

Many Instruments: Implementation in STATA

Stanislav Anatolyev
CERGE-EI
Prague, Czech Republic
stanislav.anatolyev@cerge-ei.cz

Alena Skolkova
CERGE-EI
Prague, Czech Republic
alena.skolkova@cerge-ei.cz

Abstract. During the last decades there have been developed econometric tools of handling instrumental variable regressions characterized by numerosity of instruments. We introduce a STATA command, `mivreg`, that implements consistent estimation and testing in linear IV regressions with many (possibly weak) instruments. The command `mivreg` covers both homoskedastic and heteroskedastic environments, estimators that are both non-robust and robust to error non-normality and projection matrix limit, both parameter tests and specification tests, both with and without correction for existence of moments. We also run a small simulation experiment using `mivreg` and illustrate how `mivreg` works with real data.

Keywords: `st0001`, `mivreg`, instrumental variables regression, many instruments, limited information maximum likelihood, Fuller correction, robust standard errors, specification testing

1 Introduction

Instrumental variables (IV) estimation and inference have long been a distinctive method in applied microeconomic analysis and have often spurred advances in econometric theory. The IV methods were designed to address endogeneity bias from OLS in estimating a causal/treatment effect in structural models (such as an effect of smoking on health, returns to education, or demand elasticity), see Angrist and Krueger (2001). At the dawn of the 21st century, both theory and practice were extended to accommodate such complications as weak instruments, numerous instruments, and combinations thereof. It was established that the empiricist's workhorse, the two-stage least-squares (2SLS) estimator, fails to deliver consistent estimates and results in invalid inference when such complications arise, and alternative approaches to estimation and inference were proposed. The quick progress in econometric theory did not, however, carry over to empirical practice as fast.

The seminal article by Bekker (1994) proposed an alternative asymptotic approximation for linear normal homoskedastic IV regressions with many instrumental variables, together with consistent estimation and construction of valid standard errors within the new paradigm of dimension asymptotics. Since then, there has been a significant progress in the theory of estimation and testing in IV regressions with many, possibly weak, instruments. Many new or modified versions of old estimators and tests have been proposed, including, among others, limited information maximum likelihood (LIML), bias-corrected 2SLS, several versions of jackknife IV estimators, and so on. In an important article, Hansen et al. (2008) proposed extensions of estimation and inference

methods based on LIML, when, in particular, the structural and first stage errors are not necessarily normal and when the instruments may be weak as a group. More recently, Hausman et al. (2012) showed that the leading ‘homoskedastic’ estimators fail to deliver consistency in heteroskedastic models, and proposed their ‘heteroskedastic’ modifications. Specification testing tools were developed in Anatolyev and Gospodinov (2011) and Lee and Okui (2012) for the homoskedastic case and in Chao et al. (2014) for the heteroskedastic case.

The state-of-the-art theoretical literature has converged to suggesting estimation based on LIML and its Fuller (1977)-type correction that remedies the problem of non-existence of moments. Parameter inference is based on consistent estimation of up to four terms in the asymptotic variance, while specification testing is based on asymptotically normal (or asymptotically equivalent possibly adjusted chi-squared) distribution of the overidentifying test statistic. The literature has shown that all these tools are robust to weakness of the instruments as a group (though weakness of a lesser degree than that would jeopardize identification). We describe these tools in brief in the following sections; see the recent survey Anatolyev (2019) for more technical details as well as the history of theoretical developments and suggestions of empirical strategies.

Despite the theoretical advances, practitioners rarely use appropriate tools because of their non-availability in popular econometric packages, STATA in particular. The present contribution aims at filling this void. We introduce a STATA command, `mivreg`, that implements consistent estimation and testing in linear IV regressions with many, possibly weak, instruments. This command covers both homoskedastic and heteroskedastic environments, estimators that are both non-robust and robust to error non-normality and projection matrix limit, both parameter tests and specification tests. Even though, as noted above, a number of other consistent estimators have been proposed, we build up `mivreg` around the leading LIML estimator and its Fuller (1977) correction as suggested by the state-of-the-art literature.

In Section 2, we set out the model and introduce necessary notation. In Sections 3 and 4, we describe estimation and testing tools pertaining to the homoskedastic and heteroskedastic models, respectively. In Section 5, we present the new command, `mivreg`. In Section 5, we illustrate how `mivreg` works in simulations and compare it with the classical command `ivregress` in Section 6. Finally, in Section 7, we illustrate how `mivreg` works with real data.

2 Model

The structural equation is

$$y_i = x_i' \beta_0 + e_i,$$

where β_0 is $k \times 1$ vector of structural coefficients of interest, or in matrix notation, $Y = X\beta_0 + e$, where $Y = (y_1, \dots, y_n)'$ is $n \times 1$, $X = (x_1, \dots, x_n)'$ is $n \times k$, and $e = (e_1, \dots, e_n)'$ is $n \times 1$. The first stage equation is

$$x_i = z_i' \Gamma + u_i,$$

where z_i is $\ell \times 1$ vector of instruments and Γ is $\ell \times k$ matrix of first stage coefficients, or in matrix notation, $X = Z\Gamma + U$, where $U = (u_1, \dots, u_n)'$ is $n \times k$. We assume that the rank of instrument matrix $Z = (z_1, \dots, z_n)'$ equals its column dimension ℓ . The structural and first stage errors follow

$$\begin{pmatrix} e_i \\ u_i \end{pmatrix} | z_i \sim \mathcal{D} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_i^2 & \Psi_i' \\ \Psi_i & \Omega_i \end{pmatrix} \right),$$

for some distribution \mathcal{D} , normal \mathcal{N} being a possibility. Under conditional homoskedasticity, $\sigma_i^2 = \sigma^2$, $\Psi_i = \Psi$ and $\Omega_i = \Omega$ for all $i = 1, \dots, n$.

