Instrumental variables estimation with many weak instruments using regularized JIVE [1]

Jasmine. Hao[1]

[1]University of Hong Kong

ECON 6083: Machine Learning

[1]This section is based on [Belloni et al., 2012, Hansen and Kozbur, 2014]

Motivation

- **IV regression** for treatment effect estimation in endogenous regressors
- IV estimates can be imprecise in practice
- Strategy to increase precision: include **many instruments** to capture exogenous variation
- Motivation: **nonparametrically estimate optimal instruments** (Newey, 1990; Amemiya, 1974; Chamberlain, 1987)
- **High-dimensional data** leads to increasing number of potential instruments (e.g., Belloni et al., 2010)
- Usual **GMM-type approaches** (IV & 2SLS) may have a substantial bias when number of instruments is not small relative to sample size (Bekker, 1994; Newey & Smith, 2004)
- Bias affects the performance of **asymptotic distribution** and leads to inconsistency of the **2SLS estimator**

Context and Challenges

- Studying instrumental variables regression in settings with many instruments.
- Dealing with non-sparse first-stage prediction signals.
- Addressing models with more instruments than the sample size leads to a difficult and consistent model selection.

Proposed Solutions

- Proposal of Jackknife instrumental variables estimator (JIVE) with regularization at each jackknife iteration.
- Derivation of the limiting behavior for ridge-regularized JIVE estimator (RJIVE).
- The RJIVE is consistent and asymptotically normal without requiring consistent model selection.

Performance and Application

- Simulation results show the favourable performance of RJIVE in high-dimensional settings with many weak instruments.
- Application of RJIVE to the Angrist and Krueger (1991) example, demonstrating its superiority over other robust procedures in many-instrument scenarios.

Literature Review: Many-Instrument Asymptotics I

- **Many-Instrument Asymptotics:**
  - Popularized by Bekker (1994).
  - Goal: Provide an approximate behavior of the estimators when the number of instruments K is smaller than the sample size n, but $K/n \to \rho$ where $0 \leq \rho < 1$.
  - Consistent and asymptotically normal estimators: LIML, FULL, JIV (Bekker 1994; Chao and Swanson 2005; Hansen et al. 2008; CSHNW 2012).[2]
  - Asymptotic variance differs from usual asymptotics but can be consistently estimated.

Literature Review: Many-Instrument Asymptotics II

- **Simulation Performance:**
  - Estimators perform well when the number of instruments is an appreciable fraction of the sample size.
  - Inference based on the asymptotic distribution of many instruments controls the size of the test better than the usual asymptotic approximation.
- **Limitations:**
  - Requires the number of instruments to be less than the sample size.
  - Performance tends to decrease when $K/n \approx 1$.
- **Our Contribution:**
  - We consider cases where $K > n$ and regularization or instrument selection is necessary.
  - We show that the RJIVE retains the desirable asymptotic features derived in CSHNW.

[2] Chao, J. C., N. R. Swanson, J. A. Hausman, W. K. Newey, and T. Woutersen (2012): Asymptotic Distribution of JIVE in a Heteroskedastic IV Regression with Many Instruments, Econometric Theory, 28(1), 4286.

Context and Existing Methods

- Our research complements existing methods for robust estimation and inference strategies concerning many instruments.
  - Many-instrument asymptotics aims to outline the approximated behavior of estimators.
- This approach identifies estimators that perform well where the number of instruments is **less** than the sample size, but $K/n \to \rho$ where $0 \leq \rho < 1$.

Performance of IV Estimators

- Other IV estimators such as **LIML**, Fuller's (1977) modification of LIML (FULL), and jackknife instrumental variables (JIV) remain consistent and asymptotically normal.
  - Under the many-instrument sequence, the asymptotic variance of these estimators **deviates** from the variance under the usual asymptotic conditions,
  - can be **consistently** estimated.
- Simulation studies indicate that these estimators perform well when the number of instruments is a notable fraction of the sample size.

