kinkyreg: Instrument-free inference for linear regression models with endogenous regressors

Sebastian Kripfganz
University of Exeter Business School
Exeter, United Kingdom

Jan F. Kiviet
University of Amsterdam
Amsterdam, The Netherlands
and Stellenbosch University
Stellenbosch, South Africa

S.Kripfganz@exeter.ac.uk

j.f.kiviet@uva.nl

February 13, 2021

Abstract. In models with endogenous regressors, a standard regression approach is to exploit just- or overidentifying orthogonality conditions by using instrumental variables. In just-identified models, the identifying orthogonality assumptions cannot be tested without the imposition of other non-testable assumptions. While formal testing of overidentifying restrictions is possible, its interpretation still hinges on the validity of an initial set of untestable just-identifying orthogonality conditions. We present the `kinkyreg` *Stata* program for kinky least squares inference that adopts an alternative approach to identification. By exploiting non-orthogonality conditions in the form of bounds on the admissible degree of endogeneity, feasible test procedures can be constructed that do not require instrumental variables. The kinky least squares confidence bands can be more informative than confidence intervals obtained from instrumental variables estimation, especially when the instruments are weak. Moreover, the approach facilitates a sensitivity analysis for standard instrumental variables inference. In particular, it allows to assess the validity of previously untestable just-identifying exclusion restrictions. Further instrument-free tests include linear hypotheses, functional form, heteroskedasticity, and serial correlation tests.

Keywords: kinkyreg, instrumental variables, confidence intervals, sensitivity analysis, specification tests, graphical inference

1 Introduction

The empirical literature on causal inference in linear regression models with endogenous regressors is dominated by instrumental variables (IV) based estimation methods. For valid inference under conventional asymptotic theory, instruments must be relevant and exogenous. The former condition requires that the instruments are sufficiently strongly correlated with the endogenous regressors. If this correlation is weak, coefficient estimates can be severely biased, finite-sample distributions are poorly approximated with conventional asymptotic theory, and statistical tests using conventional standard error estimates can exhibit large size distortions. To address these concerns, an extensive literature emerged on detecting instrument weakness and conducting robust statistical

inference under the presence of weak instruments. The latter methods, however, usually lead to wide confidence intervals that may not be very informative.[1] In *Stata*, tests for weak instruments and methods for weak-instruments robust inference are implemented in the community-contributed packages `ivreg2` (Baum et al. 2003, 2007), `condivreg` (Moreira and Poi 2003; Mikusheva and Poi 2006), `weakiv` as an extension of `rivtest` (Finlay and Magnusson 2009), `weakivtest` (Pflueger and Wang 2015), `twostepweakiv` (Sun 2018), and `boottest` (Roodman et al. 2019).

A noteworthy complication of the quest for good instruments is that the same features that make an instrument relevant can also be a source of a violation of the exogeneity condition (Hall et al. 1996). To be exogenous, an instrumental variable needs to be uncorrelated with the regression error term. This necessitates that the instrument is validly excluded from the structural model, i.e. that the instrument only has an indirect effect on the dependent variable via the instrumented endogenous regressors. If the model is just-identified, i.e. there are as many excluded instruments as endogenous regressors, the exclusion restriction is untestable in the standard IV framework. Intuitively, we cannot use the same instrument to identify the effect of an endogenous regressor and its own direct effect on the dependent variable. For identification of the former, IV-based estimators assume that the latter is known to be zero. Even in overidentified models, the validity of all instruments cannot be jointly tested. Routinely used overidentification tests still rely on the maintained (and untested) assumption that at least as many instruments are validly excluded from the model as there are endogenous regressors, and even then may not be informative about the instruments' ability to identify the parameters of interest (Parente and Santos Silva 2012).

In this article, we discuss an identification strategy that does not rely on such exclusion restrictions but instead imposes assumptions on the degree of regressor endogeneity, which is left unrestricted in an IV world. The kinky least squares (KLS) approach developed by Kiviet (2013, 2020a,b) achieves set identification of the regression coefficients by confining the admissible correlation of the regressors with the error term within plausible bounds. No excluded instruments are needed. Instead, the bias of the ordinary least squares (OLS) estimator is analytically corrected for all values on a grid of endogeneity correlations. This provides a set of consistent coefficient estimates in accordance with the postulated endogeneity range. Asymptotically conservative confidence intervals can be obtained as the union of the confidence intervals over the considered grid.

For a reasonably narrow range of postulated endogeneity correlations, these KLS confidence intervals are – as a general rule – narrower than those from IV / two-stage least squares (2SLS) estimations, in particular if the instruments are relatively weak. KLS inference is thus often more informative, and it avoids the problems associated with the search for strong and valid instruments. On top of that, the KLS approach enables testing of any potential exclusion restrictions. Because instrumental variables are not needed for identification, their direct effect is (set-)identifiable by adding them to the KLS regression (Kiviet 2020a,b).

---

1. For an overview on the weak-instruments literature, see Stock et al. (2002), Andrews and Stock (2007), and Andrews et al. (2019).

Similar approaches with the aim to bound a causal effect of a single endogenous regressor in the absence of suitable instruments have been recently proposed by Krauth (2016) and Oster (2019), and implemented in their *Stata* packages `rcr` and `psacalc`, respectively. Instead of an interval assumption about the correlation of the endogenous variable with the structural error term, they impose nontrivial restrictions on the magnitude of this correlation relative to the correlation of the endogenous regressor with other control variables.[2] In practice, bounding this sensitivity parameter may be a less intuitive task than placing bounds directly on the endogeneity correlation itself. When the model includes relevant control variables, the KLS estimator can be replicated with the relative correlation restriction (RCR) estimator of Krauth (2016) and the similar estimator of Oster (2019) by matching the respective sensitivity parameters. Yet, unlike the KLS estimator, the RCR estimator does not support models without control variables and is not immediately applicable to models with multiple endogenous regressors.

Undeniably, instrument-free inference is not a panacea to the problems of instrument-based methods. It replaces one set of possibly strong though speculative assumptions with another set of hopefully less restrictive conjectural assumptions. In many applications, it might be easier to specify a credible range for the correlation of an endogenous regressor with the error term than to convincingly present strong and valid instruments. For example, theoretical considerations might plausibly inform us about the sign of the endogeneity. Yet, if the chosen endogeneity range is too narrow, it may not include the true correlation value, potentially leading to serious bias. If it is too wide, the resulting confidence intervals could be less informative than those from a 2SLS estimation with strong and valid instruments.

Assuming that we have reasonable prior information about the range of endogeneity correlations, KLS confidence intervals and test procedures can provide reliable inference even in the absence of valid and strong instrumental variables. If instruments are available, the KLS inference can facilitate sensitivity checks for IV-based procedures. Because the different methods have different strengths and weaknesses, it is often reasonable to consider the instrument-free approach as a complement rather than a substitute to instrument-based procedures, possibly in addition to other methods that relax some of the assumptions underlying the traditional instrument-based inference. For instance, Conley et al. (2012) propose the construction of conservative confidence intervals that allow for a mild violation of the exclusion restrictions, assuming a plausible range of direct effects for the instruments. Nevo and Rosen (2012) derive bounds for the effect size in the presence of imperfect instruments, making assumptions about the sign and the maximum strength of the correlation of the instruments with the error term. These two procedures can be applied with the community-contributed *Stata* commands `plausexog` and `imperfectiv` by Clarke and Matta (2018), respectively.

The KLS approach to statistical inference under confined regressor endogeneity is implemented in the new `kinkyreg` *Stata* package. We review the methodology in Section 2. After introducing the syntax of `kinkyreg` and its postestimation commands in

---

2. A closely related procedure was proposed earlier by Altonji et al. (2005).

Sections 3 and 4, we illustrate the approach with an empirical example in Section 5. The main output is graphical. KLS point estimates and confidence intervals are plotted for selected variables over a user-specified range of endogeneity correlations. The results are compared to the traditional 2SLS estimates if the user specifies any instrumental variables. The exclusion restrictions can then be tested with a postestimation command which plots the p-values of the test over the endogeneity range. Similarly, instrument-free tests for linear hypotheses, correct functional form specification, heteroskedasticity, and serial correlation are implemented as postestimation commands as well. Another postestimation command calculates the RCR sensitivity parameters that can be used to replicate the KLS results with the estimators of Krauth (2016) and Oster (2019).

2 Kinky least squares inference

2.1 Coefficient estimates and confidence intervals

Consider the linear regression model with $i = 1, 2, \ldots, N$ observations, an endogenous regressor x_{1i}, and a column vector of exogenous (or predetermined) variables $\mathbf{x}_{2i}$:

$$y_i = \beta_1 x_{1i} + \mathbf{x}'_{2i}\boldsymbol{\beta}_2 + \varepsilon_i \tag{1}$$

All variables are transformed into deviations from their means.[3] The restriction to a single endogenous regressor is mainly for expositional purposes. The methodology can be applied to any number of endogenous variables.

The standard approach to estimating models with endogenous regressors is by using instrumental-variables techniques. However, instruments-based inference can be unreliable if the instrumental variables $\mathbf{z}_i$ are only weakly correlated with the endogenous regressor x_{1i} or if they are potentially endogenous themselves. To obtain consistent estimates, this approach exploits orthogonality conditions for the instruments: $E[\mathbf{z}_i\varepsilon_i] = \mathbf{0}$.

Kiviet (2020a,b) suggests an alternative instrument-free approach that makes use of a non-orthogonality condition for the endogenous regressor in model (1): $E[x_{1i}\varepsilon_i] = \rho\,\sigma_1\sigma_\varepsilon$, where ρ denotes the correlation coefficient between x_{1i} and ε_i, and σ_1 and σ_ε are the standard deviations of x_{1i} and ε_i.[4] Clearly, this approach is infeasible unless ρ, σ_1, and σ_ε are known or can be estimated consistently. For the moment, assume that ρ is indeed known. σ_1 can be easily estimated from the observed data as the square root of $\widehat{\sigma}_1^2 = N^{-1}\sum_{i=1}^{N} x_{1i}^2$. As shown by Kiviet (2020a,b), σ_ε can be consistently estimated as the square root of

$$\widehat{\sigma}_\varepsilon^2(\rho) = \widehat{\sigma}_{\varepsilon,OLS}^2 \left(1 - \rho^2 \frac{\widehat{\sigma}_1^2}{\widehat{\sigma}_1^2 - \widehat{\boldsymbol{\sigma}}'_{12}\widehat{\boldsymbol{\Sigma}}_2^{-1}\widehat{\boldsymbol{\sigma}}_{12}}\right)^{-1} \tag{2}$$

---

3. In other words, the intercept is partialled out from this model. This is done automatically by the `kinkyreg` command.

4. In the case of multiple endogenous variables, $E[\mathbf{x}_{1i}\varepsilon_i] = \sigma_\varepsilon \mathrm{dg}(\boldsymbol{\Sigma}_1)\boldsymbol{\rho}$, where $\mathrm{dg}(\boldsymbol{\Sigma}_1)$ is a diagonal matrix of standard deviations and $\boldsymbol{\rho}$ is a column vector of endogeneity correlations. The subsequent formulas need to be adjusted accordingly; see Kiviet (2020a,b) for the general case.

where $\widehat{\sigma}_{\varepsilon,OLS}^2 = N^{-1}\sum_{i=1}^N \widehat{\varepsilon}_{i,OLS}^2$ is the familiar variance estimate from OLS residuals $\widehat{\varepsilon}_{i,OLS}$. Because OLS is inconsistent when $\rho \neq 0$, we need to adjust this estimate. The adjustment term requires the covariance estimates $\widehat{\boldsymbol{\sigma}}_{12} = N^{-1}\sum_{i=1}^N x_{1i}\mathbf{x}_{2i}$ and $\widehat{\boldsymbol{\Sigma}}_2 = N^{-1}\sum_{i=1}^N \mathbf{x}_{2i}\mathbf{x}_{2i}'$ that are readily obtained from the observed data. The KLS estimator then corrects the inconsistency of the OLS estimator as follows:

$$\begin{pmatrix} \widehat{\beta}_1(\rho) \\ \widehat{\boldsymbol{\beta}}_2(\rho) \end{pmatrix} = \begin{pmatrix} \widehat{\beta}_{1,OLS} \\ \widehat{\boldsymbol{\beta}}_{2,OLS} \end{pmatrix} - \frac{\rho\,\widehat{\sigma}_1\widehat{\sigma}_\varepsilon(\rho)}{\widehat{\sigma}_1^2 - \widehat{\boldsymbol{\sigma}}_{12}'\widehat{\boldsymbol{\Sigma}}_2^{-1}\widehat{\boldsymbol{\sigma}}_{12}} \begin{pmatrix} 1 \\ -\widehat{\boldsymbol{\Sigma}}_2^{-1}\widehat{\boldsymbol{\sigma}}_{12} \end{pmatrix} \tag{3}$$

Notice that the KLS estimator is point-symmetric around $\rho = 0$, and $\widehat{\beta}_1(\rho)$ is a monotonically decreasing function in ρ.[5] $\widehat{\boldsymbol{\beta}}_2(\rho)$ can be monotonically increasing or decreasing, depending on the covariance terms, and $\widehat{\boldsymbol{\beta}}_2(\rho) = \widehat{\boldsymbol{\beta}}_{2,OLS}$ if $\rho = 0$ or if the exogenous regressors $\mathbf{x}_{2i}$ are uncorrelated with the endogenous regressor x_{1i}, i.e. $\widehat{\boldsymbol{\sigma}}_{12} = \mathbf{0}$.