Introduce the projection matrices associated with the instruments

$$P = Z(Z'Z)^{-1}Z', \quad M = I_n - P.$$

The $(i, j)^{th}$ element of P is denoted P_{ij} . Let us also denote by D the diagonal matrix with diagonal elements of P on the main diagonal: $D = \text{diag} \{P_{ii}\}_{i=1}^n$. By $\overline{P_{ii}^2}$ we denote an average of diagonal elements of P squared: $\overline{P_{ii}^2} = n^{-1} \text{tr}(D^2)$.

3 Homoskedastic case

In the conditionally homoskedastic case, correct parameter estimation and inference was developed in Bekker (1994) and Hansen et al. (2008). Specification testing was dealt with in Anatolyev and Gospodinov (2011) and Lee and Okui (2012).

3.1 Point estimation

Under many instruments, 2SLS estimation is inconsistent. The leading consistent estimator is the limited information maximum likelihood (LIML) estimator

$$\hat{\beta}_{LIML} = \arg \min_{\beta} \frac{(Y - X\beta)' P (Y - X\beta)}{(Y - X\beta)' (Y - X\beta)}.$$

Numerically, instead of the above optimization problem, it can be found via the eigenvalue problem:

$$\hat{\beta}_{LIML} = \bar{H}^{-1} X' \hat{P} Y,$$

where

$$\bar{H} = X' \hat{P} X,$$

and $\hat{P} = P - \bar{\alpha} I_n$, and $\bar{\alpha}$ is the smallest eigenvalue of the matrix $(\hat{X}' \hat{X})^{-1} \hat{X}' \hat{P} \hat{X}$, where $\hat{X} = (Y, X)$.

The LIML estimator has a disadvantage that even its low order moments do not exist. A simple Fuller (1977) adjustment solves the moment problem:

$$\tilde{\alpha} = \frac{\bar{\alpha} - (1 - \bar{\alpha}) \varsigma/n}{1 - (1 - \bar{\alpha}) \varsigma/n}. \quad (1)$$

This adjustment leads to the FULL estimator, where $\bar{\alpha}$ is replaced by $\tilde{\alpha}$ everywhere. It is usually advised to use the value $\varsigma = 1$ in practice.

Denote the vector of LIML or FULL residuals by \hat{e} , then

$$\hat{\sigma}^2 = \frac{\hat{e}'\hat{e}}{n-k}$$

is the residual variance.

3.2 Variance estimation

Under error normality and/or asymptotically constant diagonal of P , the asymptotic variance is estimated by

$$\bar{V} = n\bar{H}^{-1}\bar{\Sigma}_0\bar{H}^{-1},$$

where

$$\bar{\Sigma}_0 = \hat{\sigma}^2 \left((1 - \tilde{\alpha})^2 \bar{X}'P\bar{X} + \tilde{\alpha}^2 \bar{X}'(I_n - P)\bar{X} \right),$$

and

$$\bar{X} = X - \hat{e} \frac{\hat{e}'X}{\hat{e}'\hat{e}}$$

(Bekker 1994, Hansen et al. 2008).

Under error non-normality and asymptotically variable diagonal of P , the asymptotic variance is estimated by

$$\bar{V}_R = n\bar{H}^{-1}(\bar{\Sigma}_0 + \bar{\Sigma}_A + \bar{\Sigma}'_A + \bar{\Sigma}_B)\bar{H}^{-1},$$

where the subscript R stands for ‘robust’, and in addition

$$\bar{\Sigma}_A = \left(\sum_{i=1}^n \left(P_{ii} - \frac{\ell}{n} \right) (PX)_i \right) \left(\frac{1}{n} \sum_{i=1}^n \hat{e}_i^2 (M\bar{X})_i \right)'$$

and

$$\bar{\Sigma}_B = \frac{\overline{P_{ii}^2} - (\ell/n)^2}{1 - 2\ell/n + \overline{P_{ii}^2}} \sum_{i=1}^n (\hat{e}_i^2 - \hat{\sigma}^2) (M\bar{X})_i (M\bar{X})'_i$$

(Hansen et al. 2008).

The variance estimates \bar{V} and \bar{V}_R are a basis of parameter inference. For example, the standard error for j^{th} parameter can be computed as $\sqrt{\bar{V}_{jj}/n}$.

3.3 Specification testing

Consider the conventional J statistic

$$J = \frac{\hat{e}'P\hat{e}}{\hat{\sigma}^2} = (n-k)\tilde{\alpha},$$

and the bias-corrected J statistic

$$J_R = J - \frac{\ell}{n} \frac{\hat{e}'\hat{e}}{\hat{\sigma}^2} = (n-k) \left(\bar{\alpha} - \frac{\ell}{n} \right),$$

where the subscript R stands for ‘robust’.

