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: 30 December, 2014

JEL-code: C10; C31; C36

Keywords: endogeneity, instrumental variables, interaction term, OLS

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. 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 show that an implementation assuming exogeneity of the interaction term is valid under fairly weak conditions. In the more general case, we derive that instruments need to be interacted with the exogenous part of the interaction to achieve identification. Finally, we propose several specification tests to empirically assess the validity of OLS and IV inference for the interaction model. 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 Otilia Boldea, Stephane Bonhomme, Simon Broda, Arturas Juodis, Jan Kiviet, Frank Kleibergen, James MacKinnon, Andrew Pua and seminar participants 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. Of primary interest in regression models including interaction terms are the coefficients of those interactions. More specifically, one would like to verify whether the interaction term $x \cdot w$ is significant and economically important and thus should be included in the empirical model.

In this study we analyze the interaction model in which 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 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.

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 (2014), who analyze consistency of the OLS coefficient estimator in an interaction model with treatment effects under more restrictive conditions than in our setting. They also 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 can still rely on OLS estimation and inference. We first establish that, in some important cases, the interaction term actually can be classified as exogenous. We next show that, under a further 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 possibly 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 provide guidance on the optimal set of instruments given the presence of an endogenous regressor. Given a vector of valid instrumental variables z for x, we show that under reasonable conditions, the interaction term itself is exogenous and can therefore be included as part of the instrument set. In the more general case, the interaction term is treated as a second endogenous regressor. We show that in this case 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 documenting that (i) exploiting z alone leads to underidentification, even when the order condition is satisfied; and (ii) the interaction term itself may be a valid instrument. Finally, we propose a series of specification tests to enable the researcher to distinguish between IV and OLS and also the optimal set of instruments.

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 non-normality of the IV estimator when only the linear instruments are employed. Our results also validate the use of the interaction term as an instrument in a wide variety of cases. In addition, we 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 often 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]

$$y_i = \beta_\iota + \beta_w w_i + \beta_x x_i + \beta_{xw} x_i w_i + u_i. \tag{2.1}$$

Furthermore, a set of instruments (labeled z) is available to correct for the endogeneity of x. The endogenous regressor 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 effects of financial development depend on the initial GDP of the country. Stacking the observations $(i = 1, ..., n)$ 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 $\beta = (\beta_\iota, \beta_w, \beta_x, \beta_{xw})'$. For analysis of OLS, we do not specify a particular functional form or relation between the regressors. Of particular concern, x and w can be correlated.

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 regressors and instruments is without loss of generality. Because a constant is always included, all theoretical results below will continue to hold with rescaling of these variables (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.

[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 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 this study we will investigate to what extent this is also the case in the interaction model (2.1). Regarding the correlation between the interaction term and the structural error we can write:

$$cov(x_i w_i, u_i) = E\left(w_i E(x_i u_i | w_i)\right). \tag{2.4}$$

Now even if x_i and w_i are dependent random variables, the following assumption can still be satisfied:

Assumption 2. *The conditional expectation of $x_i u_i$ given w_i does not depend on w_i:*

$$E(x_i u_i | w_i) = E(x_i u_i). \tag{2.5}$$

Assumption 2 states that the degree of endogeneity does not depend on w.[2] Under this assumption we have that:

$$cov(x_i w_i, u_i) = E(w_i)E(x_i u_i) = 0, \tag{2.6}$$

because without loss of generalization we assumed that $E(w_i) = 0$. In other words, under Assumption 2 the interaction term is not an endogenous regressor after all.

Assumption 2 is a fairly weak condition implied by a variety of statistical data generating processes. An obvious special case is that of joint normality of x_i, w_i and u_i. In this case the conditional distribution of (x_i, u_i) is normal with conditional covariance not depending on w_i. Note that this condition holds because we assume that w is exogenous. More generally, in the linear IV model it is common to impose a Kronecker product form covariance matrix for the errors (see e.g. Staiger and Stock, 1997) implying Assumption 2. But even in the case of heteroskedasticity it still is a valid assumption as long as the conditional covariance between the structural and reduced form errors is not depending on w.

Whether Assumption 2 is plausible in relevant economic models is context-dependent. For example, consider the treatment regression model analyzed in Nizalova and Murtazashvili (2014) in which the treatment indicator w_i is randomly assigned. The treatment effect, however, may depend on a pre-treatment characteristic x_i, which potentially is correlated with other unobserved factors in u_i. Nevertheless, due to the random assignment w_i is independently distributed from both x_i and u_i, hence Assumption 2 is satisfied.

[2] The conditional covariance in (2.5) may still depend on other exogenous variables.

Assumption 2 may be reasonable as well in the case of relevant omitted variables. In the case of the returns to schooling application, for example, it will depend on how unobserved ability is related to wages. Suppose we model the returns to education being gender specific, i.e. y_i is log wage, x_i is schooling and w_i is gender in (2.1). The level of schooling x_i is endogenous, i.e. correlated with unobserved ability a_i. Suppose we have for the unobserved heterogeneity the following:

$$u_i = \beta_a a_i + \beta_{aw} a_i w_i + u_i^*, \tag{2.7}$$

where we allow for the possibility that the returns to ability may also be gender specific.[3] Note that in order to maintain exogeneity of w_i, we also need the additional assumption that $E(a_i|w_i) = 0$.[4] In the returns to schooling application this is not a strong assumption. However, in other circumstances this additional assumption will be violated, rendering w_i endogenous and therefore in need of IV estimation techniques as well. Maintaining exogeneity of w_i, we then have:

$$\begin{aligned} E\left[x_i u_i | w_i\right] &= E\left[\beta_a x_i a_i + \beta_{aw} x_i a_i w_i + x_i u_i^* | w_i\right] \\ &= (\beta_a + \beta_{aw} w_i) E\left[x_i a_i | w_i\right] + E\left[x_i u_i^* | w_i\right], \end{aligned} \tag{2.8}$$

so for Assumption 2 to hold we need at least that $\beta_{aw} = 0$.[5] In other words, the returns to ability need to be the same for males and females. Furthermore, we need that

$$E\left[x_i a_i | w_i\right] = E\left[x_i a_i\right], \tag{2.9}$$

$$E\left[x_i u_i^* | w_i\right] = E\left[x_i u_i^*\right]. \tag{2.10}$$

The second condition mimics Assumption 2, and let's assume it holds. Then for consistency of IV3 and OLS the correlation between schooling and ability should not depend on gender.

While Assumption 2 may still be reasonable in the case of relevant omitted variables, it is not when endogeneity of x_i is caused by simultaneity. Suppose the structural equation for x is:

$$x_i = \delta_\iota + \delta_w w_i + \delta_y y_i + \delta_z' z_i + \varepsilon_i. \tag{2.11}$$

As we show in the Appendix, in this case Assumption 2 only holds when either $\delta_y = 0$ (no simultaneity) or $\beta_{xw} = 0$ (no 'endogenous' interaction), which are both trivial cases.

Because its validity is context-dependent, in applied research one may therefore not *a priori* want to impose Assumption 2, and treat the interaction term as exogenous. It is relatively straightforward, however, to empirically test condition (2.5) with a Hausman

[3] Note that the specification of u_i in (2.7) rules out the validity of joint normality of x_i, w_i and u_i as discussed above. The reason is that the structural error u_i in general has a non-normal distribution, because u_i is (partly) a product of a_i and w_i.

[4] With this assumption $E(u_i|w_i) = 0$ since $E\left[u_i|w_i\right] = \beta_a E\left[a_i|w_i\right] + \beta_{aw} w_i E\left[a_i|w_i\right] + E\left[u_i^*|w_i\right]$.

[5] Alternatively, we assume that $E\left[x_i a_i|w_i\right] = 0$, but obviously this is a questionable assumption in applied research.

(1978) test comparing two different IV estimators, of which one assumes endogeneity of both x_i and $x_i w_i$ and one imposes exogeneity of $x_i w_i$. In Section 2.4 we will provide more details on this and other specification tests, but first we discuss the asymptotic properties of the OLS estimator.

