Re-thinking about instrumental variables

Domenico Giannone
IMF and University of Washington

Michele Lenza
European Central Bank

Giorgio Primiceri
Northwestern University

75th European meeting of the Econometric Society
Rotterdam, August 2024

Re-thinking about instrumental variables

Chris Sims

Domenico Giannone

IMF and University of Washington

Michele Lenza

European Central Bank

Giorgio Primiceri

Northwestern University

75th European meeting of the Econometric Society

Rotterdam, August 2024

This paper

- Traditional IV inference is distorted by an implicit prior
  - It favors instrument strength
  - Unintended consequence: Standard errors might be unrealistically tight

This paper

- Traditional IV inference is distorted by an implicit prior
  - It favors instrument strength
  - Unintended consequence: Standard errors might be unrealistically tight

- A simple *agnostic prior* on instrument strength solves the problem
  - Bayesian inference robust to weak instruments

Outline

- A refresher on IV regressions
  - The challenging case of weak instruments
- A Bayesian perspective to
  - deepen our understanding of the problem
  - propose simple and effective solution
- Empirics
  - The classic problem of estimating the return to education

IV regression

$$y = x\beta + e \qquad\qquad cov(x, e) \neq 0$$

IV regression

$$x = \qquad u$$

$$y = x\beta + \underbrace{u\delta + \varepsilon}_{e}, \qquad cov(x, e) \neq 0$$

IV regression

$$x = \qquad u$$

$$y = x\beta + u\delta + \varepsilon$$

- x is endogenous ➡ $\hat{\beta}_{OLS}$ is biased

IV regression

$$x = z\pi + u$$

$$y = x\beta + u\delta + \varepsilon$$

- x is endogenous → $\hat{\beta}_{OLS}$ is biased

- IV idea: If there is an exogenous z that is correlated with x, can exploit the exogenous variation of x to estimate β

IV regression

$$x = z\pi + u$$

$$y = x\beta + u\delta + \varepsilon$$

- x is endogenous → $\hat{\beta}_{OLS}$ is biased
- IV idea: If there is an exogenous z that is correlated with x, can exploit the exogenous variation of x to estimate β
- Standard estimator: $\hat{\beta}_{TSLS} = \frac{\hat{x}'y}{\hat{x}'\hat{x}}$

A famous example: Estimating the return to education

$$x = z\pi + u$$

$$y = x\beta + u\delta + \varepsilon$$

Log-wages

Years of schooling

A famous example: Estimating the return to education

Quarter of birth

$$x = z\pi + u$$

$$y = x\beta + u\delta + \varepsilon$$

Log-wages

Years of schooling

A famous example: Estimating the return to education

Quarter of birth → z

$$x = z\pi + u$$

$$y = x\beta + u\delta + \varepsilon$$

Log-wages → y

Years of schooling → x

- Empirical finding

	AK	
TSLS	**.083 (.009)**	

A puzzling result

- Crazy idea: run the AK regression with fake instruments
  - Randomly assigned quarters of birth

A puzzling result

- Crazy idea: run the AK regression with fake instruments
  - Randomly assigned quarters of birth

- Intuitively, we should get estimates with infinite standard errors
  - If the instrument is irrelevant, there is nothing we can say

A puzzling result

- Crazy idea: run the AK regression with fake instruments
  - Randomly assigned quarters of birth
- Intuitively, we should get estimates with infinite standard errors
  - If the instrument is irrelevant, there is nothing we can say
- But Bound-Jaeger-Baker (1995) obtained estimates similar to AK

	AK	Fake Zs
TSLS	.083 (.009)	.060 (.016)

A puzzling result

- Crazy idea: run the AK regression with fake instruments
  - Randomly assigned quarters of birth
- Intuitively, we should get estimates with infinite standard errors
  - If the instrument is irrelevant, there is nothing we can say
- But Bound-Jaeger-Baker (1995) obtained estimates similar to AK

	AK	Fake Zs
TSLS	.083 (.009)	.060 (.016)

- **HUGE PROBLEM: SE of TSLS unable to detect if IV is irrelevant**

A puzzling result

- Is this a small-sample problem?

  NO. AK used a huge sample of 300k+ observations

A puzzling result

- Is this a small-sample problem?

  NO. AK used a huge sample of 300k+ observations

- Does this happen only with a very large number of instruments?