Under error normality and/or asymptotically constant diagonal of P , the Anatolyev and Gospodinov (2011) test prescribes rejecting correct model specification at significance level ϕ when the value of J exceeds $q_{\phi^*}^{\chi^2(\ell-k)}$, the $(1 - \phi^*)$ -quantile of the chi-squared with $\ell - k$ degrees of freedom, where

$$\phi^* = \Phi \left(\sqrt{1 - \frac{\ell}{n}} \cdot \Phi^{-1}(\phi) \right).$$

Under error non-normality and asymptotically variable diagonal of P , the Lee and Okui (2012) test prescribes rejecting correct model specification at significance level ϕ when the value of

$$\frac{J_R}{\sqrt{n\hat{V}^J}}$$

exceeds $q_{\phi}^{\mathcal{N}(0,1)}$, the $(1 - \phi)$ -quantile of the standard normal. Here,

$$\hat{V}^J = 2\frac{\ell}{n} \left(1 - \frac{\ell}{n} \right) + \left(\overline{P_{ii}^2} - \left(\frac{\ell}{n} \right)^2 \right) \left(\frac{\overline{\hat{e}_i^4}}{\hat{\sigma}^4} - 3 \right).$$

4 Heteroskedastic case

In the conditionally heteroskedastic case, correct parameter estimation and inference were developed in Hausman et al. (2012). Specification testing was dealt with in Chao et al. (2014).

4.1 Point estimation

The HLIM (‘heteroskedastic LIML’) estimator is

$$\hat{\beta}_{HLIM} = \arg \min_{\beta} \frac{(Y - X\beta)'(P - D)(Y - X\beta)}{(Y - X\beta)'(Y - X\beta)}$$

Numerically, it can be found via the eigenvalue problem:

$$\hat{\beta}_{HLIM} = \bar{H}^{-1} X' \hat{P} Y,$$

where

$$\bar{H} = X' \hat{P} X,$$

and $\hat{P} = P - D - \bar{\alpha} I_n$, and $\bar{\alpha}$ is the smallest eigenvalue of the matrix $(\hat{X}' \hat{X})^{-1} \hat{X}' (P - D) \hat{X}$, where $\hat{X} = (Y, X)$. Similarly to FULL, the Fuller (1977) adjustment (1) leads to HFUL ('heteroskedastic FULL') estimation.

Denote the vector of HLIM or HFUL residuals by \hat{e} , then

$$\hat{\sigma}^2 = \frac{\hat{e}' \hat{e}}{n - k}$$

is the residual variance.

4.2 Asymptotic variance estimation

Hausman et al. (2012) provide a valid and robust variance estimator for the HLIM estimator:

$$\bar{V} = n \bar{H}^{-1} \bar{\Sigma} \bar{H}^{-1},$$

where

$$\bar{\Sigma} = \sum_{i=1}^n ((P\bar{X})_i (P\bar{X})'_i - P_{ii} \bar{X}_i (P\bar{X})'_i - P_{ii} (P\bar{X})_i \bar{X}'_i) \hat{e}_i^2 + \sum_{i=1}^n \sum_{j=1}^n P_{ij}^2 \bar{X}_i \bar{X}'_j \hat{e}_i \hat{e}_j, \quad (2)$$

where

$$\bar{X} = X - \hat{e} \frac{\hat{e}' X}{\hat{e}' \hat{e}}.$$

The variance estimate \bar{V} is a basis of parameter inference. For example, the standard error for j^{th} parameter can be computed as $\sqrt{\bar{V}_{jj}/n}$.

4.3 Specification testing

Chao et al. (2014) generalize the specification J test for the heteroskedastic case. Their statistic is based on the jackknife modification of J statistic's quadratic form:

$$J = \frac{\hat{e}' (P - D) \hat{e}}{\sqrt{\hat{V}^J}} + \ell,$$

where

$$\hat{V}^J = \frac{1}{\ell} \sum_{i \neq j} \hat{e}_i^2 P_{ij}^2 \hat{e}_j^2 = \frac{1}{\ell} \left(\sum_{i=1}^n \sum_{j=1}^n \hat{e}_i^2 P_{ij}^2 \hat{e}_j^2 - \sum_{i=1}^n P_{ii}^2 \hat{e}_i^4 \right) \quad (3)$$

is an estimate of the variance of the modified quadratic form.

The test is one-sided, and the decision rule is reject the null of instrument validity if the value of J exceeds $q_{\phi}^{\chi^2(\ell-k)}$, the $(1 - \phi)$ -quantile of the $\chi^2(\ell - k)$ distribution.

5 Command `mivreg`

5.1 Functionality

The command `mivreg` implements estimation, inference on individual parameters and specification testing under many, possibly weak, instruments. The default ‘hom’ (for ‘homoskedastic’) option is based on the LIML or FULL estimators, the ‘het’ (for ‘heteroskedastic’) option is based on the HLIM or HFUL estimators. Within the ‘hom’ version, the ‘robust’ option leads to the Hansen–Hausman–Newey variance estimator and Lee–Okui specification test, while the default non-robust variation computes the Bekker variance estimator and Anatolyev–Gospodinov specification test. The ‘hetero’ version implements the Hausman–Newey–Woutersen–Chao–Swanson variance estimator and Chao–Hausman–Newey–Swanson–Woutersen specification test. By default, the estimators used are LIML or HLIM; the ‘fuller’ option makes the Fuller correction with parameter $\varsigma = 1$, and so the FULL or HFUL estimators are used instead.

5.2 Syntax

```
mivreg devar [indepvars] (varlist1 = varlist2) [if] [in] [, hom het robust
    fuller level(#)]
```

by, rolling, statsby and xi are allowed; see [U] 11.1.10 Prefix commands.