Limitations and Our Contributions

- Many instrument robust estimators tend to underperform in simulations when $K/n \approx 1$.
  - Our contribution includes considering cases where $K > n$
  - Regularization or instrument selection is necessary.
- We show that the RJIVE maintains the desirable asymptotic features derived in [Chao et al., 2012].

RJIVE and Other Many-Instrument Estimation Strategies

- The RJIVE complements many-instrument estimation strategies using first-stage regularization.
- First-stage regularization for estimating IV models has a long history, including the
  - **instrument selection** dating back to Kloek and Mennes (1960) and Amemiya (1966). Bai and Ng (2009) introduced modern variable selection techniques for the selection of instruments at the first stage.
  - [Belloni et al., 2012] analyzed IV estimators with **first-stage fit** using methods for fitting high-dimensional-sparse models such as **LASSO or Boosting**.
  - Related ideas also appear in the works of Bai and Ng (2010), Kapetanios and Marcellino (2010), Kapetanios, Khalaf, and Marcellino (2011), and Caner (2009).

Model Overview I

- **Model Basis:** We consider a model similar to a conventional linear instrumental variables model. Our focus is on the estimation of structural parameters from a single equation.
  - Single Equation Focus: Emphasis on estimating parameters within a single equation framework.
  - Linear IV Model: Draws parallels with traditional linear instrumental variables models.
- **Departure from Tradition:** Differing from the traditional framework, we do not make the assumption that the number of instruments is smaller than the available sample size. Instead, we address the need for regularization when dealing with an extensive set of instruments.
  - Instrument Abundance: Moves away from the constraint of having fewer instruments than observations.
  - Need for Regularization: Highlights the importance of regularization techniques in handling a large instrument set.

Model Overview II

- **High-Dimensional Approach:** Unlike other models that allow for a high-dimensional instrument set, our model does not assume sparsity. We **allow for the relationship between the instruments and the endogenous variables to be dense**, accommodating a broader range of real-world scenarios.
  - Dense Relationships: Does not impose sparsity constraints on the relationship between instruments and endogenous variables.
  - Real-World Applicability: Designed to accommodate complex relationships found in real-world data.

Model Specification

- **Basic Model:**
  - Two equations: one describing the relationship between an endogenous variable Y_i and endogenous regressors X_i, and another capturing the relationship between X_i and instruments Z_i.
- **Equations:**

$$Y_i = X_i\beta + \epsilon_i,$$
$$X_i = Z_i\pi + \eta_i.$$

- **Dimensions:**
  - X_i: L-dimensional, Z_i: K-dimensional, with $K \geq L$.
  - Number of over-identifying restrictions: $K - L$.
- **Assumptions:**
  - Conditional on Z_i, the disturbance ϵ_i has expectation zero and variance σ^2.
  - $E[Z_i' Z_i | Z] = S_Z$ with rank $L - M$, where M is the number of common elements in X and Z.
  - Observations (Y_i, X_i, Z_i) are i.i.d.
- **Estimators:**
  - OLS estimator for β: $\hat{\beta}_{OLS} = (X'X)^{-1}X'Y$, not consistent.
  - 2SLS estimator: $\hat{\beta}_{2SLS} = (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'Y$.

Jackknife Instrumental Variables Estimation

Angrist, Joshua D., Guido W. Imbens, and Alan B. Krueger. "Jackknife instrumental variables estimation." Journal of Applied Econometrics 14.1 (1999): 57-67.