2.2. OLS estimation and inference

In this Section we show that, under further reasonable conditions related to higher-order moments between x_i and w_i, 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:

$$\hat{\beta} = (X'X)^{-1}X'y. \tag{2.12}$$

Taking the probability limit we have

$$\begin{aligned} \operatorname{plim}\hat{\beta} &= \beta + \left(\operatorname{plim}\frac{1}{n}X'X\right)^{-1}\operatorname{plim}\frac{1}{n}X'u \\ &= \beta + \Sigma_{XX}^{-1}\Sigma_{Xu}, \end{aligned} \tag{2.13}$$

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 Assumptions 1 and 2, the inconsistency of the OLS estimator of model (2.1) equals:*

$$\Sigma_{XX}^{-1}\Sigma_{Xu} = \frac{\sigma_{xu}}{\det\left(\Sigma_{XX}\right)}\begin{bmatrix} E(x_i w_i)\left(E(x_i w_i)E(x_i w_i^2) - E(w_i^2)E(x_i^2 w_i)\right) \\ E\left(x_i^2 w_i\right)E\left(x_i w_i^2\right) + \left(E\left(x_i w_i\right)\right)^3 - E\left(x_i w_i\right)E\left(x_i^2 w_i^2\right) \\ E(w_i^2)\left(E\left(x_i^2 w_i^2\right) - E\left(x_i w_i^2\right)\right) - \left(E\left(x_i w_i^2\right)\right)^2 \\ E(x_i w_i)E(x_i w_i^2) - E(w_i^2)E(x_i^2 w_i) \end{bmatrix}. \tag{2.14}$$

Proof. see the Appendix. ■

The last element in the inconsistency (2.14) is interesting because quite often it will be zero, irrespective of the degree of endogeneity σ_{xu}. For example, when x_i and w_i are bivariate normal distributed we have that $E(x_i^2 w_i) = E(x_i w_i^2) = 0$. More generally, multivariate elliptical distributions are sufficient, but not necessary, for these higher-order dependencies to vanish.

The possible consistency of OLS for the coefficient β_{xw} is not limited to continuous random variables x and w. Suppose we can specify the relation between x and w by a linear (reduced form) regression:

$$x_i = \pi_w w_i + \xi_i, \tag{2.15}$$

with $\xi_i \sim i.i.d.(0, \sigma_\xi^2)$ independent from w_i. This implies:

$$E\left(x_i w_i\right) = \pi_w E\left(w_i^2\right) \tag{2.16}$$

$$E\left(x_i w_i^2\right) = \pi_w E\left(w_i^3\right), \tag{2.17}$$

which follows from the implied linear relation between x and w. Furthermore, we have:

$$E\left(x_i^2 w_i\right) = \pi_w^2 E\left(w_i^3\right), \tag{2.18}$$

Plugging these three components into the last element of (2.14), the terms will cancel. OLS is therefore consistent for the coefficient of the interaction term, but now irrespective of the further nature of the distribution of x and w. In particular, w could be discrete or continuous and have any distribution, as long as the conditional expectation of x_i given w_i is linear.

Depending on the statistical assumptions, the OLS estimators of other coefficients may be consistent as well. For example, in the treatment regression model discussed earlier the treatment indicator w_i is randomly assigned, hence independently distributed of x_i. From expression (2.14) we learn that in this case also β_w, the coefficient of w_i, can be consistently estimated. Therefore, as shown already by Nizalova and Murtazashvili (2014) one can consistenly estimate by OLS the full heterogeneous treatment effect, which depends on both β_w and β_{xw}. Below, we further that analysis by also deriving the limiting distribution, which is also applicable more generally.

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

$$y_i = X_i' \beta_* + \varepsilon_i, \tag{2.19}$$

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

$$\varepsilon_i = u_i - \Sigma_{Xu}' \Sigma_{XX}^{-1} X_i, \tag{2.20}$$

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 (2.12) is simply a method of moments estimator exploiting the following moment equation:

$$E\left[f_i\left(\beta\right)\right] = E\left[X_i\left(y_i - X_i'\beta\right)\right] = 0. \tag{2.21}$$

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 (2.12) is:*

$$\sqrt{n}\left(\hat{\beta}-\beta_*\right) \xrightarrow{d} \mathcal{N}(0,V), \tag{2.22}$$

where

$$\begin{aligned} V &= A_*^{-1}B_*A_*^{\prime -1}, && (2.23)\\ A_* &= \operatorname{plim}\frac{1}{n}\sum_{i=1}^{n}\left.\frac{\partial f_i(\beta)}{\partial \beta'}\right|_{\beta_*} = -\Sigma_{XX}, && (2.24)\\ B_* &= \operatorname{plim}\frac{1}{n}\sum_{i=1}^{n} f_i(\beta_*)f_i(\beta_*)' = E\left[\varepsilon_i^2 X_i X_i'\right]. && (2.25)\end{aligned}$$

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.[6]

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

$$\sqrt{n}\left(\hat{\beta}_{xw}-\beta_{xw}\right) \xrightarrow{d} \mathcal{N}\left(0,V_{xw}\right), \tag{2.26}$$

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 and asymptotically normal. It should be noted, however, that this consistency is restricted to β_{xw} only and not the full marginal effect of x on y (i.e., $\beta_x + \beta_{xw}w$) because the OLS estimator of β_x is inconsistent.

The 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 3. *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. We differ from Kiviet and Niemczyk (2012) in 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:

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

Proposition 2: *Under Assumptions 1 and 3, the asymptotic distribution of the OLS estimator of β in Lemma 1 simplifies to:*

$$\sqrt{n}\left(\hat{\beta}-\beta_*\right) \xrightarrow{d} \mathcal{N}\left(0,V\right), \tag{2.27}$$

with

$$\begin{aligned} \beta_* &= \beta + \frac{\sigma_{xu}}{\sigma_w^2\sigma_x^2-\sigma_{xw}^2}\begin{pmatrix} 0 & -\sigma_{xw} & \sigma_w^2 & 0 \end{pmatrix}' \qquad (2.28)\\ V &= \sigma_u^2\left(1-\rho_{xu}^2\right)\Sigma_{XX}^{-1}. \qquad (2.29)\end{aligned}$$

Proof. see the Appendix. ■

The result on the asymptotic variance in (2.29) 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

$$\operatorname{plim} s_u^2 = \operatorname{plim}\frac{\hat{u}'\hat{u}}{n} = \sigma_u^2\left(1-\rho_{xu}^2\right), \tag{2.30}$$

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

$$\hat{V}_{ho} = s_u^2\left(X'X\right)^{-1}, \tag{2.31}$$

is asymptotically valid. Thus, 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.

$$\frac{\hat{\beta}_{xw}-\beta_{xw,0}}{SE_{ho}(\hat{\beta}_{xw})} \xrightarrow{d} \mathcal{N}\left(0,1\right), \tag{2.32}$$

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:

$$\hat{V}_{hr} = \left(X'X\right)^{-1}\sum_{i=1}^{n}\frac{\hat{\varepsilon}_i^2}{\left(1-\hat{h}_i\right)^2}X_iX_i'\left(X'X\right)^{-1}, \tag{2.33}$$

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 (2.33) has been proposed by Davidson and MacKinnon (1993, p554), and is a slight modification of the jackknife estimator in MacKinnon and White (1985).[7] The robust variance estimator in (2.33) will be sufficient to warrant asymptotically valid inference for β_{xw}, i.e.

$$\frac{\hat{\beta}_{xw} - \beta_{xw,0}}{SE_{hr}(\hat{\beta}_{xw})} \xrightarrow{d} \mathcal{N}(0,1), \tag{2.34}$$

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, 2013).

Summarizing, the OLS estimator of the interaction term in (2.1) is, under many reasonable conditions, consistent and asymptotically normal, and standard inference can be applied. An important implication of these results is that in some applications, if the main empirical result only depends on the interaction variable, OLS based inference is sufficient and one does not need to resort to IV techniques. In other words, OLS inference can be performed without worry about strength and exogeneity of instruments. Still certain assumptions are needed for valid OLS inference, and in Section 2.4 we will discuss several ways of verifying these conditions in practice.

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 three 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. Compared with IV2, the third implementation (labeled IV3) adds the original interaction term $x \cdot w$ as an instrumental variable.

In applied research, IV3 is never exploited because typically the interaction term is treated as an endogenous regressor too. As we showed above, however, under mild conditions, $x \cdot w$ is actually an exogenous regressor. Furthermore, typically 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 instruments for $x \cdot w$ when z are valid instruments

[7] In the literature this is known as the HC3 estimator, see MacKinnon (2013) for an overview of heteroscedasticity-robust covariance matrix estimators.

for x (Wooldridge, p122, 2002). 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 even when we have more instruments than endogenous variables.