  NO

A MC simulation with irrelevant instruments

- 1000 simulations from

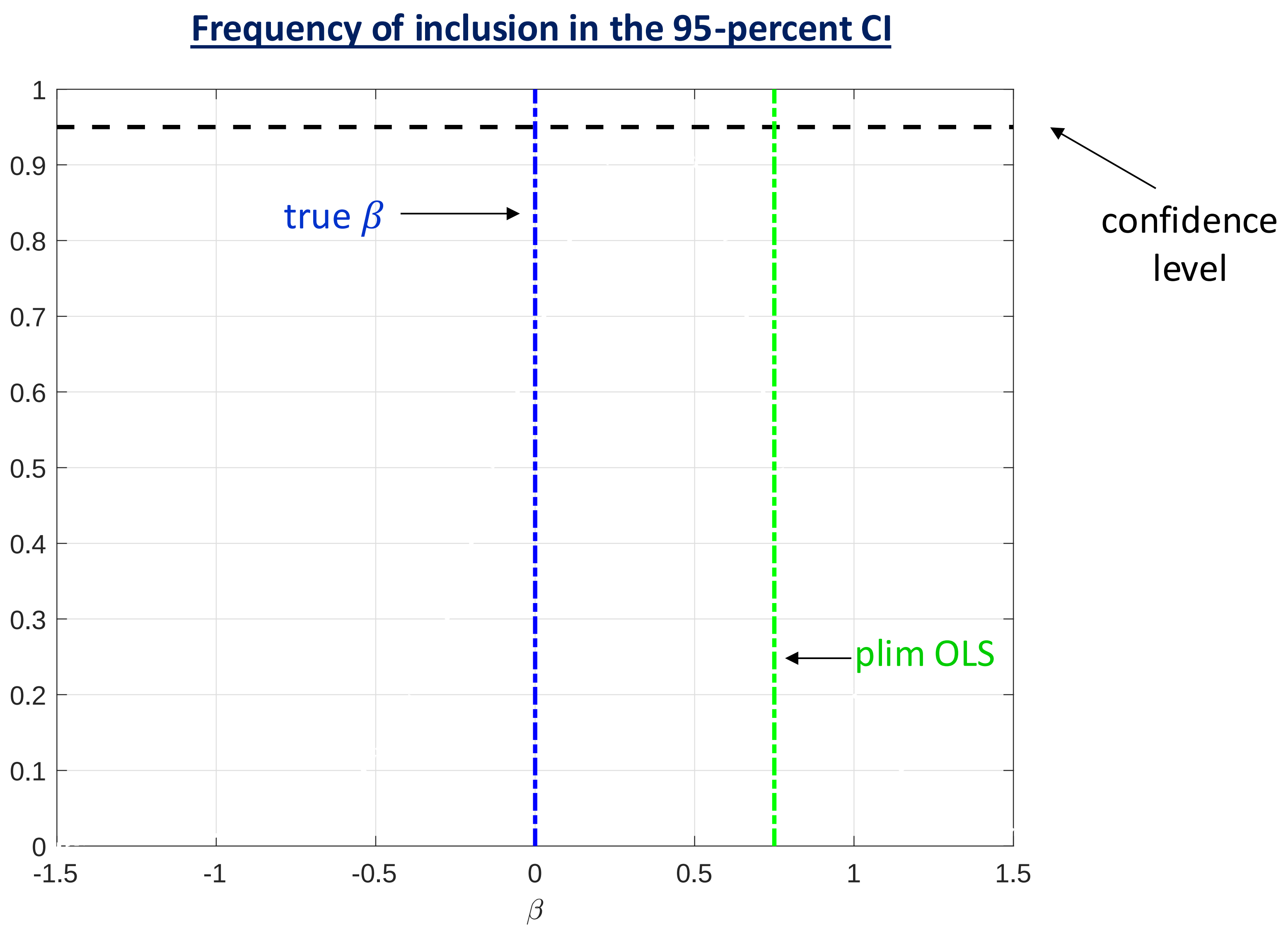
$$x = z\pi + u$$

$$y = x\beta + u\delta + \varepsilon$$

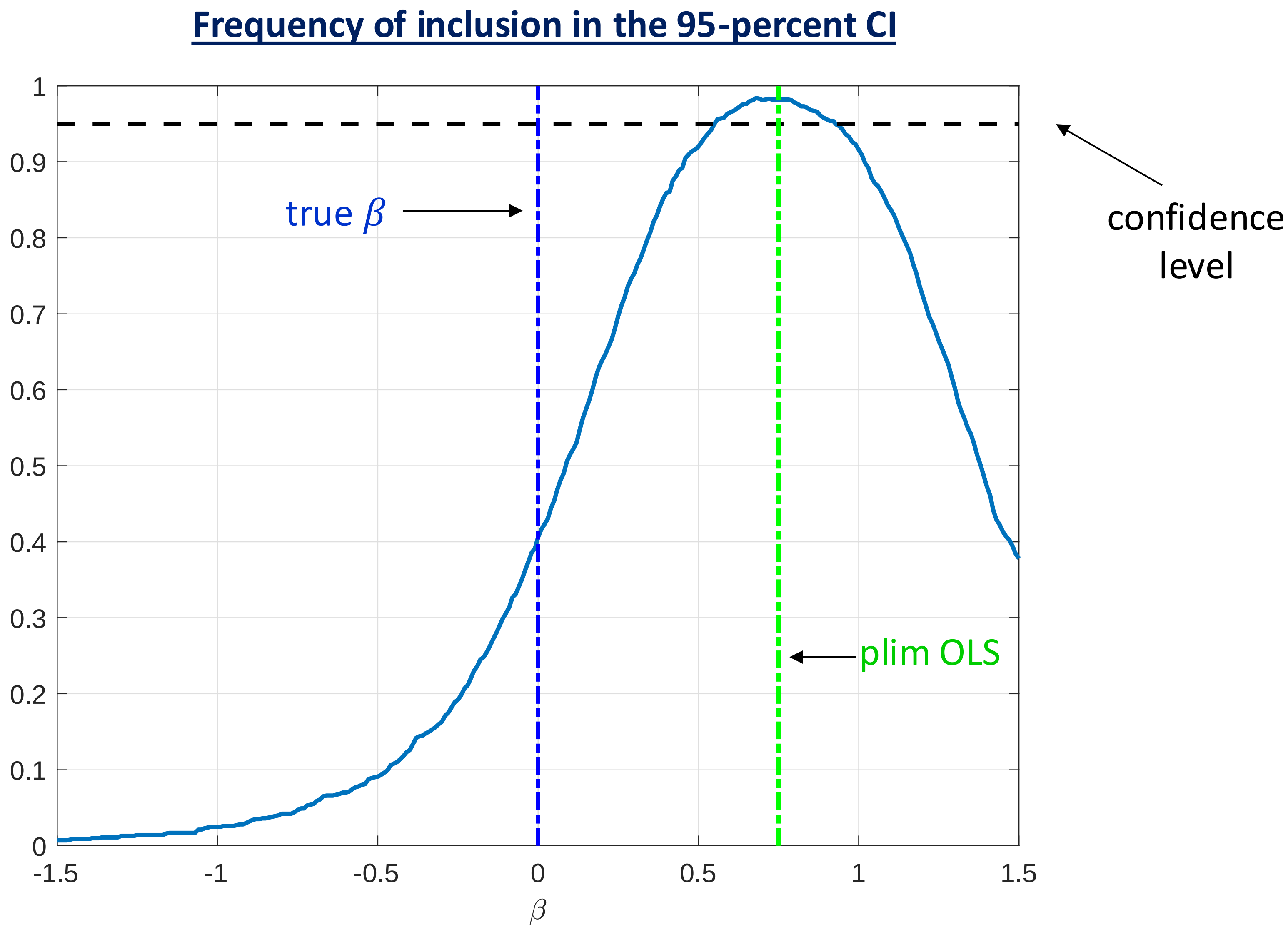
- 10 irrelevant instruments, i.e., $\pi = 0_{10\times 1}$

- $\beta = 0$

TSLS with irrelevant instruments



TSLS with irrelevant instruments



