal{C}_{k_{21},\mu'\mu}^2+\frac{1}{k_{21}}\left(\frac{\mathcal{A}_{k_{21},\mu'\mu}^2}{\sigma_{vv}^2}\right)\left(\frac{v_1(\mu'\mu,k_{21})}{k_{21}}\right)^{-1}$x

$$\text{x}\left[\left(\frac{\sigma_{vv}\mathcal{B}_{k_{21},\mu'\mu}-\mathcal{A}_{k_{21},\mu'\mu}^2}{\mathcal{A}_{k_{21},\mu'\mu}^2}\right)+\left(1-7\left(\frac{v_1(\mu'\mu,k_{21})}{k_{21}}\right)^{-1}+12\left(\frac{v_1(\mu'\mu,k_{21})}{k_{21}}\right)^{-2}-6\left(\frac{v_1(\mu'\mu,k_{21})}{k_{21}}\right)^{-3}\right)\right]=\mathcal{F}_{k_{21},\mu'\mu},$$

say. In addition, note that, using Lemma A9 and the proof of parts (c) and (d) of Lemma 6.1.1, it is easy to show that $\mathcal{F}_{k_{21},\mu'\mu}=\mathcal{C}_{k_{21},\mu'\mu}^2+O_p\left(\frac{1}{k_{21}}\right)$. Moreover, $\mathcal{C}_{k_{21},\mu'\mu}^2\xrightarrow{p}\sigma_{uu}\sigma_{vv}^{-1}\rho^2\left(\frac{1}{1+\tau^2}\right)^2$ under Assumption 4, as was shown in part (c) of this theorem. Thus, we readily deduce that, as $k_{21}\to\infty$ and $\mu'\mu\to\infty$ under Assumption 4, $\mathcal{F}_{k_{21},\mu'\mu}\xrightarrow{p}\sigma_{uu}\sigma_{vv}^{-1}\rho^2\left(\frac{1}{1+\tau^2}\right)^2$, as required.

To show part (e) for the case $i=1$, we note that by comparing the above expression for $\widetilde{MSE}_1$ with the above expression for $\overline{MSE}_1$, the only difference between these two alternative estimators for the MSE, as explained in Subsection 6.1, is that $\widetilde{MSE}_1$ estimates the quantity $\sigma_{uu}\sigma_{vv}-\sigma_{uv}$ using the consistent estimator $\widehat{\sigma}_{uu,1}\widehat{\sigma}_{vv,1}-\widehat{\sigma}_{uv,1}^2$, whereas $\overline{MSE}_1$ estimates the quantity $g_{11}g_{22}-g_{12}^2$ using the estimator $\widehat{g}_{11}\widehat{g}_{22}-\widehat{g}_{12}^2$. Since it is easy to verify that $g_{11}g_{22}-g_{12}^2=\sigma_{uu}\sigma_{vv}-\sigma_{uv}$, all that is left to show is the consistency of the estimator $\widehat{g}_{11}\widehat{g}_{22}-\widehat{g}_{12}^2$. However, given that $\widehat{g}_{11}=\frac{y_1'M_{(Z,X)}y_1}{T}$, $\widehat{g}_{12}=\frac{y_1'M_{(Z,X)}y_2}{T}$, and $\widehat{g}_{22}=\frac{y_2'M_{(Z,X)}y_2}{T}$, we see immediately from parts (f) and (g) of Lemma A8 that, as $T\to\infty$, $\widehat{g}_{11}\xrightarrow{p}g_{11}$, $\widehat{g}_{12}\xrightarrow{p}g_{12}$, and $\widehat{g}_{11}\xrightarrow{p}\sigma_{vv}=g_{22}$; and, thus, by the Slutsky Theorem, $\widehat{g}_{11}\widehat{g}_{22}-\widehat{g}_{12}^2\xrightarrow{p}g_{11}g_{22}-g_{12}^2=\sigma_{uu}\sigma_{vv}-\sigma_{uv}$. Since these limits do not depend on k_{21} and $\mu'\mu$, the desired result follows as a direct consequence.

Proof of Lemma 6.1.4: To begin, we note that since $\widehat{\sigma}_{vv,1}=\frac{y_2'M_{(Z,X)}y_2}{T}$ and $\widehat{\sigma}_{vv,2}=\frac{y_2'M_Xy_2}{T}$, parts (a) and (b) of this lemma follow immediately from parts (b) and (a), respectively, of Lemma A10.

To show part (c), write $W_{k_{21},T}=\left[\frac{y_2'(P_{(Z_1,X)}-P_X)y_2}{\widehat{\sigma}_{vv,1}}\right]k_{21}^{-1}=\frac{T}{\widehat{\sigma}_{vv,1}k_{21}}\left[\frac{y_2'M_XZ_1}{T}\left(\frac{Z_1'M_XZ_1}{T}\right)^{-1}\frac{Z_1'M_Xy_2}{T}\right]$. Thus, it follows directly from part (c) of Lemma A10, part (a) of this lemma, and the Slutsky Theorem, that $W_{k_{21},T}=O_p(T)$.

For part (d), we will only prove the result for estimator $\widehat{\sigma}_{uv,1}$, since the proof for $\widehat{\sigma}_{uv,2}$ is similar. Note that, given part (b) of Lemma A10, the well-known consistency of $\widehat{\beta}_{IV}$ under Assumption 1*, and the Slutsky Theorem, we deduce that as $T\to\infty$, $s_{uv,1}=\frac{(y_1-y_2\widehat{\beta}_{IV})'M_{(Z,X)}y_2}{T}=\frac{u'M_{(Z,X)}y_2}{T}+\left(\beta_0-\widehat{\beta}_{IV}\right)\frac{y_2'M_{(Z,X)}y_2}{T}\xrightarrow{p}\sigma_{uv}$. It follows immediately from part (c) of this lemma and the Slutsky Theorem that $\widehat{\sigma}_{uv,1}=s_{uv,1}\left(\frac{1}{1-\frac{1}{W_{k_{21},T}}}\right)\xrightarrow{p}\sigma_{uv}$, as $T\to\infty$.