To analyze the IV estimators, we supplement the structural equation (2.1) with the following reduced form[8] for the endogenous regressor x:

$$x_i = \pi + \pi_w w_i + z_i' \pi_z + v_i, \tag{2.35}$$

where z_i is the k_z dimensional vector of instrumental variables excluded from (2.1). We analyze the relevance of the IV1, IV2 and IV3 sets of instruments defined as:

$$\begin{aligned} z_i^{(1)} &= \begin{bmatrix} 1 & w_i & z_i' \end{bmatrix}', & (2.36) \\ z_i^{(2)} &= \begin{bmatrix} 1 & w_i & z_i' & z_i' \cdot w_i \end{bmatrix}', & (2.37) \\ z_i^{(3)} &= \begin{bmatrix} 1 & w_i & z_i' & z_i' \cdot w_i & x_i \cdot w_i \end{bmatrix}'. & (2.38) \end{aligned}$$

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), (2.35) and Assumption 1, we have that:*

$$\begin{aligned} \text{rank } E\left[z_i^{(1)} X_i'\right] &= 3 \text{ if } E\left[w_i^2 z_i\right] = 0 \text{ and } E\left[w_i z_i z_i'\right] = 0, & (2.39) \\ \text{rank } E\left[z_i^{(2)} X_i'\right] &= 4. \end{aligned}$$

Proof. see the Appendix. ■

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

$$E\left[w_i^2 z_i\right] = 0 \text{ and } E\left[w_i z_i z_i'\right] = 0. \tag{2.40}$$

In practice it is relatively easy to test the reduced rank for IV1 by applying the rank statistic of Kleibergen and Paap (2006), which is a generalization of the Anderson (1951) rank statistic to the case of a nonscalar error covariance structure. Note that it is often the case that when condition (2.40) holds, the last element in the OLS inconsistency (2.14) is zero, and vice versa (see e.g. the last Monte Carlo design in Section 3.4 below).

Proposition 3 shows that, 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

[8] Note that the reduced form specified in (2.15) is more restrictive in the sense that independence is assumed between w and the other stochastic components (z and v) as well as homoskedasticity.

the structural parameters. As we will show in the Monte Carlo experiments, we tend to find bimodality for the IV1 t-statistic. 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 (2.40) holds:

Proposition 4: *Under (2.1), (2.35), Assumption 1, condition (2.40) and homoskedastic errors* u_i and v_i, *the asymptotic variances of the IV2 estimator of* β_x *and* β_{xw} *are equal to:*

$$V_{(2),x} = \frac{\sigma_u^2}{E\left[z_i^2\right]\pi_z'\pi_z}, \tag{2.41}$$

$$V_{(2),xw} = \frac{\sigma_u^2}{\pi_w^2 E\left[w_i^3 z_i'\right]\left(E\left[w_i^2 z_i z_i'\right]\right)^{-1} E\left[z_i w_i^3\right] + \pi_z' E\left[w_i^2 z_i z_i'\right]\pi_z}. \tag{2.42}$$

Proof. see the Appendix. ■

Proposition 4 shows that when instruments become weak, i.e. $\pi_z \approx 0$, the IV2 estimator is subject to the usual weak instruments problem and its variance will become large.[9] 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).

Until now we assumed that we have two endogenous regressors, i.e. x and $x \cdot w$. We have shown, however, that under Assumption 2 the correlation between interaction term and structural error is zero. In this case we can classify $x \cdot w$ as exogenous, and use IV3 instead of IV2.[10] We will now discuss how to empirically assess when we can use IV2, IV3 or even OLS.

2.4. empirical comparison of IV and OLS estimators

In applied research one may not *a priori* want to impose the assumptions leading to consistency of IV3 and OLS. In this Section we therefore discuss various ways to empirically

[9] 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 $E\left[w_i^2\right] = \sigma_w^2$.

[10] An alternative implementation of IV3 is without $z \cdot w$ as instruments. It will be necessary, however, to include these instruments when we apply the specification test of Hahn et al. (2011), see Section 2.4 below.

assess the validity of IV2, IV3 and OLS estimators defined as:

$$\begin{aligned}\hat{\beta}_2 &= \left(X'P_{Z_2}X\right)^{-1}X'P_{Z_2}y \\ \hat{\beta}_3 &= \left(X'P_{Z_3}X\right)^{-1}X'P_{Z_3}y \\ \hat{\beta} &= \left(X'X\right)^{-1}X'y,\end{aligned}$$

where $Z_m = \left(z_1^{(m)}, ..., z_n^{(m)}\right)'$ and $z_i^{(m)}$, $m = \{2,3\}$, as in (2.37) and (2.38) respectively. More specifically, we limit ourselves to testing the coefficient of the interaction term β_{xw} because this is the parameter of interest.

The two prerequisites for consistency of the OLS estimator of β_{xw} are Assumption 2 and furthermore the condition:

$$E(x_iw_i)E(x_iw_i^2) - E(w_i^2)E(x_i^2w_i) = 0, \tag{2.43}$$

which is the last element in the OLS inconsistency (2.14). The condition (2.43) can be easily checked by a Wald test. Defining

$$\theta = \left[\begin{array}{cccc} E(x_iw_i) & E(x_iw_i^2) & E(w_i^2) & E(x_i^2w_i) \end{array}\right]', \tag{2.44}$$

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:

$$\frac{1}{n}\sum_{i=1}^{n} m(w_i, x_i, \hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\left(\left[\begin{array}{c} x_iw_i \\ x_iw_i^2 \\ w_i^2 \\ x_i^2w_i \end{array}\right] - \left[\begin{array}{c} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \hat{\theta}_3 \\ \hat{\theta}_4 \end{array}\right]\right) = 0. \tag{2.45}$$

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

$$\sqrt{n}\left(\hat{\theta} - \theta\right) \xrightarrow{d} \mathcal{N}\left(0, C_0\right), \tag{2.46}$$

where C_0 can be estimated consistently by

$$\hat{C} = \frac{1}{n}\sum_{i=1}^{n} m_i(\hat{\theta})m_i(\hat{\theta})'. \tag{2.47}$$

Defining now $h(\theta) = \theta_1\theta_2 - \theta_3\theta_4$, we have $\frac{\partial h(\theta)}{\partial \theta'} = \left[\begin{array}{cccc} \theta_2 & \theta_1 & -\theta_4 & -\theta_3 \end{array}\right]$ and we can test $H_0 : h(\theta) = 0$ with the following Wald t-test statistic:[11]

$$W_c = \frac{h(\hat{\theta})}{\sqrt{n^{-1}\hat{r}\hat{C}\hat{r}'}}, \tag{2.48}$$

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

where $\hat{r} = \left.\frac{\partial h(\theta)}{\partial \theta'}\right|_{\hat{\theta}}$. Under H_0 the statistic in (2.48) is asymptotically standard normal distributed. In the Appendix we describe how to adapt the statistic (2.48) to the case of nonzero means.

If the applied researcher does not reject condition (2.43) with the W_c statistic (2.48), one can proceed with testing Assumption 2. For this we can compare IV2 and IV3 estimators exploiting the conventional Hausman (1978) test, because IV3 is consistent only under Assumption 2. The Hausman (1978) statistic comparing IV2 and IV3 estimators reads:

$$\left(\hat{\beta}_2 - \hat{\beta}_3\right)' \hat{V}_H^{-1} \left(\hat{\beta}_2 - \hat{\beta}_3\right), \tag{2.49}$$

where $\hat{V}_H$ is an estimate of the asymptotic variance of the contrast vector $\hat{\beta}_2 - \hat{\beta}_3$. Because we limit ourselves to testing the coefficient of the interaction term β_{xw}, we only test the last element $\hat{\beta}_{xw,2} - \hat{\beta}_{xw,3}$. The Hausman test statistic (2.49) then simply becomes:

$$H_{23} = \frac{\left(\hat{\beta}_{xw,2} - \hat{\beta}_{xw,3}\right)^2}{\hat{V}_{H,xw}}, \tag{2.50}$$

with $\hat{V}_{H,xw}$ the last diagonal element of $\hat{V}_H$. Under homoskedasticity we use:

$$\hat{V}_H = s_{u,2}^2 \left(X'P_{Z_2}X\right)^{-1} - s_{u,3}^2 \left(X'P_{Z_3}X\right)^{-1}. \tag{2.51}$$

where

$$s_{u,j}^2 = \frac{1}{n}\left(y - X\hat{\beta}_j\right)'\left(y - X\hat{\beta}_j\right), \quad j = 2, 3. \tag{2.52}$$

is the IV estimate of the error variance. In the case of heteroskedasticity we use the following modification to estimate the asymptotic variance of the contrast vector (see e.g. Cameron and Trivedi, 2005):

$$\hat{V}_H = \hat{V}_{22} + \hat{V}_{33} - 2\hat{V}_{23}, \tag{2.53}$$

where

$$\begin{aligned}
\hat{V}_{jj} &= n\left(X'P_{Z_j}X\right)^{-1} X'Z_j\left(Z_j'Z_j\right)^{-1}\hat{S}_{jj}\left(Z_j'Z_j\right)^{-1}Z_j'X\left(X'P_{Z_j}X\right)^{-1}, \quad j = 2, 3, \\
\hat{V}_{23} &= n\left(X'P_{Z_2}X\right)^{-1} X'Z_2\left(Z_2'Z_2\right)^{-1}\hat{S}_{23}\left(Z_3'Z_3\right)^{-1}Z_3'X\left(X'P_{Z_3}X\right)^{-1},
\end{aligned}$$

$$\begin{aligned}
\hat{S}_{jj} &= \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{u}_{j,i}^2}{\left(1 - h_{j,i}\right)^2} z_i^{(j)} z_i^{(j)\prime}, \quad j = 2, 3, \\
\hat{S}_{23} &= \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{u}_{2,i}^2}{\left(1 - h_{2,i}\right)^2} z_i^{(2)} z_i^{(3)\prime},
\end{aligned}$$

and $h_{j,i}$ is the i^{th} diagonal element of P_{Z_j}, $j = 2, 3$. Using (2.53) in (2.50) results in a heteroskedasticity-robust Hausman test.