- **Background:**
  - Two-stage-least-squares (2SLS) estimates can be biased towards OLS estimates, especially with many instruments
- **Jackknife IV Estimation:**
  - Proposed as an alternative to 2SLS and LIML for models with more instruments than endogenous regressors
  - Uses a 'leave-one-out' jackknife-type fitted value in place of the usual first-stage equation
- **Advantages:**
  - First-order equivalent to 2SLS but with superior finite-sample properties in terms of bias and coverage rate of confidence intervals
  - Less sensitive than LIML to deviations from the linear reduced form used in classical simultaneous equations models

Calculation of JIVE1 and JIVE2

- **Estimated Instrument for 2SLS:**
  - $Z_i\hat{\pi} = Z_i(Z'Z)^{-1}(Z'X)$.
- **JIVE1 Estimation:**
  - Removes dependence of the constructed instrument on the endogenous regressor for observation i.
  - Estimated optimal instrument: $Z_i\tilde{\pi}^{(i)} = Z_i(Z'_{-i}Z_{-i})^{-1}(Z'_{-i}X_{-i})$.
  - Estimator for β: $\hat{\beta}_{JIVE1} = (X'_{JIVE1}X)^{-1}(X'_{JIVE1}Y)$.
- **JIVE2 Estimation:**
  - Adjusts only the $Z'X$ component of $\hat{\pi}$.
  - Estimated optimal instrument: $Z_i\tilde{\pi}^{(i)} = Z_i(Z'Z)^{-1}(\frac{N}{N-1})(Z'X - Z'_iX_i)$.
  - Estimator for β: $\hat{\beta}_{JIVE2} = (X'_{JIVE2}X)^{-1}(X'_{JIVE2}Y)$.
- **Consistency:**
  - Both JIVE1 and JIVE2 estimators are consistent.
  - Probability limits and first-order asymptotic distribution are the same as those of $\hat{\beta}_{opt}$ and $\hat{\beta}_{2SLS}$.

Model Specification

$$y_i = X_i^{'} \delta_0 + \varepsilon_i \quad (2.1)$$
$$X_i = \Upsilon_i + U_i \quad (2.2)$$

- Here, y_i is a scalar outcome of interest, X_i is a L-dimensional treatment variable, and δ_0 is the L-dimensional structural effect of interest.
- We assume that Υ captures the part of X that is orthogonal to ε; that is, we assume $E[\Upsilon_i|\varepsilon_i] = 0$.
- Further, we assume that $\mathrm{Var}(\Upsilon_i) = 0$ and that $E[U_i|\Upsilon_i] = 0$.
- Estimation of δ_0 could be achieved by a straightforward application of classical instrumental variables methods if Υ were observed.

Instrumental Variables Assumptions

- **Unobserved Optimal Instrument:** It is often unrealistic to assume that the optimal instrument, Υ, is known or observed.
- **Observed Variables:** We assume that the outcome y_i and the treatment variable X_i are observed, but Υ_i is not.
- **Approximation with Available Instruments:** Estimation is based on a K-dimensional instrument Z_i, which provides a signal about Υ_i.
- **Linear Signal Assumption:** We focus on the case where $\Upsilon_i \approx Z_i^{'}\pi$, which is a relatively weak restriction.
- **Instrument Transformation:** The vector Z_i can consist of transformations of some exogenous variables W_i, such as orthogonal polynomials or splines, forming a basis function dictionary.

High-Dimensional Instruments

- **Our Focus:** We focus on the case where the number of instruments in Z_i, denoted by K, is large relative to the number of observations in the data, n.
- **Instrument Dimensions:** The set of available instruments may be high-dimensional, or one may be interested in approximating Υ through basis expansions.
- **Need for Regularization:** With many instruments, regularization becomes desirable to avoid overfitting of the relationship between the instruments and endogenous variables.
- **Regularization Strategies:** Strategies like LIML or JIV implicitly use regularization to avoid overfitting. When K is larger than n, these strategies also become inconsistent, requiring further dimension reduction or regularization.

Estimation - Hansen, Hausman, and Newey (2008)

- **Class of Estimators:** Includes all k-class estimators except for OLS.
- **Estimator Equation:**

$$\hat{\delta} = (X^{'}PX - \hat{\alpha}X^{'}X)^{-1}(X^{'}PY - \hat{\alpha}X^{'}Y)$$

- **Notation:**
  - $P = Z(Z^{'}Z)^{-}Z^{'}$
  - X is an $n \times L$ matrix, Z is an $n \times K$ matrix, Y is an $n \times 1$ vector.