For part (e), we will also prove the result only for $\widehat{\sigma}_{uu,1}$, since the proof for $\widehat{\sigma}_{uu,2}$ is again similar. To proceed, note first that $\widehat{\sigma}_{uu,1}$ depends on s_{uu}. Note further that $s_{uu}=\frac{(y_1-y_2\widehat{\beta}_{IV})'M_X(y_1-y_2\widehat{\beta}_{IV})}{T}=\frac{u'M_Xu}{T}+2\left(\beta_0-\widehat{\beta}_{IV}\right)\frac{y_2'M_Xu}{T}+\left(\beta_0-\widehat{\beta}_{IV}\right)^2\frac{y_2'M_Xy_2}{T}\xrightarrow{p}\sigma_{uu}$, as $T\to\infty$, as a direct consequence of part (a) of Lemma A10, the consistency of $\widehat{\beta}_{IV}$ under Assumption 1*, and the Slutsky Theorem. Next, write $\widehat{\sigma}_{uu,1}=s_{uu}+2\frac{\widehat{\sigma}_{uv,1}^2}{\widehat{\sigma}_{vv,1}}\left(\frac{1}{W_{k_{21},T}}\right)-\frac{\widehat{\sigma}_{uv,1}^2}{\widehat{\sigma}_{vv,1}}\left(\frac{1}{W_{k_{21},T}}\right)^2$. In view of parts (a), (c), and (d) of this lemma and the Slutsky Theorem, it is apparent that $\widehat{\sigma}_{uu,1}=s_{uu}+Op\left(\frac{1}{T}\right)$, so that $\widehat{\sigma}_{uu,1}\xrightarrow{p}\sigma_{uu}$, as $T\to\infty$.

Proof of Theorem 6.1.5: The results of parts (a)-(d) follow as a direct consequence of the results of Lemma 6.1.4 and the Slutsky Theorem. Moreover, to show (e) note that, under Assumption 1*, $\widehat{g}_{11}=\frac{y_1'M_{(Z,X)}y_1}{T}\xrightarrow{p}$

g_{11}, $\widehat{g}_{12} = \frac{y_1' M_{(Z,X)} y_2}{T} \xrightarrow{p} g_{12}$, and $\widehat{g}_{22} = \frac{y_2' M_{(Z,X)} y_2}{T} \xrightarrow{p} g_{22}$, as $T \to \infty$, by standard arguments. Hence, applying Lemma 6.1.4 and the Slutsky Theorem, we deduce that $\overline{MSE}_1 = Op\left(\frac{1}{T}\right)$ and $\overline{MSE}_2 = Op\left(\frac{1}{T}\right)$.

Proof of Theorem 6.2.1: We will only prove consistency results for $\widetilde{\beta}_{OLS,1}$, $\widetilde{\beta}_{IV}$, and $\widetilde{\beta}_{IV,1}$ since the results for $\widetilde{\beta}_{OLS,2}$ and $\widetilde{\beta}_{IV,2}$ can be shown in a manner similar to those for $\widetilde{\beta}_{OLS,1}$ and $\widetilde{\beta}_{IV,1}$, respectively. To prove part (a) for the estimator $\widetilde{\beta}_{OLS,1}$, write $\widetilde{\beta}_{OLS,1} = \widehat{\beta}_{OLS} - \frac{\widehat{\sigma}_{uv,1}}{\widehat{\sigma}_{vv,1}} = \beta_0 + (y_2' M_X y_2)^{-1} (y_2' M_X u) - \frac{\widehat{\sigma}_{uv,1}}{\widehat{\sigma}_{vv,1}}$. Note first that, as $T \to \infty$, $\widehat{\sigma}_{vv,1} \xrightarrow{p} \sigma_{vv}$ and $\widehat{\sigma}_{uv,1} \Longrightarrow \mathcal{A}_{k_{21},\mu'\mu}$, as was shown in the proofs of parts (a) and (c) of Lemma 6.1.1. Hence, making use of part (a) of Lemma A8 and the continuous mapping theorem, we see that as $T \to \infty$, $\widetilde{\beta}_{OLS,1} \Longrightarrow \beta_0 + \frac{\sigma_{uv}}{\sigma_{vv}} - \frac{\mathcal{A}_{k_{21},\mu'\mu}}{\sigma_{vv}} = \mathcal{L}_{k_{21},\mu'\mu}$, say. Moreover, the proof of part (c) of Lemma 6.1.1 shows that $\mathcal{A}_{k_{21},\mu'\mu} \xrightarrow{p} \sigma_{uv}$, under Assumption 4. It follows immediately by the Slutsky Theorem that, as $k_{21} \to \infty$ and $\mu'\mu \to \infty$ under Assumption 4, $\mathcal{L}_{k_{21},\mu'\mu} \xrightarrow{p} \beta_0$, as required.

To show part (b), write $\widetilde{\beta}_{IV} = \widehat{\beta}_{IV} - \widehat{BIAS}_1$. Note first that the proof of part (a) of Theorem 6.1.2 shows that $\widehat{BIAS}_1 \Longrightarrow \frac{\mathcal{A}_{k_{21},\mu'\mu}}{\sigma_{vv}} \left(\frac{v_1(\mu'\mu,k_{21})}{k_{21}} \right)^{-1} = \mathcal{C}_{k_{21},\mu'\mu}$, as $T \to \infty$. It then follows from Lemma A1 and the continuous mapping theorem that, as $T \to \infty$, $\widetilde{\beta}_{IV} \Longrightarrow \beta_0 + \sigma_{uu}^{\frac{1}{2}} \sigma_{vv}^{-\frac{1}{2}} \left[\frac{v_2(\mu'\mu,k_{21})}{v_1(\mu'\mu,k_{21})} \right] - \mathcal{C}_{k_{21},\mu'\mu} = \mathcal{M}_{k_{21},\mu'\mu}$, say. Moreover, note that $\mathcal{C}_{k_{21},\mu'\mu} \xrightarrow{p} \sigma_{uu}^{\frac{1}{2}} \sigma_{vv}^{-\frac{1}{2}} \rho \left(\frac{1}{1+\tau^2} \right)$ under Assumption 4, as was shown in the proof of Theorem 6.1.2 part (a). Hence, by applying Lemma A9 and the Slutsky Theorem, we see that $\mathcal{M}_{k_{21},\mu'\mu} \xrightarrow{p} \beta_0 + \sigma_{uu}^{\frac{1}{2}} \sigma_{vv}^{-\frac{1}{2}} \rho \left(\frac{1}{1+\tau^2} \right) - \sigma_{uu}^{\frac{1}{2}} \sigma_{vv}^{-\frac{1}{2}} \rho \left(\frac{1}{1+\tau^2} \right) = \beta_0$, as $k_{21} \to \infty$ and $\mu'\mu \to \infty$ under Assumption 4.