The statistic in (2.50) will, under the null hypothesis that $\text{plim}(\hat{\beta}_{xw,2} - \hat{\beta}_{xw,3}) = 0$, be asymptotically distributed as χ_1^2. When the applied researcher rejects this null hypothesis, Assumption 2 is not valid. In this case OLS (and also IV3) inference is not valid, and the IV2 estimator is to be preferred.

As is well known, the particular version of the Hausman test above, i.e. exploiting the variance estimator (2.51), is not robust to weak instruments (Staiger and Stock, 1997; Hahn et al., 2011). In the case of weak instruments, however, one can still test Assumption 2 exploiting a weak instrument robust version of the Hausman test proposed by Hahn et al. (2011). In the case of exact identification, i.e. $k_z = 1$, the resulting Hausman statistic can be written as (2.50), but replacing $s_{u,2}^2$ in (2.51) by the IV3 variance estimator $s_{u,3}^2$:

$$\hat{V}_H = s_{u,3}^2 \left(X' P_{Z_2} X\right)^{-1} - s_{u,3}^2 \left(X' P_{Z_3} X\right)^{-1}. \tag{2.54}$$

Hahn et al. (2011) also provide an extension to heteroskedasticity.

Summarizing, to empirically evaluate the validity of OLS, IV3 and IV2 inference on the coefficient of the interaction term β_{xw} we perform the following sequence of tests.[12] We first test condition (2.43) with the W_c statistic in (2.48). Note that the W_c statistic is robust to identification issues originating from weak instruments, because it only depends on the regressors x and w. Therefore, this is the best starting point for checking consistency of OLS, because we don't have to worry about the validity of external instrumental variables. If we reject condition (2.43), then OLS is inconsistent. If we do not reject, we continue with the H_{23} statistic in (2.50) to check Assumption 2. Note that the H_{23} statistic is not robust to weak instruments, so depending on a pre-test for weak instruments one can exploit (2.51) or (2.54) as variance estimator. If we also do not reject Assumption 2, then OLS inference for β_{xw} is valid. If we reject Assumption 2, then we prefer the IV2 estimator. For both strong and weak instrument scenarios, we investigate in the Monte Carlo study below the accuracy of the moments based test W_c and the various H_{23} tests in finite samples.

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, IV2 and IV3 estimators. Once again, the difference between the three IV estimators is due to the choice of different instrument sets; see (2.36), (2.37) and (2.38). On top of the instruments of the IV1 estimator, the IV2 estimator also exploits instruments interacted with the exogenous part of the interaction term. IV3 additionally exploits exogeneity of the interaction term.

[12] An alternative Hausman test would be to compare OLS and IV2, but this is superfluous given that we already perform the W_c and H_{23} tests.

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 (2.31) and (2.33) for OLS, and equivalent expressions for IV estimators (see for details Steinhauer and Wuergler, 2010). Furthermore, for the IV estimators we calculate Cragg-Donald (1993) and Kleibergen-Paap (2006) statistics. The latter (labeled KP) tests the rank condition and can therefore be viewed as a first check on identification. Note that in the case of homoskedasticity it is equal to the Anderson (1951) rank statistic. Cragg-Donald statistics (labeled CD) indicate weak instruments in the case of (multiple) endogenous regressors. In the case of a single endogenous regressor it reduces to the first stage F statistic testing the joint significance of the excluded instruments. Cragg-Donald statistics indicate strength of identification and it is well known that very weak instruments can lead to huge standard errors and non-normality of finite sample distributions.

Finally, we report the actual rejection percentage of the W_c statistic (2.48), testing the condition (2.43) underlying consistency of OLS, as well as the Hausman H_{23} test checking Assumption 2 by comparing IV2 and IV3 estimators. We report outcomes from homoskedasticity-only and heteroskedasticity-robust Hausman tests[13]. Furthermore, we report results for the strong IV Hausman statistic using (2.51) as the variance estimator as well as the weak instrument robust Hausman statistic using (2.54) as the variance estimator (labeled H_{23}^s and H_{23}^w respectively).

We generate data for y and x according to (2.1) and (2.35). Although most of the designs fulfill Assumption 2, we also specify a Monte Carlo DGP in which the interaction term is actually endogenous and, hence, OLS and also IV3 are inconsistent. Given that conditions (2.43) and (2.40) are crucial for OLS and IV1 respectively, we furthermore specify a DGP in which we can vary the nonlinear dependencies between x, w and z.

Choosing values and distributions for the various elements in (2.1) and (2.35) 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 are not affected.

Furthermore, for all designs, we choose $\beta_\iota = \beta_w = \beta_x = \beta_{xw} = 1$ in (2.1) and the unconditional error variances are standardized at $\sigma_u^2 = \sigma_v^2 = 1$. We also generate normally distributed errors in all experiments. Finally, we set $k_z = 5$ and $n = 100$ in all experiments. These choices are not without loss of generalization, but unreported simulation experiments[14] show that these choices are relatively innocuous for the main conclusions.

[13] By construction the W_c test is heteroskedasticity-robust.

[14] Among other things, we have simulated also with χ^2 or t distributed errors, $n = \{200, 500, 1000\}$ and

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 homoskedastic 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 ~i.i.n.$(0,1)$, z_i ~i.i.n.$(0, I_{k_z})$, (u_i, v_i) ~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 inference for β_{xw} 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) for the first-stage F statistic. 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 probabilities (rp) of nominal 5% t-tests. Furthermore, we report the actual rejection percentage of the Anderson (1951) and Kleibergen-Paap (2006) statistics checking the rank condition for all IV estimators, as well as average Cragg-Donald (1993) statistics. Finally, we report actual rejection percentages of the W_c and H_{23} tests.

<Table 1 about here>

We can see that while the OLS estimator for β_x (and also β_w) remains biased, we observe neglible bias in estimating β_{xw} showing the consistency result of Section 2. Biases in IV estimators are substantially smaller for strong instruments, as expected. Comparing IV1 and IV2/IV3, we find that the standard deviation of IV1 is much larger than that of IV2/IV3 reflecting the lack of identification discussed previously. The large variance is particularly acute for the IV1 estimator of β_{xw}. Dispersion of IV2/IV3 is also larger than for OLS, but the difference in variance is much smaller than that of IV1 versus IV2/IV3. IV1 inference is

---

$k_z = \{2, 10\}$.

almost always conservative, i.e. rejection frequencies are well below the nominal significance level. IV2/IV3 inference is valid in the case of strong instruments. OLS coefficient bias (and lack of bias as it may be) 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.

The several specification tests to discriminate between OLS, IV3 and IV2 estimators are all reasonably well behaved. Note that in this Monte Carlo design both condition (2.43) and Assumption 2 are satisfied, hence we simulate under the null hypothesis of validity of both OLS and IV3. Therefore, in this design it can be checked what the actual significance level is for nominal 5% W_c and H_{23} tests. From Table 1 it can be seen that the various Hausman tests are slightly conservative in the case of weak instruments, while the W_c test has almost no size distortion already at a small sample size of $n = 100$.

