UvA ECONOMETRICS

Discussion Paper: 2014/02

OLS and IV estimation of regression models including endogenous interaction terms

Maurice J.G. Bun and Teresa D. Harrison

www.ase.uva.nl/uva-econometrics

Amsterdam School of Economics
Department of Economics & Econometrics
Valckenierstraat 65-67
1018 XE AMSTERDAM
The Netherlands

UvA UNIVERSITEIT VAN AMSTERDAM

OLS and IV estimation of regression models including endogenous interaction terms*

Maurice J.G. Bun †
Department of Economics and Econometrics
Faculty of Economics and Business
University of Amsterdam

Teresa D. Harrison ‡
Department of Economics
LeBow College of Business
Drexel University

this version: 28 January, 2014

JEL-code: C01; C31; C36

Keywords: endogeneity, instrumental variables, interaction term, ordinary least squares

Abstract

We analyze a class of linear regression models including interactions of endogenous regressors and exogenous covariates. We show that, under typical conditions regarding higher-order dependencies between endogenous and exogenous regressors, the OLS estimator of the coefficient of the interaction term is consistent and asymptotically normally distributed. Although not a necessary condition, we demonstrate that multivariate symmetrically distributed data are sufficient for OLS consistency. In general, we propose a Wald test to test for the validity of these higher-order moments. Applying heteroskedasticity-consistent covariance matrix estimators, we then show that standard inference based on OLS is valid for the coefficient of the interaction term. Furthermore, we analyze several IV estimators, and conclude that an implementation exploiting instruments interacted with the exogenous part of the interaction term is to be preferred. Using our theoretical results we confirm recent empirical findings on the nonlinear causal relation between financial development and economic growth.

---

*We are especially grateful to Peter Boswijk for his many useful suggestions. Furthermore, we like to thank Simon Broda, Arturas Juodis, Jan Kiviet, Frank Kleibergen and Andrew Pua for helpful comments.

†Amsterdam, Netherlands, M.J.G.Bun@uva.nl. The research of the author has been funded by the NWO Vernieuwingsimpuls research grant nr. 016.115.320.

‡Philadelphia, PA 19104, tharrison@drexel.edu.

1. Introduction

In applied research, it is common to use interaction terms to investigate the multiplicative effect of two variables, labeled x and w, on a dependent variable y. In this study we analyze the case where x is endogenous and w is exogenous. For example, analyzing the returns to schooling one generally regresses wages on education, gender, and other covariates (i.e., ethnicity, age, marital status, etc.). A researcher might interact education and gender in the regression to investigate the gender gap in returns to schooling (see e.g. Dougherty, 2005). At the same time one may want to correct for endogeneity of education due to e.g. selection bias or measurement error. In this case it is expected that the interaction variable, i.e. the product of education and gender, is also an endogenous regressor. Two other examples of empirical studies, in which interactions of endogenous and exogenous regressors appear, are Rajan and Zingales (1998) and Aghion et al. (2005). Both studies analyze the relation between financial development and economic growth, allowing the impact of financial development on growth to be nonlinear.

Of primary interest in regression models including interaction terms are the coefficients of those interactions. More specifically, one would like to verify whether interaction terms are significant and economically important and thus should be included in the empirical model. Furthermore, it is well known that in interaction models the magnitude and significance of lower-order coefficients are not invariant to straightforward scaling of the data (Braumoeller, 2004; Balli and Sørensen, 2013). Hence, statistical inference on these coefficients is only meaningful in combination with their corresponding interaction term.

In the presence of endogenous regressors it is expected that ordinary least squares (OLS) is inconsistent and that instrumental variables (IV) estimation is required instead. Although for linear models the properties of OLS and IV estimators and corresponding inference have received much attention the last decades, little research exists on OLS and IV estimation in models with (partly) endogenous interactions. An exemption is Nizalova and Murtazashvili (2013), who analyze consistency of the OLS coefficient estimator in an interaction model subject to a specification error (omitted relevant variables). They do not consider OLS (and IV) inference, however. Our study aims at providing a comprehensive analysis of the relative merits of OLS versus IV based inference.

We analyze to what extent one still can rely on OLS estimation and inference. We show that, under a particular condition regarding higher-order dependencies in the data, the OLS estimator for the coefficient of the interaction term is consistent and asymptotically normally distributed. These results are derived under quite general conditions, allowing for continuous and discrete interaction terms, correlation between endogenous and exogenous regressors, conditional heteroskedasticity and non-normality. Thus, the researcher can perform valid statistical inference for the interaction term using OLS prior to implementing IV estimation. Such knowledge can avoid unnecessary and more complex IV estimation, particularly in cases where validity of instrumental variables is questionable.

OLS estimation of the other coefficients, however, is not consistent. Hence, IV estimation is necessary if one is interested in the full marginal effect of the endogenous regressor on the dependent variable. We then provide guidance on the optimal set of instruments given the presence of an endogenous regressor. Given a vector of valid instrumental variables z, we show that the instrument set should include interactions of the elements in z with w in order to satisfy the necessary rank condition for IV estimation. Although *a priori* this seems the most natural set of instruments (e.g. Wooldridge, 2002, p121-122), we are unaware of any prior study that documents the underidentification by exploiting z alone. This result explains puzzling estimates found in earlier applied research. For example, Aghion et al. (2005) note that "using z without $z \cdot w$ resulted in too much collinearity between the fitted values of x and $x \cdot w$ to identify the crucial coefficients" (p193, adapted to our notation).

We demonstrate our theoretical results both through Monte Carlo experiments and by an empirical analysis. The Monte Carlo experiments show the favorable statistical properties of OLS inference on the interaction term and also bimodality of IV estimation when only the linear instruments are employed. We also partly reproduce and extend the empirical analysis of Aghion et al. (2005), who analyze the relation between financial development and convergence. In a cross-sectional growth regression they test the significance of an interaction effect between initial income and financial development. They allow financial development to be an endogenous regressor. We provide further support for their instrument set choice, and additionally report valid OLS based inference. We show that for the parameter of interest, i.e. the interaction coefficient, OLS inference is at least as credible as IV inference. Our empirical results reinforce their conclusion that low financial development makes growth convergence less likely.

In the next section, we describe the interaction model and investigate the asymptotic properties of the OLS and IV estimators. In Section 3 we report the Monte Carlo simulations, while Section 4 contains the growth application. Section 5 concludes.

2. Model and Asymptotic Properties

2.1. Basic Set-up

We consider the following model with only one endogenous regressor (labeled x) and one additional exogenous regressor (labeled w):[1]

$$y_i = \beta_\iota + \beta_w w_i + \beta_x x_i + \beta_{xw} x_i w_i + u_i, \quad i = 1, ..., n. \tag{2.1}$$

[1] The presence of additional exogenous regressors in (2.1) does not change the theoretical results. The analysis below holds exactly when we replace y, w and x by the residuals of their projection on these additional exogenous regressors.

The endogenous regressor x interacts with an exogenous variable w. One relevant application could be where y is wage, x is schooling and w is gender. In our application in Section 4, the variables y, x, and w represent country specific growth rates, a measure for the financial development of the country, and log of initial GDP per capita respectively. The parameter of interest is β_{xw}, i.e. we want to test whether the returns to education is homogeneous or depends on gender or whether the growth effect of financial development depends on the initial GDP of the country. Stacking the observations we get

$$y = X\beta + u, \tag{2.2}$$

where $y = (y_1, ..., y_n)'$ and $u = (u_1, ..., u_n)'$. Furthermore, $X = (X_1', ..., X_n')'$ with $X_i = \begin{bmatrix} 1 & w_i & x_i & x_i w_i \end{bmatrix}'$ and $\beta = (\beta_\iota, \beta_w, \beta_x, \beta_{xw})'$.

To establish the sampling properties of OLS and IV estimators, we make the following assumption regarding the data and errors:

Assumption 1. *The data (y_i, x_i, w_i, z_i) are i.i.d. across i with nonzero finite fourth moments and $E(u_i|w_i, z_i) = 0$. Furthermore, we assume $E(x_i) = 0$, $E(w_i) = 0$ and $E(z_i) = 0$.*

Although this simple random sampling assumption rules out most time series applications, it is general enough to allow for conditional heteroskedasticity and non-normality. The assumption of zero means for all regressors is without loss of generality. Because a constant is always included, all theoretical results below will continue to hold with rescaling of the regressors (Kiviet and Niemczyk, 2012). Also note that we do not specify a particular functional form or relation between the regressors x and w. Hence, x and w can be collinear as is usually the case in applied work.

The structure in (2.1) and Assumption 1 results in classic endogeneity bias if

$$cov(x_i, u_i) = \sigma_{xu} \neq 0. \tag{2.3}$$

In the linear model such endogeneity affects consistent OLS estimation of all regression coefficients. In the next section we will investigate to what extent this is also the case in the interaction model (2.1).

2.2. OLS estimation and inference

In this section we show that, under reasonable conditions related to higher-order moments that includes but is not limited to symmetric distributions, the OLS estimator of the coefficient β_{xw} of the interaction term is consistent and asymptotically normally distributed. Additionally, we show that standard heteroskedasticity-robust OLS inference can be applied to test the significance of the interaction effect and for constructing a confidence interval.

The OLS estimator of the full parameter vector β is equal to:

$$\hat{\beta} = (X'X)^{-1}X'y. \tag{2.4}$$

Taking the probability limit we have

$$\begin{aligned} \operatorname{plim}\hat{\beta} &= \beta + \left(\operatorname{plim}\frac{1}{n}X'X\right)^{-1}\operatorname{plim}\frac{1}{n}X'u \\ &= \beta + \Sigma_{XX}^{-1}\Sigma_{Xu}, \end{aligned} \tag{2.5}$$

where Assumption 1 implies that $\Sigma_{XX} = E\left[X_i X_i'\right]$ and $\Sigma_{Xu} = E\left[X_i u_i\right]$. The vector $\Sigma_{XX}^{-1}\Sigma_{Xu}$ is the OLS inconsistency. For the interaction model (2.1) the following result holds:

Proposition 1: *Under Assumption 1, the inconsistency of the OLS estimator of model (2.1) equals:*

$$\Sigma_{XX}^{-1}\Sigma_{Xu} = \frac{\sigma_{xu}}{\det\left(\Sigma_{XX}\right)}\begin{bmatrix} E(x_i w_i)\left(E(x_i w_i)E(x_i w_i^2) - E(w_i^2)E(x_i^2 w_i)\right) \\ E\left(x_i^2 w_i\right)E\left(x_i w_i^2\right) + \left(E\left(x_i w_i\right)\right)^3 - E\left(x_i w_i\right)E\left(x_i^2 w_i^2\right) \\ E(w_i^2)\left(E\left(x_i^2 w_i^2\right) - E\left(x_i w_i^2\right)\right) - \left(E\left(x_i w_i^2\right)\right)^2 \\ E(x_i w_i)E(x_i w_i^2) - E(w_i^2)E(x_i^2 w_i) \end{bmatrix}. \tag{2.6}$$

Proof. see the Appendix. ■

The last element in the inconsistency (2.6) is interesting because quite often it will be zero. For example, when x_i and w_i are bivariate normal or t distributed we have that $E(x_i^2 w_i) = E(x_i w_i^2) = 0$. More generally, multivariate symmetric disributions are sufficient, but not necessary, for these higher-order dependencies to vanish.