5.3 Description

The command `mivreg` performs estimation, inference on individual parameters and specification testing under many possibly weak instruments. The dependent variable *devar* is modeled as a linear function of *indepvars* and *varlist1*, using *varlist2* (along with *indepvars*) as instruments for *varlist1*.

5.4 Options

hom uses the LIML (default) or FULL (in combination with **full** option) estimator.

het uses the HLIM (default) or HFUL (in combination with **full** option) estimator.

robust leads, under **hom** option, to the Hansen–Hausman–Newey variance estimator and the Lee–Okui specification test, while the default non-robust variation computes the Bekker variance estimator and the Anatolyev–Gospodinov specification test; under **het** option, to the Hausman–Newey–Woutersen–Chao–Swanson variance estimator and the Chao–Hausman–Newey–Swanson–Woutersen specification test.

fuller makes the Fuller correction with parameter $\varsigma = 1$, which leads to the FULL (in combination with **hom** option) or HFUL (in combination with **het** option) estimator.

`level(#)` sets the confidence level; the default is `level(95)`.

5.5 Saved results

`mivreg` saves the following in `e()`:

Scalars

<code>e(N)</code>	number of observations	<code>e(F1)</code>	first-stage F statistic
<code>e(rmse)</code>	root mean squared error	<code>e(df_mF1)</code>	first-stage model degrees of freedom
<code>e(F)</code>	model F statistic	<code>e(df_rF1)</code>	first-stage residual degrees of freedom
<code>e(df_m)</code>	model degrees of freedom	<code>e(r2_1)</code>	first-stage R^2
<code>e(df_r)</code>	residual degrees of freedom	<code>e(jval)</code>	model J statistic
<code>e(r2)</code>	R^2	<code>e(jpv)</code>	J -test p-value
<code>e(r2.a)</code>	adjusted R^2		

Macros

<code>e(model)</code>	<code>hom</code> or <code>het</code>	<code>e(instd)</code>	instrumented variables
<code>e(title)</code>	title in estimation output	<code>e(insts)</code>	instruments
<code>e(depvar)</code>	name of dependent variable	<code>e(properties) b V</code>	
<code>e(depvar)</code>	name of dependent variable	<code>e(properties) b V</code>	

Matrices

<code>e(b)</code>	coefficient vector	<code>e(V)</code>	variance–covariance matrix of the estimators
-------------------	--------------------	-------------------	--

Functions

<code>e(sample)</code>	marks estimation sample
------------------------	-------------------------

5.6 Computational notes

First, throughout we avoid storing $n \times n$ matrices like P and I_n in the memory. For example, we compute $\bar{H} = X'(P - \bar{\alpha}I_n)X$ as

$$\bar{H} = X'Z(Z'Z)^{-1}Z'X - \bar{\alpha}X'X.$$

Second, the last term in (2) can be alternatively computed without double summations over n observations (Hausman et al. 2012):

$$\sum_{p=1}^{\ell} \sum_{r=1}^{\ell} \left(\sum_{i=1}^n \tilde{Z}_{ip} \tilde{Z}_{ir} \bar{X}_i \hat{e}_i \right) \left(\sum_{j=1}^n Z_{jp} Z_{jr} \bar{X}_j \hat{e}_j \right)',$$

where $\tilde{Z} = Z(Z'Z)^{-1}$. Similarly, the full double summation in (3) can analogously be computed as

$$\sum_{p=1}^{\ell} \sum_{r=1}^{\ell} \left(\sum_{i=1}^n \tilde{Z}_{ip} \tilde{Z}_{ir} \hat{e}_i^2 \right) \left(\sum_{j=1}^n Z_{jp} Z_{jr} \hat{e}_j^2 \right).$$

6 Simulations

6.1 Artificial data

We demonstrate how `mivreg` works with two sets of artificial data. The artificial data are generated from the Monte-Carlo setup in Hausman et al. (2012). The estimated equation is

$$y = \beta_1 + \beta_2 x_2 + e,$$

and the first stage equation is

$$x_2 = \gamma z_1 + u_2,$$

where $z_1 \sim N(0, 1)$ and $u_2 \sim N(0, 1)$. The instrument vector is

$$z = (1, z_1, z_1^2, z_1^3, z_1^4, z_1 d_1, \dots, z_1 d_{\ell-5})',$$

where $d_j \in \{0, 1\}$ with $\Pr\{d_j = 1\} = \frac{1}{2}$ independent of z_1 . The structural disturbance is given by

$$e = 0.30u_2 + \sqrt{\frac{1 - 0.30^2}{\phi^2 + 0.86^4}} (\phi v_1 + 0.86v_2),$$

with $v_1 \sim N(0, 1)$ in the homoskedastic case and $v_1 \sim N(0, z_1^2)$ in the heteroskedastic case, and $v_2 \sim N(0, 0.86^2)$, both v_1 and v_2 being independent of u_2 . Samples of size $n = 400$ are generated, with $\ell = 30$ instruments, the instrument strength γ is chosen so that the concentration parameter equals $n\gamma^2 = 32$. The parameter ϕ is set at the value 0.8 which in the heteroskedastic case corresponds to $R^2 \approx 0.25$ in the skedastic regression. The true values of β_1 and β_2 are set at 1.