A MC simulation with irrelevant instruments

- 1000 simulations from

$$x = z\pi + u$$

$$y = x\beta + u\delta + \varepsilon$$

- 10 irrelevant instruments, i.e., $\pi = 0_{10\times1}$

- $\beta = 0$

- The TSLS CIs are **centered around OLS** and **WAY too tight**

incredibly strong opinion
around a false statement

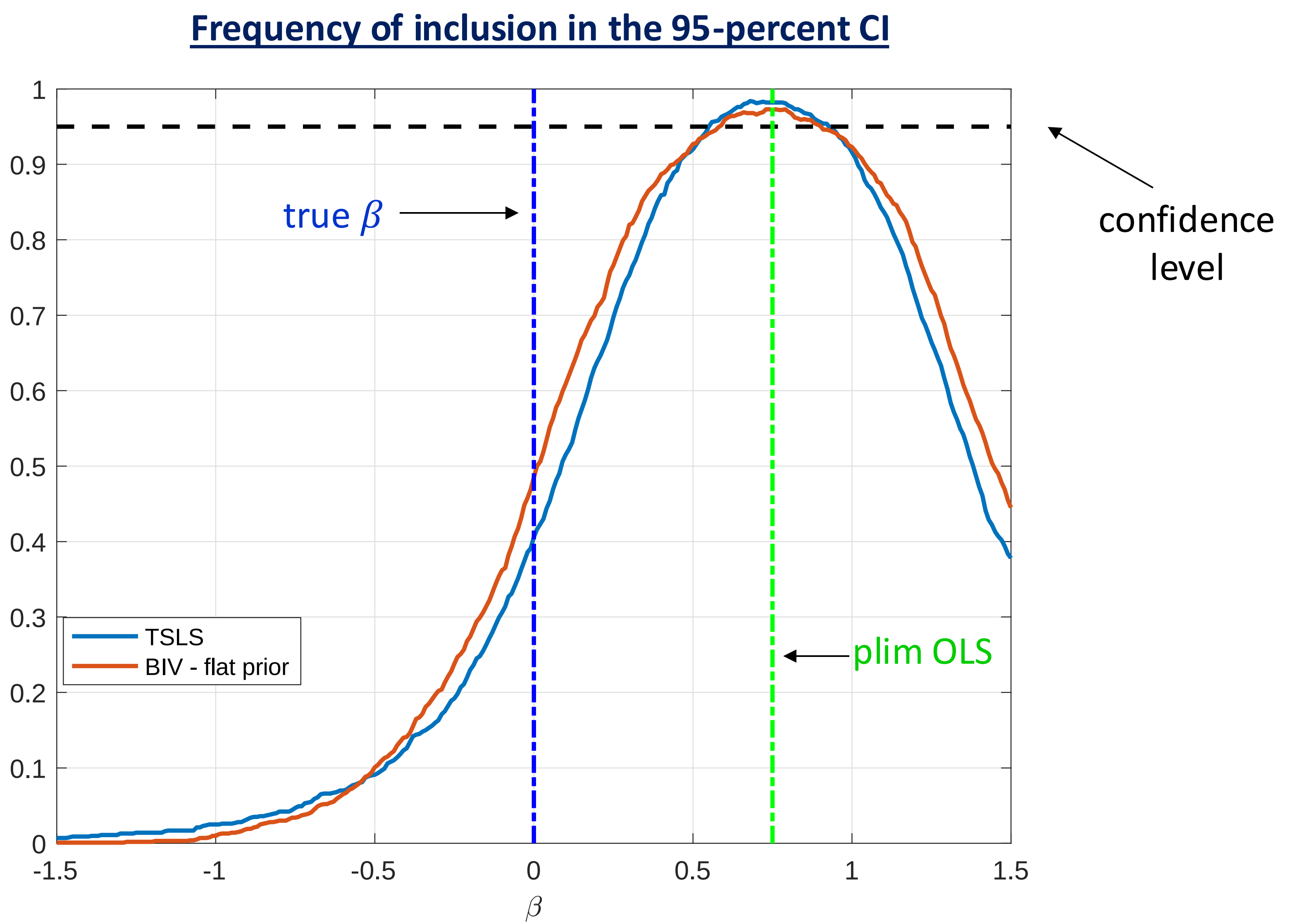
What about LH-based methods?