The KP rank statistics checking the condition for IV1 inconsistency are also accurate under the null hypothesis. In this design the rank condition for identification is not satisfied for IV1, irrespective of the strength of the instruments. When instruments are strong the actual rejection percentage for IV1 is close to the nominal level of 5%, while for IV2 and IV3 KP has large power. As a result of this, IV2 and IV3 clearly outperform IV1 with IV3 being superior to IV2. Average CD statistics in the weak instruments experiment are of similar magnitude indicating an identification problem for all IV implementations. Average CD statistics in the strong instruments experiment are of different magnitude indicating that the identification problem remains for the IV1 estimator. This corresponds to the theoretical results of Proposition 3.

<Figure 1 about here>

Figure 1 further illustrates this by plotting empirical densities for the OLS, IV1, IV2 and IV3 estimation errors of β_{xw} (left panels) and corresponding t-statistics (right panels) for testing $H_0 : \beta_{xw} = \beta_{xw,0}$.[15] 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, IV3 and OLS estimators of β_{xw}, however, are very close to a normal distribution, as expected.[16]Corresponding t-statistics should be distributed as standard normal, which we plot as well for reference in each of the four 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, IV3 and OLS t-statistics, however, are very close to a standard normal distribution, as expected.

[15] For reference we plot in the left panel of Figure 1 also normal densities with a variance that matches with the empirical distribution, while in the right panel we plot densities from the standard normal distribution.

[16] 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.

Summarizing, OLS inference on β_{xw}, the coefficient of the interaction term, is excellent reflecting the theoretical results of the previous Section. Furthermore, the several specification tests to discriminate between IV2, IV3 and OLS estimators are all reasonably well behaved under the null hypothesis of validity of both OLS and IV3. Finally, the KP and CD statistics correctly diagnose the identification problem for IV1.

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.[17] Now we generate heteroskedastic errors according to (u_i, v_i) ∼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/IV3, we again find that the standard deviation of IV1 is much larger than that of IV2/IV3 reflecting lack of identification.

IV2 or IV3 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 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

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

to nominal level. Regarding the other coefficients, however, OLS t-statistics should not be used. Finally, we find in Table 2 that the heteroskedasticity-robust specification tests are reasonably well behaved under the null hypothesis. Somewhat remarkable is that even homoskedasticity-only tests are not seriously size distorted under heteroskedasticity.

3.3. nonlinear dependence

In our third design, we assume homoskedasticity again, but 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:

$$z_{1i} = \frac{1}{\sqrt{2}} \left(w_i^2 - 1 \right), \tag{3.1}$$

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_1^2 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

$$E(z_{1i}w_i^2) = \frac{1}{\sqrt{2}} \left(E(w_i^4) - E(w_i^2) \right), \tag{3.2}$$

which is nonzero. Hence, the rank condition failure as described in Proposition 3 is not present, and IV1 identifies the structural parameters in (2.1).

It should be noted that in this design the condition (2.43) for consistency of OLS is violated, because

$$E(x_i w_i^2) = \frac{\pi_j}{\sqrt{2}} \left(E(w_i^4) - E(w_i^2) \right), \tag{3.3}$$

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

We report results from the strong instruments experiment only. This is because 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 the case of weak instruments.

Table 3 reports simulation results for this design. To maximize the effect of the alternative construction of z_{1i}, we also report results for $k_z = 1$ (left panel) in addition to $k_z = 5$ (right panel). For $k_z = 1$ the IV1 estimator is underidentified and IV2 exactly identified.[18]

<Table 3 about here>

[18] For completeness we report mean and standard deviation for IV2 in the left panel of Table 4. These numbers should be treated with care, however, as corresponding population moments do not exist in case of an exactly identified model.

The right panel of Table 3 shows that the IV1 estimator performs much better compared with other designs. Although its efficiency is still lower than IV2 and IV3, it has greatly improved compared with other designs and size distortions are smaller. In the case $k_z = 5$ (right panel) 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. However, the left panel of Table 3 ($k_z = 1$) provides a situation where there is a substantial OLS inconsistency for β_{xw}, which is reflected also in a large size distortion of its corresponding t-test.

In this design condition (2.43) is violated, hence contrary to earlier designs we now simulate the power of the W_c specification test. Table 3 shows that it has reasonable power against the null hypothesis that OLS is a consistent estimator. Furthermore, Assumption 2 is still valid in this design, hence we still simulate actual size of the various Hausman specification tests. As can be seen from Table 3, also in this design size distortions are small.

3.4. omitted variables bias

In our last design we quantify the implications of omitted variables bias as described in Section 2.1. More in particular, we investigate the consequences of violating Assumption 2. We assume that the structural errors u_i obey equation (2.7) with $a_i \sim$i.i.n.$(0,1)$, while we use a similar specification for the reduced form errors v_i:

$$v_i = \pi_a a_i + \pi_{aw} a_i w_i + v_i^*. \tag{3.4}$$

We furthermore assume that $(u_i^*, v_i^*)' \sim$i.i.n.$(0, \Sigma^*)$ with Σ^* such that the variance covariance matrix Σ of $(u_i, v_i)'$ stays unaltered compared with the benchmark design. This implies the following for the elements in Σ^*:

$$\Sigma^* = \begin{pmatrix} 1 - \beta_a^2 - \beta_{aw}^2 & \rho_{uv} - \beta_a \pi_a - \beta_{aw} \pi_{aw} \\ 1 - \beta_a \pi_a - \beta_{aw} \pi_{aw} & 1 - \pi_a^2 - \pi_{aw}^2 \end{pmatrix}. \tag{3.5}$$

The discussion in Section 2.1 predicts that the OLS and IV3 coefficient estimators of β_{xw} are inconsistent when $\beta_{aw} \neq 0$. It can be shown that this is also the case when $\pi_{aw} \neq 0$. We choose the parameters in (3.5) such that Σ^* is a positive definite matrix. For simplicity we set $\pi_{aw} = 0$. We furthermore set $\beta_a = \pi_a = 0.5$, hence the omitted variable a_i is correlated with x_i. Finally, we choose $\beta_{aw} = \{0, 0.5\}$. In the case of $\beta_{aw} = 0$ Assumption 2 is not violated and OLS/IV3 is consistent. However, for $\beta_{aw} = 0.5$ biases will occur. Note that in this latter case the resulting errors are conditionally heteroscedastic, hence valid IV2 inference only results with a heteroscedasticity-robust covariance estimator.

<Table 4 about here>

Table 4 shows the simulation results when $\beta_{aw} = 0.5$.[19] Compared with Table 1 it can be seen that omitted variables bias may affect the finite sample properties substantially. Because Assumption 2 is violated, biases are seen in estimating β_{xw} by OLS and IV3. And in the case of weak instruments bias in estimating this coefficient is present for all estimators. Naturally the bias in estimating β_{xw} carries over to the t-statistics based on OLS/IV3 resulting in substantial size distortions for these tests. As predicted by the theoretical results, only heteroscedasticity-robust IV2 inference is reasonably accurate in this case.

Because for this design both IV3 and OLS are inconsistent estimators, the rejection frequencies of the H_{23} specification test statistic in Table 4 actually measure power. It is clear that especially in the case of weak instruments power is low, i.e. rejection frequencies of the Hausman statistics are around the nominal significance level of 5%. It is only in the case of strong instruments that some power results for discriminating between IV2 and IV3/OLS. Furthermore, note that in this design the condition (2.43) is fulfilled again. From Table 4 it is seen that actual size of the W_c statistic is again close to its nominal significance level.

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").[20] 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 Scandi-

[19] We don't report the results for $\beta_{aw} = 0$, which are virtually equal to those in Table 1 (as expected).

[20] 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.

navian 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, IV2, and IV3 estimators of (2.1) using private credit as the measure of financial development.[21] The first specification does not include any further control variables; AHM (p193) consider this specification to be representative of their main result. The second set of results uses their full set of control variables.[22] Note that AHM only use IV2, and the reported IV2 coefficient estimates in columns (3) and (7) of Table 5 indeed correspond exactly to AHM's results in columns 1 and 3 of their Table 1. Also note that AHM used homoskedasticity-only standard errors, which we used as well for the IV estimation.[23] Regarding OLS we report heteroskedasticity-robust standard errors, because only under particular distributional assumptions (i.e. see Proposition 2) is homoskedasticity-only inference is valid for this estimator.

<Table 5 about here>

Table 5 shows the similarity of the OLS, IV2, and IV3 results, particularly for the interaction coefficient.[24] We also note the difference in the point estimate and the huge standard errors for IV1. Both empirical facts corroborate our theoretical and simulation results. 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 clearly seen from Table 5. The Cragg-Donald statistics (0.07 and 0.62 respectively) indicate very weak instruments for IV1 corroborating with our theoretical results.[25] Furthermore, the Kleibergen-Paap (2006) rank statistic does not reject the null hypothesis of underidentification for IV1 in both specifications of Table 5. Weak instruments are less of a problem in the case of IV2 and IV3, although the CD statistics are not very large either.