To show part (c) for the estimator $\widetilde{\beta}_{IV,1}$, write $\widetilde{\beta}_{IV,1} = \widehat{\beta}_{IV} - \widetilde{BIAS}_1$. Next, note that the proof of Theorem 6.1.2(b) shows that, as $T \to \infty$, $\widetilde{BIAS}_1 \Longrightarrow \mathcal{E}_{k_{21},\mu'\mu}$. It follows then from Lemma A1 and the continuous mapping theorem that, as $T \to \infty$, $\widetilde{\beta}_{IV,1} \Longrightarrow \beta_0 + \sigma_{uu}^{\frac{1}{2}} \sigma_{vv}^{-\frac{1}{2}} \left[\frac{v_2(\mu'\mu,k_{21})}{v_1(\mu'\mu,k_{21})} \right] - \mathcal{E}_{k_{21},\mu'\mu} = \mathcal{N}_{k_{21},\mu'\mu}$, say. Note further that $\mathcal{E}_{k_{21},\mu'\mu} \xrightarrow{p} \sigma_{uu}^{\frac{1}{2}} \sigma_{vv}^{-\frac{1}{2}} \rho \left(\frac{1}{1+\tau^2} \right)$, as was shown in the proof of Theorem 6.1.2, part (b). Hence, by applying Lemma A9 and the Slutsky Theorem, we readily deduce that $\mathcal{N}_{k_{21},\mu'\mu} \xrightarrow{p} \beta_0 + \sigma_{uu}^{\frac{1}{2}} \sigma_{vv}^{-\frac{1}{2}} \rho \left(\frac{1}{1+\tau^2} \right) - \sigma_{uu}^{\frac{1}{2}} \sigma_{vv}^{-\frac{1}{2}} \rho \left(\frac{1}{1+\tau^2} \right) = \beta_0$, as $k_{21} \to \infty$ and $\mu'\mu \to \infty$ under Assumption 4.

Proof of Theorem 6.2.2: Note that under Assumption 1*, the SEM described in Section 2 is fully identified in the usual sense. Hence, it is well-known by standard arguments that $\widehat{\beta}_{OLS} = \beta + (y_2 M_X y_2/T)^{-1} (y_2 M_X u/T) \xrightarrow{p} \beta_0 + \frac{\sigma_{uv}}{\Pi'\Omega\Pi + \sigma_{vv}}$ and $\widehat{\beta}_{IV} = \beta + \left(\frac{y_2 M_X Z_1}{T} \left(\frac{Z_1' M_X Z_1}{T} \right)^{-1} \frac{Z_1' M_X y_2}{T} \right)^{-1} \left(\frac{y_2 M_X Z_1}{T} \left(\frac{Z_1' M_X Z_1}{T} \right)^{-1} \frac{Z_1' M_X u}{T} \right) \xrightarrow{p} \beta_0$, as $T \to \infty$, as can be seen from direct application of parts (a) and (c) of Lemma A10 and the Slutsky Theorem. Parts (a) and (b) of this theorem then follow as a direct consequence of parts (a), (b), and (d) of Lemma 6.1.4, the probability limit of $\widehat{\beta}_{OLS}$ given above, and the Slutsky Theorem. Parts (c) and (d) follow as a direct consequence of parts (a)-(d) of Lemma 6.1.4, the consistency of $\widehat{\beta}_{IV}$ under full identification, and the Slutsky's Theorem. The arguments are standard, and so we omit the details.

References

Angrist, J. D. and A. B. Krueger (1995). "Split-Sample Instrumental Variables Estimates of the Return to Schooling," *Journal of Business and Economic Statistics*, 13, 225-235.

Anderson, T.W. and T. Sawa (1979). "Evaluation of the Distribution Function of the Two-Stage Least Squares Estimate," *Econometrica*, 47, 163-182.

Bekker, P.A. (1994). "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica*, 62, 657-681.

Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley & Sons, New York.

Bound, J., D.A. Jaeger and R. Baker (1995). "Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variables is Weak," *Journal of the American Statistical Association*, 90, 443-450.

Buse, A. (1992). "The Bias of Instrumental Variable Estimators," *Econometrica*, 60, 173-180.

Choi, I. and P.C.B. Phillips (1992). "Asymptotic and Finite Sample Distribution Theory for IV Estimators and Tests in Partially Identified Structural Equations," *Journal of Econometrics*, 51, 113-150.

Chao, J.C. and N.R. Swanson (2000). "Bias and MSE Analysis of the IV Estimator Under Weak Identification with Application to Bias Correction," University of Maryland and Rutgers University Working Paper.

Donald, S.G. and W.K. Newey (2001). "Choosing the Number of Instruments", *Econometrica*, 69, 1161-1191.

Dufour, J.- M. (1997). "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica*, 68, 1365-1388.

Forchini, G. and G. Hillier (1999). "Conditional Inference for Possibly Unidentified Structural Equations," University of Southampton Working Paper.

Hahn, J. and A. Inoue (2000). "A Monte Carlo Comparison of Various Asymptotic Approximations to the Distribution of Instrumental Variables Estimators," University of Brown Working Paper.

Hall, A.R., G.D. Rudebusch, and D.W. Wilcox (1996). "Judging Instrument Relevance in Instrumental Variables Estimation," *International Economic Review*, 37, 283-298.

Hall, A.R. and F. P. M. Peixe (2000). "A Consistent Method for the Selection of Relevant Instruments," North Carolina State University Working Paper.