To derive the limiting distribution of the OLS estimator, rewrite model (2.1) as:

$$y_i = X_i'\beta_* + \varepsilon_i, \tag{2.7}$$

with $\beta_* = \beta + \Sigma_{XX}^{-1}\Sigma_{Xu}$ the pseudo-true value and

$$\varepsilon_i = u_i - \Sigma_{Xu}'\Sigma_{XX}^{-1}X_i, \tag{2.8}$$

such that $E\left[X_i\varepsilon_i\right] = 0$. Note that ε_i depends on X_i and, hence, in general will be heteroskedastic.

The OLS estimator (2.4) is simply a method of moments estimator exploiting the following moment equation:

$$E\left[f_i\left(\beta\right)\right] = E\left[X_i\left(y_i - X_i'\beta\right)\right] = 0. \tag{2.9}$$

These moment conditions are satisfied in $\beta = \beta_*$. Standard asymptotic theory for method of moments estimators then gives the following large sample distribution of the OLS estimator:

Lemma 1. *Given model (2.1) and Assumption 1, the large sample distribution of the OLS estimator (2.4) is:*

$$\sqrt{n}\left(\hat{\beta}-\beta_*\right) \xrightarrow{d} \mathcal{N}(0,V), \tag{2.10}$$

where

$$\begin{aligned} V &= A_*^{-1}B_*A_*'^{-1}, && (2.11)\\ A_* &= \operatorname{plim}\frac{1}{n}\sum_{i=1}^{n}\left.\frac{\partial f_i(\beta)}{\partial\beta'}\right|_{\beta_*} = -\Sigma_{XX}, && (2.12)\\ B_* &= \operatorname{plim}\frac{1}{n}\sum_{i=1}^{n} f_i(\beta_*)f_i(\beta_*)' = E\left[\varepsilon_i^2 X_i X_i'\right]. && (2.13)\end{aligned}$$

From Lemma 1 it can be seen that, although normally distributed, the limiting distribution of the OLS estimator is centered around its pseudo-true value $\beta_* = \beta + \Sigma_{XX}^{-1}\Sigma_{Xu}$, but with a standard sandwich-type expression for the asymptotic variance.[2]

If the last element in the vector $\Sigma_{XX}^{-1}\Sigma_{Xu}$ is zero, Lemma 1 and Proposition 1 imply the following interesting result for the OLS estimator of the interaction coefficient β_{xw}:

$$\sqrt{n}\left(\hat{\beta}_{xw}-\beta_{xw}\right) \xrightarrow{d} \mathcal{N}\left(0,V_{xw}\right), \tag{2.14}$$

where V_{xw} is the diagonal element of V corresponding to the interaction term. In other words, we have the remarkable fact that, even if we have an endogenous regressor x, the OLS estimator of the coefficient β_{xw} is consistent. It should be noted, however, that this consistency is restricted to β_{xw} only and not the full marginal effect of x on y $(\beta_x + \beta_{xw}w)$ because the OLS estimator of β_x is inconsistent.

It is relative easy to test the condition for consistency of the OLS estimator of β_{xw} by checking the condition:

$$E(x_iw_i)E(x_iw_i^2) - E(w_i^2)E(x_i^2w_i) = 0. \tag{2.15}$$

Defining $\theta = \left[\begin{array}{cccc} E(x_iw_i) & E(x_iw_i^2) & E(w_i^2) & E(x_i^2w_i)\end{array}\right]'$ we want to verify $H_0: \theta_1\theta_2 - \theta_3\theta_4 = 0$. The four elements in θ can be estimated by their sample counterparts resulting in the following sample moment equations:

$$\frac{1}{n}\sum_{i=1}^{n} m(w_i,x_i,\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\left(\left[\begin{array}{c} x_iw_i\\ x_iw_i^2\\ w_i^2\\ x_i^2w_i\end{array}\right] - \left[\begin{array}{c}\hat{\theta}_1\\ \hat{\theta}_2\\ \hat{\theta}_3\\ \hat{\theta}_4\end{array}\right]\right) = 0. \tag{2.16}$$

Standard asymptotic theory for method of moments estimators shows that the limiting distribution of $\hat{\theta}$ is equal to:

$$\sqrt{n}\left(\hat{\theta}-\theta\right) \xrightarrow{d} \mathcal{N}\left(0,C_0\right), \tag{2.17}$$

[2] Lemma 1 can also be interpreted as a multivariate extension of Lemma 3.1 in Kiviet (2013).

where C_0 can be estimated consistently by

$$\hat{C} = \frac{1}{n}\sum_{i=1}^{n} m_i(\hat{\theta})m_i(\hat{\theta})'. \tag{2.18}$$

Defining now $h(\theta) = \theta_1\theta_2 - \theta_3\theta_4$, we have $\frac{\partial h(\theta)}{\partial \theta'} = \begin{bmatrix} \theta_2 & \theta_1 & -\theta_4 & -\theta_3 \end{bmatrix}$ and we can test $H_0 : h(\theta) = 0$ with the following Wald t-test statistic:[3]

$$W_c = \frac{h(\hat{\theta})}{\sqrt{n^{-1}\hat{r}\hat{C}\hat{r}'}}, \tag{2.19}$$

where $\hat{r} = \left.\frac{\partial h(\theta)}{\partial \theta'}\right|_{\hat{\theta}}$. Under H_0 the statistic in (2.19) is asymptotically standard normal distributed. We will investigate in the Monte Carlo study below how accurate this moments based test is in finite samples.

It should be noted that asymptotic normality of the OLS estimator holds under relatively weak conditions. In Assumption 1 we only specified finite fourth moments, but they do not need to coincide with those of the normal distribution. In other words, non-normality and also conditional heteroskedasticity are allowed for. However, we get interesting simplifications when we further impose normality and homoskedasticity as in Kiviet and Niemczyk (2012):

Assumption 2. *In model (2.1), u_i, w_i and x_i are jointly normally distributed with mean zero and variance matrix* $\begin{pmatrix} \sigma_u^2 & 0 & \sigma_{xu} \\ 0 & \sigma_w^2 & \sigma_{xw} \\ \sigma_{xu} & \sigma_{xw} & \sigma_x^2 \end{pmatrix}$ *for all $i = 1, ..., n$.*

Kiviet and Niemczyk (2012) analyze the asymptotic distribution of the OLS estimator for linear models with endogenous regressors. A difference with Kiviet and Niemczyk (2012) is that, although we assume normality of x_i and w_i, the interaction term $x_i w_i$ is still non-normal. Assuming normality of the data and errors, the following holds:

Proposition 2: *Under Assumptions 1 and 2, the asymptotic distribution of the OLS estimator of β in Lemma 1 simplifies to:*

$$\sqrt{n}\left(\hat{\beta} - \beta_*\right) \xrightarrow{d} \mathcal{N}(0, V), \tag{2.20}$$

with

$$\beta_* = \beta + \frac{\sigma_{xu}}{\sigma_w^2\sigma_x^2 - \sigma_{xw}^2}\begin{pmatrix} 0 & -\sigma_{xw} & \sigma_w^2 & 0 \end{pmatrix}' \tag{2.21}$$

$$V = \sigma_u^2\left(1 - \rho_{xu}^2\right)\Sigma_{XX}^{-1}. \tag{2.22}$$

[3]See e.g. Cameron and Trivedi (2005, p226).

Proof. see the Appendix. ■

The result on the asymptotic variance in (2.22) is similar to equation (32) of Kiviet and Niemczyk (2012). From Proposition 2 it also can be seen that endogeneity actually decreases the asympotic variance of the OLS estimator as $0 < \rho_{xu}^2 < 1$.

As described by Kiviet and Niemczyk (2012), under normality and homoskedasticity standard OLS inference exploiting homoskedasticity-only (ho) standard errors makes sense. The reason for this is that

$$\text{plim}\, s_u^2 = \text{plim}\, \frac{\hat{u}'\hat{u}}{n} = \sigma_u^2 \left(1 - \rho_{xu}^2\right), \tag{2.23}$$

with $\hat{u} = y - X\hat{\beta}$ the OLS residuals. Hence, the variance estimator s_u^2 is a consistent estimator of $\sigma_u^2 (1 - \rho_{xu}^2)$ and estimating V by

$$\hat{V}_{ho} = s_u^2 \left(X'X\right)^{-1}, \tag{2.24}$$

is asymptotically valid. Thus, we have the remarkable result that, under Assumptions 1 and 2, the t-statistic for testing $H_0 : \beta_{xw} = \beta_{xw,0}$ based on the OLS coefficient and variance estimators is approximately standard normal distributed, viz.

$$\frac{\hat{\beta}_{xw} - \beta_{xw,0}}{SE_{ho}(\hat{\beta}_{xw})} \xrightarrow{d} \mathcal{N}(0, 1), \tag{2.25}$$

with $SE_{ho}(\hat{\beta}_{xw})$ the square root of the last diagonal element of the standard OLS variance estimator $s_u^2 \left(X'X\right)^{-1}$.

For the more general case of non-normality and/or heteroskedasticity, however, one should base inference on the asymptotic variance provided in Lemma 1 above. The expression of the asymptotic variance is of the usual sandwich form. Therefore, a standard heteroskedasticity-robust (hr) covariance estimator (White, 1980) can be used. Here we exploit:

$$\hat{V}_{hr} = \left(X'X\right)^{-1} \sum_{i=1}^{n} \frac{\hat{\varepsilon}_i^2}{\left(1 - \hat{h}_i\right)^2} X_i X_i' \left(X'X\right)^{-1}, \tag{2.26}$$

with $\hat{h}_i = X_i \left(X'X\right)^{-1} X_i'$ and $\hat{\varepsilon}_i = y_i - X_i'\hat{\beta}$ the OLS residuals. The estimator in (2.26) has been proposed by Davidson and MacKinnon (1993, p554), and is a slight modification of the jackknife estimator in MacKinnon and White (1985).[4] The robust variance estimator in (2.26) will be sufficient to warrant asymptotically valid inference for β_{xw}, i.e.

$$\frac{\hat{\beta}_{xw} - \beta_{xw,0}}{SE_{hr}(\hat{\beta}_{xw})} \xrightarrow{d} \mathcal{N}(0, 1), \tag{2.27}$$

[4] In the literature this is known as the HC3 estimator, see MacKinnon (2012) for an overview of heteroscedasticity-robust covariance matrix estimators.

with $SE_{hr}(\hat{\beta}_{xw})$ the square root of the last diagonal element of the robust variance estimator $\hat{V}_{hr}$. Further improvements in finite sample accuracy can possibly be achieved by exploiting a robust covariance matrix estimator in combination with a (wild) bootstrap procedure (MacKinnon, 2012).