Because AHM treat financial development as an endogenous regressor, they do not consider OLS estimation. Our theoretical results, however, show that when both Assumption 2 and the further condition (2.43) are satisfied OLS inference is valid for the coefficient of

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

[22] Empirical results for the policy set of control variables are similar.

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

[24] All of these results use the raw, non-mean centered data. As Balli and Sørensen (2013) note one must be careful in interpretation when scaling data that contains interaction terms. In particular, if data are mean-centered, one must add $X * \bar{w}$ to our interaction term in order to produce correct estimates. With this correction, inference and our specification tests are invariant to mean centering.

[25] We note that the anomalous IV1 estimation results seem to originate precisely from rank issues discussed in proposition 3, because the CD statistic for the IV1 estimator is much larger when estimating a linear specification excluding the interaction term.

the interaction term. Applying the W_c test for the empty specification, we do not reject the null (p-value is 0.07). Furthermore, checking Assumption 2 with the H_{23} test we also do not reject (p-values are 0.11 and 0.73 for H_{23}^s and H_{23}^w respectively). These outcomes suggest that OLS inference for the interaction term is valid.

The results become somewhat different, however, when considering the specification including the full set of control variables. The differences between OLS, IV2 and IV3 interaction estimates become larger, as can be seen from the right panel of Table 5. Although the W_c test still does not reject the condition (2.43), the H_{23} Hausman tests reject Assumption 2 for this specification. Hence, the validity of both IV3 and OLS inference is questionable. In unreported results using liquid liabilities or bank assets as alternative measures[26] for financial development, we tend to find qualitatively similar results as in Table 5. In general, however, the differences between IV2 and IV3/OLS are larger and specification tests are more in favor of IV2.

<Table 6 about here>

In Table 6 we report additional estimates using productivity growth as an alternative dependent variable. We continue to report the "empty" and "full" specifications and use private credit again as measure of financial development. The results in Table 6 show that, irrespective of the set of control variables, one can rely on OLS regarding inference for β_{xw}. At the same time IV1 standard errors and CD and KP statistics show again the failure of this estimator to identify the coefficients of the structural relation between financial development and productivity growth.

This additional empirical evidence confirms the conclusion of AHM that the interplay between financial development and initial conditions impacts future growth. For both dependent variables, point estimates for the interaction term are quite similar between the "empty" and "full" specifications. In general, the IV2 estimates provide the largest point estimate, while the OLS and IV3 results suggest a significant but smaller impact. Furthermore, our specification tests are able to discriminate between IV2, IV3 and OLS. For some specifications IV3/OLS are rejected, but quite often the empirical evidence tends to support OLS (and/or IV3) in lieu of IV2.

5. Concluding remarks

In this study we have analyzed OLS and IV inference for regression models including interactions between endogenous regressors and exogenous covariates. We show that endogeneity bias is reduced for the OLS estimator as far as the interaction term is concerned. Under

[26] Following AHM we omit the commercial-central bank measure from the analysis, because this variable is a priori considered to be an inferior measure of financial development.

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 may still be 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.

Regarding IV estimation we note that, under a fairly weak condition, the interaction term can be classified as exogenous. We therefore propose an IV estimator instrumenting only the endogenous regressor, but treating the interaction term as an exogenous covariate. In the more general case we 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, we provide specification tests to allow the researcher to discriminate between the various IV procedures and OLS, including procedures robust to weak instruments.

Monte Carlo experiments corroborate our theoretical findings. In particular, we show that OLS inference on the coefficient of the interaction term is valid for a wide variety of designs. Even if strong instruments are available, OLS might be still beneficial in terms of efficiency. Furthermore, including interacted instruments does lead to identification and is the preferred IV implementation, but has the usual caveats in the case of weak instruments. A further issue is that Hausman specification tests proposed to discriminate between OLS and IV have low power in the case of weak instruments. But note that precisely in this case our simulation results also show that the finite sample performance of OLS is not much worse than that of IV estimators.

Finally, we 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

Assumption 2 and simultaneity Assuming that the structural equations for y_i and x_i are as in (2.1) and (2.11), the reduced form for x_i then becomes:

$$\begin{aligned}
x_i &= \delta_\iota + \delta_w w_i + \delta_y y_i + \varepsilon_i \\
&= \delta_\iota + \delta_w w_i + \delta_y \left(\beta_\iota + \beta_w w_i + \beta_x x_i + \beta_{xw} x_i w_i + u_i\right) + \varepsilon_i \\
&= \frac{\delta_\iota + \delta_y \beta_\iota + \left(\delta_w + \delta_y \beta_w\right) w_i + \delta_z' z_i + \delta_y u_i + \varepsilon_i}{1 - \delta_y \beta_x - \delta_y \beta_{xw} w_i}.
\end{aligned}$$

We therefore have:

$$x_i u_i = \frac{\delta_\iota + \delta_y \beta_\iota + \left(\delta_w + \delta_y \beta_w\right) w_i u_i + \delta_z' z_i u_i + \delta_y u_i^2 + u_i \varepsilon_i}{1 - \delta_y \beta_x - \delta_y \beta_{xw} w_i},$$

and

$$\begin{aligned}
E\left[x_i u_i | w_i\right] &= \frac{\delta_\iota + \delta_y \beta_\iota + \left(\delta_w + \delta_y \beta_w\right) w_i E\left[u_i | w_i\right] + \delta_z' E\left[z_i u_i | w_i\right] + \delta_y E\left[u_i^2 | w_i\right] + E\left[u_i \varepsilon_i | w_i\right]}{1 - \delta_y \beta_x - \delta_y \beta_{xw} w_i} \\
&= \frac{\delta_\iota + \delta_y \beta_\iota + \delta_y E\left[u_i^2 | w_i\right] + E\left[u_i \varepsilon_i | w_i\right]}{1 - \delta_y \beta_x - \delta_y \beta_{xw} w_i},
\end{aligned}$$

because w_i and z_i are exogenous. Now even if $E\left[u_i^2 | w_i\right]$ and $E\left[u_i \varepsilon_i | w_i\right]$ do not depend on w_i it is still the case that $E\left[x_i u_i | w_i\right]$ depends on w_i via the denominator. This non-linear dependency only disappears when either $\delta_y = 0$ (no simultaneity) or $\beta_{xw} = 0$ (no 'endogenous' interaction), which are both trivial cases.

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

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

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

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

Note that the last element of Σ_{Xu} is zero because of Assumption 2.

Furthermore, we have that

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

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:

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

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

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

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

and the inconsistency is equal to

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

which is equal to expression (2.14).

Proof of Proposition 2. Assumption 3 implies that Σ_{XX} further simplifies to:

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

which results in:

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

Using properties of the multivariate normal distribution, we also have:

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

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

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

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

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

where $\rho_{xu}^2 = \frac{\sigma_{xu}^2}{\sigma_x^2\sigma_u^2}$. Regarding

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

we now have that

$$\varepsilon_i|X_i \sim i.i.n.\left(0,\sigma_\varepsilon^2\right), \qquad \sigma_\varepsilon^2 = \sigma_u^2\left(1-\rho_{xu}^2\right).$$

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

$$\begin{aligned} V &= \sigma_\varepsilon^2\left(\operatorname{plim}\frac{1}{n}X'X\right)^{-1} \\ &= \sigma_u^2\left(1-\rho_{xu}^2\right)\Sigma_{XX}^{-1}. \end{aligned}$$

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) instruments without affecting again the IV estimator. Summarizing, without loss of generalization we can asume:

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

with $E\left[z_i^2\right]$ 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 (2.35) we get the following:

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

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

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

Now assuming condition (2.40) we get:

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

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

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:

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

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:

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

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

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

Extension of Wald test checking OLS consistency. In practice x and w may have nonzero means, but one can rescale the data by taking deviations from sample mean. These sample means are estimators, and will affect the asymptotic variance of the moment conditions in (2.46). Therefore, the Wald test has to be adapted slightly to capture the effect of centering the data.