- Problems of TSLS with weak instruments are well-known

What about LH-based methods?

- Problems of TSLS with weak instruments are well-known
- Not just TSLS: Same issue with LH and flat-prior Bayesian methods

TSLS and flat-prior BIV with irrelevant instruments



What about LH-based methods?

- Problems of TSLS with weak instruments are well-known
- Not just TSLS: Same issue with LH and flat-prior Bayesian methods
- This is even more puzzling because
  1. LH principle: all the sample evidence relevant to parameters is in the LH
  2. LH is correctly specified in this controlled experiment

What about LH-based methods?

- Problems of TSLS with weak instruments are well-known
- Not just TSLS: Same issue with LH and flat-prior Bayesian methods
- This is even more puzzling because
  1. LH principle: all the sample evidence relevant to parameters is in the LH
  2. LH is correctly specified in this controlled experiment
- Understanding the "Bayesian IV puzzle" will
  - deepen our understanding of the problem
  - help us suggest viable solutions

The root of the problem

$$x = z\pi + u$$
$$y = x\beta + u\delta + \varepsilon$$

- Suppose that we know with certainty that $\pi = 0$

The root of the problem

$$x = z\pi + u$$
$$y = x\beta + u\delta + \varepsilon$$

- Suppose that we know with certainty that $\pi = 0$
- x and u are perfectly collinear → CI of β have infinite length

The root of the problem

$$x = z\pi + u$$
$$y = x\beta + u\delta + \varepsilon$$

- Suppose that we know with certainty that $\pi = 0$
- x and u are perfectly collinear → CI of β have infinite length

☞ With irrelevant instruments, SE are too small because of overfitting
  - $\pi = 0$, but $\hat{\pi} \neq 0$

The root of the problem

$$x = z\pi + u$$
$$y = x\beta + u\delta + \varepsilon$$

- Suppose that we know with certainty that $\pi = 0$
- x and u are perfectly collinear ➡ CI of β have infinite length

☞ With irrelevant instruments, SE are too small because of overfitting

- $\pi = 0$, but $\hat{\pi} \neq 0$

- More formally: $var(\beta|\pi, data) \approx \frac{\sigma_\varepsilon^2}{\pi' z' z\pi}$ ← $\propto$ **concentration parameter**

The root of the problem

$$x = z\pi + u$$
$$y = x\beta + u\delta + \varepsilon$$

- Suppose that we know with certainty that $\pi = 0$
- x and u are perfectly collinear ⟹ CI of β have infinite length

☞ With irrelevant instruments, SE are too small because of overfitting

  - $\pi = 0$, but $\hat{\pi} \neq 0$

- More formally: $var(\beta|\pi, data) \approx \frac{\sigma_\varepsilon^2}{\pi' z' z\pi}$ ← $\propto$ **concentration parameter**
- Therefore, $\pi' z' z\pi$ must be estimated to be too large