For inference on the coefficients $\boldsymbol{\beta} = (\beta_1, \boldsymbol{\beta}_2')'$, we need to calculate confidence bands that rely on consistent estimates of the estimator's variance. Kiviet (2020a,b) shows that the KLS estimator (3) is asymptotically normally distributed with variance-covariance matrix $\sigma_\varepsilon^2 \mathbf{V}(\rho, \kappa_x, \kappa_\varepsilon)$, where κ_x is the kurtosis of the regressors x_{1i} and $\mathbf{x}_{2i}$, and κ_ε is the kurtosis of the error term ε_i.[6] To arrive at an analytical expression for $\mathbf{V}(\rho, \kappa_x, \kappa_\varepsilon)$, Kiviet (2020a) assumes that κ_x is identical for all regressors. Usually, this will not be the case, but a conservative variance estimate is obtained by taking for κ_x the largest kurtosis estimate across all regressors.[7] For a given regressor, say x_{1i}, the kurtosis can be estimated as $\widehat{\kappa}_x = N^{-1}\sum_{i=1}^N (x_{1i}/\widehat{\sigma}_1)^4$. Similarly, $\hat{\kappa}_\varepsilon(\rho) = N^{-1}\sum_{i=1}^N [\widehat{\varepsilon}_i(\rho)/\widehat{\sigma}_\varepsilon(\rho)]^4$, with KLS residuals $\widehat{\varepsilon}_i(\rho) = y_i - \widehat{\beta}_1(\rho)x_{1i} - \mathbf{x}_{2i}'\widehat{\boldsymbol{\beta}}_2(\rho)$.

However, the correlation coefficient ρ is unknown and without imposing additional restrictions a consistent estimate of ρ is unattainable. Instead of tying oneself to a particular value $\rho = r$, we can assume that the true value is contained within a set $\rho \in [r_l, r_u]$. Often, there might be prior information about the magnitude or the sign of the endogeneity that allows us to pin down reasonable boundaries for this interval. We can then obtain the KLS estimator $\widehat{\boldsymbol{\beta}}(r)$ for a range of values $r \in [r_l, r_u]$. Corresponding confidence intervals can be constructed with variance estimates $\widehat{\sigma}_\varepsilon^2(r)\mathbf{V}(r, \widehat{\kappa}_x, \widehat{\kappa}_\varepsilon(r))$. For a significance level α, the union of these confidence intervals over the range $r \in [r_l, r_u]$ has asymptotic coverage of at least $1 - \alpha$.

As a more illuminating approach, we can also plot the coefficient estimates with corresponding confidence intervals over the chosen range of endogeneity correlations. This shows immediately for which values of ρ we can reject (or not reject) the null hypothesis that a coefficient of interest equals a certain value, most prominently whether the coefficient is statistically significantly different from zero. Such graphs are the main output of the new `kinkyreg` command and examples can be seen in Section 5.

---

5. $\widehat{\sigma}_1^2 - \widehat{\boldsymbol{\sigma}}_{12}'\widehat{\boldsymbol{\Sigma}}_2^{-1}\widehat{\boldsymbol{\sigma}}_{12} > 0$ as a consequence of the variance-covariance matrix being positive definite.
6. See Kiviet (2020a) for the full formula of $\mathbf{V}(\rho, \kappa_x, \kappa_\varepsilon)$, and Kiviet (2020b) for the special case $\mathbf{V}(\rho, 3, 3)$ that holds in particular when the regressors and the error term are normally distributed.
7. Simulations by Kiviet (2020a) show that this procedure works reasonably well in samples with 100 observations.

It is important to note that the choice of r_l and r_u is restricted by certain feasibility bounds. In order to rule out a negative estimate of $\widehat{\sigma}^2_\varepsilon(r)$, it follows from equation (2) that r must satisfy

$$|r| < \sqrt{1 - \frac{\widehat{\boldsymbol{\sigma}}'_{12}\widehat{\boldsymbol{\Sigma}}^{-1}_2\widehat{\boldsymbol{\sigma}}_{12}}{\widehat{\sigma}^2_1}} \leq 1 \tag{4}$$

Thus, unless the endogenous regressor is uncorrelated (in the sample) with the exogenous regressors, i.e. $\widehat{\boldsymbol{\sigma}}_{12} = \mathbf{0}$, the interval $[r_l, r_u]$ cannot be expanded arbitrarily close to -1 or 1. The closer we get to these feasibility bounds, the wider the confidence intervals become. For informative inference, we need to use some initial information or prior belief to restrict the admissible endogeneity to a reasonably narrow range.

2.2 Specification tests

Just like after OLS or 2SLS estimation, we usually want to scrutinize our model specification. Based on the KLS coefficient and variance estimates, we can calculate and visualize the p-values for any desired test statistic over the range $r \in [r_l, r_u]$. Such tests can be conventional tests of linear hypotheses $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{c}$, implemented by the `kinkyreg` postestimation command `estat test`. These tests are based on the Wald statistic,

$$\widehat{W}(r) = \left[\mathbf{R}\widehat{\boldsymbol{\beta}}(r) - \mathbf{c}\right]' \left[\widehat{\sigma}^2_\varepsilon(r)\mathbf{V}(r, \widehat{\kappa}_x, \widehat{\kappa}_\varepsilon(r))\right]^{-1} \left[\mathbf{R}\widehat{\boldsymbol{\beta}}(r) - \mathbf{c}\right] \tag{5}$$

or alternatively the corresponding F-statistic if small-sample statistics are desired. It is then straightforward to test the valid exclusion of a set of variables $\mathbf{x}_{3i}$ from model (1) by testing for joint statistical insignificance, $H_0 : \boldsymbol{\beta}_3 = \mathbf{0}$, in the auxiliary KLS regression

$$y_i = \beta_1 x_{1i} + \mathbf{x}'_{2i}\boldsymbol{\beta}_2 + \mathbf{x}'_{3i}\boldsymbol{\beta}_3 + \varepsilon_i \tag{6}$$

The results inform us which values of $r \in [r_l, r_u]$ are compatible with the valid exclusion of $\mathbf{x}_{3i}$.

Testing exclusion restrictions is particularly useful in the context of IV / 2SLS estimation. Instead of estimating model (1) by KLS, we might choose $\mathbf{x}_{3i}$ as external instrumental variables for the endogenous regressor x_{1i}, assuming that those instruments are indeed validly excluded from the model and that they are sufficiently correlated with x_{1i}. The predictive power of the instruments for the endogenous regressor can be assessed with conventional first-stage diagnostics. If the model is overidentified, i.e. $\mathbf{x}_{3i}$ contains more than one excluded variable, we can use overidentifying-restrictions tests to assess the validity of the instruments, maintaining the assumption that at least one of the variables in $\mathbf{x}_{3i}$ (or a linear combination of them) is valid.[8] However, the joint validity of all instruments under unconstrained endogeneity of x_{1i} is untestable in this context, and the maintained assumption of valid exclusion for a subset of the instruments requires expert justification.

---

8. For a discussion of first-stage diagnostic tests and overidentification tests, see Baum et al. (2003, 2007) and the references therein.

This is where KLS comes into play. By constraining the endogeneity of x_{1i}, we can test the valid exclusion of $\mathbf{x}_{3i}$. This does not come for free but requires expert judgment on the admissible degree of endogeneity. Yet, it may often be easier to argue that the correlation of the endogenous regressor with the error term falls into a certain interval, than to justify that there is no direct effect of (some of) the instruments. Such KLS exclusion restrictions tests can be performed with the postestimation command `estat exclusion`.

The KLS approach can also be applied to other specification tests. Closely related to the exclusion restrictions test is the Ramsey (1969) regression equation specification error test (RESET). By testing the valid exclusion of polynomials in the fitted values or right-hand side variables, insights are provided whether we used the correct functional form. However, the presence of the endogenous regressor x_{1i} will cause the fitted values, $\widehat{y}_i(r) = \widehat{\beta}_1(r)x_{1i} + \mathbf{x}'_{2i}\widehat{\boldsymbol{\beta}}_2(r)$, to be endogenous as well. To circumvent this problem, we can apply an endogeneity correction. Following Kiviet (2020a), we can decompose the endogenous regressor into an exogenous and an endogenous part, $x_{1i} = \xi_{1i} + \rho\,\sigma_1\varepsilon_i/\sigma_\varepsilon$. The exogenous part ξ_{1i} is unobserved, but we can consistently estimate it for any postulated degree of endogeneity r as

$$\widehat{\xi}_{1i}(r) = x_{1i} - r\frac{\widehat{\sigma}_1}{\widehat{\sigma}_\varepsilon(r)}\widehat{\varepsilon}_i(r) \tag{7}$$

An operationalized RESET version, implemented by `estat reset`, then uses adjusted fitted values $\widetilde{y}_i(r) = \widehat{\beta}_1(r)\widehat{\xi}_{1i}(r) + \mathbf{x}'_{2i}\widehat{\boldsymbol{\beta}}_2(r)$ in the auxiliary KLS regression (6). Notice that the added regressors now vary with r, i.e. $\mathbf{x}_{3i}(r) = (\widetilde{y}_i^2(r), \widetilde{y}_i^3(r), \ldots, \widetilde{y}_i^p(r))$ for some polynomial order $p \geq 2$. Alternatively, $\mathbf{x}_{3i}(r)$ can be the respective powers of $(\widehat{\xi}_{1i}(r), \mathbf{x}'_{2i})'$. The test statistic is again the Wald statistic (5) (or its F-statistic analogue) for the null hypothesis $H_0 : \boldsymbol{\beta}_3 = \mathbf{0}$.

The KLS estimator (3) is derived by assuming a constant variance σ_ε^2. It is thus desirable to lay out a test for this assumption. We can follow the Breusch and Pagan (1979) approach and run an auxiliary KLS regression of the squared residuals $\widehat{\varepsilon}_i^2(r)$ on the endogeneity-corrected fitted values $\widetilde{y}_i(r)$ or the (exogenous variation of the) right-hand side variables $(\widehat{\xi}_{1i}(r), \mathbf{x}'_{2i})'$.[9] The null hypothesis of no conditional heteroskedasticity then corresponds to joint irrelevance of all variables in this auxiliary regression. This test is available with the postestimation command `estat hettest`.

In a time series setting, the KLS approach rests on the assumption that there is no serial error correlation. If we suspect serial correlation, we could add lags of the dependent variable and the right-hand side variables to the regression model to obtain a dynamically complete model (Wooldridge 2020, Chapter 11.4). To be adequate for a model with a lagged dependent variable, a test for serial correlation should allow for regressors that are not strictly exogenous. This is the case for the "alternative test" of Durbin (1970), implemented by `estat durbinalt` as an exclusion restrictions test for the lagged residuals $\mathbf{x}_{3i} = (\widehat{\varepsilon}_{i-1}(r), \widehat{\varepsilon}_{i-2}(r), \ldots, \widehat{\varepsilon}_{i-p}(r))'$, up to some lag order $p \geq 1$,

---

9. Occasionally, it can be desirable to just use a subset of the right-hand side variables or to include additional variables in the conditional heteroskedasticity model.

in the auxiliary equation (6). The null hypothesis of no serial correlation is not rejected if the coefficients of the lagged residuals are jointly statistically insignificant.[10]

3 The kinkyreg command

3.1 Syntax

kinkyreg *depvar* $\left[\,varlist1\,\right]$ (*varlist2* $\left[\,\text{= }\,varlist\_iv\,\right]$) $\left[\,if\,\right]$ $\left[\,in\,\right]$ $\left[\,\text{,}\right.$
   <u>endog</u>eneity(*numlist*) <u>range</u>(#$_1$ #$_2$) <u>stepsize</u>(#) ekurtosis(#)
   xkurtosis(#) <u>nocons</u>tant <u>corr</u>elation(#) <u>level</u>(#) <u>small</u>
   <u>inference</u>(*varlist*) lincom(#:*exp*) <u>twoway</u>($\left[\,varname\,|\,\#\,\right]$ $\left[\,\text{, }\,twoway\_options\right.$
   order(*orderinfo*) yrange(#$_1$ #$_2$) addplot(*plotinfo*)$\left.\right]$) <u>coefplot</u>(kls|iv
   $\left[\,varname\,|\,\#\,\right]$ $\left[\,\text{, }\,line\_options\,\text{recast}(newplottype)\right]$) <u>ciplot</u>(kls|iv
   $\left[\,varname\,|\,\#\,\right]$ $\left[\,\text{, }\,fitarea\_options\,\text{recast}(newplottype)\right]$) <u>names</u>tub(*namestub*)
   <u>nogr</u>aph novstore *display_options*$\left.\right]$

kinkyreg2dta *depvar* $\left[\,varlist1\,\right]$ (*varlist2* $\left[\,\text{= }\,varlist\_iv\,\right]$) $\left[\,if\,\right]$ $\left[\,in\,\right]$ $\left[\,\text{,}\right.$
   <u>range</u>(*numlist*) <u>stepsize</u>(*numlist*) ekurtosis(#) xkurtosis(#) <u>nocons</u>tant
   <u>level</u>(#) <u>small</u> lincom(#:*exp*) coef($\left[\,\text{b}\,\right]$ $\left[\,\text{se}\,\right]$ $\left[\,\text{ciub}\,\right]$ $\left[\,\text{cilb}\,\right]$:$\left[\,varlist\,\right]$
   $\left[\,numlist\,\right]$) estat(# $\left[\,\text{chi2|F}\,\right]$ $\left[\,\text{p}\,\right]$:*estat_cmdline*) <u>frame</u>(*framename* $\left[\,\text{,}\right.$
   replace$\left.\right]$) replace saving(*filename* $\left[\,\text{, }\,\text{replace}\right]$) double$\left.\right]$

    kinkyreg2dta is a wrapper for kinkyreg that creates a data set with the KLS estimation and postestimation results. It does not produce any graphs or estimation output. While kinkyreg can only vary the endogeneity correlation of one endogenous regressor at a time, kinkyreg2dta allows to vary these correlations for multiple endogenous regressors jointly.