Estimation - Simplification and Consistency

- **Simplification when $K \geq n$:**
  - $X^{'}PX = X^{'}X$ and $X^{'}PY = X^{'}Y$.
  - For a fixed $\hat{\alpha} \neq 1$, $\hat{\delta}$ reduces to the OLS estimator.
- **Consistency:**
  - The estimator is inconsistent for estimating δ_0 unless $E[X_i\varepsilon_i] = 0$.

Further Regularization

- **JIVE Estimators:** Require $(Z'Z)^{-1}$ in their construction and rely on $K \leq n$ in practice.
- **Further Regularization:** Desirable when $K < n$ to improve estimator behavior in finite samples, especially when K/n is close to one.
- **Generic IV Estimators:**

$$\hat{\delta}_{IV} = (\hat{\Upsilon}' X)(\hat{\Upsilon}' Y)$$

  for $\hat{\Upsilon} = (P - \hat{\alpha} I_n)X$ where I_n is the identity matrix $n \times n$.
- **Formal Requirement by CSHNW:** $(Z^\top Z/n)$ must have a minimum eigenvalue bounded away from 0 for large n.

Regularization with a Dense Signal: Options and Approaches

- **Regularization Options:**
  - Existence of many options when instruments outnumber endogenous variables.
  - Common approach: reduce the number of instruments.
    - Bai and Ng (2009), BCCH , and Gautier and Tsybakov (2011) provide examples of this approach.
  - Reduction can be intuitive or through formal mechanisms.
- **Common Approaches:**
  - Selecting instruments at random.
  - Performing factor decomposition and choosing first few factors as instruments.

Regularization with a Dense Signal: Drawbacks of Common Approaches

- **Drawbacks:**
  - Potential discarding of significant signal between instruments and endogenous variables.
  - Can result in efficiency loss.
  - May lead to identification issues.

Regularization with a Dense Signal: Alternative Approaches

- **Alternative Approaches:**
  - Estimation of first-stage relationship using shrinkage devices instead of instrument selection.
- **Regularization Devices:**
  - Carrasco (2012) and Carrasco and Tchuente Nguembu (2012) explored regularization devices for $(Z^{'}Z/n)$, the inverse of the covariance matrix of instruments.

First-Stage Variable Selection

These approaches make use of a model where $\Gamma_i = \tilde{Z}_i' \Pi$ up to a small approximation error where $\tilde{Z}_i$ is an s-dimensional set of the "relevant" instruments whose identities are unknown and estimated from the data.

- [Belloni et al., 2012] show that valid inference for δ_0 can be obtained after first-stage variable selection if $s^2/n \to 0$.
- This condition can be weakened to $s/n \to 0$ if a split-sample procedure is used.
- Sparsity requires that the signal be concentrated among a small set of factors within the set of considered instruments.
- Sparsity seems like a reasonable assumption in many cases.

Other Shrinkage Devices

- Other approaches estimate the first-stage relationship using other shrinkage devices.
- Carrasco (2012) and Carrasco and Tchuente Nguembu (2012) propose regularization devices to directly regularize the inverse of the covariance matrix of the instruments, $(Z'Z/n)$.
- Okui (2010) and Chamberlain and Imbens (2004) take similar approaches.
- These methods differ by relying on conditions that effectively rule out the case of a dense signal and by using the full-sample in constructing the first-stage fit for each observation.

Contrast to Existing Approaches

- Unlike other models, our estimation method allows for a dense signal.
- Consider an array of models with only one endogenous variable with an exactly linear and homoscedastic first-stage.
- The concentration parameter measures the information available in the instruments and determines the rate of convergence of IV estimators.
- Our model allows for cases where the concentration parameter based on any subset of the instruments satisfies certain conditions, and we refer to this as a "dense first-stage signal" or a "dense signal".