The root of the problem: Implicit priors

- LH principle +
- LH correctly specified

→ The pathology must be due to the prior

The root of the problem: Implicit priors

- LH principle +
- LH correctly specified

→ The pathology must be due to the prior

$$x = z\pi + u$$
$$y = x\beta + u\delta + \varepsilon$$

The root of the problem: Implicit priors

- LH principle +
- LH correctly specified

} ➡ The pathology must be due to the prior

$$x = z\pi + u$$
$$y = x\beta + u\delta + \varepsilon$$

- Gaussian prior on π

$$\pi \sim N(0\,, \gamma^2 \cdot \sigma_u^2 (z'z/T)^{-1})$$

The root of the problem: Implicit priors

- LH principle +
- LH correctly specified

→ The pathology must be due to the prior

- Gaussian prior on π in the first stage

$$\pi \sim N(0\,, \gamma^2 \cdot \sigma_u^2 (z'z/T)^{-1})$$

The root of the problem: Implicit priors

- LH principle +
- LH correctly specified

→ The pathology must be due to the prior

- Gaussian prior on π in the first stage

$$\pi \sim N(0\,,\gamma^2 \cdot \sigma_u^2 (z'z/T)^{-1})$$

- implies a scaled χ^2 prior on the strength of the instruments

$$\frac{1}{T}\frac{\pi' z' z\pi}{\sigma_u^2} \sim \gamma^2 \cdot \chi_k^2$$

The root of the problem: Implicit priors

- LH principle +
- LH correctly specified

→ The pathology must be due to the prior

- Gaussian prior on π in the first stage

$$\pi \sim N(0\,, \gamma^2 \cdot \sigma_u^2 (z'z/T)^{-1})$$

- implies a scaled χ^2 prior on the strength of the instruments

$$\frac{1}{T}\frac{\pi' z' z\pi}{\sigma_u^2} \sim \gamma^2 \cdot \chi_k^2 \quad \Rightarrow \quad E\left(\frac{1}{T}\frac{\pi' z' z\pi}{\sigma_u^2}\right) = k \cdot \gamma^2$$

The root of the problem: Implicit priors

- LH principle +
- LH correctly specified

→ The pathology must be due to the prior

- Gaussian prior on π in the first stage

$$\pi \sim N(0\,, \gamma^2 \cdot \sigma_u^2 (z'z/T)^{-1})$$

- implies a scaled χ^2 prior on the strength of the instruments

$$\frac{1}{T}\frac{\pi' z' z \pi}{\sigma_u^2} \sim \gamma^2 \cdot \chi_k^2 \quad \Rightarrow \quad E\left(\frac{1}{T}\frac{\pi' z' z \pi}{\sigma_u^2}\right) = k \cdot \gamma^2$$

- As $\gamma^2 \to \infty$

The root of the production problem: Implicit priors

- LH principle +
- LH correctly specified

→ The pathology must be due to the prior

- Gaussian prior on π in the first stage

$$\pi \sim N(0\,, \gamma^2 \cdot \sigma_u^2 (z'z/T)^{-1})$$

- implies a scaled χ^2 prior on the strength of the instruments

$$\frac{1}{T}\frac{\pi'z'z\pi}{\sigma_u^2} \sim \gamma^2 \cdot \chi_k^2 \quad \Rightarrow \quad E\left(\frac{1}{T}\frac{\pi'z'z\pi}{\sigma_u^2}\right) = k \cdot \gamma^2$$

- As $\gamma^2 \to \infty$, inference on β is distorted towards high precision

$$var(\beta|\pi, data) \approx \frac{\sigma_\varepsilon^2}{\pi'z'z\pi} \quad \Rightarrow \quad 0$$

The root of the problem: Implicit priors

- LH principle +
- LH correctly specified

$\Rightarrow$ The pathology must be due to the prior

- Gaussian prior on π in the first stage

$$\pi \sim N(0\,, \gamma^2 \cdot \sigma_u^2 (z'z/T)^{-1})$$

- implies a scaled χ^2 prior on the strength of the instruments

$$\frac{1}{T}\frac{\pi' z' z \pi}{\sigma_u^2} \sim \gamma^2 \cdot \chi_k^2 \quad \Rightarrow \quad E\left(\frac{1}{T}\frac{\pi' z' z \pi}{\sigma_u^2}\right) = k \cdot \gamma^2$$

