

Heteroskedasticity-consistent instrument-free inference in linear regressions with endogenous regressors

Sebastian Kripfganz¹

and

Jan F. Kiviet²

 ¹ University of Exeter Business School, Department of Economics, Exeter, UK
 www.kripfganz.de

 ² University of Amsterdam, Amsterdam School of Economics, The Netherlands

UK Stata Conference, 3rd September 2026, London



University of Exeter
Business School

Instrument-free inference

- The `kinkyreg` Stata package (Kripfganz and Kiviet, 2021) enables instrument-free [kinky least-squares](#) (KLS) estimation (Kiviet, 2020, 2023):
 - KLS analytically corrects the bias of OLS for all values of the endogeneity correlations on a specified grid.
 - Set identification is achieved by confining the admissible degree of regressor endogeneity within plausible bounds.
 - For a reasonably narrow range of endogeneity correlations, KLS confidence intervals are often narrower than those from 2SLS, in particular if instruments are weak.
 - Exclusion restrictions are testable within the KLS framework.
- The KLS bias correction relies on a homoskedasticity assumption, which is often hard to justify.
- In the spirit of KLS, we present a [kinky method-of-moments](#) (KMM) approach that can accommodate heteroskedasticity.

Kinky least-squares estimation

- Linear regression model:

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + u_i$$

where some of the regressors $\mathbf{x}_i$ are endogenous; i.e., $\text{Corr}(\mathbf{x}_i, u_i) = \rho \neq 0$.

- KLS analytically corrects the inconsistency of OLS for given values of ρ :

$$\hat{\boldsymbol{\beta}}_{KLS}(\rho) = \hat{\boldsymbol{\beta}}_{OLS} - B(\hat{\sigma}_{KLS}(\rho); \rho)$$

where $\hat{\sigma}_{KLS}^2(\rho)$ is a consistent estimator of $\text{Var}(u_i) = \sigma^2$ under homoskedasticity.

- In practice, $\hat{\sigma}_{KLS}^2(\rho)$ is obtained by bias correcting the OLS variance estimator $\hat{\sigma}_{OLS}^2$ for given values of ρ .
- Basic Stata syntax; see Kripfganz and Kiviet (2021) and the help file:
`kinkyreg depvar exovars (endovars), range(# #) options`

Kinky method-of-moments estimation

- Under heteroskedasticity, $\text{Var}(u_i) = \sigma_i^2$, using robust standard errors is insufficient because the bias expression itself is a function of $\bar{\sigma} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sigma_i$.
- Importantly, a consistent estimator of $\overline{\sigma^2} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sigma_i^2$ is not of help because $\sqrt{\sigma^2} \neq \sqrt{\bar{\sigma}^2} = \bar{\sigma}$ (unless $\sigma_i = \sigma$ for all i).
- When $u_i \sim \mathcal{N}(0, \sigma_i^2)$ (and similarly for t -distributed errors), we can use results for folded distributions: $E[|u_i|] = \sigma_i \sqrt{\frac{2}{\pi}}$

Kinky method-of-moments estimation

- To jointly estimate $\bar{\sigma}$ and β , we can then resort to nonlinear method of moments estimation for given values of ρ :

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N E[\mathbf{x}_i u_i] - b(\bar{\sigma}; \rho) = \mathbf{0}$$

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N E[|u_i|] - \bar{\sigma} \sqrt{\frac{2}{\pi}} = 0$$

- Basic Stata syntax:
`kinkyreg depvar exovars (endovars), range(# #) mm robust options`

Kinky method-of-moments estimation

- Unlike KLS, inference with KMM is straightforward as the asymptotic variance matrix of $(\hat{\beta}_{KMM}(\rho)', \hat{\sigma}_{KMM}(\rho))'$ can easily be obtained from the gradient matrix and the variance matrix of the moment functions.
- Like KLS, the KMM estimates and confidence intervals can be plotted over a grid of postulated endogeneity correlations ρ . The union of those confidence intervals enables conservative inference over the whole set of values $\rho \in [\underline{r}, \bar{r}]$.