Summarizing, the OLS estimator of the interaction term in (2.1) is, under reasonable conditions, consistent and asymptotically normal, and standard inference can be applied. An important implication of these results is that one can conduct OLS inference first without having to resort to instrumental variables techniques. In other words, OLS inference can be performed very early in the empirical investigation, without worry about strength and exogeneity of instruments or even before suitable instruments have been found. Moreover, in some applications, if the main empirical result only depends on the interaction variable, there is no need to move to IV estimation at all. We will provide one of such an example in Section 4 below, providing empirical evidence (from our Wald-based test) that OLS is sufficient for valid inference on the relation between financial development and growth.

2.3. IV estimation and the optimal set of instruments

We showed that OLS can provide asymptotically valid inference of the interaction coefficient β_{xw} even in the presence of endogeneity. However, because of the remaining endogeneity bias in estimating β_x, the main effect of the endogenous regressor x, we still need to instrument for x if the full marginal effect of x on y is of importance. In this section we therefore compare two different implementations of IV assuming that we have k_z instrumental variables z available satisfying Assumption 1. The first IV estimator (labeled IV1) only exploits these k_z excluded instrumental variables z. The second implementation (labeled IV2) also uses the k_z additional instruments $z \cdot w$. This approach parallels Kelejian (1971) and Amemiya (1974), who demonstrate that employing polynomial series as instruments provides consistent IV estimators for nonlinear structural models.

In applied research, IV2 has been preferred over IV1, and here we provide a theoretical background for this choice. A priori one would prefer IV2 over IV1 as $z \cdot w$ are natural candidate IVs for $x \cdot w$ when z are valid instruments for x (Wooldridge, p122, 2002). Also Aghion et al. (2005) dismiss IV1 because of collinearity problems between the two first stage regressions. Exploiting only z as an instrument set, the resulting IV1 standard errors are large, and hence estimates are imprecise and not significant. Below we will provide an explanation for this anomalous behaviour by showing that IV1 is likely to suffer from underidentification.

To analyze the IV estimators, we supplement the structural equation (2.1) with the following reduced form for the endogenous regressor x:

$$x_i = \pi + \pi_w w_i + z_i' \pi_z + v_i, \tag{2.28}$$

where z_i is the k_z dimensional vector of instrumental variables excluded from (2.1). We analyze the relevance of the IV1 and IV2 sets of instruments defined as:

$$\begin{aligned} z_i^{(1)} &= \begin{bmatrix} 1 & w_i & z_i' \end{bmatrix}', \\ z_i^{(2)} &= \begin{bmatrix} 1 & w_i & z_i' & z_i' \cdot w_i \end{bmatrix}'. \end{aligned}$$

The lack of identification of the IV1 estimator can be explained by the fact that, under certain conditions, the rank condition is not satisfied:

Proposition 3: *Under (2.1), (2.28) and Assumption 1, we have that:*

$$\begin{aligned} \text{rank } E\left[z_i^{(1)} X_i'\right] &= 3 \text{ if } E\left[w_i^2 z_i\right] = 0 \text{ and } E\left[w_i z_i z_i'\right] = 0, && (2.29) \\ \text{rank } E\left[z_i^{(2)} X_i'\right] &= 4. \end{aligned}$$

Proof. see the Appendix. ■

Thus, the IV1 estimator is subject to an identification problem when:

$$E\left[w_i^2 z_i\right] = 0 \text{ and } E\left[w_i z_i z_i'\right] = 0. \tag{2.30}$$

Just like for the condition in (2.15), it is relatively easy to test the moments in (2.30) by simple moments based Wald tests. The $k_z + \frac{1}{2}k_z(k_z+1)$ population moments in (2.30) can simply be estimated by their sample counterparts and asymptotically valid Wald t statistics result. It should be noted that these tests are robust to weak identification originating from weak instruments, i.e. π_z in (2.28) close to zero. The reason is simply that the conditions in (2.30) do not depend on π_z. This robustness property certainly enhances the use of such Wald statistics in applied research.

Summarizing, irrespective of the strength of the instruments z, the IV1 estimator may not meet the rank condition for identification and, hence, it does not identify the structural parameters. As we will show in the Monte Carlo experiments, we tend to find bimodality of the IV1 estimator. However, the IV2 estimator does fulfill the rank condition for identification. Under standard, strong instruments asymptotics the IV2 estimator is consistent and asymptotically normal, and conventional IV based inference can be applied. The following proposition gives its asymptotic variance in terms of model parameters when condition (2.30) holds:

Proposition 4: *Under (2.1), (2.28), Assumption 1, condition (2.30) and homoscedastic errors* u_i and v_i, *the asymptotic variances of the IV2 estimator of* β_x *and* β_{xw} *are equal to:*

$$V_{(2),x} = \frac{\sigma_u^2}{E\left[z_i^2\right]\pi_z'\pi_z}, \tag{2.31}$$

$$V_{(2),xw} = \frac{\sigma_u^2}{\pi_w^2 E\left[w_i^3 z_i'\right]\left(E\left[w_i^2 z_i z_i'\right]\right)^{-1} E\left[z_i w_i^3\right] + \pi_z' E\left[w_i^2 z_i z_i'\right]\pi_z}. \tag{2.32}$$

Proof. see the Appendix. ■

Proposition 4 shows that, as long as $\pi_z \neq 0$, the IV2 estimator identifies the structural parameters β_x and β_{xw} (and also β and β_w). Of course, when instruments become weak, i.e. $\pi_z \approx 0$, also the IV2 estimator is subject to the usual weak instruments problem and its variance will become large.[5] Furthermore, with weak instruments the IV2 estimator is biased in the direction of the OLS estimator, and its distribution is non-normal affecting inference (Staiger and Stock, 1997).

3. Monte Carlo experiments

In order to further illustrate the theoretical results from the previous section, we perform a number of Monte Carlo experiments. We simulate finite sample distributions of the OLS, IV1 and IV2 estimators. We calculate for each replication the OLS estimator:

$$\hat{\beta} = (X'X)^{-1} X'y, \tag{3.1}$$

and the IV estimators:

$$\hat{\beta}_{IV}^{(m)} = \left(X'P_{Z_{(m)}}X\right)^{-1} X'P_{Z_{(m)}}y, \tag{3.2}$$

with $Z_{(m)} = \left(z_1^{(m)}, ..., z_n^{(m)}\right)'$ and

$$\begin{aligned} X_i' &= \begin{bmatrix} 1 & w_i & x_i & x_i w_i \end{bmatrix}, \\ z_i^{(1)\prime} &= \begin{bmatrix} 1 & w_i & z_i' \end{bmatrix}, \\ z_i^{(2)\prime} &= \begin{bmatrix} 1 & w_i & z_i' & z_i' \cdot w_i \end{bmatrix}. \end{aligned}$$

Once again, the difference between the two IV estimators is due to the choice of different instrument sets. On top of the instruments of the IV1 estimator, the IV2 estimator also exploits instruments interacted with the exogenous part of the interaction term.

Apart from analyzing bias and variance of these coefficient estimators, we also report actual rejection probabilities of corresponding t-statistics. We report both homoskedasticity-only and heteroskedasticity-robust inference exploiting (2.24) and (2.26) for OLS, and equivalent expressions for both IV estimators (see for details Steinhauer and Wuergler, 2010). Furthermore, we report the actual rejection percentage of the Wald t-statistic (2.19), testing the condition (2.15) underlying consistency of OLS (labeled W_c). Finally, we also calculate Cragg-Donald (1993) statistics for both IV estimators. Cragg-Donald statistics (labeled CD) indicate weak instruments in case of multiple endogenous regressors. And it is

[5] Note that, assuming either independence or joint normality of w_i and z_i, we have $E\left[z_i w_i^3\right] = 0$ and $E\left[w_i^2 z_i z_i'\right] = E\left[w_i^2\right] E\left[z_i^2\right] I_{k_z}$. In this case, the above expressions further simplify to $V_{(2),x} = \frac{\sigma_u^2}{\pi_z' \pi_z \sigma_z^2}$ and $V_{(2),xw} = \frac{1}{\sigma_w^2} V_{(2),x}$, where $E\left[z_i^2\right] = \sigma_z^2$ and $E\left[w_i^2\right] = \sigma_w^2$.

well known that very weak instruments can lead to (1) huge standard errors; (2) bimodality of finite sample distributions.

We generate data for y and x according to (2.1) and (2.28). Given that conditions (2.15) and (2.30) are crucial for OLS and IV1 respectively, we specify a Monte Carlo DGP in which we can vary the dependencies between x, w and z. Furthermore, choosing values and distributions for the various elements in (2.1) and (2.28) we have to keep in mind any invariance properties of our theoretical results. Because we include a constant in our model, without loss of generalization we will always center all variables so that they have mean zero. Additionally, without loss of generalization we can generate exogenous variables (w_i and z_i) such that they are mutually uncorrelated and have unit variance. The reason is that we always can provide a nonsingular transformation of the data, such that OLS and IV inference is not affected.

Furthermore, we choose $\beta_\iota = \beta_w = \beta_x = \beta_{xw} = 1$ in (2.1) always and unconditional error variances are standardized at $\sigma_u^2 = \sigma_v^2 = 1$. Finally, we set $k_z = 5$ and $n = 100$ in all experiments. These choices are not without loss of generalization, but unreported simulation experiments show that these choices are relatively innocuous for the main conclusions. All simulation results are based on 10,000 replications.