- As $\gamma^2 \to \infty$, inference on β is distorted towards high precision

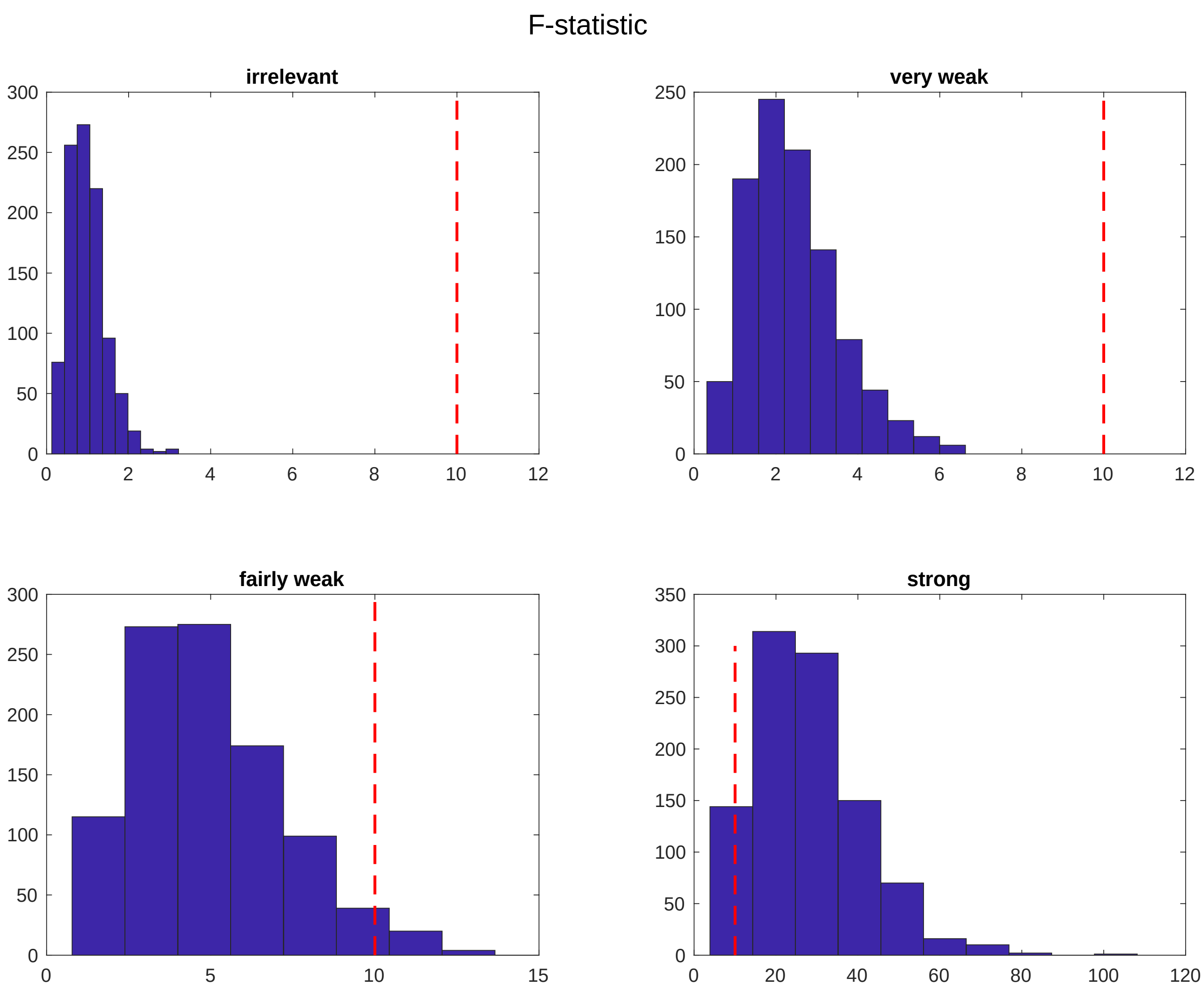
$$var(\beta|\pi, data) \approx \frac{\sigma_\varepsilon^2}{\pi' z' z \pi} \quad \Rightarrow \quad 0$$

- Solution: Flat prior on concentration parameter

Next few slides

- Simulation evidence
  - Instruments:
    - strong
    - fairly weak
    - very weak
    - irrelevant

F-stats of simulations



Next few slides

- Simulation evidence
  - Instruments:
    - strong
    - fairly weak
    - very weak
    - irrelevant
  - Compare methods:
    - TSLS
    - BIV—Gaussian prior
    - pre-testing