Limitations of Sparse-based Estimators

- **IV Exclusion Failure:**
  - Occurs when model selection mistakes select instruments with a population coefficient of zero.
- **Dense Signal Challenge:**
  - Consistent variable selection is not feasible, as individual instruments' signals are not well-separated from being uninformative.
- **Parameter Selection Issues:**
  - Default parameter choices often result in the selection of no variables.
  - Relaxing parameters leads to selecting a random set of instruments, mixing informative and uninformative instruments most correlated with first-stage noise.

Introduction: Regularized JIVE

Ridge Regression Coefficient Estimate:

- The estimate from regressing variable X onto set Z is given by:

$$\Pi = \arg\min_{\Pi} \|X - Z\Pi\|_{2,I}^2 + \|\Pi\|_{2,\Lambda}^2$$

  where $\|W\|_{2,A}^2 = W^\top A^\top A W$ for vector W and positive definite matrix A.
- Penalty Design: Favors models with small coefficients to avoid overfitting.
- Closed Form Solution:

$$\Pi = (Z^\top Z + \Lambda^\top \Lambda)^{-1}(Z^\top X)$$

Impact of Penalty Term

- **Stabilization:** The penalty term stabilizes the inverse of the sample covariance matrix of regressors, $Z^\top Z$. The addition of $\Lambda^\top \Lambda$ ensures the inverse in Π is always well-defined.
- **Contrast with OLS:** The usual OLS estimator $\Pi_{\text{OLS}} = (Z^\top Z)^{-1}(Z^\top X)$ is ill-defined when $K > n$ or near-singular, especially for high-dimensional Z relative to n.
- **Implementation:** Commonly, $\Lambda = \gamma^{1/2} I_K$ for a scalar penalty parameter γ. Our theory allows for more general cases, but this implementation is used in simulations and empirical examples.

Regularized IV Estimator and Carrasco's Approach

- In principle, one could use $\hat{\Upsilon} = \hat{\Pi}X$ for $\hat{\Pi}$ in (3.4) to define a *regularized IV estimator*, which corresponds to one of the regularization strategies pursued in Carrasco (2012).
- Carrasco (2012) provides conditions under which this approach is consistent and asymptotically normal for estimating δ_0. However, the drawback is that the estimated instrument Υ is *correlated with the structural error* ε by construction in finite samples.

Our Complementary Approach and Ridge-Regularized JIVE

- We offer a complementary approach that applies in the case of a *high-dimensional dense signal*. We define $\Pi^{\Lambda}_{-i} = (Z^\top Z - Z_i Z_i^\top + \Lambda^\top \Lambda)^{-1}(Z^\top X - Z_i X_i^\top)$, which is the ridge regression coefficient from running a ridge regression of X on Z with regularization matrix Λ using all but the ith observation.
- The leave-one-out estimator $\hat{\Upsilon}_i$ for the value of the instrument for the ith individual is then defined as $\hat{\Upsilon}_i = Z_i \Pi^{\Lambda}_{-i}$.
- Using the constructed $\hat{\Upsilon}_i$, we define the ridge-regularized JIVE as:

$$\tilde{\delta} = \left(\frac{1}{n} \sum_{i=1}^{n} \Pi^{\Lambda}_{-i} Z_i X_i^\top \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} \Pi^{\Lambda}_{-i} Z_i y_i \right). \tag{3.5}$$

By using the sample excluding the ith observation, the RJIVE breaks the correlation between Υi and ε_i in the case of a dense first-stage signal.