Empirical example

- Mroz (1987) data set from the Panel Study of Income Dynamics (PSID) on the labor force participation of 428 working, married women:

```
. bcuse mroz
```

- Textbook wage regression on education and labor market experience:

$$\ln(\text{wage})_i = \alpha + \beta_{\text{educ}} \text{educ}_i + \beta_{\text{exper}} \text{exper}_i + \beta_{\text{exper}^2} \text{exper}_i^2 + \varepsilon_i$$

- Education might be correlated with omitted wage determinants (e.g., unobserved “ability”). We assume $\rho = \text{Corr}(\text{educ}, \varepsilon_i) \in [0, 0.7]$.

Empirical example: KLS estimation

```
. local ylabel "ylabel(-0.25(0.05)0.15, format(%3.2f))"
. local xlabel "xlabel(0(0.1)0.7, format(%2.1f))"

. kinkyreg lwage exper expersq (educ = motheduc fatheduc),
> range(0 0.7) stepsize(0.005) correlation(0.5) small namestub(kls)
> twoway(, title("kinky least-squares estimation") 'ylabel' 'xlabel')
```

Kinky least-squares estimation

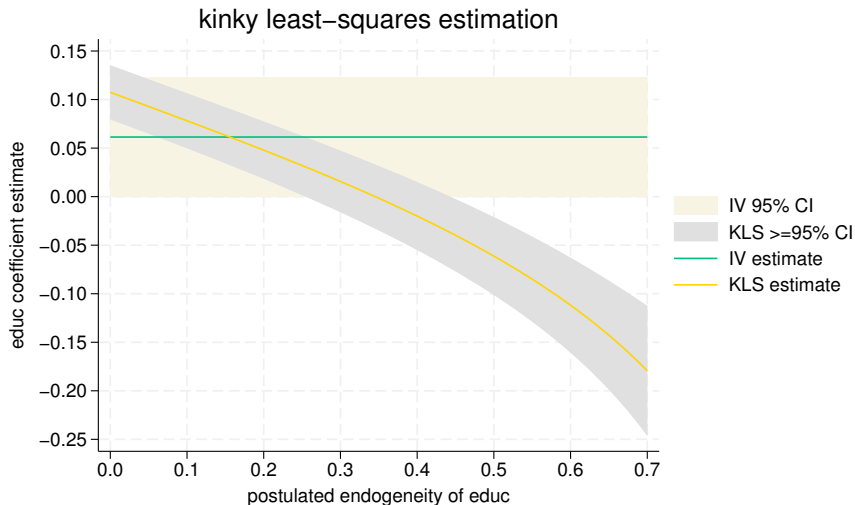
Number of obs = 428

Postulated endogeneity of educ = 0.5000

	lwage	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
educ		-.0612436	.0203417	-3.01	0.003	-.1012269	-.0212604
exper		.0510986	.01526	3.35	0.001	.0211039	.0810932
expersq		-.0011325	.0004557	-2.48	0.013	-.0020283	-.0002367
_cons		1.565081	.2728938	5.74	0.000	1.028688	2.101475

```
. estimates store kls
```

Empirical example: KLS inference



Empirical example: KMM estimation

```
. kinkyreg lwage exper expersq (educ = motheduc fatheduc), mm robust
> range(0 0.7) stepsize(0.005) correlation(0.5) small namestub(kmm)
> twoway(, title("robust kinky method-of-moments estimation") 'ylabel' 'xlabel')
```