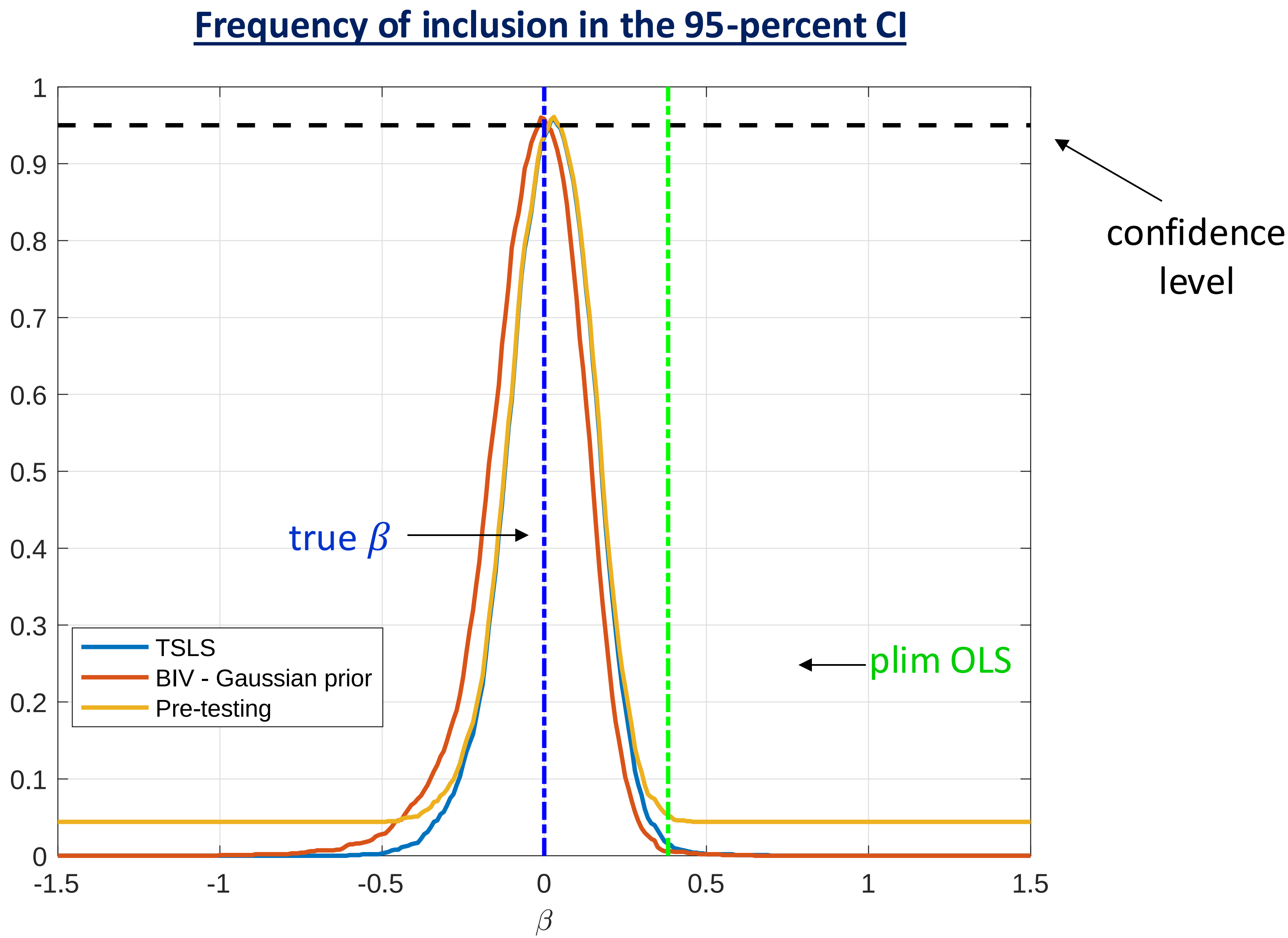
Next few slides

- Simulation evidence
  - Instruments:
    - strong
    - fairly weak
    - very weak
    - irrelevant
  - Compare methods:
    - TSLS
    - BIV—Gaussian prior
    - pre-testing, CLR