Hahn, J. and J. Hausman (2002). "Notes on Bias in Estimators for Simultaneous Equation Model", *Economic Letters, 75, 237-241.*

Hahn, J., J. Hausman, and G. Kuersteiner (2002). "Estimation With Weak Instruments: Accuracy of Higher Order Bias and MSE Approximations", M.I.T. Working Paper.

Hahn, J. and G. Kuersteiner (2002). "Discontinuities of Weak Instrument Limiting Distributions", *Economic Letters, 75, 325-331.*

Hausman J., J. Hahn, and G. Kuersteiner (2001). "Bias Corrected Instrumental Variables Estimation for Dynamic Panel Models with Fixed Effects," M.I.T. Working Paper.

Hillier, G.H., T.W. Kinal and V.K. Srivastava (1984). "On the Moments of Ordinary Least Squares Instrumental Variables Estimators in a General Structural Equation," *Econometrica*, 52, 185-202.

Kitamura, Y. (1994). "Specification Tests with Instrumental Variables and Rank Deficiency," University of Minnesota Working Paper.

Kleibergen, F. (2002). "Pivotal Statistics for testing Structural Parameters in Instrumental Variables Regression", *Econometrica*, 70, 1781-1804.

Kleibergen, F. (2004). "Testing Subsets of Structural Parameters in the IV Regression Model", *Review of Economics and Statistics*, 86, 418-423.

Lebedev, N. N. (1972). *Special Functions and Their Applications* (Dover Publications, Inc.; New York).

Mariano, R.S. and 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.

Mariano, R.S. and J.B. McDonald (1979). "A Note on the Distribution Functions of LIML and 2SLS Structural Coefficient in the Exactly Identified Case," *Journal of the American Statistical Association*, 74, 847-848.

Moreira, M.J. (2001). "Tests with Correct Size when Instruments Can Be Arbitrarily Weak," U.C. Berkeley Working Paper.

Moreira, M.J. (2002). "A Conditional Likelihood Ratio Test for Structural Models," U.C. Berkeley Working Paper.

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.

Nelson, C. and R. Startz (1990a). "Some Further Results on the Exact Small Sample Properties of the Instrumental Variables Estimator," *Econometrica*, 58, 967-976.

Nelson, C. and R. Startz (1990b). "The Distribution of the Instrumental Variables Estimator and Its t-ratio when the Instrument is a Poor One," *Journal of Business*, 63, S125-S140.

Phillips, P.C.B. (1984). "The Exact Distribution of LIML: I," *International Economic Review*, 25, 249-261.

Phillips, P.C.B. (1985). "The Exact Distribution of LIML: II," *International Economic Review*, 26, 21-36.

Phillips, P.C.B. (1989). "Partially Identified Econometric Models," *Econometric Theory*, 5, 181-240.

Richardson, D. H. (1968). "The Exact Distribution of a Structural Coefficient Estimator," *Journal of the American Statistical Association*, 63, 1214-1226.

Richardson, D. H. and D.-M. Wu (1971). "A Note on the Comparison of Ordinary and Two-Stage Least Squares Estimators," *Econometrica*, 39, 973-982.

Sawa, T. (1969). "The Exact Finite Sampling Distribution of OLS and 2SLS Estimators," *Journal of the American Statistical Association*, 64, 923-937.

Shea, J. (1997). "Instrumental Relevance in Multivariate Linear Models: A Simple Measure," *Review of Economics and Statistics*, 79, 348-352.

Slater, L.J. (1960). *Confluent Hypergeometric Functions* (Cambridge University Press, London).

Staiger, D. and J.H. Stock (1997). "Instrumental Variables Regression with Weak Instruments." *Econometrica*, 65, 557-586.

Stock, J.H., J. Wright, and M. Yogo (2002). "GMM, Weak Instruments, and Weak Identification", *Journal of Business and Economic Statistics*, 20, 518-529.

Stock, J.H. and M. Yogo (2002). "Testing for Weak Instruments in Linear IV Regression", Harvard University Working Paper.

Stock, J.H. and M. Yogo (2003). "Asymptotic Distributions of Instrumental Variables Statistics with Many Weak Instruments," Harvard University Working Paper.

Wang, J. and E. Zivot (1998). "Inference on Structural Parameters in Instrumental Variables Regressions with Weak Instruments," *Econometrica*, 66, 1389-1404.

Table 1: Approximation Accuracy for Various Values of $\mu'\mu$ [*]