3.2 Options

endogeneity(*numlist*) specifies values for the correlations of the endogenous variables with the error term. The order of the values corresponds to the order of the variables in *varlist2*. A missing value (.) must be specified for the variable for which the endogeneity correlation should be varied over the range specified with option range(). All other endogeneity correlations are held fixed.[11] This option is required if *varlist2*

---

10. The serial correlation test is usually constructed by using the residuals $\widehat{\varepsilon}_i(r)$ instead of y_i as the dependent variable in equation (6). However, this modification does not affect the test statistic.
11. Users can create their own loops with repeated calls to kinkyreg or use the kinkyreg2dta command if the endogeneity correlation shall be varied for multiple variables. An example is provided in Section 5.2.

contains multiple variables, and it is redundant otherwise.

range($\#_1$ $\#_2$) requests to compute the KLS estimator for all feasible endogeneity correlations in the interval $[\#_1, \#_2]$. The default is range(-1 1).[12]

With kinkyreg2dta, range($\#_1$ $\#_2$ $\big[$ $\#_3$ $\#_4$ ... $\big]$) specifies the endogeneity intervals $[\#_1 \#_2]$ for the first endogenous variable, $[\#_3 \#_4]$ for the second endogenous variable, and so on. range($\#_1$ $\#_2$) with only two elements yields identical intervals for all endogenous variables.

stepsize($\#$) sets the step size for the interval over which the KLS estimator is computed. The default is stepsize(0.01).

With kinkyreg2dta, stepsize($\#_1$ $\big[$ $\#_2$ ... $\big]$) specifies separate step sizes for each endogenous variable in the order in which they appear in *varlist2*. stepsize($\#_1$) with only one element yields identical step sizes for all endogenous variables.

ekurtosis($\#$) specifies a value for the kurtosis of the error term to be used in the variance calculation. By default, the kurtosis is estimated based on the KLS estimates.

xkurtosis($\#$) specifies a value for the kurtosis of the right-hand side variables to be used in the variance calculation. By default, the maximum of the estimated kurtosis for all variables in *varlist1* and *varlist2* is used.

noconstant suppresses the constant term; see [R] **estimation options**.

correlation($\#$) requests to display estimation results for the specified endogeneity correlation and to return the results in e(b) and e(V). If $\#$ does not match a value on the grid specified with options range() and stepsize(), the estimation results for the closest grid point to $\#$ are displayed. By default, a regression table is not displayed and estimation results are not returned in e(b) and e(V).

level($\#$) sets the confidence level in %; see [R] **estimation options**. The default is level(95).

small requests that a degrees-of-freedom adjustment be made to the variance-covariance matrix and that small-sample t and F statistics be reported. The adjustment factor is $N/(N - K)$, where N is the number of observations and K the number of coefficients, including the intercept. By default, no degrees-of-freedom adjustment is made and z and Wald statistics are reported.

inference(*varlist*) specifies variables for which KLS inference graphs are generated. By default, KLS inference is only carried out for the endogenous regressors, i.e. inference(*varlist2*), unless option lincom() is specified. In the latter case, the default is to produce KLS inference only for the specified linear combinations.

lincom($\#$:*exp*) specifies linear combinations *exp* of the regression coefficients for which KLS inference graphs are generated; see [R] **lincom**. You may specify as many sets of linear combinations, with different reference number $\#$ (an integer number between 1 and 1,999), as you need.

---

12. The feasible range is determined by the bounds in equation (4).

twoway([*varname* | #] [, order(*orderinfo*) yrange(#₁ #₂) addplot(*plotinfo*) *twoway_options*]) specifies options allowed by graph twoway; see [G-3] ***twoway options***. *varname* must be a variable name in *varlist1* or *varlist2*. # must be the reference number for a linear combination specified with option lincom(). If neither *varname* nor # are specified, then all twoway graphs are addressed.

The twoway options name() and saving() require *varname* or # to be specified; see [G-3] ***name option*** and [G-3] ***saving option***. If name() is not specified, name(*namestub_varname* | #, replace) is assumed. The prefix is set with option namestub(*namestub*). If *varname* is specified and the addressed variable contains factor variable or time series operators, the symbols "." and "#" are replaced by "_".

order(*orderinfo*) allows to change the order in which the plots are drawn. *orderinfo* is a list containing one or more of the following graph elements in the order in which they shall be drawn: kls for the KLS coefficient estimate, kls_ci for the KLS confidence interval, iv for the IV coefficient estimate, and iv_ci for the IV confidence interval. The default is order(iv_ci iv kls_ci kls). This option also affects the order of the graph elements in the graph legend; see [G-3] ***legend options***.

yrange(#₁ #₂) specifies that the coefficient and confidence interval plots be restricted to the interval [#₁, #₂] on the *y* axis. A missing value for #₁ or #₂ refers to minus or plus infinity, respectively.

addplot(*plot* [, before(*orderinfo*)]) allows to overlay the twoway graphs with additional plots; see [G-3] ***addplot option***. before(*orderinfo*) allows to change the order of the graph elements by drawing the additional plots immediately before the specified element. *orderinfo* is one of the graph elements kls, kls_ci, iv, iv_ci as specified with suboption order(). By default, the additional plots are ordered last.

coefplot(kls | iv [*varname* | #] [, *line_options* recast(*newplottype*)]) determines the look of the KLS and IV coefficient plots. *line_options* are options allowed by graph twoway line; see [G-3] ***line options***. *varname* must be a variable name in *varlist1* or *varlist2*. # must be the reference number for a linear combination specified with option lincom(). If neither *varname* nor # are specified, then all coefficient plots are addressed.

recast(*newplottype*) allows to treat the plot as *newplottype* instead of a line plot; see [G-3] ***advanced options***.

ciplot(kls | iv [*varname* | #] [, *fitarea_options* recast(*newplottype*)]) determines the look of the KLS and IV confidence interval plots. *fitarea_options* are options allowed by graph twoway rarea; see [G-3] ***fitarea options***. *varname* must be a variable name in *varlist1* or *varlist2*. # must be the reference number for a linear combination specified with option lincom(). If neither *varname* nor # are specified, then all confidence interval plots are addressed.

recast(*newplottype*) allows to treat the plot as *newplottype* instead of a range plot

with area shading; see [G-3] *advanced options*.[13]

namestub(*namestub*) sets the prefix for the names of all graphs being created unless a name is explicitly specified with option twoway(*varname*|#, name(*name*)). The default is namestub(kinkyreg). This option also affects the graphs created by the postestimation commands.

nograph suppresses the creation of graphs for KLS inference.

novstore requests not to store the variance-covariance matrices for each grid point in order to consume less memory; seldom used. By default, these matrices are stored as hidden estimation results.[14] They are required by some postestimation commands.

display_options: <u>coefl</u>egend, <u>nohea</u>der, <u>notabl</u>e; see [R] **estimation options**.

The following options are specific to kinkyreg2dta:

coef($\big[$b$\big]$ $\big[$se$\big]$ $\big[$ciub$\big]$ $\big[$cilb$\big]$:$\big[$*varlist*$\big]$ $\big[$*numlist*$\big]$) specifies the kinkyreg estimation results to be saved. These can be coefficient estimates (b), standard errors (se), confidence interval upper bounds (ciub), and confidence interval lower bounds (cilb). The respective results are saved for the coefficients of all variables in *varlist* and linear combinations with reference numbers # in *numlist*. If neither *varlist* nor *numlist* are specified, the results are saved for all endogenous variables in *varlist2*. You may specify as many sets of estimation results as you need.

estat(# $\big[$chi2|F$\big]$ $\big[$p$\big]$:*estat_cmdline*) specifies the kinkyreg postestimation estimation results to be saved. These can be the values of the test statistic (chi2 or F) and the *p*-values (p). *estat_cmdline* is the full syntax of the estat subcommand, including any options. The word estat is optional. You may specify as many postestimation results, with different reference number #, as you need.

frame(*framename* $\big[$, replace$\big]$) requests to create a new frame with name *framename* in which the new variables are generated. The new frame is made the current frame; see [D] **frames**. replace specifies that the frame may be replaced if it already exists.

replace specifies that the data in memory shall be replaced with the newly generated data, even if the current data have not been saved to disk.

saving(*filename* $\big[$, replace$\big]$) requests to save the newly generated data to disk under the name *filename*. replace permits to overwrite an existing data set.

double requests to use the storage type double for the variables in the new data set.

---

13. For example, recast(rline) might be useful to replace the shaded areas with lines for the boundaries of the confidence intervals.
14. These matrices are stored as e(V_#), where # are consecutive integers from 1 to the number of grid points, with e(V_1) the variance-covariance matrix for the lowest endogeneity correlation. Storing these matrices also allows to replay estimation results for different endogeneity correlations without reestimation by typing kinkyreg, correlation(#).

3.3 Stored results

kinkyreg stores the following results in e():

Scalars
e(N)	number of observations	e(grid_min)	lower bound for the endogeneity
e(df_r)	residual degrees of freedom	e(grid_max)	upper bound for the endogeneity
e(rank)	rank of e(V)	e(grid_step)	step size for the endogeneities
e(xkurtosis)	maximum kurtosis of right-hand side variables	e(corr)	postulated endogeneity for e(b); not always saved

Macros
e(cmd)	kinkyreg	e(endovars)	endogenous variables
e(cmdline)	command as typed	e(exovars)	exogenous variables
e(estat_cmd)	kinkyreg_estat	e(ivvars)	instrumental variables
e(predict)	kinkyreg_p	e(namestub)	prefix for graph names
e(klsvar)	endogenous variable with varied endogeneity	e(properties)	b V; not always saved
		e(depvar)	name of dependent variable

Matrices
e(b)	coefficient vector; not always saved	e(ciub_kls)	KLS CI upper bounds
		e(cilb_kls)	KLS CI lower bounds
e(V)	variance-covariance matrix; not always saved	e(endogeneity)	fixed endogeneity correlations; not always saved
e(b_kls)	KLS coefficient vectors	e(sigma2)	KLS estimates of σ_ε^2
e(se_kls)	KLS standard errors	e(ekurtosis)	kurtosis of KLS residuals

Functions
e(sample)	marks estimation sample

4 Postestimation commands

The kinkyreg package provides the following special-interest postestimation commands: estat test for tests of linear hypotheses, estat exclusion for tests of exclusion restrictions, estat reset for RESET, estat hettest for heteroskedasticity tests, estat durbinalt for Durbin's alternative serial correlation test,[15] and estat rcr for the calculation of the RCR sensitivity parameters.

4.1 Syntax

estat test (*test_spec*) [(*test_spec*) ...] [, *test_options* <u>correl</u>ation(#)

    <u>tw</u>oway([, *twoway_options* yrange(#$_1$ #$_2$) addplot(*plot*)])

    <u>pvalplot</u>([*varname*] [, *line_options* recast(*newplottype*)]) <u>nograph</u>]

where *test_spec* is a coefficient list or expression and *test_options* are standard options allowed by the test command; see [R] **test**.

estat <u>exclu</u>sion [, <u>nojoint</u> <u>noindi</u>vidual <u>ekurtosis</u>(#) <u>xkurtosis</u>(#)

    <u>corr</u>elation(#) <u>l</u>evel(#) <u>tw</u>oway([, *twoway_options* yrange(#$_1$ #$_2$)

---

15. estat durbinalt requires formally ordered data, typically time series data.

addplot(*plot*)]) <u>pvalp</u>lot([*varname*] [, *line_options* recast(*newplottype*)])
<u>nogr</u>aph <u>nota</u>ble]

estat reset [, xb rhs <u>or</u>der(*numlist*) <u>ek</u>urtosis(#) <u>xk</u>urtosis(#)
<u>corr</u>elation(#) <u>tw</u>oway([, *twoway_options* yrange(#$_1$ #$_2$) addplot(*plot*)])
<u>pvalp</u>lot(# [, *line_options* recast(*newplottype*)]) <u>nogr</u>aph]

estat <u>het</u>test [(*varlist*) (*varlist*) ...] [, xb rhs minp <u>corr</u>elation(#)
<u>tw</u>oway([, *twoway_options* yrange(#$_1$ #$_2$) addplot(*plot*)]) <u>pvalp</u>lot(# [,
line_options recast(*newplottype*)]) <u>nogr</u>aph]

estat <u>dur</u>binalt [, <u>or</u>der(*numlist*) <u>ek</u>urtosis(#) <u>xk</u>urtosis(#)
<u>corr</u>elation(#) <u>tw</u>oway([, *twoway_options* yrange(#$_1$ #$_2$) addplot(*plot*)])
<u>pvalp</u>lot(# [, *line_options* recast(*newplottype*)]) <u>nogr</u>aph]

estat rcr [, lambda delta <u>corr</u>elation(#) <u>tw</u>oway([, *twoway_options*
yrange(#$_1$ #$_2$) addplot(*plot*)]) <u>pvalp</u>lot(# [, *line_options*
recast(*newplottype*)]) <u>nogr</u>aph]

4.2 Options

nojoint requests not to compute the joint exclusion test of all variables.

noindividual requests not to compute the individual exclusion tests for each variable.

xb requests to use the fitted values. Only the exogenous variation of the endogenous
right-hand side variable, equation (7), is used to compute the fitted values.

With estat reset, powers of the fitted values are used. This is the default.

With estat hettest, a test with fitted values only is computed, in addition to
tests with other specified varlists, if any. This option is the default if no varlists are
specified.

rhs requests to use the right-hand side variables of the fitted regression model. Only
the exogenous variation of the endogenous variable, equation (7), is used.

With estat reset, powers of the individual right-hand side variables are used in-
stead of the fitted values.

With estat hettest, the right-hand side variables are added to each *varlist*. This
option allows *varlist* to be empty but parentheses are still required if multiple varlists
are specified.

order(*numlist*) specifies the orders to be used for the test. A separate test is computed for each value in *numlist*.

With `estat reset`, these are the polynomial orders of the fitted values or right-hand side variables. The default is `order(2 3 4)`.