Note that the instrument vector is such that the diagonal of P is asymptotically heterogeneous (see Anatolyev and Yaskov 2017). In the homoskedastic case, simplifications due to error normality pertaining to variance estimation and specification testing (see subsections 3.2 and 3.3) are applicable.

6.2 Simulation results

In this section, we report output statistics resulting in simulations from using `mivreg` and compare it with that when the STATA command `ivregress` was used.¹ The reported results are obtained from 10,000 simulations.

First, we focus on point estimates. Table 1 collects percentiles of simulated distributions of 2SLS, LIML and GMM estimators produced by `ivregress`, and LIML, FULL, HLIM and HFUL estimators produced by `mivreg`. Naturally, the LIML rows coincide.

The 2SLS and GMM estimators (whose results are very similar) are always rightward biased, as expected. In the homoskedastic case, all the other estimators deliver unbiased estimation. The LIML estimator is a bit more concentrated towards the center than

1. For example, to compute 2SLS-related statistics, `ivregress 2sls y one (x = z*)`, `nocons robust` was used.

Table 1. Percentiles of simulated distribution of various estimators.

Estimator	Homoskedastic case					Heteroskedastic case				
	5%	25%	50%	75%	95%	5%	25%	50%	75%	95%
command <code>ivregress</code>										
2SLS	0.93	1.06	1.14	1.23	1.35	0.85	1.02	1.14	1.26	1.43
GMM	0.91	1.05	1.14	1.23	1.37	0.85	1.02	1.14	1.26	1.42
LIML	0.47	0.83	1.00	1.16	1.42	-4.08	-0.27	0.49	1.07	4.48
command <code>mivreg</code>										
LIML	0.47	0.83	1.00	1.16	1.42	-4.08	-0.27	0.49	1.07	4.48
FULL	0.52	0.84	1.01	1.17	1.41	-1.14	-0.03	0.56	1.09	2.77
HLIM	0.43	0.82	1.00	1.17	1.43	0.15	0.76	1.01	1.22	1.62
HFUL	0.52	0.84	1.01	1.17	1.43	0.30	0.79	1.02	1.22	1.60

Note: The true value of the parameter is unity.

HLIM, which reflects higher efficiency of the former. The Fuller versions are more concentrated away from the tails, which reflects their resistance to outliers. In the heteroskedastic case, LIML and FULL have severe negative biases, which reflects their inconsistency. Their ‘heteroskedastic’ versions, HLIM and HFUL, are both median unbiased. While the HLIM estimator is susceptible to outliers, especially in the left tail, its Fuller version, HFUL, exhibits much tighter and more symmetric distribution.

Table 2 contains actual rejection rates corresponding to the 5% nominal rate for the two sided t-test of the null $H_0 : \beta_2 = 1$ marked as $t_{\beta_2=1}$, the Wald test of the null $H_0 : \beta_1 = \beta_2 = 1$ marked as $W_{\beta_1=\beta_2=1}$, and the specification test marked as $J_{E[z\epsilon]=0}$. The 2SLS and LIML tests produced by `ivregress` come in two forms: non-robust and robust to heteroskedasticity. In the specification tests (which are available only for efficient estimators), the Basmann (1957) variance estimator is used. The test statistics produced by `mivreg` use the following estimators and robustness regimes:² non-robust LIML, non-robust FULL, robust LIML, robust FULL, HLIM, and HFUL.

As expected, severe size distortions are exhibited by conventional parameter tests based on 2SLS, GMM and LIML.³ In the homoskedastic case, all the `mivreg` tests exhibit similar behavior, with much smaller distortions, though the ‘heteroskedastic’ versions seem to be more reliable. In the heteroskedastic case, the latter are the only valid ones theoretically, and do deliver rejection rates close to nominal. The Fuller correction does

2. Note again the different use of the term ‘robust’: the classical tests produced by `ivregress` may be robust to heteroskedasticity; of course, they are not robust to instrument numerosity. The tests produced by `mivreg` may or may not be robust, within natural robustness to many possibly weak instruments, to error non-normality and asymptotically variable diagonal of the projection matrix.

3. The conventional specification tests do not exhibit too much of distortions in this particular design; however, in general they may well do; see Anatolyev and Gospodinov (2011).

Table 2. Actual rejection rates for parameter and specification tests

Estimator	Homoskedastic case			Heteroskedastic case		
	$t_{\beta_2=1}$	$W_{\beta_1=\beta_2=1}$	$J_{E[ze]=0}$	$t_{\beta_2=1}$	$W_{\beta_1=\beta_2=1}$	$J_{E[ze]=0}$
command <code>ivregress</code>						
non-robust 2SLS	22.0%	17.7%	6.2%			
robust 2SLS				14.9%	13.1%	–
GMM	33.9%	31.8%	2.5%	26.8%	24.4%	2.1%
non-robust LIML	12.0%	9.6%	3.0%			
robust LIML				1.6%	1.3%	–
command <code>mivreg</code>						
non-robust LIML	4.1%	4.3%	3.0%	9.4%	4.6%	60.1%
non-robust FULL	4.2%	4.5%	2.4%	9.3%	4.7%	56.8%
robust LIML	4.0%	4.3%	2.1%	9.2%	4.5%	54.2%
robust FULL	4.2%	4.5%	1.7%	9.2%	4.6%	50.9%
HLIM	4.7%	4.9%	2.8%	5.4%	4.9%	3.5%
HFUL	5.0%	5.2%	2.9%	5.7%	5.1%	3.4%

Note: The nominal significance level of all tests is 5%.

not significantly affect these rejection rates. The results of specification testing point at huge distortions if one relies on ‘homoskedastic’ specification tests when in fact the homoskedasticity assumption is violated. One must avoid using them in heteroskedastic environments as one is too much likely to receive a signal of instrument invalidity when in fact the instruments are valid.