Define $\theta^* = (\mu_w, \mu_x, \theta)'$ with μ_w and μ_x the expectation of of w and x respectively. Centering the sample moment equations leads to:

$$\frac{1}{n}\sum_{i=1}^{n} m^*(w_i, x_i, \hat{\theta}^*) = \frac{1}{n}\sum_{i=1}^{n}\left(\begin{bmatrix} x_i \\ w_i \\ (x_i - \bar{x})(w_i - \bar{w}) \\ (x_i - \bar{x})(w_i - \bar{w})^2 \\ (w_i - \bar{w})^2 \\ (x_i - \bar{x})^2 (w_i - \bar{w}) \end{bmatrix} - \begin{bmatrix} \hat{\mu}_x \\ \hat{\mu}_w \\ \hat{\theta}_1 \\ \hat{\theta}_2 \\ \hat{\theta}_3 \\ \hat{\theta}_4 \end{bmatrix}\right) = 0,$$

where we have added two moments to estimate the means of x and w. Standard asymptotic theory for method of moments estimators shows that the limiting distribution of $\hat{\theta}^*$ is equal to:

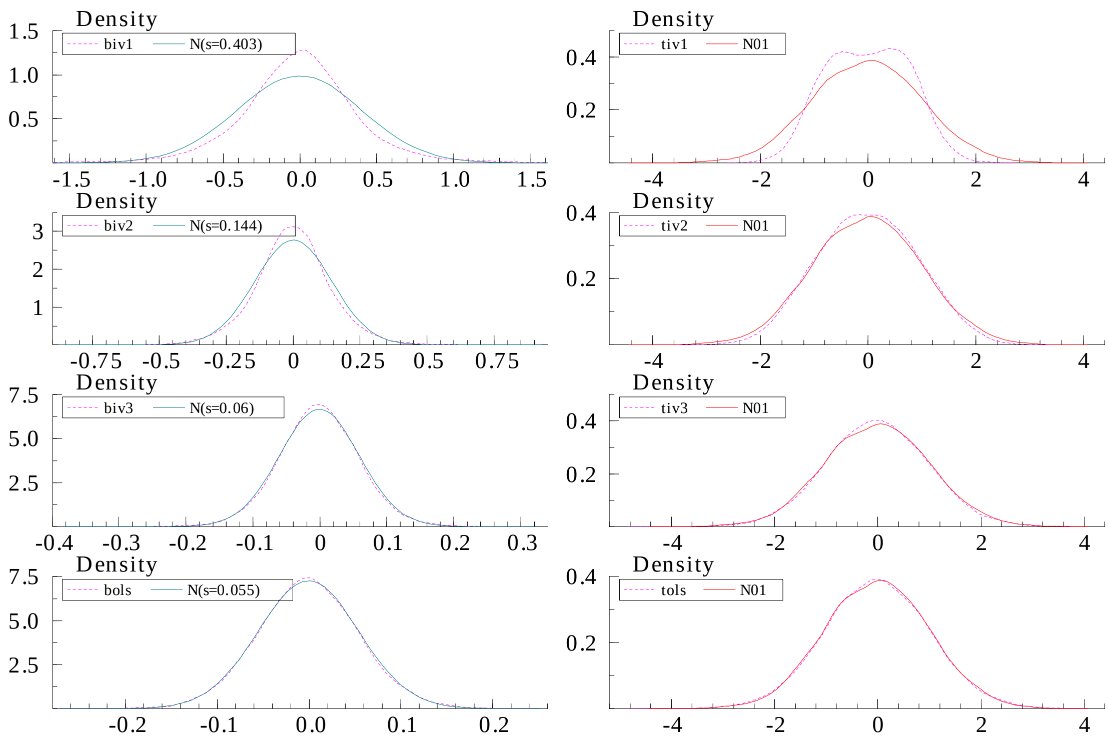
$$\sqrt{n}\left(\hat{\theta}^* - \theta^*\right) \xrightarrow{d} \mathcal{N}\left(0, G_0^{-1} S_0 \left(G_0'\right)^{-1}\right),$$

with

$$\begin{aligned} G_0 &= \operatorname{plim} \frac{1}{n}\sum_{i=1}^{n} \left.\frac{\partial m_i^*}{\partial \theta^{*\prime}}\right|_{\theta_0^*} = -\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ var(w_i) & 2cov(x_i, w_i) & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 2cov(x_i, w_i) & var(x_i) & 0 & 0 & 0 & 1 \end{bmatrix}, \\ S_0 &= \operatorname{plim} \frac{1}{n}\sum_{i=1}^{n} m_i^* m_i^{*\prime}|_{\theta_0^*}. \end{aligned}$$

Taking the lower right 4×4 block from $G_0^{-1} S_0 \left(G_0'\right)^{-1}$ delivers the asymptotic variance of θ, i.e. C_0 in (2.46), in the case of centered data.

Figure 1: finite sample distributions of OLS and IV coefficient estimators of β_{xw} and corresponding t-statistics, strong instruments case



Note: Left panel are coefficient estimators, while right panel contains corresponding t-statistics; Solid lines are normal densities with a variance matching the empirical distribution (left panel) or standard normal densities (right panel)

Table 1: homoskedasticity and normality

		weak instruments				strong instruments			
		OLS	IV1	IV2	IV3	OLS	IV1	IV2	IV3
bias	β_ι	-0.002	-0.001	-0.000	-0.002	-0.001	-0.001	-0.000	-0.001
	β_w	-0.478	-0.232	-0.323	-0.333	-0.085	-0.006	-0.008	-0.011
	β_x	0.477	0.228	0.321	0.329	0.083	0.003	0.006	0.007
	β_{xw}	0.002	0.000	-0.000	0.002	0.001	0.000	0.000	0.001
sd	β_ι	0.101	0.426	0.168	0.108	0.104	0.276	0.110	0.106
	β_w	0.125	0.529	0.294	0.275	0.110	0.172	0.114	0.113
	β_x	0.087	0.490	0.277	0.260	0.041	0.067	0.046	0.046
	β_{xw}	0.055	0.432	0.147	0.060	0.038	0.264	0.047	0.039
rp ho	β_ι	4.81	0.55	3.02	4.07	4.89	0.82	4.80	5.00
	β_w	95.51	7.13	24.52	29.66	12.84	2.54	5.52	5.72
	β_x	99.96	7.49	26.51	31.67	52.81	2.31	5.47	5.73
	β_{xw}	5.23	0.33	2.61	4.37	5.52	0.27	5.19	5.64
rp hr	β_ι	4.50	0.80	2.41	3.61	4.42	0.99	4.51	4.59
	β_w	95.79	6.88	22.53	28.06	11.99	1.90	5.19	5.25
	β_x	99.96	7.41	24.18	30.04	50.25	1.76	5.08	5.26
	β_{xw}	5.68	0.22	1.55	4.42	5.48	0.19	4.81	5.37
CD			0.63	1.15	3.06		0.81	21.66	100.58
KP ho			1.41	12.73	23.68		5.09	100.00	100.00
KP hr			1.34	6.33	15.64		3.33	83.38	100.00
specification tests			H^s_{23}	H^w_{23}	W_c		H^s_{23}	H^w_{23}	W_c
ho			2.03	3.07			5.32	4.50	
hr			3.64	3.40	5.04		5.96	4.51	4.47

Notes: based on 10,000 MC replications; rp is actual rejection % of nominal 5% t-statistics; ho is homoskedasticity-only, hr means heteroskedasticity-robust; KP and various specification tests (H^s_{23}, H^w_{23} and W_c) are rp of nominal 5% tests, while CD are average values. H^s_{23} tests IV2 versus IV3 assuming strong IVs, H^w_{23} does the same allowing for weak IVs, and W_c verifies condition (2.43) for OLS consistency.

Table 2: heteroskedasticity

		weak instruments				strong instruments			
		OLS	IV1	IV2	IV3	OLS	IV1	IV2	IV3
bias	β_ι	-0.002	-0.001	-0.001	-0.002	-0.001	-0.000	-0.001	-0.001
	β_w	-0.478	-0.221	-0.365	-0.376	-0.086	-0.010	-0.013	-0.017
	β_x	0.475	0.210	0.360	0.370	0.078	-0.002	0.003	0.005
	β_{xw}	0.002	0.001	0.002	0.002	0.001	-0.001	0.001	0.001
sd	β_ι	0.088	0.323	0.170	0.096	0.095	0.266	0.096	0.097
	β_w	0.201	0.464	0.336	0.317	0.110	0.163	0.111	0.110
	β_x	0.142	0.460	0.312	0.297	0.064	0.088	0.073	0.072
	β_{xw}	0.087	0.346	0.182	0.096	0.062	0.267	0.077	0.066
rp ho	β_ι	3.88	0.51	6.01	3.63	3.57	0.60	2.63	3.41
	β_w	87.98	8.77	39.98	46.44	13.42	2.25	5.54	5.76
	β_x	97.89	8.51	41.73	48.34	50.16	12.27	21.92	21.93
	β_{xw}	31.97	1.34	17.72	28.50	25.20	0.80	23.28	27.36
rp hr	β_ι	3.97	1.04	3.07	3.46	4.08	1.02	4.10	4.12
	β_w	66.07	8.07	23.58	28.34	10.06	1.64	3.87	4.32
	β_x	86.90	8.04	24.99	29.83	24.81	3.08	5.40	5.59
	β_{xw}	7.59	1.08	4.63	6.62	7.06	0.71	6.67	7.46
CD			0.64	1.54	4.98		0.82	20.49	111.64
KP ho			1.49	30.90	62.33		5.25	100.00	100.00
KP hr			1.84	7.83	25.93		3.05	81.23	99.98
specification tests			H_{23}^{s}	H_{23}^{w}	W_c		H_{23}^{s}	H_{23}^{w}	W_c
ho			9.32	9.84			13.05	8.01	
hr			4.19	4.32	5.46		5.60	4.36	4.56

Note: see Table 1.