With `estat durbinalt`, these are the maximum lag orders of the residuals. The default is `order(1)`.

ekurtosis(#) specifies a value for the kurtosis of the error term to be used in the variance calculation. By default, the kurtosis is estimated based on the KLS estimates.

xkurtosis(#) specifies a value for the kurtosis of the right-hand side variables to be used in the variance calculation. By default, the maximum of the estimated kurtosis for all right-hand side variables is used.

lambda and delta request to either compute Krauth's λ or Oster's δ for the replication of the KLS estimates with the respective RCR estimator. By default, both sensitivity parameters are computed.

minp returns for each endogeneity correlation the minimum p-value of individual significance tests among all variables in the respective variable list. By default, `estat hettest` computes joint significance tests of all variables in the auxiliary regression.

correlation(#) requests to display test results or parameter values for the specified endogeneity correlation. If # does not match a value on the estimation grid, the results for the closest grid point to # are displayed.

level(#) sets the confidence level; see [R] **estimation options**.

twoway([, *twoway_options* yrange($\#_1$ $\#_2$) addplot(*plot*)]) specifies options allowed by `graph twoway`; see [G-3] **twoway options**.

If the twoway option `name()` is not specified, `name(`*namestub_test*`, replace)` is assumed, where *test* is either `test`, `excl`, `reset`, `hett`, `dur`, or `rcr`, according to the minimum abbreviation of the respective `estat` subcommand. The prefix is set with the `kinkyreg` option `namestub(`*namestub*`)`.

yrange($\#_1$ $\#_2$) specifies that the *p*-value or parameter value plots be restricted to the interval [$\#_1$, $\#_2$] on the *y* axis. A missing value for $\#_1$ or $\#_2$ refers to minus or plus infinity, respectively.

addplot(*plot*) allows to overlay the twoway graphs with additional plots; see [G-3] **addplot option**.

pvalplot([*name*|#] [, *line_options* recast(*newplottype*)]) determines the look of the *p*-value or parameter value plots. *line_options* are options allowed by `graph twoway line`; see [G-3] **line options**.

With `estat test`, neither *name* nor # must be specified.

With `estat exclusion`, *name* must be a variable name for the individual exclusion tests. For the joint exclusion test, *name* must not be specified.

14

With `estat reset` or `estat durbinalt`, # must be the integer value of an order specified with option `order()`.

With `estat hettest`, # must be the integer value referring to the #-th specified *varlist*. If option `xb` was specified, the corresponding test is ordered last.

With `estat rcr`, *name* must be `lambda` for Krauth's λ or `delta` for Oster's δ.

`recast(`*newplottype*`)` allows to treat the plot as *newplottype* instead of a line plot; see [G-3] ***advanced options***.

`nograph` suppresses the creation of the graph for KLS inference.

`notable` suppresses display of the results table.

4.3 Stored results

All postestimation commands besides `estat rcr` store the values of the test statistics in the matrices `r(chi2_kls)` or `r(F_kls)`, and the corresponding p-values in the matrix `r(p_kls)`. The command `estat exclusion` furthermore stores in matrix `r(rho)` the values of the endogeneity correlation and corresponding confidence bounds that are implied under validity of the exclusion restrictions.[16] The command `estat rcr` stores in matrix `r(rcr_kls)` the RCR parameter values λ and δ. Various additional scalars are returned by each postestimation command if the option `correlation(`#`)` is specified.

5 Example

5.1 KLS estimation with a single endogenous regressor

We reanalyze data from the National Longitudinal Survey of Young Men used by Griliches (1976) to estimate the returns to schooling while accounting for individual differences in ability. Further control variables are labor market experience, job-specific tenure, location in the South, residence in a metropolitan area, and a set of year dummies.

```
. use http://www.stata-press.com/data/imeus/griliches
(Wages of Very Young Men, Zvi Griliches, J.Pol.Ec. 1976)

. describe lw s expr tenure rns smsa _I* iq age mrt kww

              storage   display    value
variable name   type    format     label       variable label
-----------------------------------------------------------------------
lw              float   %9.0g                  log wage
s               float   %9.0g                  completed years of schooling
expr            float   %9.0g                  experience, years
tenure          float   %9.0g                  tenure, years
rns             float   %9.0g                  residency in South
smsa            float   %9.0g                  reside metro area = 1 if urban
_Iyear_67       byte    %8.0g                  year==67
```

---

16. See Section 5.1 for an example of such an inversion of the exclusion restrictions test.

```
_Iyear_68       byte     %8.0g                    year==68
_Iyear_69       byte     %8.0g                    year==69
_Iyear_70       byte     %8.0g                    year==70
_Iyear_71       byte     %8.0g                    year==71
_Iyear_73       byte     %8.0g                    year==73
iq              float    %9.0g                    iq score
age             float    %9.0g
mrt             float    %9.0g                    marital status = 1 if married
kww             float    %9.0g                    score on knowledge in world of
                                                    work test
```

Because ability as a joint predictor of men's wages and the achieved level of schooling is unobserved, the returns to schooling cannot be consistently estimated by OLS. This omitted-variable bias can be mitigated by using a proxy variable for ability. In the following, it is assumed that by controlling for an individual's IQ score we can account for the relationship between the completed years of schooling and the unobserved ability. However, being an imperfect measure of ability, such a proxy variable usually suffers from measurement error and thus needs to be treated as endogenous.[17] The standard approach is to find instrumental variables that are both relevant and exogenous, i.e. sufficiently correlated with the endogenous variable, validly excluded from the model, and uncorrelated with the measurement error. Such candidate instruments might be the age and the marital status of the individuals.[18]

```
. ivregress 2sls lw s expr tenure rns smsa _I* (iq = age mrt), small
Instrumental variables (2SLS) regression
      Source |       SS       df       MS              Number of obs =      758
-------------+------------------------------           F( 12,   745) =     3.95
       Model | -894.146506    12 -74.5122089           Prob > F      =   0.0000
    Residual |  1033.43266   745  1.38715793           R-squared     =        .
-------------+------------------------------           Adj R-squared =        .
       Total |   139.28615   757  .183997556           Root MSE      =   1.1778

----------------------------------------------------------------------------
          lw |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+--------------------------------------------------------------
          iq |  -.0948902   .0436835    -2.17   0.030    -.1806475   -.0091329
           s |   .3397121   .1266165     2.68   0.007     .0911445    .5882797
        expr |   -.006604   .0288202    -0.23   0.819    -.0631824    .0499745
      tenure |   .0848854   .0330404     2.57   0.010     .0200221    .1497487
         rns |  -.3769393   .1598202    -2.36   0.019    -.6906908   -.0631878
        smsa |   .2181191   .1031496     2.11   0.035     .0156207    .4206175
   _Iyear_67 |   .0077748   .1748639     0.04   0.965    -.3355098    .3510595
   _Iyear_68 |   .0377993   .1631148     0.23   0.817    -.2824202    .3580188
   _Iyear_69 |   .3347027   .1681069     1.99   0.047      .004683    .6647224
   _Iyear_70 |   .6286425   .2507784     2.51   0.012     .1363261    1.120959
   _Iyear_71 |   .4446099   .1843204     2.41   0.016     .0827607     .806459
   _Iyear_73 |    .439027     .15558     2.82   0.005     .1335997    .7444544
       _cons |   10.55096   2.845916     3.71   0.000     4.963995    16.13793
----------------------------------------------------------------------------
```

---

17. See Hayashi (2000, Chapter 3.9) for a discussion of the omitted-variable and ability biases in the context of this application. Griliches (1976) emphasizes that schooling itself might be correlated with the measurement error of the ability proxy. We ignore this complication and treat schooling as exogenous once ability is controlled for.

18. We choose these instruments for illustrative purposes, following Baum et al. (2007). In his original study, Griliches (1976) uses different sets of instruments.

```
        Instrumented:  iq
        Instruments:   s expr tenure rns smsa _Iyear_67 _Iyear_68 _Iyear_69
                       _Iyear_70 _Iyear_71 _Iyear_73 age mrt
      . estat overid
        (output omitted )
      . estat firststage
        (output omitted )
```

The 2SLS estimates yield a relatively high wage return of 34% to one additional year of schooling, while the significantly negative ability effect seems odd. To economize on space, we do not show the detailed output of the postestimation commands. The key statistics of interest are the Sargan test, 1.39 with a p-value of 0.238, and the first-stage F-statistic, 2.72. While the overidentification test seems to indicate that the instruments are valid,[19] their relevance is questionable given a first-stage F-statistic well below 10.

Baum et al. (2007) use this example to illustrate how their `ivreg2` command suite can be used for further weak-instruments diagnostics. Instead, we resort to instrument-free inference with the `kinkyreg` command. Let us focus on the KLS inference for the endogenous regressor, the IQ score, and the main variable of interest, completed years of schooling. For the model specification in Baum et al. (2007), which we label specification A, we obtain the respective graphs shown in Figures 1 and 2 by specifying the option `inference(iq s)`. With the `range(-0.75 0.75)` option, we request to compute and graph the KLS estimates for 151 potential correlations of IQ with the error term in the interval $[-0.75, 0.75]$, given a default step size of 0.01. The appearance of the `kinkyreg` graphs can be fine-tuned with the `twoway()`, `coefplot()`, and `ciplot()` options, enabling the full flexibility of *Stata*'s `graph` command suite, but for simplicity we just start with the factory settings of the *Stata Journal* scheme. We will illustrate some of the graph options further below.

```
      . set scheme sj
      . kinkyreg lw s expr tenure rns smsa _I* (iq = age mrt), range(-0.75 0.75)
      > small inference(iq s)
      Kinky least squares estimation                Number of obs   =       758
```

The wide confidence intervals of the 2SLS estimates immediately strike the eye. This is a well-known consequence of weak instruments. The KLS confidence intervals for a given endogeneity correlation are much narrower.[20] However, the true correlation is unknown and we should consider the union of the confidence intervals over a reasonable range of correlations. In our example, over the whole range from -0.75 to 0.75, the union of KLS confidence intervals is about as wide as the 2SLS confidence interval, although the former is inconclusive regarding the sign of the effect. The KLS and 2SLS

---

19. Keep in mind that overidentification tests cannot actually test the joint validity of both instruments without maintaining the assumption that a linear combination of them is already valid to begin with.

20. Recall that the KLS confidence intervals are (asymptotically) conservative. The coverage asymptotically equals 95% for a given value of ρ if the kurtosis of all regressors is identical and consistently estimated, for example if all regressors are normally distributed. Nevertheless, in finite samples, based on simulation evidence by Kiviet (2020a), the asymptotic confidence intervals may turn out to be slightly liberal.
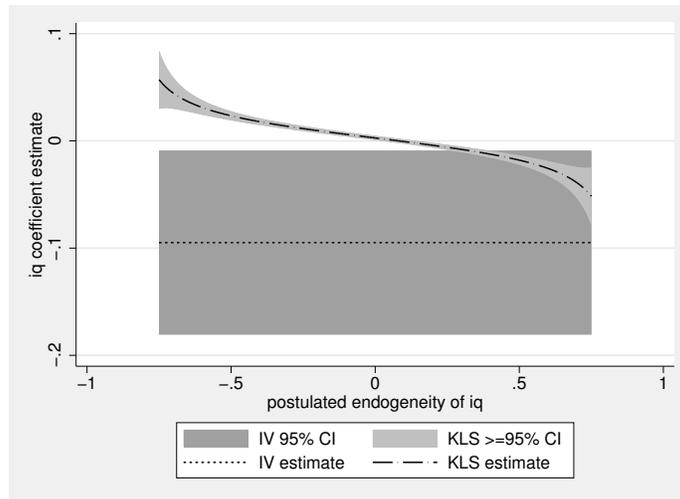
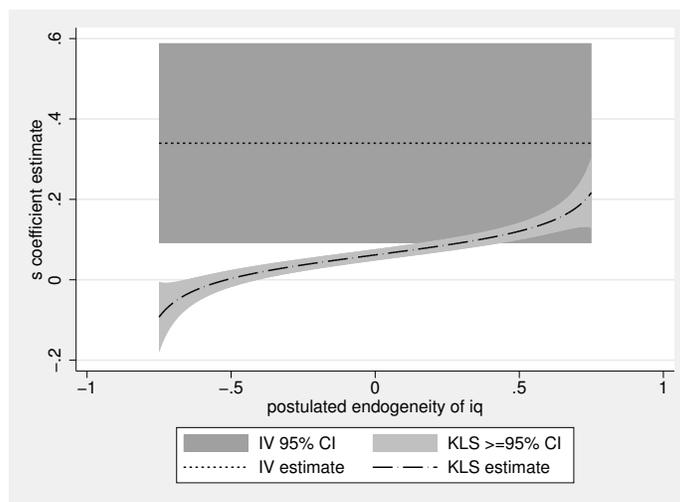Figure 1: KLS and 2SLS coefficient estimates and confidence intervals for `iq` in specification A

Figure 2: KLS and 2SLS coefficient estimates and confidence intervals for `s` in specification A

confidence intervals only overlap for relatively large positive endogeneity correlations and, noteworthy, the 2SLS point estimates are always outside of the KLS intervals over the whole considered range. This observation casts serious doubt on the appropriateness of the chosen instrumental variables.

With prior information on the reasonable range of the endogeneity, we can substan-

tially sharpen the KLS inference. For example, we might be confident that it is less than 0.4 in absolute terms. Moreover, if measurement error is the only source of endogeneity, the correlation of the IQ score with the error term is negative by construction. Due to the resulting attenuation bias, the OLS estimates of the IQ coefficient (which are the KLS estimates with an endogeneity correlation of 0) are biased towards zero. Moreover, we would generally expect the effect of ability on wages to be nonnegative, which is incompatible with positive endogeneity correlations given our KLS estimates, but also at odds with the 2SLS estimate.