3.1. normality and homoskedasticity

In our benchmark design we assume normality of the data and homoscedastic error terms. These are ideal conditions for the OLS estimator of β_{xw}, which is consistent and asymptotically normal and inference using the standard homoskedasticity-only variance estimator can be applied. In this design we choose $w_i \sim$i.i.n.$(0,1)$, $z_i \sim$i.i.n.$(0, I_{k_z})$, $(u_i, v_i) \sim$i.i.n.$(0, \Sigma)$ and

$$\Sigma = \begin{pmatrix} 1 & \rho_{uv} \\ \rho_{uv} & 1 \end{pmatrix}.$$

The parameter ρ_{uv} determines the degree of endogeneity. Regarding the reduced form, without loss of generalization we can set $\pi_w = 1$ because in this design OLS and IV estimators are invariant to this parameter. We fix the concentration parameter

$$\mu = \frac{\pi_z' Z' Z \pi_z}{\sigma_v^2},$$

across experiments. We set $\mu / k_z = \{1, 100\}$; these values are well below and above the rule of thumb of 10 proposed by Staiger and Stock (1997). Hence, a value of 1 corresponds to weak instruments, while 100 indicates strong instruments. Choosing all π_j's equal this implies for each element in the vector of reduced form coefficients that

$$\pi_j = \sqrt{\frac{\sigma_v^2}{n} \frac{\mu}{k_z}}.$$

Simulation results for this first design are in Table 1. We report bias, standard deviation (sd) and actual rejection percentage (rp) of nominal 5% t-tests. Furthermore, we report the actual rejection percentage of the Wald t-statistic testing the consistency condition for OLS, as well as average Cragg-Donald statistics for both IV estimators.

<Table 1 about here>

We can see that while the OLS estimator for β_x remains biased, we observe neglible bias in estimating β_{xw} showing the consistency result of Section 2.2. In fact, the β_{xw} bias of IV1 and IV2 is at least equal to or larger than for OLS. Biases in both IV1 and IV2 are smaller for strong instruments, as expected. Comparing IV1 and IV2, we find that the standard deviation of IV1 is much larger than that of IV2 reflecting the lack of identification discussed previously. The large variance is particularly acute for the IV1 estimator of β_{xw}. Dispersion of IV2 is also larger than for OLS, but the difference in variance is much smaller than that of IV1 versus IV2. IV1 inference is almost always conservative, i.e. rejection frequencies are well below the nominal significance level. IV2 inference is valid in case of strong instruments. OLS coefficient bias carries over to t-tests: actual rejection frequency for testing β_{xw} is close to nominal level, while that for testing β_x can be close to 100%. Difference in finite-sample performance between homoskedasticity-only or heteroskedasticity-robust test statistics is negligible.

Summarizing, OLS inference on β_{xw}, the coefficient of the interaction term, is excellent reflecting the theoretical results of the previous section. Also the Wald t-test for checking the condition for OLS consistency is accurate under the null hypothesis. Furthermore, IV2 clearly outperforms IV1 in the case of strong instruments. Average CD statistics in the weak instruments experiment are of similar magnitude (0.63 and 1.13 for IV1 and IV2 respectively) indicating for both IV implementations an identification problem. Average CD statistics in the strong instruments experiment are of different magnitude (0.81 and 21.65 for IV1 and IV2 respectively) indicating that the identification problem remains for the IV1 estimator. This corresponds to the theoretical results of proposition 3. In this design the rank condition for identification is not satisfied for IV1, irrespective of the strength of the instruments.

<Figure 1 about here>

Figure 1 further illustrates this by plotting empirical densities for the OLS, IV1 and IV2 estimation errors of β_{xw} (left panels) and corresponding t-statistics (right panels) for testing $H_0 : \beta_{xw} = \beta_{xw,0}$. It is clearly seen that the IV1 coefficient estimator is not normally distributed and has large tails consistent with the large standard deviations reported in Table 1. The IV2 and OLS estimators of β_{xw}, however, are very close to a normal distribution, as expected.[6]Corresponding t-statistics should be distributed as standard normal, which density is shown as well for reference in each of the three right panels. Striking is the bimodality of the distribution of the IV1 t-statistic, which is caused by the lack of identification facing this implementation. IV2 and OLS t-statistics, however, are very close to a standard normal distribution, as expected.

3.2. heteroskedasticity

In our second design we continue to assume normality of the data, but allow for conditional heteroskedasticity in the errors. We specify the following scedastic function:

$$\omega_i = \phi \left(\pi + \pi_w w_i + z_i' \pi_z\right)^2 ,$$

where ϕ is a scaling factor to ensure that on average the variance of u_i and v_i are one. This specification implies substantial heteroskedasticity determined by w_i and z_i.[7] Now we generate heteroskedastic errors according to $(u_i, v_i) \sim$i.i.n.$(0, \Sigma_i)$ with

$$\Sigma = \begin{pmatrix} \omega_i & \rho_{uv} \\ \rho_{uv} & \omega_i \end{pmatrix}.$$

In this second design, the OLS coefficient estimator of β_{xw} continues to be consistent, but the standard variance estimator is incorrect and we have to exploit the heteroskedasticity-robust covariance estimator instead. Regarding the IV estimators we also show results for both homoskedasticity-only and heteroskedasticity-robust covariance matrix estimators where obviously the latter is necessary for asymptotically valid inference.

<Table 2 about here>

Table 2 reports simulation results for this heteroskedastic design, again distinguishing between weak and strong instruments cases. Focusing on the estimation of the interaction coefficient β_{xw}, we again observe neglible OLS and IV bias. Comparing IV1 and IV2, we again find that the standard deviation of IV1 is much larger than that of IV2 reflecting lack of identification.

IV2 inference exploiting heteroskedasticity-robust standard errors is valid in the case of strong instruments, while using homoskedasticity-only standard errors creates large size distortions for testing β_x and β_{xw} as can be seen. In the case of weak instruments, however, uncertainty is much larger compared with OLS, and large size distortions may result

[6]A similar figure (not reported here) holds for β_x with the only difference that the OLS estimator and t-statistic for β_x is shifted to the right reflecting the positive coefficient bias.

[7]For simplicity we choose the coefficients in the scedastic function equal to those in the reduced form for x_i.

irrespective of the type of standard errors used. OLS inference on β_{xw} exploiting robust standard errors is valid as the actual rejection frequency for testing β_{xw} is close to nominal level. Regarding the other coefficients, however, OLS t-statistics should not be used.

3.3. non-normality

In our third design we assume homoskedasticity again, but choose data and error distributions different from normality. More specifically, we choose w_i, z_i and (u_i, v_i) from t_4 and χ^2_2 distributions. Compared with the normal distribution, the former has relatively fat tails, while the latter is skewed. To make the comparison with our first design easier, we standardize these t and χ^2 distributions such that the resulting variables have mean zero and unit variance again.

In these cases the OLS coefficient estimator of β_{xw} is consistent, but the standard variance estimator may be incorrect and we have to exploit the heteroskedasticity-robust covariance estimator again. Given our small sample size of $n = 100$ it is interesting to see how accurate asymptotics are for distributions other than normal.

<Table 3 about here>

Table 3 reports simulation results for χ^2 errors, those for t errors are qualitatively similar. Compared with Table 1 it can be seen that non-normality does not affect the simulation results a lot. Coefficient biases and standard deviations of all three estimators are more or less the same, and size distortions for the IV t-tests are similar too. The only notable difference is that exploiting the usual OLS variance estimator for inference leads to a (modestly) size distorted t-test. However, heteroskedasticity-robust inference is able to solve this.

3.4. nonlinear dependence

Finally, in our last design we specify the dependency between w_i and z_i such that IV1 is no longer subject to failure of the rank condition, hence also for this estimator strong identification may result. For that, we change the specification for z_{1i} into:

$$z_{1i} = \frac{1}{\sqrt{2}}\left(w_i^2 - 1\right), \tag{3.3}$$

while maintaining all other choices of our benchmark design. Given $w_i \sim$i.i.n.$(0,1)$, this alternative construction of z_{1i} implies that it has a standardized χ^2_1 distribution, i.e. mean zero and variance one. It also implies that $E(z_{1i}w_i) = 0$ and $E(z_{1i}^2 w_i) = 0$, but

$$E(z_{1i}w_i^2) = \frac{1}{\sqrt{2}}\left(E(w_i^4) - E(w_i^2)\right), \tag{3.4}$$

which is nonzero. Hence, the rank condition failure as described in (2.30) is not present, and IV1 identifies the structural parameters in (2.1).

It should be noted that in this design the condition (2.15) for consistency of OLS is violated, because

$$E(x_i w_i^2) = \frac{\pi_j}{\sqrt{2}} \left(E(w_i^4) - E(w_i^2) \right), \tag{3.5}$$

is non-zero as long as $\pi_j \neq 0$.

<Table 4 about here>

Table 4 reports simulation results for this design. Compared with Table 1 it can be seen that only in the strong instruments case differences can be seen. In the weak instruments experiment the nonlinear dependence between w_i and z_i has much less effect on both IV1 and OLS estimators. For IV1 this is so because in general there is a weak correlation between instruments and endogenous regressor. Regarding OLS in this design the dependence between x_i and w_i^2 is parametrized by π_j, which is small in case of weak instruments.

In the strong instruments case the IV1 estimator performs much better compared with other designs. Although its efficiency is still lower than IV2, it has greatly improved compared with other designs and size distortions are smaller.

The OLS estimator is not much affected, although theoretically it is inconsistent for β_{xw}. The finite sample bias for the OLS estimator of β_{xw} is small in magnitude, and the t-test has a small size distortion only. At the same time the rejection probability of the Wald t-test checking (2.15) is close to its nominal significance level of 5%, indicating no power. Probably the deviation from (2.15) or sample size or both are just too small. We therefore increased sample size to $n = 1000$ and repeated this simulation experiment. The larger sample size does increase the OLS bias in estimating β_{xw} slightly from -0.006 to -0.025. In combination with the increased precision, this results in a significance t-test rejection frequency of around 22%. At the same time, power of the Wald t-test checking (2.15) has increased to 27%.

4. Economic growth and financial development

Aghion, Howitt and Mayer-Foulkes (2005), hereafter AHM, develop a theory implying that economic growth convergence depends on the level of financial development. They test their theory in a cross-country growth regression including an interaction term between initial GDP per capita and an indicator of financial development. In our notation y_i is the average growth rate of GDP per capita in the period 1960-1995, w_i is initial (1960) per capita GDP and x_i is the average level of financial development. Sample size is $n = 71$ countries. Their specifications include different sets of control variables (labeled "empty", "policy" and "full").[8] The data are taken from Levine, Loayza and Beck (2000) and include four different measures of financial development ("private credit", "liquid liabilities", "bank assets" and "commercial-central bank").

AHM conjecture that financial development is an endogenous regressor because of feedback from growth to finance, or because of relevant omitted variables. They acknowledge that the interaction between financial development and initial income may be an endogenous regressor too. They follow La Porta et al. (1997, 1998) and use legal origin as source of exogenous variation in financial development to construct instrumental variables. Legal origin is a categorical variable with 4 categories, i.e. French, English, German and Scandinavian traditions. La Porta et al. (1997, 1998) construct three binary indicators (omitting Scandinavion), which serve as instrumental variables for financial development.