7 Example with real data

We illustrate the use of `mivreg` using real data from a well-known application to the married female labor supply of Mroz (1987). The number of observations is 428.⁴

The left-side variable is working hours `hours`, the only endogenous right-side variable is log wages `lwage`; there are also 6 exogenous controls: `nwifeinc`, `educ`, `age`, `kidslt6`, `kidsge6`, and the constant `one`. The list of basic instruments includes, in addition to the 6 exogenous controls, 8 exogenous variables: `exper`, `expersq`, `fatheduc`, `motheduc`, `hushrs`, `husage`, `huseduc`, and `mtr`, resulting in 14 instruments in total. The basic instruments are pretty strong as a group: the first-stage F statistic equals 183.5. We also consider an extended set of instruments – the basic instruments plus all their cross-products (‘interactions’), the total numerosity amounting to 92. The use of the

4. The data can be found at <http://www.stata.com/data/jwooldridge/eacsap/mroz.dta>. We use only the records that correspond to women in labor force.

Table 3. Various estimates of wage coefficient for married female labor supply

Options	Estimator	Instruments	Estimate	(Standard error)
command <code>reg</code>				
<code>robust</code>	OLS	–	–17.4	(81.4)
command <code>ivregress</code>				
<code>robust</code>	2SLS	basic only	1179.1	(185.2)
<code>robust</code>	2SLS	extended	536.4	(101.5)
command <code>mivreg</code>				
<code>hom</code>	LIML	extended	1120.6	(195.3)
<code>hom robust fuller</code>	FULL	extended	1110.0	(197.2)
<code>het robust fuller</code>	HFUL	extended	1058.3	(170.5)

extended instrument set is meant to possibly enhance estimation efficiency by exploiting information in the instruments more actively. However, while the conventional tools are suitable for the basic set of instruments, the extended instrument set evidently requires handling via many-instrument asymptotics: the ratio of the number of instruments to the sample size is sizable: $\ell/n \approx 0.215$.

Table 3 presents various estimates for the slope coefficient of log wages: OLS, heteroskedasticity-robust 2SLS (employing the basic and extended instrument sets), as well as three many-instrument-robust estimators – LIML, FULL and HFUL (employing the extended instrument set) – whose STATA output will appear below.

Evidently, due to unaccounted endogeneity, OLS estimation from applying the `reg` command is inconsistent; the numerical value of the OLS estimate is even negative revealing a big endogeneity bias. The (more than twofold!) difference between the two 2SLS estimates points at invalidity of conventional tools and the `ivregress` command when instruments are many. The LIML, FULL and HFUL point estimates produced by the `mivreg` command are quite in line with the 2SLS estimate that uses only the basic instruments.⁵ There is a small difference between ‘homoskedastic’ LIML and FULL point estimates and the ‘heteroskedastic’ HFUL point estimate. Though not too big, this difference makes the HFUL estimate more trustworthy.⁶ The smaller standard error of HFUL compared to that of 2SLS may be interpreted as a gain in efficiency from using the extended instrument set.

The STATA outputs produced by the command `mivreg` to deliver the three many-instrument-robust estimators appear next.

5. Note also from the STATA outputs that all three corresponding specification tests produce very high p-values and agree on the model validity.

6. Mroz (1987) reports a similar 2SLS estimate using a short list of instruments (line 2 in his Table IV), but 2SLS estimates also get a lot smaller with longer lists of instruments (lines 3–6 in Table IV). Eventually, Mroz (1987) adopts smaller estimates than ones seeming correct from our experiments.

► **Example**

The STATA output for LIML estimation with option hom:

```
. mivreg hours nwifeinc educ age kidslt6 kidsge6 one (lwage = nwifeinc educ ///
> age kidslt6 kidsge6 one exper expersq fatheduc motheduc hushrs husage ///
> huseduc mtr *X*) if inlf==1 , hom
```

MIVREG estimation (HOM)

First-stage summary

F(86, 336) = 2.07
 Prob > F = 0.0000
 R-squared = 0.8471

Number of obs = 428
 F(7, 421) = 95.74
 Prob > F = 0.0000
 R-squared = -0.5157
 Adj R-squared = -0.5373
 Root MSE = 1.1e+03

LIML estimation

Bekker variance estimation

hours	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lwage	1120.595	195.3494	5.74	0.000	736.6134	1504.577
nwifeinc	-7.890468	5.261348	-1.50	0.134	-18.23225	2.451317
educ	-133.1851	31.79141	-4.19	0.000	-195.6748	-70.69543
age	-9.954741	7.918058	-1.26	0.209	-25.51859	5.609111
kidslt6	-246.5892	143.8619	-1.71	0.087	-529.3663	36.18793
kidsge6	-65.87681	44.77805	-1.47	0.142	-153.8932	22.13958
one	2345.98	487.9451	4.81	0.000	1386.868	3305.092