Sticking to the measurement error story with an endogeneity range $[-0.4, 0]$, the unions of KLS confidence intervals span the bands $[0.001, 0.021]$ for the IQ coefficient and $[0.001, 0.076]$ for the return to schooling. Instead of reading these numbers from the graphs, we can also display regression output with the confidence intervals for specific endogeneity correlations by replaying the `kinkyreg` command with the `correlation()` option.[21]

```
. kinkyreg, correlation(-0.4)
Kinky least squares estimation                    Number of obs   =        758
Postulated endogeneity of iq = -0.4000
```

lw	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
iq	.0178505	.0015908	11.22	0.000	.0147275	.0209735
s	.018874	.0090115	2.09	0.037	.001183	.036565
expr	.036647	.0073454	4.99	0.000	.0222269	.0510672
tenure	.0355367	.0084409	4.21	0.000	.018966	.0521074
rns	−.0527647	.0312384	−1.69	0.092	−.1140905	.0085611
smsa	.1196815	.0299368	4.00	0.000	.060911	.178452
_Iyear_67	−.0638234	.0538705	−1.18	0.236	−.1695794	.0419327
_Iyear_68	.0872164	.0505387	1.73	0.085	−.0119988	.1864316
_Iyear_69	.1878763	.0494006	3.80	0.000	.0908953	.2848573
_Iyear_70	.1661179	.055196	3.01	0.003	.0577597	.2744761
_Iyear_71	.1882715	.048602	3.87	0.000	.0928583	.2836846
_Iyear_73	.3048592	.0457922	6.66	0.000	.214962	.3947564
_cons	3.255792	.1407933	23.12	0.000	2.979394	3.532191

```
. kinkyreg, correlation(0)
Kinky least squares estimation                    Number of obs   =        758
Postulated endogeneity of iq = 0.0000
```

lw	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
iq	.0027121	.0010225	2.65	0.008	.0007047	.0047195
s	.0619548	.0072159	8.59	0.000	.0477889	.0761207
expr	.0308395	.006454	4.78	0.000	.0181692	.0435097
tenure	.0421631	.0074168	5.68	0.000	.0276028	.0567233
rns	−.0962935	.0273095	−3.53	0.000	−.1499061	−.0426808
smsa	.1328993	.026347	5.04	0.000	.0811762	.1846224
_Iyear_67	−.0542095	.0474401	−1.14	0.254	−.1473416	.0389226
_Iyear_68	.0805808	.0445084	1.81	0.071	−.006796	.1679577

---

21. We could also have directly restricted the KLS estimates in the graphs to this interval with the `range()` option, but the wider range provides a more complete picture that also allows to assess the sensitivity to different assumptions about the endogeneity correlation.

_Iyear_69	.2075915	.0434827	4.77	0.000	.1222282	.2929548
_Iyear_70	.2282237	.0483791	4.72	0.000	.1332481	.3231994
_Iyear_71	.2226915	.0427241	5.21	0.000	.1388176	.3065654
_Iyear_73	.3228747	.0403073	8.01	0.000	.2437453	.4020041
_cons	4.235357	.1123727	37.69	0.000	4.014752	4.455962

The second output are simply the OLS results. Both the IQ and schooling effects are statistically significantly positive, as we would generally expect, but the KLS estimate of the return to schooling is substantially smaller than the 2SLS point estimate. Also, these KLS intervals do not overlap with the corresponding 2SLS confidence intervals, further reducing the confidence in the 2SLS approach, provided our assumptions on the model and the postulated endogeneity range are correct. Admittedly, and as a word of caution, our choice for the lower bound of the endogeneity range is quite arbitrary. If we relax that restriction, the KLS return-to-schooling estimate would turn statistically insignificant. Yet, the confidence interval would expand in the opposite direction from the 2SLS estimate.

While we have seen above that the conventional overidentification tests after the 2SLS regression did not reject the null hypothesis, the weakness of the instruments or the nonexistence of a valid linear combination of the instruments might have been detrimental to the reliability of the test. The KLS approach instead allows us to perform instrument-free inference on the exclusion restrictions with the `estat exclusion` postestimation command. This produces the graph in Figure 3, showing p-values for F-tests (or Wald tests if we did not specify the `small` option in the `kinkyreg` command line) for the significance of the instruments' coefficients when added (jointly or individually) to the regression model, again treating the IQ score as endogenous. To facilitate the interpretation, the y-axis labels are amended using the `twoway()` option to indicate the conventional significance levels.

```
. estat exclusion, twoway(, ylabel(0.01 0.05 0.1)) notable
```

The KLS exclusion restriction tests presented in Figure 3 substantiate our claim that age and marital status are unlikely to be valid instruments. Only for very large positive endogeneity correlations do we not reject the null hypothesis that the instruments are validly excluded from the model. Besides questioning the reliability of the 2SLS estimates, this result also has implications for the KLS approach. If age and martial status are not validly excluded from the model, the KLS estimates would suffer from omitted-variables bias if any of the included regressors is correlated with the excluded variables. In this sample of very young men aged between 16 and 30 years, it is in particular age that is substantially correlated with the schooling regressor, but also with experience and tenure. Clearly, the youngest men in the sample cannot be among those with highest years of schooling or experience. In the following specification B, we have added age and marital status as regressors. If we were to apply 2SLS again, we would have to find another instrument for the endogenous IQ score. The advantage of the KLS approach is that we can obtain valid inference without any instruments. For a compact presentation of the results, we combine all the graphs of interest in the single Figure 4. To improve the visibility of the axis titles and labels, we also manipulate a few of the
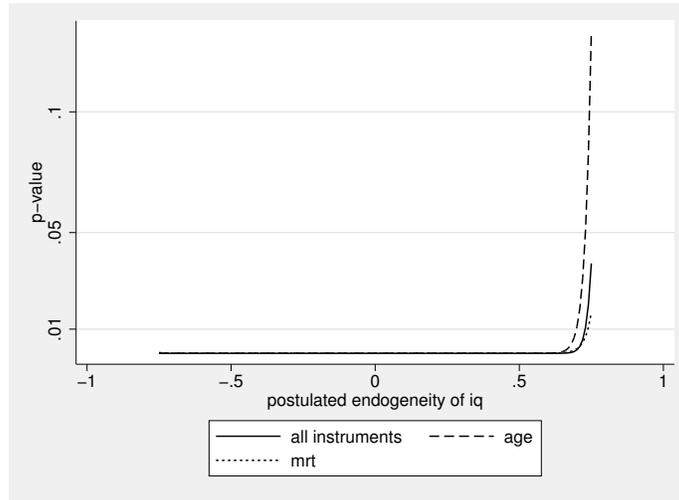
20

Figure 3: *p*-values for three KLS exclusion restriction tests in specification A

graph settings with standard `twoway` options.

```
. local tw_combine_options "xtitle(, size(vlarge)) ytitle(, size(vlarge))
> xlabel(, labsize(vlarge)) ylabel(, labsize(vlarge)) legend(off size(vlarge))
> nodraw"
. kinkyreg lw s expr tenure rns smsa _I* age mrt (iq), range(-0.75 0.75) small
> inference(iq s expr tenure age mrt) twoway(, `tw_combine_options´)
> twoway(mrt, legend(on rows(2)))
Kinky least squares estimation                    Number of obs    =       758
. graph combine kinkyreg_iq kinkyreg_s kinkyreg_expr kinkyreg_tenure
> kinkyreg_age kinkyreg_mrt, altshrink
```

   Directly implied by the previous exclusion restrictions test, the coefficient of age and the marriage premium are statistically significant. Both have a positive sign over the whole range of the IQ endogeneity correlations. We could interpret this positive age effect as the wage return to being more mature, which might be associated with the ability to perform more responsible tasks. Another explanation would be legal working age restrictions for some higher-paying jobs. Focusing again on the endogeneity range $[-0.4, 0]$, the estimated ability effect remains significantly positive, hardly affected by the inclusion of the two additional variables. The schooling effect, however, now turned statistically insignificant after controlling for age and marital status. It appears that the previously found positive return to schooling resulted primarily from the fact that men with many years of school attendance are also older. Labor market experience and job tenure also no longer seem to have a significant effect at this early stage of the individuals' labor market career. The full returns to schooling or experience may only be reaped in later years, while ability makes a difference from the start.[22]

22. See Griliches (1976) for a discussion of the peculiarities in the analysis of labor market data for
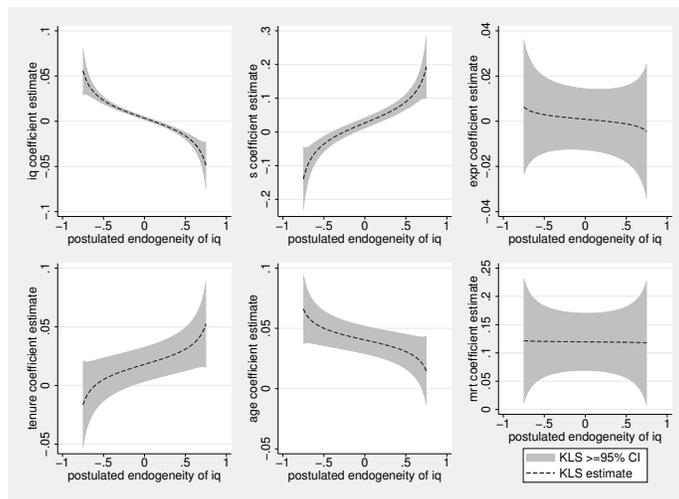
Figure 4: KLS coefficient estimates and confidence intervals in specification B

Above, we used the exclusion restrictions test to investigate whether age and marriage were validly excluded from the model. A similar model misspecification test is the RESET test (Ramsey 1969). By testing the valid exclusion of polynomials in the fitted values, it can hint towards possible functional-form misspecification. By default, the `estat reset` postestimation command computes the test for polynomials in the fitted values up to the fourth order.[23] This is shown in Figure 5 that features added grid lines for the 5% and 10% significance levels.

```
. estat reset, twoway(, ymtick(0.05 0.1, grid))
```

At the 5% significance level, we do not reject the null hypothesis of correct model specification when we use at least a third-order polynomial. However, the evidence is not too comfortable for the endogeneity correlation range that is of particular interest to us.

Because the IQ score may not be an ideal proxy for ability, let us follow Griliches (1976) by considering the knowledge in world of work (KWW) test score as an alternative proxy variable. He suggests to use one of the potential proxy variables as an instrument for the other. While we could carry out the KLS analysis again without any instrument, it is insightful to compare the results for this specification C to just-identified 2SLS estimates with the IQ score as the instrumental variable.

```
. ivregress 2sls lw s expr tenure rns smsa _I* age mrt (kww = iq), small
  (output omitted)
```

---

very young workers.

23. In the calculation of the fitted values, the endogenous IQ regressor is replaced by an estimate of its exogenous variation as in equation (7) to ensure the asymptotic exogeneity of the fitted values in the test regression.
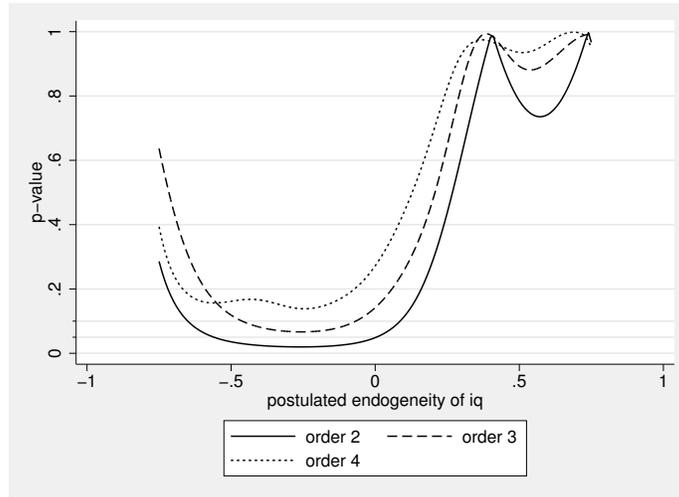
Figure 5: p-values for KLS RESET tests in specification B

```
. estat firststage
  (output omitted)
. estat endogenous
  (output omitted)
```

For brevity, detailed 2SLS results are omitted. The point estimates are 0.028 for the KWW coefficient and 0.003 for the return to schooling. The latter is neither statistically nor economically significant. These 2SLS results are now in line with our KLS evidence, and the confidence intervals are substantially smaller than with the potentially weak and invalid age and marriage instruments. A noteworthy deviation from the previous KLS results is that the 2SLS estimate of the age effect is not statistically significant.

The first-stage F-statistic is 46.1, providing confidence that the instrument is sufficiently strong. The Durbin-Wu-Hausman F-statistic of 8.68 with a p-value of 0.003 supports the assumption that KWW is endogenous. However, this conclusion relies on the validity of the instrument which is untestable in the 2SLS framework because the model is just-identified.[24] The negative sign of the t-statistic version of the Durbin-Wu-Hausman test further indicates a negative endogeneity correlation, in line with the measurement error story.[25]

---

24. The assumption that both ability proxies are measured with error does not invalidate the use of the IQ score as an instrumental variable as long as the measurement errors are uncorrelated with each other.

25. We can compute the t-statistic version of the Durbin-Wu-Hausman test manually as a significance test for the first-stage residuals kww_r in the last line of the following procedure:

```
.  regress kww s expr tenure rns smsa _I* are mrt iq
.  predict kww_r, residuals
.  regress lw kww s expr tenure rns smsa _I* age mrt kww_r
```

Given that the two ability measures are not perfect substitutes, the IQ score might still have a direct effect on wages even after controlling for the KWW test score, thus violating the exclusion restriction. Before we employ again our instrument-free machinery to test the valid exclusion of the IQ score, let us consider another instrument-based approach that has been proposed recently. Conley et al. (2012) propose to obtain interval estimates over a range of plausible values for the direct effect of the excluded instrument in the regression model. Because the support of this direct effect is in principle unbounded, forming a prior belief about a plausible range for it could generally be harder than agreeing on a reasonable range of endogeneity correlations. If this plausible range is chosen too large, the resulting confidence bands will become uninformatively wide. If the range is chosen too small, it might miss the true value.