Grid points

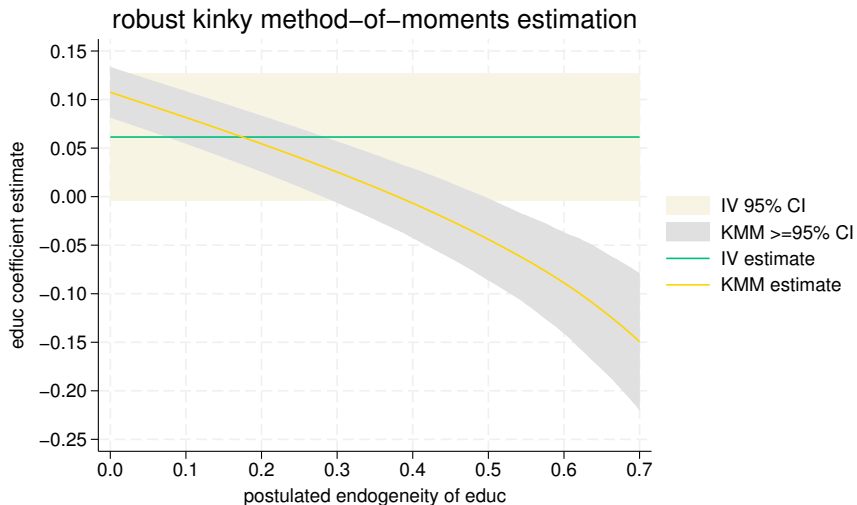
```
-----+--- 1 ----+--- 2 ----+--- 3 ----+--- 4 ----+--- 5
..... 50
..... 100
..... 141
```

Kinky method-of-moments estimation Number of obs = 428

Postulated endogeneity of educ = 0.5000

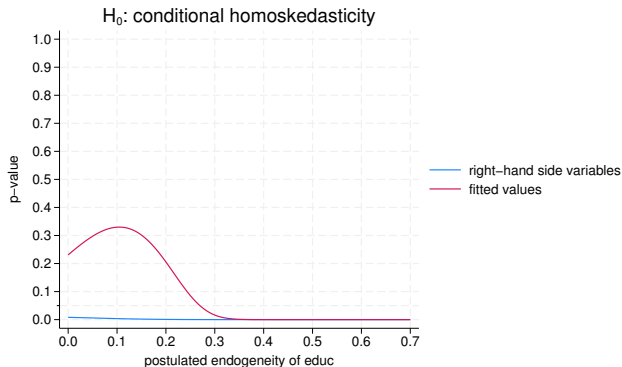
		Robust					
lwage	Coefficient	std. err.	t	P> t	[95% conf. interval]		
educ	-.0440185	.0213962	-2.06	0.040	-.0860744	-.0019626	
exper	.0501255	.0171156	2.93	0.004	.0164836	.0837674	
expersq	-.0010997	.0004859	-2.26	0.024	-.0020547	-.0001447	
_cons	1.352017	.29708	4.55	0.000	.7680843	1.93595	

Empirical example: KMM inference



Empirical example: Heteroskedasticity testing

```
. estimates store kmm  
. estimates restore kls  
. local ylabel "ylabel(0(0.1)1, format(%2.1f)) ymtick(0.05, grid)"  
  
. estat hettest (), rhs xb twoway(, title(H{sub:0}: conditional homoskedasticity) 'ylabel'  
> 'xlabel' legend(label(1 "right-hand side variables") label(2 "fitted values")))
```

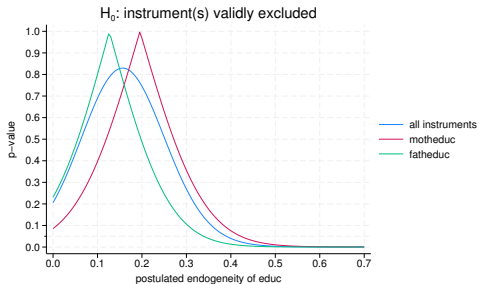


Empirical example: Testing exclusion restrictions

```
. estat exclusion, twoway(, title(H{sub:0}: instrument(s) validly excluded) 'ylabel' 'xlabel')
```

Endogeneity of educ compatible with valid exclusion

	Corr.	[95% Confid. Bounds]	
motheduc	.1955294	.	.422525
fatheduc	.1265772	.	.3389017