Table 5 reports empirical results using OLS, IV1 and IV2 estimators of (2.1) using private credit as measure of financial development.[9] The specification does not include any further control variables, but AHM (p193) consider this specification to be representative of their main result. Note that AHM only use IV2, and the reported IV2 estimates in Table 5 of the interaction model indeed correspond exactly to AHM's results in column 1 of their Table 1. Also note that AHM used homoskedasticity-only standard errors, and we supplement them with heteroskedasticity-robust ones.[10] Also we report for comparison estimation results for the linear model, i.e. omitting the interaction regressor.

<Table 5 about here>

The left panel of Table 5 shows estimation results for the interaction model. Striking is the similarity of the OLS and IV2 results, and the huge standard errors for IV1. Both empirical facts corroborate the simulation results of the previous section. We report in Table 5 also IV1 and OLS estimates. Because AHM treat financial development as an endogenous regressor, they do not consider OLS estimation. Our theoretical results, however, show that when condition (2.15) is satisfied OLS inference is valid for the coefficient of the interaction term. Applying the Wald statistic in (2.19) we indeed do not reject condition (2.15) for this application (p-value is 0.09).

Furthermore, AHM do not report IV1 estimates because this resulted in too much collinearity in the second stage regression (between the fitted values of x and $x \cdot w$) to identify the parameters β_x and β_{xw}. In other words, exploiting only legal origin as an instrument set without interactions with initial GDP per capita, the resulting IV1 standard errors are large, and hence estimates are imprecise and not significant. This can be seen

[8] The policy control variables are average years of schooling, government size, inflation, black market premium and trade openness. The full conditioning set is the policy set plus indicators for revolution and coups, political assassinations and ethnic diversity.

[9] Private credit is AHM's preferred measure of financial development.

[10] Tests for heteroskedasticity do not reject the null hypothesis of homoskedasticity.

from Table 5 when an interaction term is included in the model. Additionally, the Cragg-Donald statistic (0.072) indicates very weak instruments in this case corroborating with our theoretical results. Furthermore, the Anderson (1951) canonical correlations LM statistic, Kleibergen-Paap (2006) LM statistic and the Wald statistic testing the conditions in (2.30) all do not reject the null hypothesis of underidentification (p-values are 0.89, 0.87, and 0.64 respectively).

Weak instruments seems less of a problem in the case of IV2, although the CD statistic is not very large either (6.281). In general instruments are much stronger for the linear model, but the IV2 Sargan test for the linear model indicates invalid instruments. A reason for this could be of course that we incorrectly imposed linearity further reassuring the necessity of including the interaction term.

<Table 6 about here>

In Table 6 we report a summary of further empirical results for the interaction model. We estimated all twelve specifications[11] from Table 1 of AHM by OLS, IV1 and IV2, and report a summary of the estimation results for the interaction coefficient β_{xw}. Interesting is that a positive association can be seen between the magnitude of the IV2-CD statistic and the absolute difference between OLS and IV2 estimates (sample correlation coefficient is 0.83). For some specifications we observe that IV2 is subject to weak identification problems, and in those cases the IV2 estimate lies further away from the OLS estimate. This pattern corroborates our theoretical findings, which imply consistency for both OLS and IV2 in case of strong instruments. In case of weak instruments, however, the IV2 estimator is biased, while OLS is still consistent.

4.1. Further Monte Carlo results

To verify the validity of our theoretical results for the current application we conducted further Monte Carlo experiments, in which we choose data and parameter values as close as possible to our empirical setting. The data generating process is again (2.1) and (2.28), but this time we take for β the true values of the coefficients the OLS estimates of the interaction model in Table 5. We also use OLS estimates of the reduced form as true values for π, i.e. we specify:

$$y_i = 2.204 + 1.300w_i - 0.012x_i - 0.048x_iw_i + u_i, \tag{4.1}$$

$$x_i = 69.00 + 17.90w_i + \begin{pmatrix} -4.32 & -11.03 & 40.35 \end{pmatrix}' z_i + v_i. \tag{4.2}$$

Furthermore, we assume that (u_i, v_i) i.i.n. $(0, \Sigma)$ with

$$\Sigma = \begin{bmatrix} \hat{\sigma}_u^2 & \hat{\sigma}_{uv} \\ \hat{\sigma}_{uv} & \hat{\sigma}_v^2 \end{bmatrix} = \begin{bmatrix} 1.57 & -2.29 \\ -2.29 & 322.12 \end{bmatrix},$$

[11] Four financial development measures are distinguished as well as three different sets of control variables.

and where again the estimates are based on OLS. The estimate of the degree of endogeneity is $\hat{\rho}_{uv} = -0.10$ implying a moderate degree of endogeneity only. Finally, we take for the exogenous variables (w_i and z_i) the empirical values, which implies that they are kept fixed across replications.

<Table 7 about here>

Because we assume homoskedasticity and normality, we use the OLS coefficient estimator with the conventional variance estimator, and implement IV1 and IV2 with homoskedasticity-only standard errors. A summary of the simulation results are in Table 7, which are confirming earlier conclusions. The only difference is that apparently the endogeneity problem is less severe because OLS coefficient and test bias is small for all coefficients.

<Figure 2 about here>

Figure 2 shows empirical density functions for t statistics testing $H_0 : \beta_x = \beta_{x0}$ (left panels) and $H_0 : \beta_{xw} = \beta_{xw0}$ (right panels) with β_{x0} and β_{xw0} the true values. Striking is again the bimodality of IV1, while the null distributions of OLS and IV2 t statistics seem very close to their asymptotic $N(0, 1)$ distribution, as expected.

5. Concluding remarks

In this study we show that endogeneity bias is reduced for the OLS estimator when the endogenous regressor is interacted with an exogenous covariate. Under typical conditions regarding higher-order dependencies in the data we show that the OLS estimator of the coefficient of the interaction term is consistent, and that standard OLS inference applies. Although OLS estimators of other regression coefficients are still inconsistent, this result implies that for testing on the presence of endogenous interactions one does not necessarily have to resort to IV techniques. Given the difficulties associated with finding valid instruments, our results suggest that the researcher may produce more reliable estimates with little worry about endogeneity bias if the economic variable of interest is the interaction term.

We furthermore show that, for identification of the full marginal effect of the endogenous regressor by IV techniques, it is necessary to include in the instrument set interactions of instrumental variables with the exogenous part of the interaction term. Due to the nonlinearity of the model, exploiting linear instruments only will lead to underidentification irrespective of the strength of these instrumental variables. Furthermore, including interacted instruments does lead to identification and is the preferred IV implementation, but has the usual caveats in case of weak instruments.

Monte Carlo experiments corroborate our theoretical findings. We further partly reproduce and extend the empirical analysis of Aghion et al. (2005), who analyze the interaction between financial development and growth convergence. We provide support for their instrument set choice, but at the same time show that identification with the current set of instrumental variables is not always strong. Additionally, we are able to produce credible OLS inference for testing their hypothesis that financial development matters for convergence. Our supplementary empirical results reinforce their conclusion that low financial development makes growth convergence less likely.

Appendix

Proof of Proposition 1. The components of the OLS estimation error are:

$$X_i u_i = \begin{pmatrix} u_i \\ w_i u_i \\ x_i u_i \\ x_i w_i u_i \end{pmatrix}, \; X_i X_i' = \begin{pmatrix} 1 & w_i & x_i & x_i w_i \\ w_i & w_i^2 & w_i x_i & x_i w_i^2 \\ x_i & x_i w_i & x_i^2 & x_i^2 w_i \\ x_i w_i & x_i w_i^2 & x_i^2 w_i & x_i^2 w_i^2 \end{pmatrix}.$$

Because we can assume, without loss of generality, that $E[w_i] = 0$ and $E[x_i] = 0$, we have

$$\Sigma_{XX} = \begin{pmatrix} 1 & 0 & 0 & E[x_i w_i] \\ 0 & E[w_i^2] & E[x_i w_i] & E[x_i w_i^2] \\ 0 & E[x_i w_i] & E[x_i^2] & E[x_i^2 w_i] \\ E[x_i w_i] & E[x_i w_i^2] & E[x_i^2 w_i] & E[x_i^2 w_i^2] \end{pmatrix}, \; \Sigma_{Xu} = \begin{bmatrix} 0 \\ 0 \\ \sigma_{xu} \\ 0 \end{bmatrix}.$$

Furthermore, we have that

$$\Sigma_{XX}^{-1} = \frac{1}{\det(\Sigma_{XX})} adj(A),$$

where the transpose of $adj(A)$ is the matrix of cofactors of Σ_{XX}. Because Σ_{Xu} has only the third element nonzero, for the evaluation of $\Sigma_{XX}^{-1}\Sigma_{Xu}$ we only need cofactors corresponding to the third column of Σ_{XX}. Denoting with c_{ij} the cofactor of entry d_{ij} in matrix Σ_{XX}, we have:

$$\begin{aligned} c_{13} &= \det \begin{pmatrix} 0 & E[w_i^2] & E[x_i w_i^2] \\ 0 & E[x_i w_i] & E[x_i^2 w_i] \\ E[x_i w_i] & E[x_i w_i^2] & E[x_i^2 w_i^2] \end{pmatrix} \\ &= E(x_i w_i)\left(E(w_i^2)E(x_i^2 w_i) - E(x_i w_i)E(x_i w_i^2)\right), \end{aligned}$$

$$\begin{aligned} c_{23} &= -\det \begin{pmatrix} 1 & 0 & E[x_i w_i] \\ 0 & E[x_i w_i] & E[x_i^2 w_i] \\ E[x_i w_i] & E[x_i w_i^2] & E[x_i^2 w_i^2] \end{pmatrix} \\ &= E\left(x_i^2 w_i\right) E\left(x_i w_i^2\right) + \left(E\left(x_i w_i\right)\right)^3 - E\left(x_i w_i\right) E\left(x_i^2 w_i^2\right), \end{aligned}$$

$$\begin{aligned} c_{33} &= \det \begin{pmatrix} 1 & 0 & E[x_i w_i] \\ 0 & E[w_i^2] & E[x_i w_i^2] \\ E[x_i w_i] & E[x_i w_i^2] & E[x_i^2 w_i^2] \end{pmatrix} \\ &= E(w_i^2)\left(E\left(x_i^2 w_i^2\right) - E\left(x_i w_i^2\right)\right) - \left(E\left(x_i w_i^2\right)\right)^2, \end{aligned}$$

$$\begin{aligned} c_{43} &= -\det \begin{pmatrix} 1 & 0 & E[x_i w_i] \\ 0 & E[w_i^2] & E[x_i w_i^2] \\ 0 & E[x_i w_i] & E[x_i^2 w_i] \end{pmatrix} \\ &= E(w_i^2)E(x_i^2 w_i) - E(x_i w_i)E(x_i w_i^2), \end{aligned}$$

and the inconsistency is equal to

$$\Sigma_{XX}^{-1}\Sigma_{Xu} = \frac{\sigma_{xu}}{\det(\Sigma_{XX})} \begin{bmatrix} c_{13} \\ c_{23} \\ c_{33} \\ c_{43} \end{bmatrix},$$

which is equal to expression (2.6).