Instrumented: lwage

Instruments: nwifeinc educ age kidslt6 kidsge6 one exper expersq fatheduc motheduc hushrs husage huseduc mtr educXnwifeinc ageXnwifeinc ageXeduc kidslt6Xnwifeinc kidslt6Xeduc kidslt6Xage kidsge6Xnwifeinc kidsge6Xeduc kidsge6Xage kidsge6Xkidslt6 experXnwifeinc experXeduc experXage experXkidslt6 experXkidsge6 expersqXnwifeinc expersqXeduc expersqXage expersqXkidslt6 expersqXkidsge6 expersqXexper fatheducXnwifeinc fatheducXeduc fatheducXage fatheducXkidslt6 fatheducXkidsge6 fatheducXexper fatheducXexpersq motheducXnwifeinc motheducXeduc motheducXage motheducXkidslt6 motheducXkidsge6 motheducXexper motheducXexpersq motheducXfatheduc hushrsXnwifeinc hushrsXeduc hushrsXage hushrsXkidslt6 hushrsXkidsge6 hushrsXexper hushrsXexpersq hushrsXfatheduc hushrsXmotheduc husageXnwifeinc husageXeduc husageXage husageXkidslt6 husageXkidsge6 husageXexper husageXexpersq husageXfatheduc husageXmotheduc husageXhushrs huseducXnwifeinc huseducXeduc huseducXage huseducXkidslt6 huseducXkidsge6 huseducXexper huseducXexpersq huseducXfatheduc huseducXmotheduc huseducXhushrs huseducXhusage mtrXnwifeinc mtrXeduc mtrXage mtrXkidslt6 mtrXkidsge6 mtrXexper mtrXexpersq mtrXfatheduc mtrXmotheduc mtrXhushrs mtrXhusage mtrXhuseduc

AG specification test:
 J statistic = 0.1748
 Prob > J = 0.8059



► **Example**

The STATA output for FULL estimation with options `hom robust fuller`:

```
. mivreg hours nwifeinc educ age kidslt6 kidsge6 one (lwage = nwifeinc educ ///
> age kidslt6 kidsge6 one exper expersq fatheduc motheduc hushrs husage ///
> huseduc mtr *X*) if inlf==1 , hom robust fuller
```

MIVREG estimation (HOM)

First-stage summary

```
F( 86, 336) = 2.07
Prob > F    = 0.0000
R-squared   = 0.8471
```

```
Number of obs = 428
F( 7, 421)    = 96.46
Prob > F      = 0.0000
R-squared     = -0.5013
Adj R-squared = -0.5227
Root MSE     = 1.1e+03
```

FULL estimation

HHN variance estimation

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
hours						
lwage	1109.999	197.2334	5.63	0.000	722.314	1497.684
nwifeinc	-7.856532	5.235509	-1.50	0.134	-18.14753	2.434462
educ	-132.0795	31.83561	-4.15	0.000	-194.656	-69.50294
age	-9.934026	7.879563	-1.26	0.208	-25.42221	5.554159
kidslt6	-247.4823	143.2961	-1.73	0.085	-529.1472	34.18265
kidsge6	-66.3344	44.59569	-1.49	0.138	-153.9923	21.32355
one	2343.827	485.5647	4.83	0.000	1389.394	3298.26

Instrumented: lwage

Instruments: nwifeinc educ age kidslt6 kidsge6 one exper expersq fatheduc motheduc hushrs husage huseduc mtr educXnwifeinc ageXnwifeinc ageXeduc kidslt6Xnwifeinc kidslt6Xeduc kidslt6Xage kidsge6Xnwifeinc kidsge6Xeduc kidsge6Xage kidsge6Xkidslt6 experXnwifeinc experXeduc experXage experXkidslt6 experXkidsge6 expersqXnwifeinc expersqXeduc expersqXage expersqXkidslt6 expersqXkidsge6 expersqXexper fatheducXnwifeinc fatheducXeduc fatheducXage fatheducXkidslt6 fatheducXkidsge6 fatheducXexper fatheducXexpersq motheducXnwifeinc motheducXeduc motheducXage motheducXkidslt6 motheducXkidsge6 motheducXexper motheducXexpersq motheducXfatheduc hushrsXnwifeinc hushrsXeduc hushrsXage hushrsXkidslt6 hushrsXkidsge6 hushrsXexper hushrsXexpersq hushrsXfatheduc hushrsXmotheduc husageXnwifeinc husageXeduc husageXage husageXkidslt6 husageXkidsge6 husageXexper husageXexpersq husageXfatheduc husageXmotheduc husageXhushrs huseducXnwifeinc huseducXeduc huseducXage huseducXkidslt6 huseducXkidsge6 huseducXexper huseducXexpersq huseducXfatheduc huseducXmotheduc huseducXhushrs huseducXhusage mtrXnwifeinc mtrXeduc mtrXage mtrXkidslt6 mtrXkidsge6 mtrXexper mtrXexpersq mtrXfatheduc mtrXmotheduc mtrXhushrs mtrXhusage mtrXhuseduc

L0 specification test:

J statistic (bias-corrected) = -0.0382

Prob > J = 0.8752

◀

► **Example**

The STATA output for HFUL estimation with options `het robust fuller`:

```
. mivreg hours nwifeinc educ age kidslt6 kidsge6 one (lwage = nwifeinc educ ///
> age kidslt6 kidsge6 one exper expersq fatheduc motheduc hushrs husage ///
> huseduc mtr *X*) if inlf==1 , het robust fuller
```

MIVREG estimation (HET)