Ridge-Regularized JIVE: Advantages

- The leave-one-out estimator $\hat{\Upsilon}_i$ for the value of the instrument for the i-th individual may then be defined as $\hat{\Upsilon}_i = Z_i' \hat{\Pi}_{\Lambda}^{-i}$.
- We then define the ridge-regularized JIVE as in equation (3.5).
- This approach breaks the correlation between $\hat{\Upsilon}_i$ and ϵ_i in the case of a dense first-stage signal.
- Ridge regression regularizes the problem, allowing signal extraction from the large set of instruments while avoiding potential overfitting.

Ridge-Regularized JIVE: Link to JIVE Formulation

- The RJIVE is quite similar to the JIVE formulation from [Chao et al., 2012].
- Exploiting this similarity simplifies proofs and technical details of the RJIVE.

Implementing the RJIVE: Penalty Matrix Selection

- For the implementation of RJIVE, selection of the penalty matrix Λ is necessary.
- We follow the usual approach for ridge regression by setting $\Lambda = \gamma^{1/2} I_n$ and further set $\gamma^{1/2} = CK^{1/2}$.
- The constant of proportionality C for each first-stage equation is set to the sample standard deviation of the element of X_i being considered.

RJIVE: An Illustration with Angrist and Krueger (1991)

- We illustrate the use of the *RJIVE* by revisiting the classic example in the many-instrument literature, Angrist and Krueger (1991).
- The focus is on estimating the *causal effect of schooling* on earnings by addressing the potential endogeneity of schooling through the use of instrumental variables.
- Angrist and Krueger (1991) provide many instruments for schooling, but concerns arise about biases and inferential problems from using the full set of available instruments (Bound, Jaeger, and Baker, 1995; Angrist, Imbens, and Krueger, 1999; Staiger and Stock, 1997; Hansen, Hausman, and Newey, 2008).

Angrist and Krueger (1991) Revisited: Model

We consider the model from Angrist and Krueger (1991):

$$\log(\text{wage}_i) = \alpha \text{Schooling}_i + \mathbf{w}_i^\top \gamma + \varepsilon_i$$

$$\text{Schooling}_i = \mathbf{z}_i^\top \Pi_1 + \mathbf{w}_i^\top \Pi_2 + v_i$$

- Unobservables ε_i and v_i satisfy the exclusion restriction $E[\varepsilon_i | \mathbf{w}_i, \mathbf{z}_i] = E[v_i | \mathbf{w}_i, \mathbf{z}_i] = 0$.
- $\log(\text{wage}_i)$ is the log wage of individual i, Schooling_i is the reported years of completed schooling, $\mathbf{w}_i$ is a vector of control variables, and $\mathbf{z}_i$ is a vector of instrumental variables.

Angrist and Krueger (1991) Revisited I

- Data are drawn from the 1980 U.S. Census, consisting of 329,509 men born between 1930 and 1939.
- Control variables $\mathbf{w}_i$ include a constant, 9 year-of-birth dummies, 50 state-of-birth dummies, and 450 state-of-birth $\times$ year-of-birth interactions (total of 510 variables).
- Instruments $\mathbf{z}_i$ include three quarter-of-birth dummies and their interactions with all state-of-birth and year-of-birth controls (total of 1527 potential instruments).
- Angrist and Krueger (1991) provide an argument for the validity of $\mathbf{z}_i$ as instruments, based on compulsory schooling laws and the shape of the life-cycle earnings profile.
- The coefficient of interest is α, which summarizes the causal impact of education on earnings.

Angrist and Krueger (1991) Revisited II

- Past discussion
  - **Angrist, J. D., and Krueger, A. B. (1991):** Classic example in the many-instrument literature focusing on the causal effect of schooling on earnings and addressing the potential endogeneity of schooling through instrumental variables.
  - **Bound, J., Jaeger, D. A., and Baker, R. M. (1995):** Discusses concerns about potential biases and inferential problems introduced from using the full set of available instruments.
  - **Angrist, J. D., Imbens, G. W., and Krueger, A. B. (1999):** Further exploration of the instrumental variables approach to estimating the causal effect of schooling on earnings.
  - **Staiger, D., and Stock, J. H. (1997):** Examines the reliability of instrumental variables in the context of estimating the effect of education on earnings.
  - **Hansen, C., Hausman, J., and Newey, W. K. (2008):** Discusses estimation with many instrumental variables, addressing concerns related to the Angrist and Krueger (1991) example.