Proof of Proposition 2. Assumption 2 implies that Σ_{XX} further simplifies to:

$$\Sigma_{XX} = \begin{pmatrix} 1 & 0 & 0 & E[x_i w_i] \\ 0 & E[w_i^2] & E[x_i w_i] & 0 \\ 0 & E[x_i w_i] & E[x_i^2] & 0 \\ E[x_i w_i] & 0 & 0 & E[x_i^2 w_i^2] \end{pmatrix},$$

which results in:

$$\Sigma_{XX}^{-1}\Sigma_{Xu} = \frac{\sigma_{xu}}{\sigma_w^2\sigma_x^2 - \sigma_{xw}^2} \begin{pmatrix} 0 & -\sigma_{xw} & \sigma_w^2 & 0 \end{pmatrix}'.$$

We also have

$$\begin{aligned} E[u_i|w_i, x_i] &= \begin{pmatrix} 0 & \sigma_{xu} \end{pmatrix} \begin{pmatrix} \sigma_w^2 & \sigma_{xw} \\ \sigma_{xw} & \sigma_x^2 \end{pmatrix}^{-1} \begin{pmatrix} w_i \\ x_i \end{pmatrix} \\ &= \frac{\sigma_{xu}}{\sigma_w^2\sigma_x^2 - \sigma_{xw}^2}\left(\sigma_w^2 x_i - \sigma_{xw} w_i\right), \end{aligned}$$

$$\begin{aligned} V[u_i|w_i, x_i] &= \sigma_u^2 - \begin{pmatrix} 0 & \sigma_{xu} \end{pmatrix} \begin{pmatrix} \sigma_w^2 & \sigma_{xw} \\ \sigma_{xw} & \sigma_x^2 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \sigma_{xu} \end{pmatrix} \\ &= \sigma_u^2 - \frac{\sigma_w^2\sigma_{xu}^2}{\sigma_w^2\sigma_x^2 - \sigma_{xw}^2}. \end{aligned}$$

Furthermore, it is clear that $E[u_i|w_i, x_i] = E[u_i|X_i]$ and $V[u_i|w_i, x_i] = V[u_i|X_i]$ where $X_i = \begin{pmatrix} 1 & w_i & x_i & x_i w_i \end{pmatrix}'$ and it can be shown that:

$$E[u_i|X_i] = \Sigma_{Xu}'\Sigma_{XX}^{-1}X_i,$$

$$\begin{aligned} V[u_i|X_i] &= \sigma_u^2 - \Sigma_{Xu}'\Sigma_{XX}^{-1}\Sigma_{Xu} \\ &= \sigma_u^2\left(1 - \rho_{xu}^2\right), \end{aligned}$$

where $\rho^2_{xu} = \frac{\sigma^2_{xu}}{\sigma^2_x \sigma^2_u}$. Regarding

$$\varepsilon_i = u_i - \Sigma'_{Xu}\Sigma^{-1}_{XX}X_i,$$

we now have that

$$\varepsilon_i | X_i \sim i.i.n. \left(0, \sigma^2_\varepsilon\right), \qquad \sigma^2_\varepsilon = \sigma^2_u \left(1 - \rho^2_{xu}\right).$$

This implies that in the model

$$y = X\beta + u = X\beta_* + \varepsilon,$$

the errors obey the classical OLS assumptions, hence we have for the OLS estimator:

$$\sqrt{n}\left(\hat{\beta} - \beta_*\right) \xrightarrow{d} \mathcal{N}(0, V),$$

with

$$\begin{aligned} V &= \sigma^2_\varepsilon \left(\operatorname{plim} \frac{1}{n} X'X\right)^{-1} \\ &= \sigma^2_u \left(1 - \rho^2_{xu}\right) \Sigma^{-1}_{XX}. \end{aligned}$$

Proof of Proposition 3. Note that without loss of generalization we can assume that (1) only one of the excluded instruments z_i has a reduced form coefficient unequal to zero; (2) the excluded instruments z_i have a scalar covariance matrix; (3) the elements of w_i and z_i are uncorrelated. The reason for (1) and (2) is that any nonsingular transformation of the excluded instruments results in the same IV estimator. Hence, we can take a transformation such that only the first instrument has a reduced form coefficient unequal to zero and that the covariance matrix of the excluded instruments is scalar. The reason for (3) is that we can orthogonalize included (w_i) and excluded (z_i) IVs without affecting again the IV estimator. Summarizing, without loss of generalization we can asume:

$$E\left[z^{(1)}_i z^{(1)\prime}_i\right] = \begin{pmatrix} 1 & 0 & 0 \\ 0 & E\left[w^2_i\right] & 0 \\ 0 & 0 & E\left[z^2_i\right] I_{k_z} \end{pmatrix}, \; \pi_z = \begin{pmatrix} \pi_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

with $E\left[z^2_i\right]$ a scalar indicating the common variance of the elements in z_i and π_1 the only non-zero element of π_z.

Now exploiting the model structure in (2.1) and (2.28) we get the following:

$$\begin{aligned} E\left[z^{(1)}_i X'_i\right] &= \begin{pmatrix} 1 & E\left[w_i\right] & E\left[x_i\right] & \pi_w E\left[w^2_i\right] \\ E\left[w_i\right] & E\left[w^2_i\right] & \pi_w E\left[w^2_i\right] & \pi_w E\left[w^3_i\right] + \pi'_z E\left[z_i w^2_i\right] \\ E\left[z_i\right] & E\left[z_i w_i\right] & E\left[z_i z'_i\right]\pi_z & \pi_w E\left[z_i w^2_i\right] + E\left[w_i z_i z'_i\right]\pi_z \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 & \pi_w E\left[w^2_i\right] \\ 0 & E\left[w^2_i\right] & \pi_w E\left[w^2_i\right] & \pi_w E\left[w^3_i\right] + \pi'_z E\left[z_i w^2_i\right] \\ 0 & 0 & E\left[z^2_i\right]\pi_z & \pi_w E\left[z_i w^2_i\right] + E\left[w_i z_i z'_i\right]\pi_z \end{pmatrix}, \end{aligned}$$

Note that $E\left[z_i^{(1)}X_i'\right]$ is a $(k_z+2)\times 4$ matrix. Because $E\left[z_i z_i'\right] = E\left[z_i^2\right] I_{k_z}$ and only the first element of π_z is non-zero, a full rank has to come exclusively from the term $\pi_w E\left[z_i w_i^2\right] + E\left[w_i z_i z_i'\right]\pi_z$. There are a number of cases, but it is easily seen that the rank is only 3 when $E\left[z_i w_i^2\right] = 0$ and $E\left[w_i z_i z_i'\right] = 0$, which is the first result in Proposition 3 follows.

Regarding the IV2 estimator, we have that

$$E\left[z_i^{(2)}X_i'\right] = \begin{pmatrix} 1 & 0 & 0 & \pi_w E\left[w_i^2\right] \\ 0 & E\left[w_i^2\right] & \pi_w E\left[w_i^2\right] & \pi_w E\left[w_i^3\right] + \pi_z' E\left[z_i w_i^2\right] \\ 0 & 0 & E\left[z_i^2\right]\pi_z & \pi_w E\left[z_i w_i^2\right] + E\left[w_i z_i z_i'\right]\pi_z \\ 0 & E\left[z_i w_i^2\right] & \pi_w E\left[z_i w_i^2\right] + E\left[w_i z_i z_i'\right]\pi_z & \pi_w E\left[z_i w_i^3\right] + E\left[w_i^2 z_i z_i'\right]\pi_z \end{pmatrix},$$

which is of full rank.

Proof of Proposition 4. Given homoskedasticity and strong instruments, the asymptotic variance of the IV2 estimator is defined as:

$$V_{(2)} = \sigma_u^2 \left(plim\frac{1}{n}X'Z_{(2)} \left(plim\frac{1}{n}Z_{(2)}'Z_{(2)}\right)^{-1} plim\frac{1}{n}Z_{(2)}'X\right)^{-1},$$

where $plim\frac{1}{n}Z_{(2)}'X = E\left[z_i^{(2)}X_i'\right]$ is given above and

$$\begin{aligned} plim\frac{1}{n}Z_{(2)}'Z_{(2)} &= \begin{pmatrix} 1 & E\left[w_i\right] & E\left[z_i'\right] & E\left[w_i z_i'\right] \\ E\left[w_i\right] & E\left[w_i^2\right] & E\left[w_i z_i'\right] & E\left[w_i^2 z_i'\right] \\ E\left[z_i\right] & E\left[z_i w_i\right] & E\left[z_i z_i'\right] & E\left[w_i z_i z_i'\right] \\ E\left[z_i w_i\right] & E\left[z_i w_i^2\right] & E\left[w_i z_i z_i'\right] & E\left[w_i^2 z_i z_i'\right] \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & E\left[w_i^2\right] & 0 & E\left[w_i^2 z_i'\right] \\ 0 & 0 & E\left[z_i^2\right] I_{k_z} & E\left[w_i z_i z_i'\right] \\ 0 & E\left[z_i w_i^2\right] & E\left[w_i z_i z_i'\right] & E\left[w_i^2 z_i z_i'\right] \end{pmatrix}, \end{aligned}$$

Now assuming condition (2.30) we get:

$$E\left[z_i^{(2)}X_i'\right] = \begin{pmatrix} 1 & 0 & 0 & \pi_w E\left[w_i^2\right] \\ 0 & E\left[w_i^2\right] & \pi_w E\left[w_i^2\right] & \pi_w E\left[w_i^3\right] \\ 0 & 0 & E\left[z_i^2\right]\pi_z & 0 \\ 0 & 0 & 0 & \pi_w E\left[z_i w_i^3\right] + E\left[w_i^2 z_i z_i'\right]\pi_z \end{pmatrix},$$

$$E\left[z_i^{(2)}z_i^{(2)\prime}\right] = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & E\left[w_i^2\right] & 0 & 0 \\ 0 & 0 & E\left[z_i^2\right] I_{k_z} & 0 \\ 0 & 0 & 0 & E\left[w_i^2 z_i z_i'\right] \end{pmatrix},$$

hence

$$\begin{aligned}
\sigma_u^2 V_{(2)}^{-1} &= plim\frac{1}{n}X'Z_{(2)}\left(plim\frac{1}{n}Z_{(2)}'Z_{(2)}\right)^{-1}plim\frac{1}{n}Z_{(2)}'X \\
&= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & a_{22} & 0 & 0 \\ 0 & a_{14} & a_{33}' & 0 \\ a_{14} & a_{24} & 0 & a_{44}' \end{pmatrix} * \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & b_{22} & 0 & 0 \\ 0 & 0 & B_{33} & 0 \\ 0 & 0 & 0 & B_{44} \end{pmatrix}^{-1} * \begin{pmatrix} 1 & 0 & 0 & a_{14} \\ 0 & a_{22} & a_{14} & a_{24} \\ 0 & 0 & a_{33} & 0 \\ 0 & 0 & 0 & a_{44} \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 & 0 & a_{14} \\ 0 & \frac{a_{22}^2}{b_{22}} & \frac{a_{22}a_{14}}{b_{22}} & \frac{a_{22}a_{24}}{b_{22}} \\ 0 & \frac{a_{22}a_{14}}{b_{22}} & \frac{a_{14}^2}{b_{22}} + a_{33}'B_{33}^{-1}a_{33} & a_{14}\frac{a_{24}}{b_{22}} \\ a_{14} & \frac{a_{22}a_{24}}{b_{22}} & \frac{a_{14}a_{24}}{b_{22}} & a_{14}^2 + \frac{a_{24}^2}{b_{22}} + a_{44}'B_{44}^{-1}a_{44} \end{pmatrix},
\end{aligned}$$

where we used shorthand notation to indicate the nonzero entries in $E\left[z_i^{(2)}X_i'\right]$ and $E\left[z_i^{(2)}z_i^{(2)\prime}\right]$.