$\mu'\mu$	Bias CS Approximation		MSE CS Approximation		MSE DN Approximation	
	$\widehat{\phi}_1(t_{\widehat{\phi}_1})$	R^2	$\widehat{\pi}_1(t_{\widehat{\pi}_1})$	R^2	$\widehat{\pi}_1^*(t_{\widehat{\pi}_1^*})$	R^2
2	1.0004(1563.2)	0.9996	0.9993(701.76)	0.9980	0.0004(29.555)	0.4665
4	1.0030(836.01)	0.9986	0.9980(871.11)	0.9987	0.0017(34.233)	0.5398
6	1.0053(693.30)	0.9979	0.9982(1130.2)	0.9992	0.0037(38.209)	0.5937
8	1.0069(657.54)	0.9977	0.9985(1373.2)	0.9995	0.0064(41.777)	0.6360
10	1.0079(656.18)	0.9977	0.9986(1563.0)	0.9996	0.0097(45.084)	0.6705
12	1.0085(668.18)	0.9978	0.9986(1697.9)	0.9997	0.0135(48.213)	0.6994
14	1.0088(685.61)	0.9979	0.9985(1789.8)	0.9997	0.0179(51.212)	0.7242
16	1.0089(705.04)	0.9980	0.9983(1851.5)	0.9997	0.0226(54.110)	0.7456
18	1.0090(724.95)	0.9981	0.9981(1892.8)	0.9997	0.0276(56.929)	0.7644
20	1.0090(744.56)	0.9982	0.9978(1920.2)	0.9997	0.0330(59.684)	0.7810
22	1.0089(763.59)	0.9983	0.9976(1938.0)	0.9997	0.0386(62.385)	0.7957
24	1.0087(781.85)	0.9984	0.9973(1949.0)	0.9997	0.0444(65.043)	0.8090
26	1.0086(799.33)	0.9984	0.9970(1955.0)	0.9997	0.0504(67.663)	0.8209
28	1.0084(816.00)	0.9985	0.9967(1957.1)	0.9997	0.0566(70.250)	0.8317
30	1.0082(831.92)	0.9986	0.9964(1956.4)	0.9997	0.0629(72.810)	0.8414
32	1.0080(847.12)	0.9986	0.9961(1953.3)	0.9997	0.0692(75.346)	0.8504
34	1.0078(861.63)	0.9987	0.9958(1948.3)	0.9997	0.0757(77.861)	0.8585
36	1.0076(875.48)	0.9987	0.9954(1941.8)	0.9997	0.0822(80.357)	0.8660
38	1.0074(888.69)	0.9987	0.9951(1934.2)	0.9997	0.0888(82.838)	0.8729
40	1.0072(901.37)	0.9988	0.9948(1925.4)	0.9997	0.0954(85.304)	0.8793
42	1.0070(913.51)	0.9988	0.9945(1915.7)	0.9997	0.1020(87.758)	0.8852
44	1.0068(925.16)	0.9988	0.9942(1905.5)	0.9997	0.1086(90.200)	0.8906
46	1.0066(936.35)	0.9989	0.9940(1894.6)	0.9997	0.1152(92.633)	0.8957
48	1.0063(947.08)	0.9989	0.9937(1883.2)	0.9997	0.1218(95.057)	0.9004
50	1.0061(957.40)	0.9989	0.9934(1871.3)	0.9997	0.1283(97.473)	0.9049
52	1.0059(967.32)	0.9989	0.9931(1859.0)	0.9997	0.1349(99.882)	0.9090
54	1.0057(976.95)	0.9990	0.9928(1846.4)	0.9997	0.1414(102.28)	0.9128
56	1.0055(986.17)	0.9990	0.9926(1833.5)	0.9997	0.1479(104.68)	0.9165
58	1.0053(995.08)	0.9990	0.9923(1820.3)	0.9997	0.1543(107.07)	0.9198
60	1.0051(1003.7)	0.9990	0.9921(1806.9)	0.9997	0.1606(109.46)	0.9230
62	1.0049(1012.0)	0.9990	0.9918(1793.4)	0.9997	0.1670(111.84)	0.9260
64	1.0047(1020.0)	0.9990	0.9916(1779.7)	0.9997	0.1732(114.22)	0.9289
66	1.0045(1027.9)	0.9991	0.9913(1765.8)	0.9997	0.1795(116.60)	0.9316
68	1.0043(1035.4)	0.9991	0.9911(1751.8)	0.9997	0.1856(118.98)	0.9341
70	1.0041(1042.8)	0.9991	0.9909(1737.8)	0.9997	0.1917(121.35)	0.9365
72	1.0039(1049.9)	0.9991	0.9906(1723.7)	0.9997	0.1977(123.72)	0.9387
74	1.0037(1056.8)	0.9991	0.9904(1709.6)	0.9997	0.2037(126.09)	0.9409
76	1.0036(1063.5)	0.9991	0.9902(1695.4)	0.9997	0.2096(128.46)	0.9429
78	1.0034(1069.9)	0.9991	0.9900(1681.1)	0.9996	0.2154(130.83)	0.9449
80	1.0032(1076.3)	0.9991	0.9898(1666.9)	0.9996	0.2212(133.19)	0.9467
82	1.0030(1082.5)	0.9991	0.9896(1652.7)	0.9996	0.2269(135.56)	0.9484
84	1.0029(1088.5)	0.9992	0.9894(1638.5)	0.9996	0.2326(137.92)	0.9501
86	1.0027(1094.3)	0.9992	0.9892(1624.3)	0.9996	0.2381(140.29)	0.9517
88	1.0025(1100.0)	0.9992	0.9891(1610.1)	0.9996	0.2436(142.65)	0.9532
90	1.0024(1105.7)	0.9992	0.9889(1596.0)	0.9996	0.2491(145.01)	0.9547
92	1.0022(1111.0)	0.9992	0.9887(1581.9)	0.9996	0.2544(147.38)	0.9560
94	1.0020(1116.3)	0.9992	0.9886(1568.0)	0.9996	0.2597(149.74)	0.9574
96	1.0019(1121.5)	0.9992	0.9884(1554.1)	0.9996	0.2650(152.11)	0.9586
98	1.0017(1126.6)	0.9992	0.9882(1540.2)	0.9996	0.2702(154.47)	0.9598
100	1.0016(1131.4)	0.9992	0.9881(1526.4)	0.9996	0.2753(156.84)	0.9610

[*] Notes: 50000 actual bias and MSE values were generated using the analytical formulae given in Section 2 for various values of $\mu'\mu$, k_{21} and β (i.e. $\sigma_{uu}^{1/2}\sigma_{vv}^{-1/2}\rho$), as discussed above. For each value of $\mu'\mu$, a pseudo regression (with 1000 observations) was then run with the actual bias (MSE) regressed on an intercept and an approximate bias (MSE). Slope coefficients (with t-statistics in brackets) are reported, as well as regression R^2 values.