Earlier, we obtained KLS estimates of a direct effect of the IQ score that is positive but below 0.018, based on the 95% union of confidence intervals within the endogeneity range $[-0.4, 0]$. To treat the IQ score as "plausibly exogenous" (PE) in the sense of Conley et al. (2012), we assume that this effect is at least halved once we control for the KWW score. For direct effects of the IQ score within the interval $[0, 0.009]$, we can then use the `plausexog` package (Clarke and Matta 2018) to obtain the corresponding union of confidence intervals.

```
. plausexog uci lw s expr tenure rns smsa _I* age mrt (kww = iq), gmin(0)
> gmax(0.009)
  (output omitted)
```

Instead of showing the output from the `plausexog` command, let us instead add the resulting confidence bands, $[-0.066, 0.044]$ for KWW, $[-0.023, 0.098]$ for schooling, and $[-0.005, 0.106]$ for age, to the graphical output from the `kinkyreg` command. We can use the `addplot()` suboption to overlay the default KLS graphs with these additional confidence bands. Specifically, we use twoway line plots of a function (which is just a constant in our case) to draw horizontal lines over the endogeneity range $[-0.75, 0.75]$.[26] For aesthetic reasons, we also force these added plots to be drawn before the KLS confidence intervals and we make some adjustments to the legend. The results are shown in Figure 6.

```
. kinkyreg lw s expr tenure rns smsa _I* age mrt (kww = iq), range(-0.75 0.75)
> small inference(kww s age) twoway(, nodraw) twoway(kww, legend(off)
> addplot((function y = 0.044, range(-0.75 0.75) lpattern(shortdash)) ||
> (function y = -0.066, range(-0.75 0.75) lpattern(shortdash)), before(kls_ci)))
> twoway(s, legend(off) addplot((function y = 0.098, range(-0.75 0.75)
> lpattern(shortdash)) || (function y = -0.023, range(-0.75 0.75)
> lpattern(shortdash)), before(kls_ci))) twoway(age, legend(order(1 4 5 6 2)
> label(2 "PE >=95% CI")) addplot((function y = 0.106, range(-0.75 0.75)
> lpattern(shortdash)) || (function y = -0.005, range(-0.75 0.75)
> lpattern(shortdash)), before(kls_ci)))
Kinky least squares estimation                    Number of obs    =       758
. graph combine kinkyreg_kww kinkyreg_s kinkyreg_age, rows(3) ysize(6)
```

26. See [G-2] **graph twoway function**. We could achieve the same effect by first generating new variables for the confidence interval bounds and the endogeneity range, and then overlaying the graph with a range plot with lines; see [G-2] **graph twoway rline**.
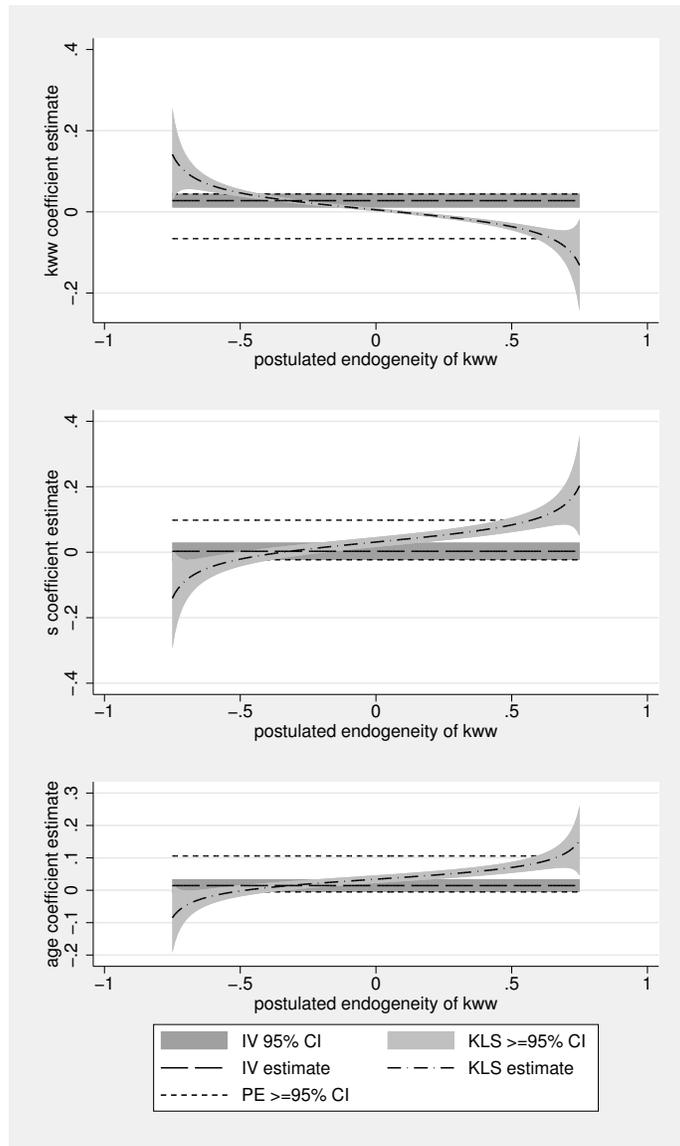
Figure 6: KLS, 2SLS, and PE coefficient estimates and confidence intervals in specification C

While the instrument-based analysis becomes more robust if we allow the IQ score to have a (small) nonnegative direct effect, the resulting widened PE confidence bands make it harder to infer meaningful implications.[27] Most notably, we would no longer

---

27. Notice that the PE upper bound for the KWW score and the lower bound for schooling and age

have conclusive evidence of a positive ability effect. In contrast, the KLS inference remains informative as long as we restrict our attention to a reasonable subset of endogeneity correlations.

```
. kinkyreg, correlation(-0.4)
  (output omitted)
. kinkyreg, correlation(0)
  (output omitted)
```

Maintaining the assumption that the endogeneity of the ability proxy is due to measurement error and therefore negative, the KLS estimate of the ability effect is still significantly positive. The schooling and age profiles over different endogeneity values are now remarkably similar, in contrast to the earlier results with the IQ score as the ability proxy. When KWW is just mildly endogenous, the returns to both schooling and age are statistically significantly positive. Over the endogeneity range $[-0.4, 0]$, the union of KLS confidence intervals covers $[0.001, 0.041]$ for the ability effect, $[-0.025, 0.046]$ for the return to schooling, and $[-0.006, 0.046]$ for the age coefficient. All three intervals encompass the respective 2SLS point estimate. This provides some indication that the IQ score could indeed be a valid and relevant instrument. As a further investigation of this matter, let us look again at the KLS exclusion restriction test. The p-values are shown in Figure 7.

```
. estat exclusion, twoway(, ymtick(0.05 0.1, grid))
Endogeneity of kww compatible with valid exclusion
```

	Corr.	[95% Confid. Bounds]	
iq	-.3183786	-.5207143	-.1120693

If the KWW score was subject to only minor measurement error, the test would still reject the hypothesis of valid exclusion of the IQ score. The output table of the `estat exclusion` command reveals that the null hypothesis is not rejected at the 5% significance level for endogeneity correlations in the interval $[-0.521, -0.112]$. Effectively, the KLS exclusion restriction test is asymptotically equivalent to a test of coefficient equality between the KLS and 2SLS estimates, assuming that our prior belief about the endogeneity correlation is correct. If the exclusion restriction holds, then both the KLS and the 2SLS estimators are consistent. If it does not hold, the 2SLS estimator becomes inconsistent while the KLS estimator remains consistent. The peak of the p-value curve occurs at a correlation of -0.318.[28] Inverting the test, i.e. starting from the assumption of instrument validity, this can be interpreted as the 2SLS-based point estimate of the endogeneity correlation.

---

coincide with the respective 2SLS bounds. This is because we have chosen zero as the lower limit for the direct effect of the IQ score, in which case it would be a valid instrument.

28. As demonstrated by Kiviet (2020a), the p-value asymptotically equals 1 at the actual peak. The estimate of -0.318 is obtained as a weighted average of the two adjacent grid points with highest p-values, -0.32 and -0.31, with weights equal to $1/(1-p)$. The reported bounds are obtained by linear interpolation between the two grid points with p-values just above and just below the significance level.
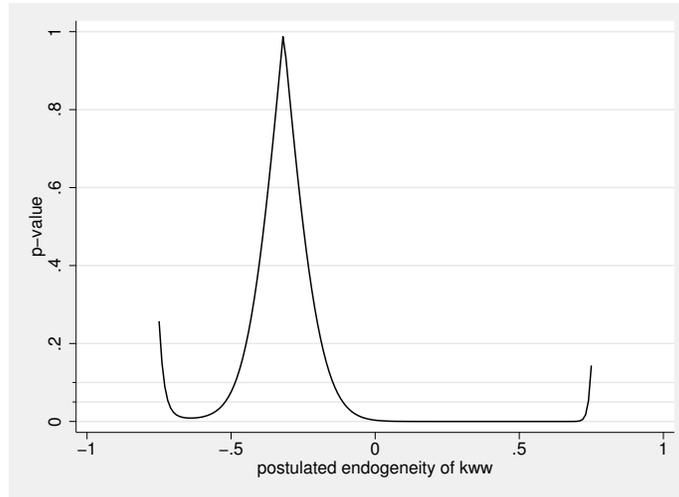
Figure 7: *p*-values for the KLS exclusion restriction test of `iq` in specification C

To reinforce the trust in our KLS results, we can look at further specification tests. For example, we might suspect that squares and interaction terms of some of the regressors have predictive power. Instead of running the less specific RESET test again, we can test the exclusion restrictions for some of these terms, one at a time. The corresponding *p*-value curves are shown in Figure 8.

```
. estat exclusion c.expr#c.expr c.tenure#c.tenure c.age#c.age c.expr#c.tenure
> c.expr#c.age c.tenure#c.age, twoway(, ymtick(0.05 0.1, grid)) nojoint notable
```

Most squares and interaction terms appear to be validly excluded, besides the interaction effect between tenure and age.[29] This indicates that the return to tenure varies with age but does not yet tell us anything about the magnitude or sign of this effect. In our specification D, we therefore include this interaction term in our regression model and compute the marginal return of tenure at three different ages, 18, 24, and 30 years:

$$\beta_{\texttt{tenure}} + 18 \cdot \beta_{\texttt{c.tenure\#c.age}}$$
$$\beta_{\texttt{tenure}} + 24 \cdot \beta_{\texttt{c.tenure\#c.age}}$$
$$\beta_{\texttt{tenure}} + 30 \cdot \beta_{\texttt{c.tenure\#c.age}}$$

We can do this with the `lincom()` option of `kinkyreg`. In addition, let us test the simple linear hypotheses whether the respective marginal effects differ from the return to experience, i.e. whether any of the above effects statistically differs from the coefficient $\beta_{\texttt{expr}}$. The latter we do with the `estat test` postestimation command. We combine all graphs in Figure 9.

---

29. To be precise, we can only conclude that the insignificant interaction terms are indeed validly excluded after re-running the exclusion restriction tests for the model augmented with the tenure-age interaction effect. The conclusions do not change in our case.
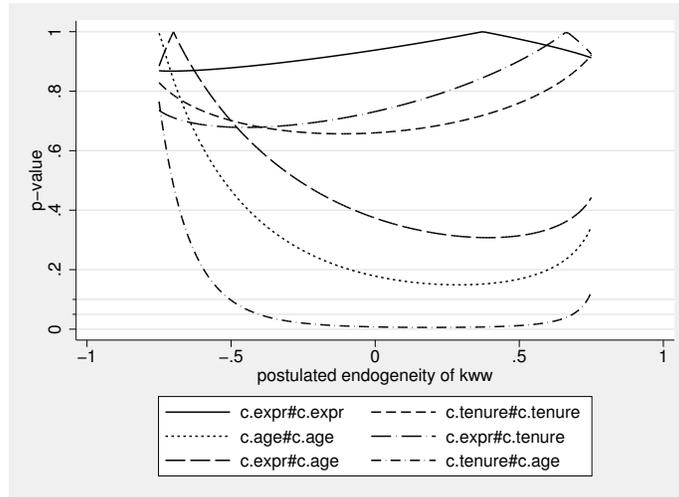
Figure 8: *p*-values for various KLS exclusion restriction tests in specification C

```
. kinkyreg lw s expr tenure rns smsa _I* age mrt c.tenure#c.age (kww),
> range(-0.75 0.75) small lincom(1: tenure+c.tenure#c.age*18)
> lincom(2: tenure+c.tenure#c.age*24) lincom(3: tenure+c.tenure#c.age*30)
> twoway(, ylabel(-0.15(0.05)0.2) `tw_combine_options´)
> twoway(1, title("age = 18")) twoway(2, title("age = 24"))
> twoway(3, title("age = 30"))
Kinky least squares estimation                    Number of obs    =        758
. estat test tenure+c.tenure#c.age*18=expr, twoway(, ylabel(0(0.2)1)
> ymtick(0.05 0.1, grid) `tw_combine_options´ name(kinkyreg_test_1))
. estat test tenure+c.tenure#c.age*24=expr, twoway(, ymtick(0.05 0.1, grid)
> `tw_combine_options´ name(kinkyreg_test_2))
. estat test tenure+c.tenure#c.age*30=expr, twoway(, ymtick(0.05 0.1, grid)
> `tw_combine_options´ name(kinkyreg_test_3))
. graph combine kinkyreg_1 kinkyreg_2 kinkyreg_3 kinkyreg_test_1
> kinkyreg_test_2 kinkyreg_test_3, altshrink
```