Defining the 4×4 matrix:

$$C = \begin{pmatrix} 1 & 0 & 0 & -a_{14} \\ 0 & 1 & -a_{14}/a_{22} & -a_{24}/a_{22} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

we have

$$\sigma_u^2 C'V_{(2)}^{-1}C = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{a_{22}^2}{b_{22}} & 0 & 0 \\ 0 & 0 & a_{33}'B_{33}^{-1}a_{33} & 0 \\ 0 & 0 & 0 & a_{44}'B_{44}^{-1}a_{44} \end{pmatrix}.$$

The inverse of C is:

$$C^{-1} = \begin{pmatrix} 1 & 0 & 0 & a_{14} \\ 0 & 1 & \frac{a_{14}}{a_{22}} & \frac{a_{24}}{a_{22}} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

and straightforward algebra now gives:

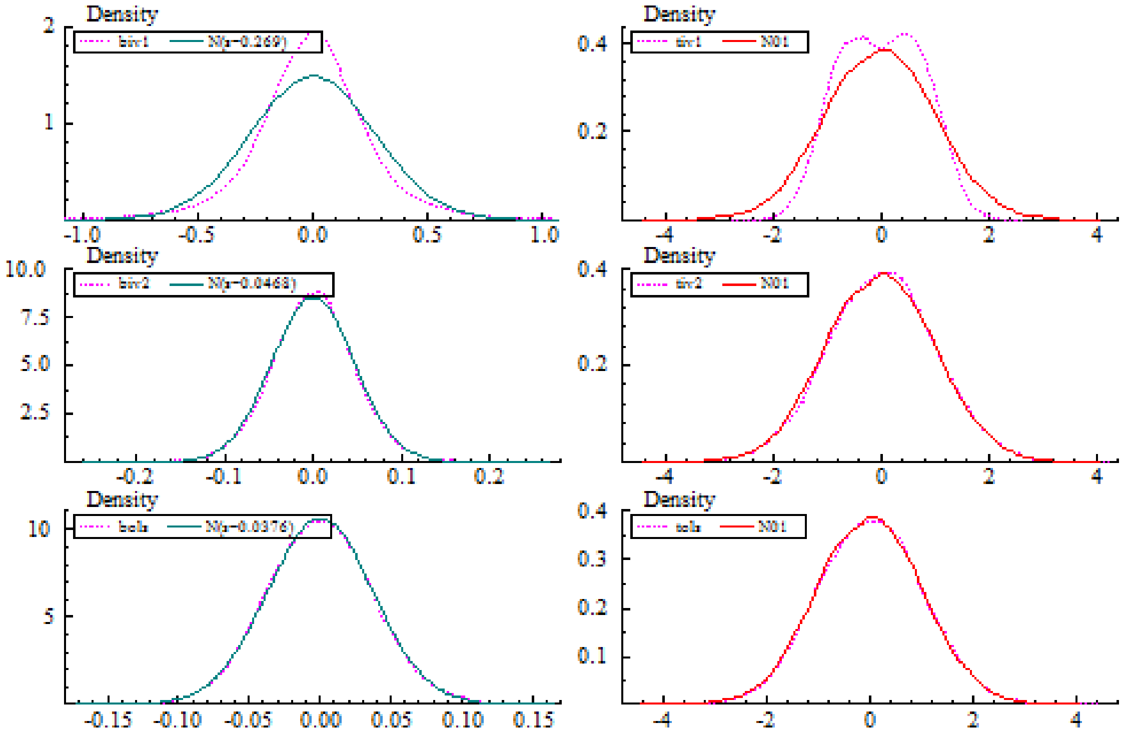
$$\begin{aligned}
V_{(2)} &= \sigma_u^2\left(C^{-1\prime}\sigma_u^2 C'V_{(2)}^{-1}CC^{-1}\right)^{-1} \\
&= \sigma_u^2 C\left(\sigma_u^2 C'V_{(2)}^{-1}C\right)^{-1}C' \\
&= \sigma_u^2 \begin{pmatrix} . & . & . & . \\ . & . & . & . \\ . & . & \left(a_{33}'B_{33}^{-1}a_{33}\right)^{-1} & 0 \\ . & . & 0 & \left(a_{44}'B_{44}^{-1}a_{44}\right)^{-1} \end{pmatrix},
\end{aligned}$$

where for brevity we only show the structure of the lower right block explicitly. From this it is seen that:

$$\begin{aligned}
V_{(2),x} &= \frac{\sigma_u^2}{E\left[z_i^2\right]\pi_z'\pi_z}, \\
V_{(2),x\cdot w} &= \frac{\sigma_u^2}{\left(\pi_w E\left[z_i w_i^3\right] + E\left[w_i^2 z_i z_i'\right]\pi_z\right)'\left(E\left[w_i^2 z_i z_i'\right]\right)^{-1}\left(\pi_w E\left[z_i w_i^3\right] + E\left[w_i^2 z_i z_i'\right]\pi_z\right)} \\
&= \frac{\sigma_u^2}{\pi_w^2 E\left[w_i^3 z_i'\right]\left(E\left[w_i^2 z_i z_i'\right]\right)^{-1} E\left[z_i w_i^3\right] + \pi_z' E\left[w_i^2 z_i z_i'\right]\pi_z}.
\end{aligned}$$

Figure 1: finite sample distributions of OLS, IV1 and IV2

coefficient estimators of β_{xw} and t-statistics, strong instruments case

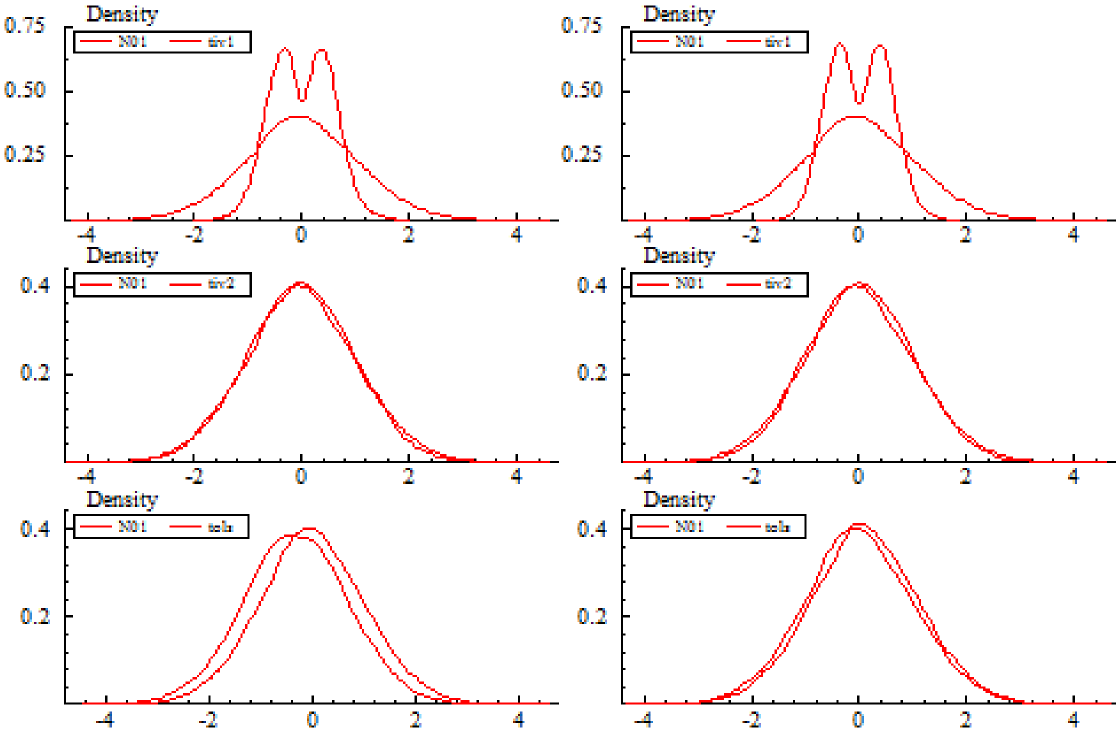


1

Note: Left panel are coefficient estimators, while right panel contains corresponding t-statistics.

Figure 2: finite sample distributions of OLS, IV1 and IV2

t-statistics in the Monte Carlo DGP based on the AHM data



Note: left (right) panels are t-statistics for testing β_x (β_{xw}).

Table 1: homoskedasticity and normality

		weak instruments			strong instruments		
		OLS	IV1	IV2	OLS	IV1	IV2
bias	β	0.000	0.000	0.001	0.001	0.001	0.001
	β_w	-0.000	-0.024	-0.033	-0.084	-0.005	-0.007
	β_x	0.476	0.228	0.321	0.083	0.003	0.006
	β_{xw}	-0.001	0.005	0.002	0.000	0.001	0.001
sd	β	0.089	0.136	0.099	0.105	0.276	0.110
	β_w	0.091	0.158	0.106	0.108	0.169	0.112
	β_x	0.089	0.453	0.284	0.041	0.067	0.046
	β_{xw}	0.091	0.667	0.245	0.038	0.269	0.047
rp ho	β	5.23	1.50	3.82	5.39	0.75	5.19
	β_w	8.62	1.97	4.77	12.01	2.47	4.98
	β_x	99.94	7.56	25.52	52.59	2.06	5.19
	β_{xw}	5.49	0.24	2.52	5.02	0.19	5.00
rp hr	β	4.85	1.43	3.61	5.03	0.97	4.93
	β_w	8.17	1.39	4.34	11.49	1.95	4.88
	β_x	99.91	7.13	23.25	50.17	1.54	4.90
	β_{xw}	5.57	0.22	1.86	5.12	0.16	4.58
W_c		3.09			3.62		
CD			0.63	1.13		0.81	21.65

Note: based on 10,000 MC replications; rp is actual rejection % of nominal 5% t-statistics; ho is homoskedasticity-only, hr means heteroskedasticity-robust; W_c is rp of nominal 5% test, while CD are average values.