Table 3: nonlinear dependence

		k_z= 1				k_z= 5			
		OLS	IV1	IV2	IV3	OLS	IV1	IV2	IV3
bias	β_ι	0.092		-0.008	0.003	0.014	0.002	0.002	0.002
	β_w	-0.225		0.020	-0.003	-0.074	-0.005	-0.008	-0.011
	β_x	0.324		-0.027	0.003	0.088	0.003	0.007	0.010
	β_{xw}	-0.083		0.008	-0.002	-0.012	-0.003	-0.002	-0.002
sd	β_ι	0.113		0.573	0.132	0.104	0.138	0.106	0.106
	β_w	0.141		0.796	0.188	0.117	0.184	0.120	0.120
	β_x	0.096		0.807	0.187	0.043	0.055	0.047	0.047
	β_{xw}	0.059		0.328	0.078	0.036	0.122	0.039	0.037
rp ho	β_ι	16.68		4.71	5.29	5.03	3.58	4.98	5.18
	β_w	46.01		4.42	5.51	10.62	3.97	5.36	5.69
	β_x	97.95		4.87	6.69	55.51	4.33	5.97	6.27
	β_{xw}	44.49		4.38	6.64	6.95	3.09	5.11	5.61
rp hr	β_ι	13.60		3.77	4.35	4.66	3.56	4.56	4.53
	β_w	33.99		4.64	5.38	10.12	3.12	5.20	5.48
	β_x	88.22		4.89	6.34	51.90	3.48	5.46	5.78
	β_{xw}	33.36		3.88	4.65	6.70	1.81	5.06	5.32
CD				24.81	64.39		6.56	44.14	97.40
KP ho				98.09	99.19		78.69	100.00	100.00
KP hr				70.81	70.91		31.07	99.91	100.00
specification tests			H^s_{23}	H^w_{23}	W_c		H^s_{23}	H^w_{23}	W_c
ho			6.61	8.76			8.03	4.75	
hr			3.29	2.27	55.85		5.13	4.05	19.00

Note: see Table 1.

Table 4: omitted variables bias

		weak instruments				strong instruments			
		OLS	IV1	IV2	IV3	OLS	IV1	IV2	IV3
bias	β_ι	-0.083	-0.081	-0.062	-0.083	-0.037	-0.038	-0.007	-0.038
	β_w	-0.484	-0.276	-0.359	-0.369	-0.107	-0.005	-0.010	-0.015
	β_x	0.482	0.272	0.357	0.365	0.105	0.004	0.009	0.010
	β_{xw}	0.085	0.083	0.063	0.085	0.038	0.040	0.008	0.039
sd	β_ι	0.096	0.407	0.176	0.104	0.102	0.299	0.115	0.105
	β_w	0.146	0.518	0.333	0.314	0.130	0.183	0.137	0.135
	β_x	0.088	0.497	0.306	0.289	0.046	0.075	0.053	0.052
	β_{xw}	0.071	0.408	0.169	0.075	0.049	0.289	0.062	0.051
rp ho	β_ι	12.37	0.87	6.48	10.33	6.02	0.77	5.79	5.87
	β_w	95.56	8.65	28.52	34.40	21.46	4.85	10.40	10.34
	β_x	99.94	8.29	30.59	36.47	64.79	2.55	6.01	6.44
	β_{xw}	42.02	1.12	11.51	36.67	23.00	0.69	10.16	22.84
rp hr	β_ι	12.84	1.08	4.26	10.15	5.98	1.29	4.72	5.88
	β_w	90.67	7.48	21.32	26.19	13.83	2.64	5.39	5.74
	β_x	99.89	7.90	24.24	29.86	61.77	2.19	5.40	5.55
	β_{xw}	26.31	0.91	5.84	22.66	13.98	0.65	5.93	13.01
CD			0.61	1.06	2.82		0.82	16.64	76.09
KP ho			1.17	9.13	18.06		5.17	99.99	100.00
KP hr			1.15	4.78	12.03		3.34	78.99	100.00
specification tests			H_{23}^s	H_{23}^w	W_c		H_{23}^s	H_{23}^w	W_c
ho			4.35	5.62			19.56	9.69	
hr			3.63	3.71	4.90		10.98	5.33	4.70

Note: see Table 1.

Table 5: empirical results for GDP growth and private credit

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Empty				Full			
	OLS	IV1	IV2	IV3	OLS	IV1	IV2	IV3
β_w	1.299	-4.294	1.507	1.390	0.191	2.739	1.131	0.719
	(0.335)	(12.298)	(0.480)	(0.458)	(0.780)	(2.783)	(0.758)	(0.691)
β_x	-0.012	0.164	-0.015	-0.015	-0.006	-0.056	-0.016	-0.025
	(0.009)	(0.379)	(0.016)	(0.015)	(0.010)	(0.068)	(0.020)	(0.019)
β_{xw}	-0.048	0.106	-0.061	-0.050	-0.037	-0.098	-0.063	-0.047
	(0.009)	(0.354)	(0.011)	(0.009)	(0.013)	(0.060)	(0.014)	(0.011)
CD		0.07	6.28	6.29		0.62	3.63	3.92
KP		0.23	26.57	26.87		2.24	19.98	21.30
		(0.89)	(0.00)	(0.00)		(0.33)	(0.00)	(0.00)
		H_{23}^s	H_{23}^w	W_c		H_{23}^s	H_{23}^w	W_c
		2.62	0.12	-1.83		4.52	5.10	-0.17
		(0.11)	(0.73)	(0.07)		(0.03)	(0.02)	(0.87)

Note: numbers in parentheses are standard errors (below estimates) or p-values (below test statistics); IV and OLS standard errors are homoskedasticity-only and heteroskedasticity-robust respectively; Empty and Full refer to the set of control variables; Financial development measure (x) is private credit.

Table 6: empirical results for productivity growth and private credit

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Empty				Full			
	OLS	IV1	IV2	IV3	OLS	IV1	IV2	IV3
β_w	0.593	-1.662	0.844	0.856	-0.574	0.110	0.496	0.308
	(0.328)	(1.871)	(0.510)	(0.401)	(0.780)	(7.258)	(0.866)	(0.762)
β_x	0.052	0.018	0.064	0.065	0.044	0.060	0.067	0.067
	(0.011)	(0.035)	(0.011)	(0.011)	(0.018)	(0.106)	(0.015)	(0.015)
β_{xw}	-0.041	0.016	-0.051	-0.052	-0.034	-0.049	-0.057	-0.054
	(0.009)	(0.049)	(0.012)	(0.011)	(0.015)	(0.152)	(0.017)	(0.015)
CD		0.97	6.23	7.15		0.06	3.94	4.94
KP		3.01	25.74	28.19		0.22	20.93	24.42
		(0.22)	(0.00)	(0.00)		(0.90)	(0.00)	(0.00)
		H_{23}^s	H_{23}^w	W_c		H_{23}^s	H_{23}^w	W_c
		0.00	0.14	-1.71		0.21	0.02	-0.51
		(0.97)	(0.71)	(0.09)		(0.65)	(0.88)	(0.61)

Note: see Table 5.

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.

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.

Durbin, J. (1954). Errors in variables. *Review of the International Statistical Institute* 22, 23-32.

Hahn, J., Ham, J.C., and H.R. Moon (2011). The Hausman test and weak instruments. *Journal of Econometrics* 160, 289-299.

Hausman, J.A. (1978). Specification tests in Econometrics. *Econometrica* 46, 1251-1271.

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 (2000). 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.Y. and I. Murtazashvili (2014). Exogenous treatment and endogenous factors: Vanishing of omitted variable bias on the interaction term. Forthcoming in *Journal of Econometric Methods*.

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.