We observe that the return to tenure increases with age. For the youngest, who just started their labor market careers, tenure does not determine the wage outcome, irrespective of the postulated endogeneity of the ability measure. At an age of 24 years, the point estimate of the return to tenure is positive throughout, although economically small and statistically significant only for a moderate endogeneity of ability. Because there is still not much difference between job-specific tenure and overall labor market experience at such an early age, it is not surprising to find no statistical difference between the two effects. For the oldest in our sample, the marginal effect of tenure rises further and is now statistically significant over the whole range of endogeneity correlations that we considered to be reasonable, $r \in [-0.4, 0]$. Moreover, we now reject the null hypothesis that the returns to tenure and experience are equal. At this age, the accumulation of job-specific knowledge and skills eventually pays off.
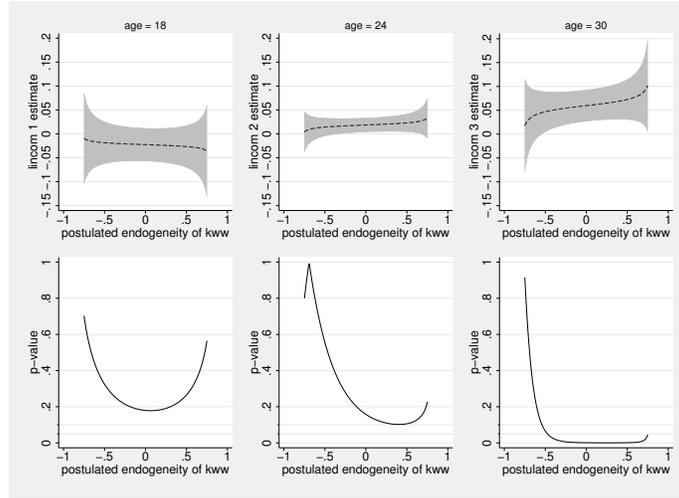
Figure 9: KLS estimates and confidence intervals in specification D of the return to `tenure`, and p-values for linear hypothesis tests of equality of the returns to `tenure` and `expr`, when `age` equals 18, 24, or 30, respectively

Let us scrutinize our regression specification D again with some specification tests. Figure 10 displays the results from RESET tests. The left subfigure shows p-value curves for tests based on polynomials in the fitted values. The right subfigure considers polynomials in all of the right-hand side variables.[30] To economize on the degrees of freedom, we only consider second- and third-order polynomials for this second variant of the test.

```
. estat reset, twoway(, ymtick(0.05 0.1, grid) `tw_combine_options´
> legend(on rows(1)) name(kinkyreg_reset_xb))
. estat reset, rhs order(2 3) twoway(, ymtick(0.05 0.1, grid) `tw_combine_options´
> legend(on) name(kinkyreg_reset_rhs))
. graph combine kinkyreg_reset_xb kinkyreg_reset_rhs, ysize(1.5) altshrink
```

The results are now much more reassuring than those for our initial model specification. The RESET tests with polynomials of the fitted values in the left-hand graph of Figure 10 would still cause some worries if we believed in a quite strong negative endogeneity correlation of the KWW score.[31]

Next, we use the `estat hettest` postestimation command to take a look at some KLS versions of Breusch and Pagan (1979) heteroskedasticity tests. Because homoskedasticity was assumed in the derivation of the KLS formulae, evidence of heteroskedasticity may cast doubt on the robustness of our results. We consider four variants of the test,

---

30. For the endogenous KWW regressor, we use again just an estimate of its exogenous variation as in equation (7).
31. While a chosen significance level provides a useful reference point, it should not be treated as a hard threshold. Low p-values above the significance level are still not entirely reassuring.
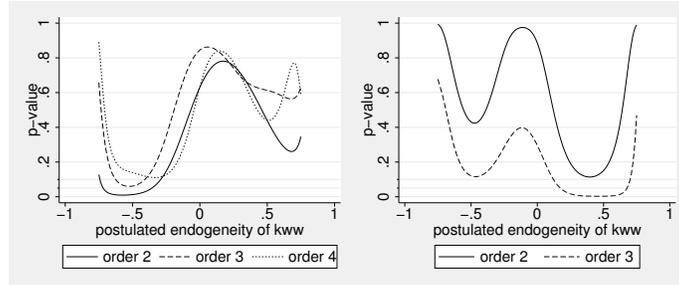
Figure 10: p-values for KLS RESET tests in specification D

in which we let different sets of variables enter the conditional heteroskedasticity model: (i) the right-hand side variables (option `rhs` with the empty variable list), (ii) the right-hand side variables and the instrument (option `rhs` and the first nonempty variable list), (iii) the right-hand side variables plus some interaction effects (option `rhs` and the second nonempty variable list), and (iv) the fitted values (option `xb`).[32] Because joint hypotheses tests with a large number of restrictions might have low power to detect a violation of just a few restrictions entailed by the null hypothesis, we also display a graph that shows the minimum of the p-values among all individual significance tests for a given variable list. This is achieved by adding option `minp`. Figure 11 shows all results in a single graph, with the joint significance tests in the left subfigure and the minimum p-values in the right subfigure.

```
. estat hettest () (iq) (c.expr#c.expr c.tenure#c.tenure c.age#c.age
> c.expr#c.tenure c.expr#c.age), xb rhs twoway(, ymtick(0.05 0.1, grid)
> `tw_combine_options´ legend(on) name(kinkyreg_hett))

. estat hettest () (iq) (c.expr#c.expr c.tenure#c.tenure c.age#c.age
> c.expr#c.tenure c.expr#c.age), rhs minp twoway(, ylabel(0(0.2)1)
> ymtick(0.05 0.1, grid) `tw_combine_options´ legend(on) name(kinkyreg_hett_minp))

. graph combine kinkyreg_hett kinkyreg_hett_minp, ysize(1.5) altshrink
```
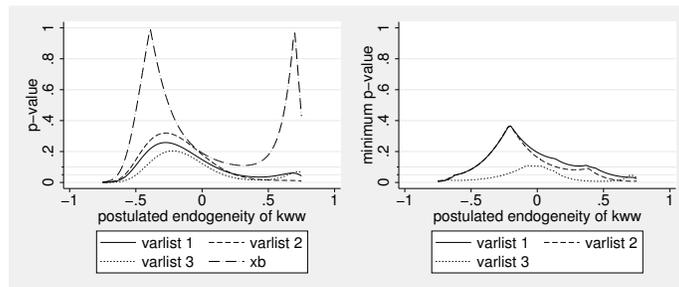

Figure 11: p-values for KLS heteroskedasticity tests in specification D

32. The fitted values in (iv) and the endogenous right-hand side variable in (i)–(iii) are again adjusted to correct for the endogeneity.

30

The joint hypotheses tests do not reject the null hypothesis of no conditional heteroskedasticity within our range of most reasonable endogeneity correlations. Just for the most flexible specification (iii), we find at least one regressor in the auxiliary regression with a statistically significant coefficient for negative endogeneity correlations of at least -0.25. While we could add further interaction terms to our regression model in an attempt to mitigate any heteroskedasticity concerns, the quantitative and qualitative conclusions would hardly change. Since most specification tests are already supportive for our chosen model, we are confident that the insights we have drawn from our KLS analysis are meaningful and statistically well grounded. Having said that, the analysis stands and falls with our maintained assumption that the ability proxy has a moderately negative correlation with the error term, consistent with a measurement error story, and that all remaining regressors are exogenous.

The KLS procedure is related to the alternative instrument-free approaches proposed by Krauth (2016) and Oster (2019). Here, we briefly illustrate that all three methods coincide by translating the endogeneity correlation into the respective sensitivity parameters of the other two approaches. Krauth (2016) places bounds on a relative correlation restriction parameter λ. This is the ratio of the endogeneity correlation to the correlation of the endogenous regressor with an index of the control variables. For a given choice of the endogeneity correlation, say our lower bound $r_l = -0.4$, we can obtain $\widehat{\lambda}(r_l) = r_l/Corr(x_{1i}, \mathbf{x}'_{2i}\widehat{\boldsymbol{\beta}}_2(r_l))$ with the respective KLS estimate $\widehat{\boldsymbol{\beta}}_2(r_l)$. Oster (2019) uses a similar measure of relative variability as her main sensitivity parameter δ. It is obtained by scaling Krauth's parameter with a ratio of standard deviations, $\widehat{\delta}(r_l) = \widehat{\lambda}(r_l) \cdot SD(\mathbf{x}'_{2i}\widehat{\boldsymbol{\beta}}_2(r_l))/SD(\widehat{\varepsilon}_i(r_l))$, where $\widehat{\varepsilon}_i(r_l)$ are the KLS residuals. The kinkyreg postestimation command estat rcr computes these RCR parameters for the whole range of considered endogeneity correlations. The resulting graphs are shown in Figure 12. To zoom in, we have truncated the y axis to the interval $[-5, 5]$ with the twoway() suboption yrange(-5 5) for the second subfigure. Furthermore, with option correlation(-0.4), the values corresponding to $r_l = -0.4$ can be displayed and stored as scalars r(lambda) and r(delta), respectively.

```
. estat rcr, twoway(, `tw_combine_options´ legend(on) name(kinkyreg_rcr))
. estat rcr, correlation(-0.4) twoway(, yrange(-5 5) `tw_combine_options´
> legend(on) name(kinkyreg_rcr_yrange))
Postulated endogeneity of kww = -0.4000
Relative correlation restriction sensitivity parameters
Krauth´s lambda =   -4.1450
Oster´s delta   =   -2.7802
. local lambda = r(lambda)
. local delta = r(delta)
. graph combine kinkyreg_rcr kinkyreg_rcr_yrange, ysize(1.5) altshrink
```

We immediately notice that the functions $\widehat{\lambda}(r)$ and $\widehat{\delta}(r)$ have a singularity. This is the point where the correlation of the endogenous regressor with the index of the control variables switches signs. The RCR estimators are not defined if this correlation equals zero, or if no control variables are present at all. When we approach this singularity from the left or right, the functions tend to $\pm\infty$. Furthermore, without additional
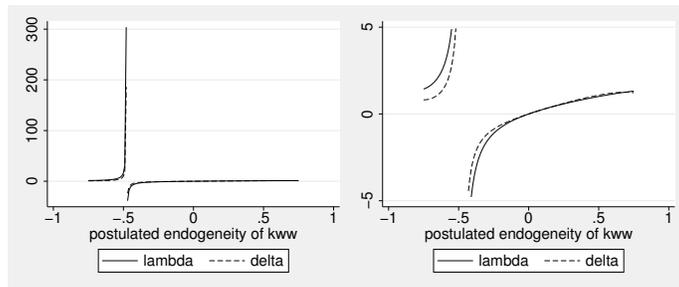
31

Figure 12: Corresponding RCR values λ and δ for specification D

restrictions, the RCR estimators may have multiple solutions for a given value of the sensitivity parameter.[33] For example, with the data at hand, a value of $\delta = 1.24$ corresponds to three different endogeneity correlations of about -0.63, 0.64, and 0.73. In contrast, the KLS estimator is monotonous in r and does not suffer from these peculiarities.

For $r_l = -0.4$, we have obtained $\widehat{\lambda}(r_l) = -4.145$ and $\widehat{\delta}(r_l) = -2.78$. With these values, we can replicate the KLS bounds for the KWW coefficient with the `rcr` and `psacalc` commands of Krauth (2016) and Oster (2019), respectively. The latter works as a postestimation command for `regress`.[34]

```
. generate tenure_age = c.tenure#c.age
. rcr lw kww s expr tenure rns smsa _I* age mrt tenure_age, lambda(`lambda´ 0)
(output omitted)
. regress lw kww s expr tenure rns smsa _I* age mrt c.tenure#c.age
(output omitted)
. psacalc beta kww, delta(`delta´)
(output omitted)
```

The KWW coefficient estimates from the KLS and both RCR procedures are identical. The standard errors reported by the `rcr` command differ slightly from those computed by `kinkyreg`. While the latter uses the analytical formula from Kiviet (2020a) for the asymptotic variance-covariance matrix, Krauth (2016) uses delta-method techniques that may provide a poor approximation of the finite-sample distribution when the slope of $\widehat{\lambda}(r)$ is either relatively large or relatively small.[35] The `psacalc` command does not calculate any standard errors due to a lack of asymptotic results. Oster (2019) proposes to instead compute bootstrap standard errors, which could be done with the `bootstrap` prefix command.

A disadvantage of the `rcr` and `psacalc` commands is that they do not report coef-

---

33. See Krauth (2016) and Oster (2019) for computational details about their methods.

34. Note that for our upper bound, $r_u = 0$ and $\widehat{\lambda}(r_u) = \widehat{\delta}(r_u) = 0$, all three approaches collapse to the OLS estimator.

35. Even for the OLS case, the standard errors reported by `rcr` and `regress` differ, with and without option `vce(robust)` for the latter.

ficient estimates for the control variables. We can manually recover them by removing the effect of the endogenous variable from the dependent variable, and then running a regression on the control variables. While the resulting coefficient estimates coincide with the KLS estimates, the standard errors are incorrect. If the RCR method is the starting point, say because there is prior information about the sensitivity parameter δ (or λ), it is therefore advisable to calculate the corresponding value r and then apply the KLS method to obtain the full set of coefficient estimates with correct standard errors. The following code lines illustrate how we can map the RCR results back into the KLS estimates.

```
. local delta = r(delta)
. generate u = lw - kww * r(beta)
. regress u s expr tenure rns smsa _I* age mrt c.tenure#c.age
  (output omitted)
. predict cb, xb
. summarize cb
  (output omitted)
. local sd_cb = r(sd)
. generate e = u - cb
. summarize e
  (output omitted)
. local sd_e = r(sd)
. corr kww cb
  (output omitted)
. local rho = `delta´ * `sd_e´ / `sd_cb´ * r(rho)
. kinkyreg lw s expr tenure rns smsa _I* age mrt c.tenure#c.age (kww),
> range(`rho´ `rho´) correlation(`rho´) small nograph
  (output omitted)
```

While we can numerically match the coefficient estimates with the different estimators, at least as long as relevant control variables are present, it should be noted that measurement error as the source of endogeneity is not the ideal example for the RCR methods. The RCR sensitivity parameters are usually interpreted as a measure for "the relative selection on observables and unobservables" (Oster 2019) in an evaluation of the OLS robustness to omitted-variables bias. In this sense, measurement error would not be seen as an omitted control variable. Oster (2019) considers an additional sensitivity parameter, the maximum R-squared that is attainable from a hypothetical regression that includes all unobserved control variables. If there is remaining unexplained variation, for instance due to measurement error, this maximum R-squared would be smaller than its default value 1. Yet, the illustrated equivalence of the three instrument-free methods only holds if this hypothetical maximum R-squared is set to 1 in Oster's approach. This does not invalidate the KLS approach, which is completely flexible regarding the source of the endogeneity, but the corresponding RCR sensitivity parameters would have to be interpreted with caution.