Table 2: heteroskedasticity and normality

		weak instruments			strong instruments		
		OLS	IV1	IV2	OLS	IV1	IV2
bias	β	0.000	0.000	0.001	0.001	0.001	0.000
	β_w	-0.046	-0.012	-0.029	-0.084	-0.007	-0.010
	β_x	0.477	0.283	0.356	0.078	-0.002	0.003
	β_{xw}	-0.002	0.001	0.003	0.001	0.001	0.002
sd	β	0.087	0.134	0.094	0.096	0.259	0.097
	β_w	0.099	0.163	0.110	0.110	0.163	0.111
	β_x	0.147	0.604	0.369	0.065	0.088	0.072
	β_{xw}	0.145	0.671	0.306	0.062	0.267	0.076
rp ho	β	5.07	1.84	3.13	3.59	0.60	2.94
	β_w	11.40	2.39	6.13	13.39	2.36	5.11
	β_x	97.36	13.82	35.60	50.24	12.44	22.26
	β_{xw}	25.34	1.83	11.74	24.84	0.87	23.54
rp hr	β	4.35	1.71	3.39	4.12	1.06	4.10
	β_w	6.71	1.61	3.84	9.28	1.73	3.86
	β_x	86.09	7.71	21.43	24.97	3.41	5.53
	β_{xw}	6.71	0.82	4.36	7.11	0.80	6.45
W_c		2.21			3.20		
CD			0.65	1.34		0.81	20.42

Note: see Table 1.

Table 3: homoskedasticity and non-normality

		weak instruments			strong instruments		
		OLS	IV1	IV2	OLS	IV1	IV2
bias	β	0.004	-0.004	-0.002	-0.002	-0.004	-0.002
	β_w	-0.049	-0.024	-0.033	-0.083	-0.006	-0.007
	β_x	0.491	0.231	0.323	0.083	0.004	0.007
	β_{xw}	0.001	-0.005	-0.000	-0.000	0.002	-0.000
sd	β	0.090	0.136	0.099	0.105	0.275	0.110
	β_w	0.094	0.159	0.107	0.110	0.169	0.114
	β_x	0.113	0.438	0.286	0.043	0.066	0.046
	β_{xw}	0.116	0.701	0.257	0.039	0.258	0.047
rp ho	β	6.49	2.26	4.57	5.88	0.93	5.61
	β_w	9.84	2.23	5.53	12.71	2.38	5.39
	β_x	100.00	8.06	26.13	53.67	2.55	5.72
	β_{xw}	10.60	0.32	2.99	6.59	0.30	5.30
rp hr	β	5.48	2.38	4.29	5.72	1.35	5.44
	β_w	7.76	1.61	4.72	11.59	1.71	5.00
	β_x	99.69	7.49	24.18	49.90	1.73	5.08
	β_{xw}	6.36	0.21	1.83	5.50	0.23	4.49
W_c		2.57			3.36		
CD			0.64	1.17		0.82	21.56

Note: see Table 1.

Table 4: nonlinear dependence

		weak instruments			strong instruments		
		OLS	IV1	IV2	OLS	IV1	IV2
bias	β	0.002	0.007	0.005	-0.015	0.004	-0.002
	β_w	-0.047	-0.022	-0.032	-0.072	-0.003	-0.007
	β_x	0.477	0.229	0.326	0.087	0.002	0.007
	β_{xw}	-0.006	-0.034	-0.016	-0.012	-0.004	-0.000
sd	β	0.089	0.118	0.098	0.105	0.141	0.107
	β_w	0.093	0.151	0.107	0.110	0.189	0.119
	β_x	0.089	0.459	0.286	0.043	0.055	0.047
	β_{xw}	0.092	0.525	0.220	0.039	0.128	0.039
rp ho	β	5.28	2.27	3.89	5.64	3.99	5.48
	β_w	8.63	2.36	5.08	10.16	3.70	5.13
	β_x	99.94	9.27	26.23	54.58	4.28	5.70
	β_{xw}	5.89	1.07	3.76	7.11	3.20	5.02
rp hr	β	4.88	2.21	3.66	5.21	3.88	5.09
	β_w	8.06	1.98	4.62	9.22	2.97	4.86
	β_x	99.90	8.18	23.90	50.62	3.65	5.59
	β_{xw}	5.84	0.57	2.83	6.57	1.63	4.88
W_c		2.90			6.94		
CD			0.91	1.20		6.47	44.22

Note: see Table 1.

Table 5: estimation results, empty conditioning set

	interaction			linear		
	OLS	IV1	IV2	OLS	IV1	IV2
β	2.204	-6.347	2.065	0.254	-1.089	-0.458
	(0.644)	(17.831)	(1.001)	(0.756)	(1.168)	(1.123)
β_w	1.300	-4.293	1.507	-0.237	-0.638	-0.449
	(0.320)	(12.298)	(0.479)	(0.306)	(0.409)	(0.394)
β_x	-0.012	0.165	-0.015	0.033	0.052	0.043
	(0.009)	(0.379)	(0.016)	(0.009)	(0.015)	(0.015)
β_{xw}	-0.048	0.106	-0.061			
	(0.007)	(0.354)	(0.011)			
CD		0.072	6.281		12.423	6.324
Sargan		0.004	3.125		0.465	20.38
		(0.952)	(0.537)		(0.792)	(0.001)

Note: standard errors/p-values between parentheses.

Table 6: distribution of estimates of interaction coefficient

estimator	min	P25	P50	P75	max	sd	iqr
OLS	-0.051	-0.043	-0.040	-0.033	-0.008	0.012	0.010
IV1	-0.109	-0.097	-0.051	0.027	0.106	0.077	0.124
IV1-CD	0.072	0.121	0.479	0.921	3.654	1.016	0.800
IV2	-0.110	-0.090	-0.079	-0.068	-0.061	0.016	0.022
IV2-CD	0.839	1.669	3.464	4.375	6.281	1.693	2.706

Note: the twelve underlying specifications are taken from Table 1 of AHM.

Table 7: Monte Carlo DGP using OLS estimates as true parameter values

	bias			sd			rp		
	OLS	IV1	IV2	OLS	IV1	IV2	OLS	IV1	IV2
$\beta_0 = 2.204$	0.263	-0.012	0.020	0.706	7.033	0.960	6.41	0.06	4.33
$\beta_{w0} = 1.300$	0.067	-0.004	0.005	0.329	4.793	0.434	5.35	0.00	4.35
$\beta_{x0} = -0.012$	-0.003	0.000	-0.000	0.011	0.160	0.015	6.15	0.00	4.17
$\beta_{xw0} = -0.048$	0.001	0.000	0.000	0.006	0.149	0.010	5.27	0.00	3.88

Note: based on 10,000 MC replications; rp is actual rejection % (nominal 5%) of homoscedasticity-only t-tests.

References

Aghion, P., Howitt, P. and D. Mayer-Foulkes (2005). The effect of financial development on convergence: theory and evidence. *The Quarterly Journal of Economics* 120(1), 173-222.

Amemiya, T. (1974). Multivariate regression and simultaneous equation models when the dependent variables are truncated normal. *Econometrica* 42, 999-1012.

Anderson, T.W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics* 22, 327-351.

Balli, H.O. and B.E. Sørensen (2013). Interaction effects in Econometrics. *Empirical Economics* 45, 583-603.

Braumoeller, B.F. (2004). Hypothesis testing and multiplicative interaction terms. *International Organization* 58, 807-820.

Cameron, A.C. and P.K. Trivedi (2005). *Microeconometrics: methods and applications*. Cambridge University Press.

Cragg, J.G. and S.G. Donald (1993). Testing identifiability and specification in instrumental variable models. *Econometric Theory* 9, 222-240.

Davidson, R. and J.G. MacKinnon (1993). *Estimation and inference in Econometrics*. Oxford University Press, New York.

Dougherty, C. (2005). Why are the returns to schooling higher for women than for men? *Journal of Human Resources* 40, 969-988.

Kelejian, H.H. (1971). Two-Stage Least Squares and econometric systems linear in parameters but nonlinear in the endogenous variables. Journal of the American Statistical Association 66, 373-374.

Kiviet, J.F. (2013). Identification and inference in a simultaneous equation under alternative information sets and sampling schemes. *The Econometrics Journal* 16, S24-S59.

Kiviet, J.F. and J. Niemczyk (2012). The asymptotic and finite sample (un)conditional distributions of OLS and simple IV in simultaneous equations. *Journal of Computational Statistics and Data Analysis* 56, 3567-3586.

Kleibergen, F. and R. Paap (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics* 133, 97-126.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A. and R.W. Vishny (1997). Legal determinants of external finance. *Journal of Finance* 52, 1131-1150.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A. and R.W. Vishny (1998). Law and finance. *Journal of Political Economy* 106, 1113-1155.

Levine, R., Loayza, N. and T. Beck. Financial intermediation and growth: causality and causes. *Journal of Monetary Economics* 46, 31-77.

MacKinnon, J.G. (2013). Thirty years of heteroskedasticity-robust inference. In: Chen, X. and N.R. Swanson (eds.), *Recent advances and future directions in causality, prediction, and specification Analysis. Essays in Honor of Halbert L. White Jr.* Springer.

MacKinnon, J.G. and H. White (1985). Some heteroskedasticity consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics* 29, 305-325.

Nizalova, O. and I. Murtazashvili (2013). Exogenous treatment and endogenous factors: Vanishing of omitted variable bias on the interaction term. Working paper, Kyiv School of Economics.

Rajan, R.G. and L. Zingales (1998). Financial dependence and growth. *American Economic Review* 88, 559-586.

Staiger, D. and J.H. Stock (1997). Instrumental Variables regression with weak instruments. *Econometrica* 65, 557-586.

Steinhauer, A. and T. Wuergler (2010). Leverage and covariance matrix estimation in finite-sample IV regressions. Working paper, University of Zurich.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48, 817-838.

Wooldridge, J.M. (2002). *Econometric analysis of cross-section and panel data.* The MIT Press, Cambridge.