Table 2: Selected Bias and MSE Approximation Comparison Based on Morimune (1983)[*]

					————Bias and MSE Approximations—					
k_2	Exact Bias	Exact MSE	M Simul Bias	M Simul MSE	CS Bias1	CS MSE1	CS Bias2	CS MSE2	M Bias	M MSE
6	-0.52	1.04	-0.50	1.04	-0.68	0.47	1.89	0.89	-0.59	1.02
11	-0.98	1.52	-0.99	1.53	-1.08	1.17	-0.59	1.45	-1.01	1.53
16	-1.31	2.17	-1.32	2.18	-1.38	1.90	-1.22	2.13	-1.33	2.17
21	-1.57	2.83	-1.50	2.62	-1.61	2.61	-1.55	2.81	-1.57	2.83
30	-2.64	7.36	NA	NA	-2.65	7.03	-2.65	7.33	-2.63	7.06
40	-2.82	8.27	NA	NA	-2.83	8.00	-2.83	8.25	-2.81	8.04
50	-2.94	8.91	NA	NA	-2.95	8.68	-2.95	8.89	-2.94	8.72
100	-3.21	10.45	NA	NA	-3.21	10.33	-3.21	10.45	-3.21	10..36

[*] Notes: As in Table 1, exact bias and MSE values are calculated using the analytical formulae given in Section 2. 'M Simul' denotes simulated values reported in Morimune (1983). Approximations reported in the last 6 columns of entries are based on the formulae given in: (i) Morimune (1983) - see columns 10 and 11 and (ii) Section 2 of this paper - see columns 6-9. Note that $\mu'\mu$ equals 25 for the first 4 rows of numerical entries in the table, and equals 10 for the last four rows (corresponding to $\mu^2 = 50$ and $\mu^2 = 20$, respectively, in Morimune's notation since, assuming the canonical setup, we have the relationship $\mu^2 = 2\mu'\mu$). Moreover, the results given above assume that $\alpha = (g_{22}\beta - g_{12})/\sqrt{|G|} = 1$. Finally, note that the bias and MSE values reported are bias and MSE for the standardized $2SLS$ estimator $\sqrt{\sigma_{vv}\sigma_{uu}^{-1}\mu'\mu}\left(\widehat{\beta}_{2SLS,n} - \beta_0\right)$. Thus, the bias values in this table are not directly comparable to median bias reported in Tables 3-5.

Table 3: Median Bias of OLS, IV, LIML and Bias Adjusted Estimators: T=500, k_{21}=20[*]

ρ_{uv}	R^2_{relev}	τ^2	$\widehat{\beta}_{OLS}$	$\widehat{\beta}_{IV}$	$\widetilde{\beta}_{LIML}$	$\widetilde{\beta}_{OLS,1}$	$\widetilde{\beta}_{OLS,2}$	$\widetilde{\beta}_{IV}$	$\widetilde{\widetilde{\beta}}_{IV,1}$	$\widetilde{\widetilde{\beta}}_{IV,2}$
						Bias				
0.30	0.01	0.25	0.642069	0.519421	0.207378	0.384340	0.399306	0.390193	0.393145	0.402227
0.30	0.05	1.32	0.617790	0.275286	0.004419	-0.050427	0.009578	-0.015563	-0.006030	0.017985
0.30	0.10	2.78	0.585699	0.168440	0.001820	-0.074615	0.013305	-0.009481	-0.000710	0.021049
0.30	0.20	6.25	0.521212	0.090011	0.001342	-0.132598	0.019505	-0.000331	0.006385	0.024548
0.50	0.01	0.25	0.989298	0.806058	0.285123	0.563522	0.587890	0.572311	0.576758	0.592036
0.50	0.05	1.32	0.950598	0.429786	-0.005849	-0.071633	0.023109	-0.014204	0.000567	0.035953
0.50	0.10	2.78	0.901002	0.255592	-0.001625	-0.115415	0.024043	-0.010886	0.003143	0.036152
0.50	0.20	6.25	0.801280	0.133479	0.000231	-0.202777	0.028636	-0.002890	0.007622	0.035981

[*] Notes: The 1st and 2nd columns report values of the correlation (ρ_{uv}) between the errors in the canonical model (the degree of endogeneity) and the correlation (R^2_{relev}) between the instruments and the endogenous explanatory variable (instrument relevance). The third column contains the numerical value of $\tau^2 = T\overline{\pi}^2 = \frac{TR^2}{k_{21}}\left(1 - R^2\right)$. The remainder of columns of numerical entries in the table report median bias for the OLS, IV, LIML, and 5 bias corrected estimators. In the 1st column, the correlations, ρ_{uv}, correspond to β = -0.65 and -1.0, respectively, in the canonical model. In the 2nd column, the R^2_{relev} values correspond to $\overline{\pi}$= 0.0225, 0.0513, 0.0745, and 0.1290, respectively, in the canonical model ($\overline{\pi} = \sqrt{(R^2/(k_{21}(1 - R^2)))}$). All entries are based on 5000 Monte Carlo trials (see above for further details).

Table 4: Median Bias of OLS, IV, LIML and Bias Adjusted Estimators: T=500, k_{21}=100[*]

ρ_{uv}	R^2_{relev}	τ^2	$\widehat{\beta}_{OLS}$	$\widehat{\beta}_{IV}$	$\widetilde{\beta}_{LIML}$	$\widetilde{\beta}_{OLS,1}$	$\widetilde{\beta}_{OLS,2}$	$\widetilde{\beta}_{IV}$	$\widetilde{\widetilde{\beta}}_{IV,1}$	$\widetilde{\widetilde{\beta}}_{IV,2}$
						Bias				
0.30	0.01	0.05	0.643391	0.618000	0.380443	0.489365	0.526551	0.500755	0.500871	0.526609
0.30	0.05	0.26	0.617056	0.511610	0.013575	0.228980	0.324097	0.268548	0.269135	0.324553
0.30	0.10	0.56	0.585197	0.414930	-0.004984	0.085967	0.229080	0.157931	0.159392	0.229910
0.30	0.20	1.25	0.519994	0.284456	-0.007453	-0.057066	0.151818	0.077828	0.079610	0.152817
0.50	0.01	0.05	0.990532	0.949783	0.522746	0.787770	0.831620	0.797758	0.798054	0.831691
0.50	0.05	0.26	0.950081	0.790227	0.006611	0.360763	0.509136	0.420137	0.421305	0.509864
0.50	0.10	0.56	0.900656	0.640729	-0.013155	0.140296	0.352060	0.242929	0.244740	0.353416
0.50	0.20	1.25	0.801072	0.440951	-0.010463	-0.079105	0.237252	0.125275	0.128004	0.238985

[*] Notes: See notes to Table 3.