First-stage summary

F(86, 336) = 2.07
 Prob > F = 0.0000
 R-squared = 0.8471

Number of obs = 428
 F(7, 421) = 124.27
 Prob > F = 0.0000
 R-squared = -0.4339
 Adj R-squared = -0.4543
 Root MSE = 1.1e+03

HFUL estimation

HNWCS variance estimation

hours	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lwage	1058.269	170.4895	6.21	0.000	723.1527	1393.386
nwifeinc	-8.041127	4.708921	-1.71	0.088	-17.29705	1.214798
educ	-133.5581	29.08721	-4.59	0.000	-190.7323	-76.38381
age	-10.71399	8.313921	-1.29	0.198	-27.05596	5.627971
kidslt6	-274.0719	166.8757	-1.64	0.101	-602.0853	53.94151
kidsge6	-81.38394	43.17962	-1.88	0.060	-166.2584	3.49055
one	2485.039	466.6137	5.33	0.000	1567.856	3402.222

Instrumented: lwage

Instruments: nwifeinc educ age kidslt6 kidsge6 one exper expersq fatheduc motheduc hushrs husage huseduc mtr educXnwifeinc ageXnwifeinc ageXeduc kidslt6Xnwifeinc kidslt6Xeduc kidslt6Xage kidsge6Xnwifeinc kidsge6Xeduc kidsge6Xage kidsge6Xkidslt6 experXnwifeinc experXeduc experXage experXkidslt6 experXkidsge6 expersqXnwifeinc expersqXeduc expersqXage expersqXkidslt6 expersqXkidsge6 expersqXexper fatheducXnwifeinc fatheducXeduc fatheducXage fatheducXkidslt6 fatheducXkidsge6 fatheducXexper fatheducXexpersq motheducXnwifeinc motheducXeduc motheducXage motheducXkidslt6 motheducXkidsge6 motheducXexper motheducXexpersq motheducXfatheduc hushrsXnwifeinc hushrsXeduc hushrsXage hushrsXkidslt6 hushrsXkidsge6 hushrsXexper hushrsXexpersq hushrsXfatheduc hushrsXmotheduc husageXnwifeinc husageXeduc husageXage husageXkidslt6 husageXkidsge6 husageXexper husageXexpersq husageXfatheduc husageXmotheduc husageXhushrs huseducXnwifeinc huseducXeduc huseducXage huseducXkidslt6 huseducXkidsge6 huseducXexper huseducXexpersq huseducXfatheduc huseducXmotheduc huseducXhushrs huseducXhusage mtrXnwifeinc mtrXeduc mtrXage mtrXkidslt6 mtrXkidsge6 mtrXexper mtrXexpersq mtrXfatheduc mtrXmotheduc mtrXhushrs mtrXhusage mtrXhuseduc

CHNSW specification test:

J statistic (bias-corrected) = 76.4821
 Prob > J = 0.7340



8 Acknowledgments

The authors thank the editor and referee whose valuable suggestions were helpful in making the article more comprehensible to broader audiences. The first author gratefully acknowledges the support by the grant 17-26535S from the Czech Science Foundation. The presentations at the European Meeting of Statisticians (Helsinki, 2017) and 10th conference of the Czech Economic Society (Prague, 2018) were also useful.

9 References

- Anatolyev, S. 2019. Many instruments and/or regressors: a friendly guide. *Journal of Economic Surveys* 33: 689–726.
- Anatolyev, S., and N. Gospodinov. 2011. Specification testing in models with many instruments. *Econometric Theory* 27: 427–441.
- Anatolyev, S., and P. Yaskov. 2017. Asymptotics of diagonal elements of projection matrices under many instruments/regressors. *Econometric Theory* 33: 717–738.
- Angrist, J., and A. Krueger. 2001. Instrumental variables and the search for identification: from supply and demand to natural experiments. *Journal of Economic Perspectives* 15: 69–85.
- Basmann, R. L. 1957. A generalized classical method of linear estimation of coefficients in a structural equation. *Econometrica* 25: 77–83.
- Bekker, P. A. 1994. Alternative approximations to the distributions of instrumental variable estimators. *Econometrica* 62: 657–681.
- Chao, J. C., J. A. Hausman, W. K. Newey, N. R. Swanson, and T. Woutersen. 2014. Testing overidentifying restrictions with many instruments and heteroskedasticity. *Journal of Econometrics* 178: 15–21.
- Fuller, W. A. 1977. Some properties of a modification of the limited information estimator. *Econometrica* 45: 939–954.
- Hansen, C., J. Hausman, and W. K. Newey. 2008. Estimation with many instrumental variables. *Journal of Business & Economics Statistics* 26: 398–422.
- Hausman, J. A., W. K. Newey, T. Woutersen, J. C. Chao, and N. R. Swanson. 2012. Instrumental variable estimation with heteroskedasticity and many instruments. *Quantitative Economics* 3: 211–255.
- Lee, Y., and R. Okui. 2012. Hahn–Hausman test as a specification test. *Journal of Econometrics* 167: 133–139.
- Mroz, T. A. 1987. The sensitivity of an empirical model of married women’s hours of work to economic and statistical assumptions. *Econometrica* 55: 765–799.

About the authors

Stanislav Anatolyev is associate professor of economics at CERGE-EI (Prague, Czech Republic) and Access Industries professor of economics at New Economic School (Moscow, Russia).

Alena Skolkova is graduate student of economics at CERGE-EI (Prague, Czech Republic).