Empirical example: Testing exclusion restrictions

```
. estimates restore kmm

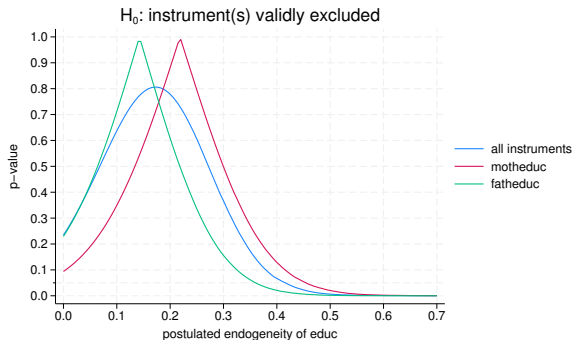
. estat exclusion, twoway(, title(H{sub:0}: instrument(s) validly excluded) 'ylabel' 'xlabel')
Grid points
-----+--- 1 ----+--- 2 ----+--- 3 ----+--- 4 ----+--- 5
..... 50
..... 100
..... 141
Grid points
-----+--- 1 ----+--- 2 ----+--- 3 ----+--- 4 ----+--- 5
..... 50
..... 100
..... 141
Grid points
-----+--- 1 ----+--- 2 ----+--- 3 ----+--- 4 ----+--- 5
..... 50
..... 100
..... 141
```

(continued on next slide)

Empirical example: Testing exclusion restrictions

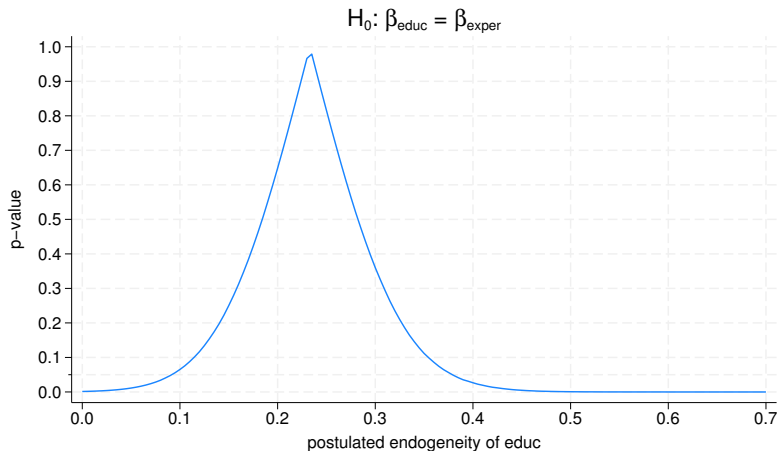
Endogeneity of educ compatible with valid exclusion

	Corr.	[95% Confid. Bounds]	
motheduc	.2184188	.	.4544897
fatheduc	.1424758	.	.3601344



Empirical example: Hypothesis testing

```
. estat test educ = exper, twoway(, title(H{sub:0}: {&beta}{sub:educ} = {&beta}{sub:exper}))  
> 'ylabel' 'xlabel')
```



Summary

- The `kinkyreg` Stata package provides tools for instrument-free inference in linear regression models with an arbitrary number of endogenous regressors.
 - For additional functionality, see our published article in the Stata Journal (2021).
- KLS/KMM inference can facilitate sensitivity checks. It is therefore often reasonable to consider KLS/KMM as a complement rather than substitute to instrument-based procedures.
- The new KMM approach enables heteroskedasticity-robust inference, although at the cost of additional computational complexity (and by assuming normally/ t -distributed errors).
- In practice, other model specification aspects and an appropriate choice for the bounds of ρ appear much more important.

References

- Kiviet, J. F. (2020). Testing the impossible: identifying exclusion restrictions. *Journal of Econometrics* 218 (2), 294–316.
- Kiviet, J. F. (2023). Instrument-free inference under confined regressor endogeneity and mild regularity. *Econometrics and Statistics* 25, 1-22.
- Kripfganz, S., and J. F. Kiviet (2021). kinkyreg: Instrument-free inference for linear regression models with endogenous regressors. *Stata Journal* 21 (3), 772–813.
- Mroz, T. A. (1987). The Sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica* 55 (4), 765–799.