Table 5: Median Bias of OLS, IV, LIML and Bias Adjusted Estimators: T=2000, k_{21}=20 [*]

ρ_{uv}	R^2_{relev}	τ^2	$\widehat{\beta}_{OLS}$	$\widehat{\beta}_{IV}$	$\widetilde{\beta}_{LIML}$	$\widetilde{\beta}_{OLS,1}$	$\widetilde{\beta}_{OLS,2}$	$\widetilde{\widetilde{\beta}}_{IV}$	$\widetilde{\widetilde{\beta}}_{IV,1}$	$\widetilde{\widetilde{\beta}}_{IV,2}$
						Bias				
0.30	0.01	1.01	0.643867	0.325961	0.036167	-0.017684	-0.004708	-0.011181	-0.003259	0.003100
0.30	0.05	5.26	0.617551	0.099999	0.006689	-0.039633	0.001360	-0.004801	0.002371	0.007929
0.30	0.10	11.1	0.585473	0.051138	0.003070	-0.068635	0.003375	-0.002443	0.002071	0.007391
0.30	0.20	25.0	0.520073	0.023744	0.003080	-0.130454	0.004848	-0.000517	0.001860	0.006704
0.50	0.01	1.01	0.990862	0.495984	0.026662	-0.044099	-0.024186	-0.034235	-0.021983	-0.012091
0.50	0.05	5.26	0.951044	0.154537	0.002772	-0.058638	0.002296	-0.007080	0.003566	0.012258
0.50	0.10	11.1	0.900966	0.080038	0.002844	-0.103521	0.004547	-0.004286	0.002330	0.010706
0.50	0.20	25.0	0.800782	0.037816	0.002363	-0.200564	0.006417	-0.001575	0.002096	0.009369

[*] Notes: See notes to Table 3.

Table 6: Concentration Probabilities of OLS, IV, LIML and Bias Adjusted Estimators: T=500, k_{21}=20 [*]

ρ_{uv}	R^2_{relev}	τ^2	$\widehat{\beta}_{OLS}$	$\widehat{\beta}_{IV}$	$\widetilde{\beta}_{LIML}$	$\widetilde{\beta}_{OLS,1}$	$\widetilde{\beta}_{OLS,2}$	$\widetilde{\widetilde{\beta}}_{IV}$	$\widetilde{\widetilde{\beta}}_{IV,1}$	$\widetilde{\widetilde{\beta}}_{IV,2}$
					$\xi = 0.5$					
0.30	0.01	0.25	0.000	0.111	0.251	0.199	0.197	0.194	0.192	0.197
0.30	0.05	1.32	0.000	0.148	0.319	0.279	0.292	0.280	0.283	0.289
0.30	0.10	2.78	0.000	0.212	0.364	0.299	0.348	0.346	0.345	0.350
0.30	0.20	6.25	0.000	0.291	0.384	0.204	0.360	0.370	0.371	0.363
0.50	0.01	0.25	0.000	0.014	0.235	0.183	0.175	0.179	0.178	0.175
0.50	0.05	1.32	0.000	0.049	0.339	0.286	0.272	0.281	0.282	0.276
0.50	0.10	2.78	0.000	0.119	0.375	0.301	0.325	0.335	0.338	0.329
0.50	0.20	6.25	0.000	0.233	0.382	0.135	0.354	0.365	0.368	0.357
					$\xi = 1.0$					
0.30	0.01	0.25	0.008	0.531	0.446	0.374	0.381	0.372	0.374	0.382
0.30	0.05	1.32	0.000	0.390	0.584	0.508	0.536	0.522	0.530	0.540
0.30	0.10	2.78	0.000	0.466	0.636	0.571	0.614	0.603	0.611	0.620
0.30	0.20	6.25	0.000	0.570	0.652	0.380	0.664	0.656	0.653	0.668
0.50	0.01	0.25	0.000	0.212	0.443	0.357	0.354	0.352	0.351	0.353
0.50	0.05	1.32	0.000	0.196	0.598	0.514	0.500	0.503	0.502	0.506
0.50	0.10	2.78	0.000	0.320	0.643	0.567	0.593	0.585	0.584	0.598
0.50	0.20	6.25	0.000	0.465	0.660	0.285	0.652	0.640	0.648	0.650
					$\xi = 2.5$					
0.30	0.01	0.25	1.000	0.999	0.734	0.700	0.718	0.702	0.705	0.719
0.30	0.05	1.32	0.226	0.981	0.917	0.857	0.881	0.869	0.870	0.886
0.30	0.10	2.78	0.000	0.974	0.958	0.906	0.953	0.935	0.944	0.957
0.30	0.20	6.25	0.000	0.980	0.977	0.842	0.974	0.968	0.970	0.975
0.50	0.01	0.25	1.000	1.000	0.755	0.701	0.708	0.701	0.702	0.711
0.50	0.05	1.32	0.000	0.951	0.913	0.862	0.879	0.867	0.870	0.884
0.50	0.10	2.78	0.000	0.958	0.958	0.897	0.944	0.932	0.937	0.948
0.50	0.20	6.25	0.000	0.970	0.974	0.757	0.972	0.968	0.968	0.977

[*] Notes: Numerical values entries are the probabilities of concentration defined as follows:

$$P\left[-\xi \le \left(\sigma_{uu}^{-1}\widehat{\mu'\mu}\right)\left(\widehat{\beta} - \beta_0\right) \le \xi\right],$$

where $\widehat{\mu'\mu} = \Pi'Z'Z\Pi$. A canonical model where the reduced form error covariance matrix is the identity matrix is assumed. All entries are based on 5000 Monte Carlo trials (see above for further details).