RJIVE: An Illustration with Angrist and Krueger (1991)

Table 3: Estimates of the Return to Schooling in Angrist and Krueger Data

	2SLS	Post-LASSO	JIVE	RJIVE
		A. 3 Instruments		
Schooling Coefficient	0.1079	0.1115	0.1091	0.1091
Estimated Standard Error	0.0196	0.0205	0.0202	0.0202
		B. 180 Instruments		
Schooling Coefficient	0.0928	0.1125	0.1096	0.1062
Estimated Standard Error	0.0097	0.0173	0.0161	0.0157
		C. 1527 Instruments		
Schooling Coefficient	0.0712	0.0862	0.0816	0.1067
Estimated Standard Error	0.0049	0.0254	0.5168	0.0171

Note: This table reports estimates of the returns-to-schooling parameter in the Angrist-Krueger 1991 data using different estimators and different numbers of instruments. In the rows, we give point estimates of the schooling coefficient and heteroskedasticity consistent standard error estimates. We report results for 2SLS, the Post-LASSO estimator of Belloni, Chen, Chernozhukov, and Hansen (2012) (Post-LASSO), JIVE, and our regularized JIVE (RJIVE). Further details are provided in the text. For comparison, the OLS estimate (standard error) of the schooling coefficient is 0.0673 (0.0004).

Result Summary: Quarter-of-birth Dummies as Instruments (Panel A)

- Three main quarter-of-birth dummies used as instruments.
- All four estimators (2SLS, JIVE, Post-LASSO, RJIVE) produced **similar results** indicating no substantial overfitting in the first stage.
- LASSO selected only two of the three possible instruments but provided similar estimates to JIVE.

Result Summary: 180 Instruments (Panel B)

- 180 instruments used formed by interactions of quarter-of-birth main effects with year-of-birth and state-of-birth effects.
- 2SLS estimates were found to have a **substantial bias towards OLS** due to overfitting.
- Procedures robust to many instruments, such as JIVE or LASSO, are more recommended.
- Many-instrument-robust estimates remained close to the value estimated using only the three main effects as instruments, indicating **additional signal available in the larger instrument set**.

Result Summary: 1527 Instruments (Panel C)

- Full set of 1527 instruments used.
- The RJIVE estimator was found to be quite stable, with its point estimates around the value estimated using only three instruments.
- Both Post-LASSO and JIVE estimates shifted towards the OLS estimate, with their standard errors now significantly larger than those from the RJIVE.
- The LASSO estimator remained stable because it selected only one variable, potentially discarding useful smaller signals and being inefficient.
- The RJIVE effectively regularized non-informative signals while capturing more signal than LASSO.

Overall Conclusion

- The RJIVE estimator showed **stable performance** across different sets of instruments.
- It demonstrated its potential as a useful **complement** to the approaches currently used in dealing with many instruments.
- Its ability to extract more signal compared to LASSO, especially when the signal is not sparse, appears to be a significant advantage.

References I

Belloni, A., Chen, D., Chernozhukov, V., and Hansen, C. (2012).
Sparse models and methods for optimal instruments with an application to eminent domain.
Econometrica, 80(6):2369–2429.

Chao, J. C., Swanson, N. R., Hausman, J. A., Newey, W. K., and Woutersen, T. (2012).
Asymptotic distribution of jive in a heteroskedastic iv regression with many instruments.
Econometric Theory, 28(1):42–86.

Hansen, C. and Kozbur, D. (2014).
Instrumental variables estimation with many weak instruments using regularized jive.
Journal of Econometrics, 182(2):290–308.