In general, it might be difficult to pick reasonable intervals for δ or λ, not least

because there are no natural bounds for these sensitivity parameters.[36] In contrast, the endogeneity correlation ρ is bounded by construction and a restriction of its sign is often credible.

5.2 KLS estimation with multiple endogenous regressors

For the exclusion restriction test that is underlying Figure 7, the attentive reader might have noticed that we implicitly assumed the variable `iq` to be uncorrelated with the error term of the auxiliary regression. However, if we follow our argumentation that the ability proxies are measured with error, this assumption is violated. When we augment the model specification D with the second proxy variable, we thus have two endogenous regressors. In this specification E, varying both endogeneity correlations simultaneously yields a three-dimensional grid of coefficient estimates and corresponding confidence intervals. While there are no theoretical limits to the number of endogenous variables in the model, implementing a general package for flexible KLS inference quickly reaches computational limitations. It also requires a different approach to the visualization of the results, already when there are just two endogenous variables.

The `kinkyreg` command does not attempt to provide a full-fledged solution to these complications. Yet, it allows to estimate the model with an arbitrary number of endogenous regressors by fixing all but one endogeneity correlations at user-specified values. The produced graphs can be regarded as two-dimensional slices through a multi-dimensional surface. The user can then call the `kinkyreg` command multiple times with different values for the correlations to produce a set of slices as desired.

To illustrate this approach, let us vary the endogeneity correlation of the KWW score automatically over the range $[-0.75, 0.75]$, but choose fixed values for the correlation of the IQ score with the error term from the set $\{-0.4, -0.2, 0\}$, one at a time. We do this with a simple loop and the option `endogeneity()` that gets filled with the respective value for the endogeneity correlation of `iq`. The second entry of that option is set to missing which indicates that the endogeneity for the second endogenous variable, `kww`, should be varied automatically. We eventually plot the coefficient estimates and confidence intervals for the two endogenous variables and the exogenous schooling variable in the single Figure 13:

```
. forvalues i = 1 / 3 {
2. local endo : word `i´ of -0.4 -0.2 0
3. quietly kinkyreg lw s expr tenure rns smsa _I* age mrt c.tenure#c.age (iq kww),
> endogeneity(`endo´ .) range(-0.75 0.75) small inference(iq kww s)
> twoway(, yrange(-0.2 0.2) ylabel(-0.2(0.1)0.2) xlabel(-1(0.5)1)
> `tw_combine_options´) twoway(iq, ylabel(-0.1(0.1)0.2)
> title("postulated endogeneity of iq fixed at `endo´") name(kinkyreg_iq_`i´))
> twoway(kww, name(kinkyreg_kww_`i´)) twoway(s, legend(on) name(kinkyreg_s_`i´))
4. }

. graph combine kinkyreg_iq_1 kinkyreg_iq_2 kinkyreg_iq_3 kinkyreg_kww_1
```

---

36. Altonji et al. (2005) and Oster (2019) suggest a bound of 1 for the sensitivity parameter δ, which they interpret as equal selection on observed and unobserved variables. Cinelli and Hazlett (2020) critically discuss this approach and conclude that δ is not easily interpretable.

```
> kinkyreg_kww_2 kinkyreg_kww_3 kinkyreg_s_1 kinkyreg_s_2 kinkyreg_s_3, altshrink
```
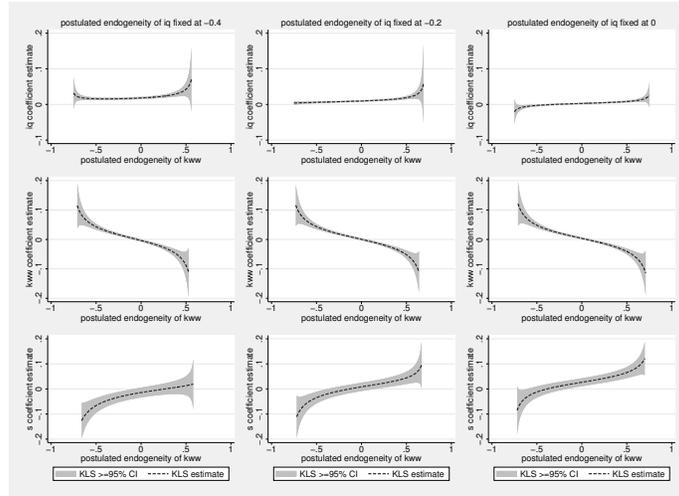

Figure 13: KLS coefficient estimates and confidence intervals in specification E

Notice that some of the graphs do not extend over the full range from -0.75 to 0.75. This is because the feasible range of endogeneity correlations becomes tighter when we have multiple endogenous variables. To avoid distorted pictures from very wide confidence intervals towards the boundaries, we have truncated the y axis with the `twoway()` suboption `yrange(-0.2 0.2)`. This has the effect that the portions of the plots are omitted where the confidence interval spans beyond ± 0.2.

The right-most column displays results when `iq` is treated as exogenous. In line with the results from the exclusion restriction test of the previous subsection, the IQ score could be excluded from the model if it was indeed exogenous, unless there is only a small negative or even positive endogeneity of the KWW score.[37] However, when we allow the IQ score to be endogenous itself, its direct effect becomes statistically significant, as evidenced in the first two columns. While there is hardly any noticeable effect of the endogeneity correlation of IQ on the coefficient estimate of the KWW score, the return to schooling decreases slightly with an increasingly negative endogeneity of the IQ score. When the endogeneity correlation of the latter is -0.4, the return to schooling even becomes statistically significantly negative for most of the plausible negative endogeneity correlations of the KWW score.

We could carry out further specification tests and redo our analysis for the returns to tenure and labor market experience, but to economize on space we leave this as an exercise to the interested reader. Instead, we illustrate how one can produce three-dimensional surface plots and contour plots, varying both endogeneity correlations. To

---

37. We only showed the exclusion restriction test for specification C. The result is virtually unchanged for specification D.

achieve this, we use the `kinkyreg2dta` command to create a new data set with the coefficient estimates for the variable of interest, `s`, and the *p*-values from the corresponding statistical significance test. Here, we restrict ourselves to the endogeneity interval $[-0.4, 0]$ for both ability proxies. Note that `kinkyreg2dta` is a simple wrapper command for `kinkyreg` that allows for multiple endogenous regressors. Internally, it calls `kinkyreg` (and the respective postestimation commands) for each value of `iq`'s endogeneity correlation to compute the desired results over the grid range of `kww`'s endogeneity correlations. We ask for the data to be generated in a new frame, and subsequently we obtain the surface and contour plots in Figure 14. For the former, we use the community-contributed `surface` package (Mander 1999).

```
. kinkyreg2dta lw s expr tenure rns smsa _I* age mrt c.tenure#c.age (iq kww),
> range(-0.4 0) coef(b: s) estat(1 p: test s) frame(kinkyreg) small
. label var _b_s "s coefficient estimate"
. label var _test_1_p_1 "p-value"
. label var _rho_iq "postulated endogeneity of iq"
. label var _rho_kww "postulated endogeneity of kww"
. surface _rho_iq _rho_kww _b_s, plotregion(color(gs15)) ytitle(endog. kww) nodraw
> name(surface_s)
. surface _rho_iq _rho_kww _test_1_p_1, plotregion(color(gs15)) ytitle(endog. kww)
> nodraw name(surface_p)
. twoway contour _b_s _rho_kww _rho_iq, ccuts(-0.06(0.01)0.03) nodraw
> name(contour_s)
. twoway contour _test_1_p_1 _rho_kww _rho_iq, ccuts(0.01 0.05 0.1) nodraw
> name(contour_p)
. graph combine surface_s surface_p contour_s contour_p, altshrink
```
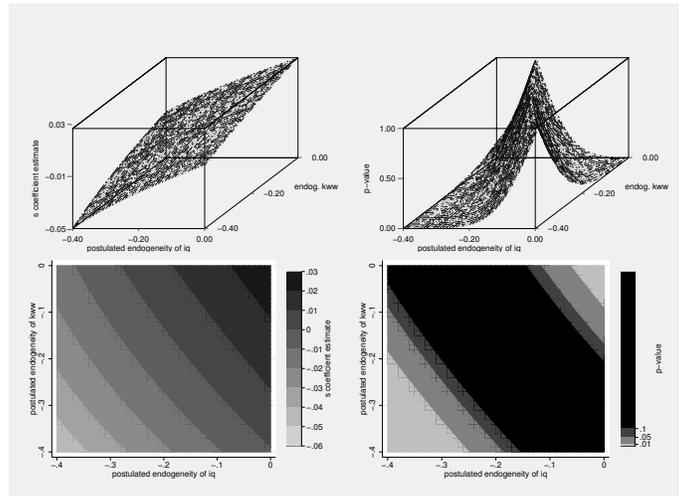

Figure 14: Surface plots for the KLS coefficient estimates of `s` and the *p*-values for the corresponding statistical significance tests in specification E

This figure highlights again the positive relationship between the return to schooling and the correlations of the ability proxies with the error term. A statistically significantly positive return to schooling is only consistent with a small negative endogeneity of both ability variables, while large negative endogeneities yield implausible statistically significantly negative returns to schooling.

6 Conclusion

In this article, we introduced the `kinkyreg` command for kinky least squares estimation of linear regression models. For models with endogenous regressors, the KLS approach provides valid confidence intervals for the regression coefficients, adopting a credible range for the endogeneity correlation. In many applications, researchers might have a strong prior belief about such a credible endogeneity range, while it is often more difficult to justify identifying exclusion restrictions. Our outlined instrument-free approach can provide more decisive evidence on the validity of exclusion restrictions than overidentification tests. Besides, the approach is not vulnerable to the familiar weak-instruments problem of instrument-based methods. Eventually, no approach strictly dominates the other. Pursuing instrument-free inference can be a reasonable standalone approach, or it can complement instrument-based methods.

7 Acknowledgement

8 References

Altonji, J. G., T. E. Elder, and C. R. Taber. 2005. Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy* 113(1): 151–184.

Andrews, D. W. K., and J. H. Stock. 2007. Inference with weak instruments. In *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, ed. R. Blundell, W. K. Newey, and T. Persson, chap. 6, 122–173. Vol. 3. Cambridge University Press.

Andrews, I., J. H. Stock, and L. Sun. 2019. Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics* 11: 727–753.

Baum, C. F., M. E. Schaffer, and S. Stillman. 2003. Instrumental variables and GMM: Estimation and testing. *Stata Journal* 3(1): 1–31.

———. 2007. Enhanced routines for instrumental variables/generalized method of moments estimation and testing. *Stata Journal* 7(4): 465–506.

Breusch, T. S., and A. R. Pagan. 1979. A simple test for heteroscedasticity and random coefficient variation. *Econometrica* 47(5): 1287–1294.

Cinelli, C., and C. Hazlett. 2020. Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society, Series B* 82(1): 39–67.

Clarke, D., and B. Matta. 2018. Practical considerations for questionable IVs. *Stata Journal* 18(3): 663–691.

Conley, T. G., C. B. Hansen, and P. E. Rossi. 2012. Plausibly exogenous. *Review of Economics and Statistics* 94(1): 260–272.

Durbin, J. 1970. Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables. *Econometrica* 38(3): 410–421.

Finlay, K., and L. M. Magnusson. 2009. Implementing weak-instrument robust tests for a general class of instrumental-variables models. *Stata Journal* 9(3): 398–421.

Griliches, Z. 1976. Wages of very young men. *Journal of Political Economy* 84(4): S69–S86.

Hall, A. R., G. D. Rudebusch, and D. W. Wilcox. 1996. Judging instrument relevance in instrumental variables estimation. *International Economic Review* 37(2): 283–298.

Hayashi, F. 2000. *Econometrics*. Princeton: Princeton University Press.

Kiviet, J. F. 2013. Identification and inference in a simultaneous equation under alternative information sets and sampling schemes. *Econometrics Journal* 16(1): S24–S59.

———. 2020a. Instrument-free inference under confined regressor endogeneity and mild regularity. Unpublished manuscript.

———. 2020b. Testing the impossible: Identifying exclusion restrictions. *Journal of Econometrics* 218(2): 294–316.

Krauth, B. 2016. Bounding a linear causal effect using relative correlation restrictions. *Journal of Econometric Methods* 5(1): 117–141.

Mander, A. 1999. 3D surface plots. *Stata Technical Bulletin* 51: 7–10.

Mikusheva, A., and B. P. Poi. 2006. Tests and confidence sets with correct size when instruments are potentially weak. *Stata Journal* 6(3): 335–347.

Moreira, M. J., and B. P. Poi. 2003. Implementing tests with correct size in the simultaneous equations model. *Stata Journal* 3(1): 57–70.

Nevo, A., and A. M. Rosen. 2012. Identification with imperfect instruments. *Review of Economics and Statistics* 94(3): 659–671.

Oster, E. 2019. Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics* 37(2): 187–204.

Parente, P. M. D. C., and J. M. C. Santos Silva. 2012. A cautionary note on tests of overidentifying restrictions. *Economics Letters* 115(2): 314–317.

Pflueger, C. E., and S. Wang. 2015. A robust test for weak instruments in Stata. *Stata Journal* 15(1): 216–225.

Ramsey, J. B. 1969. Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society, Series B* 31(2): 350–371.

Roodman, D., J. G. MacKinnon, M. Ø. Nielsen, and M. D. Webb. 2019. Fast and wild: Bootstrap inference in Stata using boottest. *Stata Journal* 19(1): 4–60.

Stock, J. H., J. H. Wright, and M. Yogo. 2002. A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics* 20(4): 518–529.

Sun, L. 2018. Implementing valid two-step identification-robust confidence sets for linear instrumental-variables models. *Stata Journal* 18(4): 803–825.

Wooldridge, J. M. 2020. *Introductory Econometrics: A Modern Approach*. 7th ed. Cengage Learning.