Table 7: Concentration Probabilities of OLS, IV, LIML and Bias Adjusted Estimators:
T=500, k_{21}=100[*]

ρ_{uv}	R^2_{relev}	τ^2	$\widehat{\beta}_{OLS}$	$\widehat{\beta}_{IV}$	$\widetilde{\beta}_{LIML}$	$\widetilde{\beta}_{OLS,1}$	$\widetilde{\beta}_{OLS,2}$	$\widetilde{\beta}_{IV}$	$\widetilde{\beta}_{IV,1}$	$\widetilde{\widetilde{\beta}}_{IV,2}$
					$\xi = 0.5$					
0.30	0.01	0.05	0.000	0.000	0.176	0.212	0.190	0.204	0.203	0.190
0.30	0.05	0.26	0.000	0.000	0.193	0.210	0.144	0.200	0.198	0.144
0.30	0.10	0.56	0.000	0.000	0.237	0.242	0.143	0.225	0.224	0.142
0.30	0.20	1.25	0.000	0.000	0.278	0.237	0.138	0.279	0.276	0.135
0.50	0.01	0.05	0.000	0.000	0.162	0.120	0.096	0.113	0.113	0.096
0.50	0.05	0.26	0.000	0.000	0.208	0.157	0.075	0.128	0.127	0.075
0.50	0.10	0.56	0.000	0.000	0.258	0.223	0.085	0.163	0.165	0.084
0.50	0.20	1.25	0.000	0.000	0.308	0.266	0.087	0.228	0.223	0.085
					$\xi = 1.0$					
0.30	0.01	0.05	0.009	0.208	0.343	0.459	0.448	0.452	0.451	0.448
0.30	0.05	0.26	0.000	0.000	0.370	0.387	0.320	0.368	0.369	0.321
0.30	0.10	0.56	0.000	0.000	0.465	0.458	0.327	0.409	0.407	0.327
0.30	0.20	1.25	0.000	0.010	0.551	0.485	0.339	0.476	0.475	0.333
0.50	0.01	0.05	0.000	0.000	0.322	0.321	0.267	0.301	0.300	0.267
0.50	0.05	0.26	0.000	0.000	0.406	0.313	0.187	0.268	0.269	0.187
0.50	0.10	0.56	0.000	0.000	0.493	0.433	0.200	0.334	0.330	0.197
0.50	0.20	1.25	0.000	0.000	0.569	0.489	0.199	0.413	0.414	0.196
					$\xi = 2.5$					
0.30	0.01	0.05	1.000	1.000	0.664	0.865	0.889	0.869	0.869	0.890
0.30	0.05	0.26	0.230	0.777	0.725	0.834	0.851	0.835	0.835	0.851
0.30	0.10	0.56	0.437	0.838	0.878	0.846	0.869	0.870	0.845	
0.30	0.20	1.25	0.000	0.406	0.942	0.887	0.858	0.930	0.931	0.854
0.50	0.01	0.05	1.000	1.000	0.654	0.856	0.883	0.860	0.860	0.883
0.50	0.05	0.26	0.000	0.161	0.756	0.812	0.747	0.786	0.785	0.747
0.50	0.10	0.56	0.000	0.044	0.865	0.877	0.712	0.825	0.823	0.709
0.50	0.20	1.25	0.000	0.071	0.948	0.881	0.713	0.904	0.903	0.710

[*] Notes: See notes to Table 6.

Table 8: Concentration Probabilities of OLS, IV, LIML and Bias Adjusted Estimators:
T=2000, k_{21}=20[*]

ρ_{uv}	R^2_{relev}	τ^2	$\widehat{\beta}_{OLS}$	$\widehat{\beta}_{IV}$	$\widetilde{\beta}_{LIML}$	$\widetilde{\beta}_{OLS,1}$	$\widetilde{\beta}_{OLS,2}$	$\widetilde{\beta}_{IV}$	$\widetilde{\beta}_{IV,1}$	$\widetilde{\widetilde{\beta}}_{IV,2}$
					$\xi = 0.5$					
0.30	0.01	1.01	0.000	0.136	0.302	0.225	0.224	0.225	0.224	0.230
0.30	0.05	5.26	0.000	0.286	0.332	0.312	0.331	0.324	0.335	0.331
0.30	0.10	11.1	0.000	0.325	0.354	0.276	0.339	0.348	0.343	0.338
0.30	0.20	25.0	0.000	0.344	0.355	0.037	0.349	0.354	0.346	0.355
0.50	0.01	1.01	0.000	0.046	0.297	0.221	0.221	0.218	0.219	0.224
0.50	0.05	5.26	0.000	0.209	0.347	0.314	0.321	0.329	0.319	0.317
0.50	0.10	11.1	0.000	0.284	0.350	0.240	0.335	0.340	0.331	0.344
0.50	0.20	25.0	0.000	0.334	0.360	0.004	0.348	0.359	0.350	0.339
					$\xi = 1.0$					
0.30	0.01	1.01	0.000	0.373	0.549	0.464	0.475	0.469	0.475	0.476
0.30	0.05	5.26	0.000	0.517	0.642	0.584	0.611	0.607	0.609	0.620
0.30	0.10	11.1	0.000	0.577	0.655	0.513	0.635	0.636	0.638	0.636
0.30	0.20	25.0	0.000	0.628	0.657	0.111	0.653	0.647	0.649	0.652
0.50	0.01	1.01	0.000	0.166	0.571	0.465	0.456	0.462	0.456	0.456
0.50	0.05	5.26	0.000	0.430	0.644	0.584	0.610	0.604	0.614	0.614
0.50	0.10	11.1	0.000	0.530	0.648	0.453	0.634	0.627	0.632	0.640
0.50	0.20	25.0	0.000	0.594	0.661	0.015	0.650	0.646	0.652	0.655
					$\xi = 2.5$					
0.30	0.01	1.01	0.795	0.982	0.888	0.816	0.823	0.819	0.823	0.828
0.30	0.05	5.26	0.000	0.980	0.977	0.950	0.965	0.960	0.967	0.967
0.30	0.10	11.1	0.000	0.983	0.984	0.918	0.979	0.975	0.978	0.983
0.30	0.20	25.0	0.000	0.988	0.987	0.512	0.986	0.986	0.986	0.987
0.50	0.01	1.01	0.000	0.952	0.897	0.814	0.818	0.815	0.818	0.821
0.50	0.05	5.26	0.000	0.964	0.974	0.937	0.959	0.955	0.961	0.963
0.50	0.10	11.1	0.000	0.979	0.982	0.886	0.981	0.976	0.980	0.981
0.50	0.20	25.0	0.000	0.984	0.987	0.273	0.986	0.983	0.987	0.986

[*] Notes: See notes to Table 6.

38