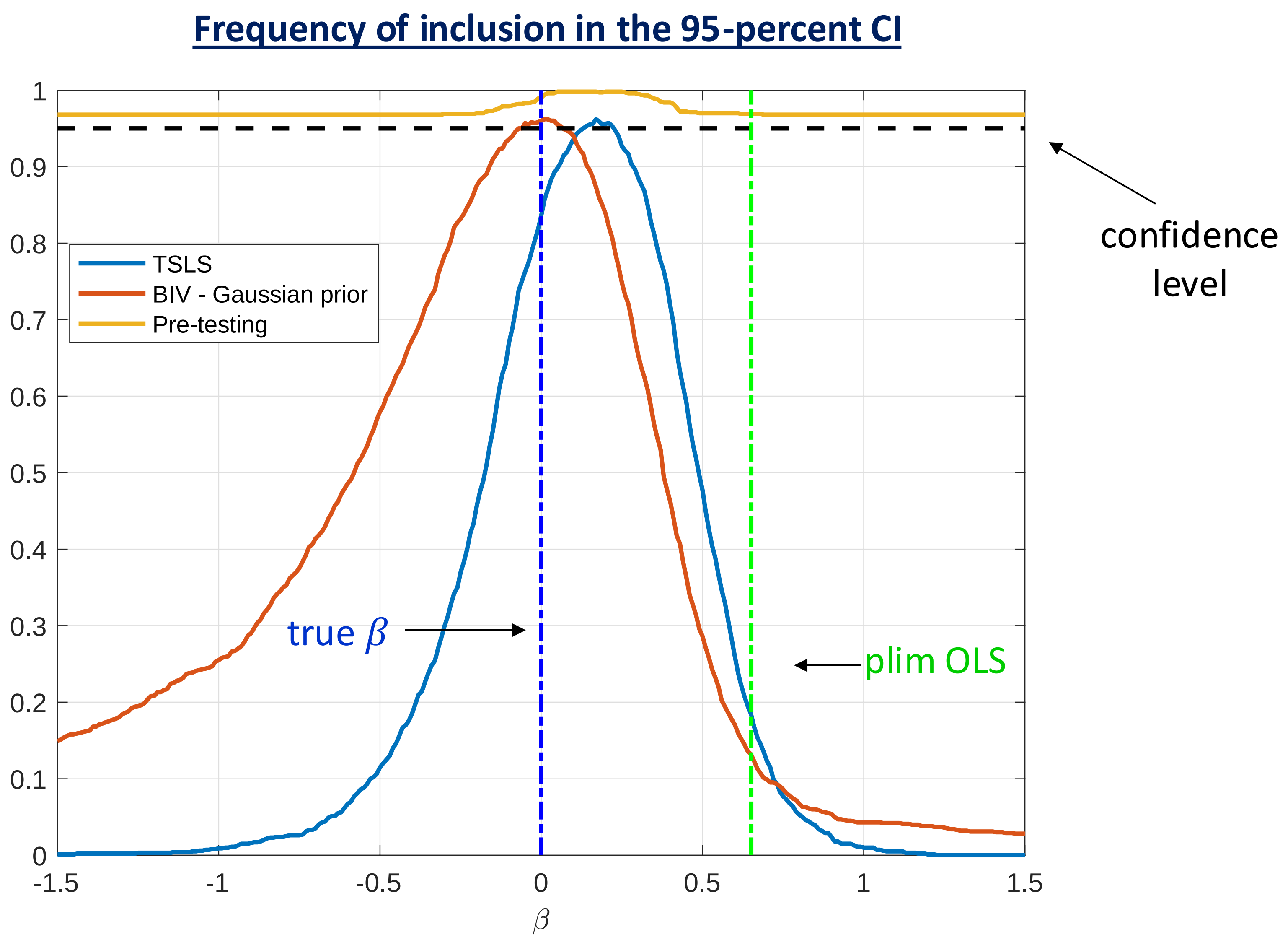
Next few slides

- Simulation evidence
  - Instruments:
    - strong
    - fairly weak
    - very weak
    - irrelevant
  - Compare methods:
    - TSLS
    - BIV—Gaussian prior
    - pre-testing, CLR
- Back to empirics

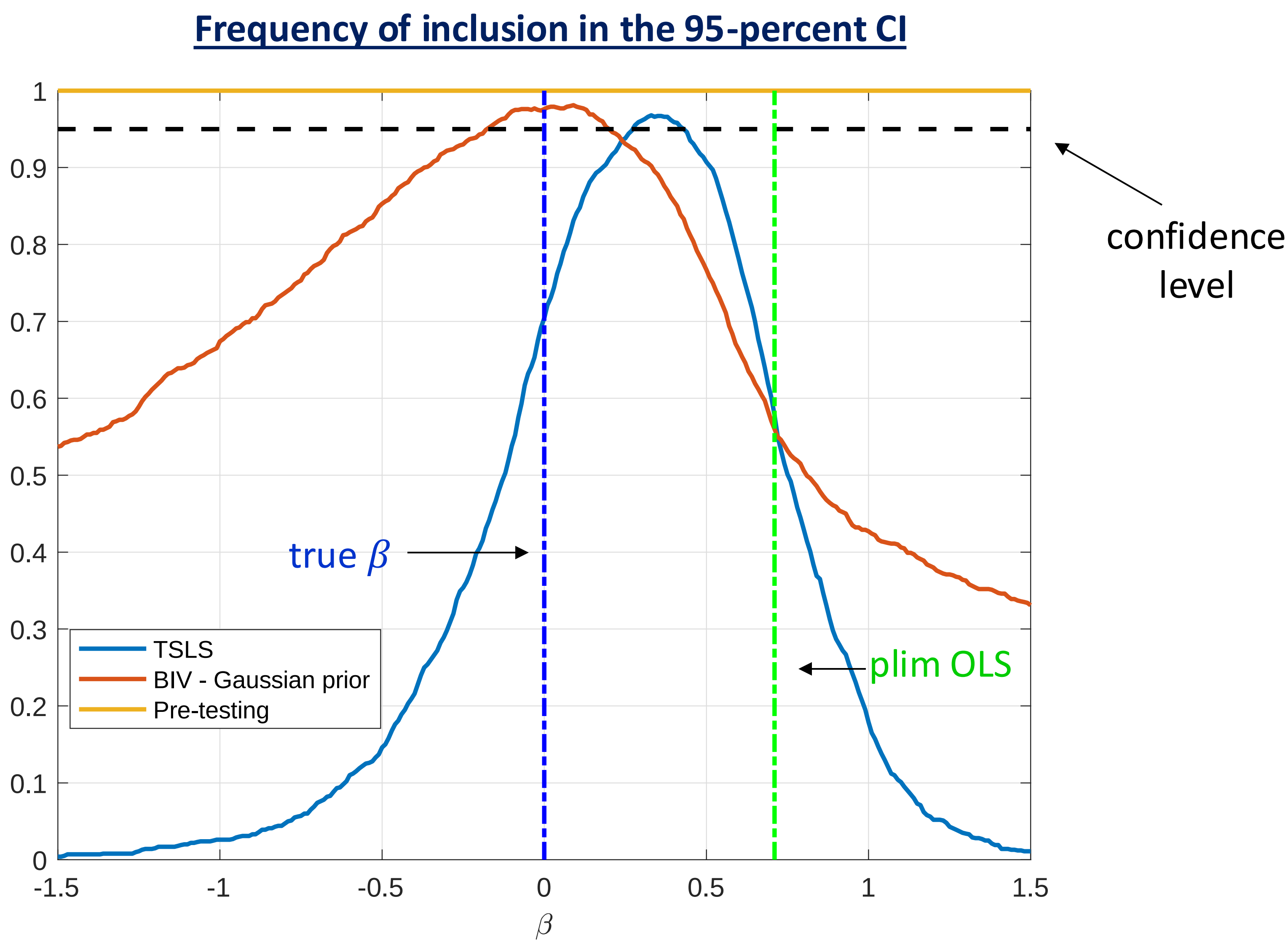
TSLS, BIV and pre-testing with strong instruments



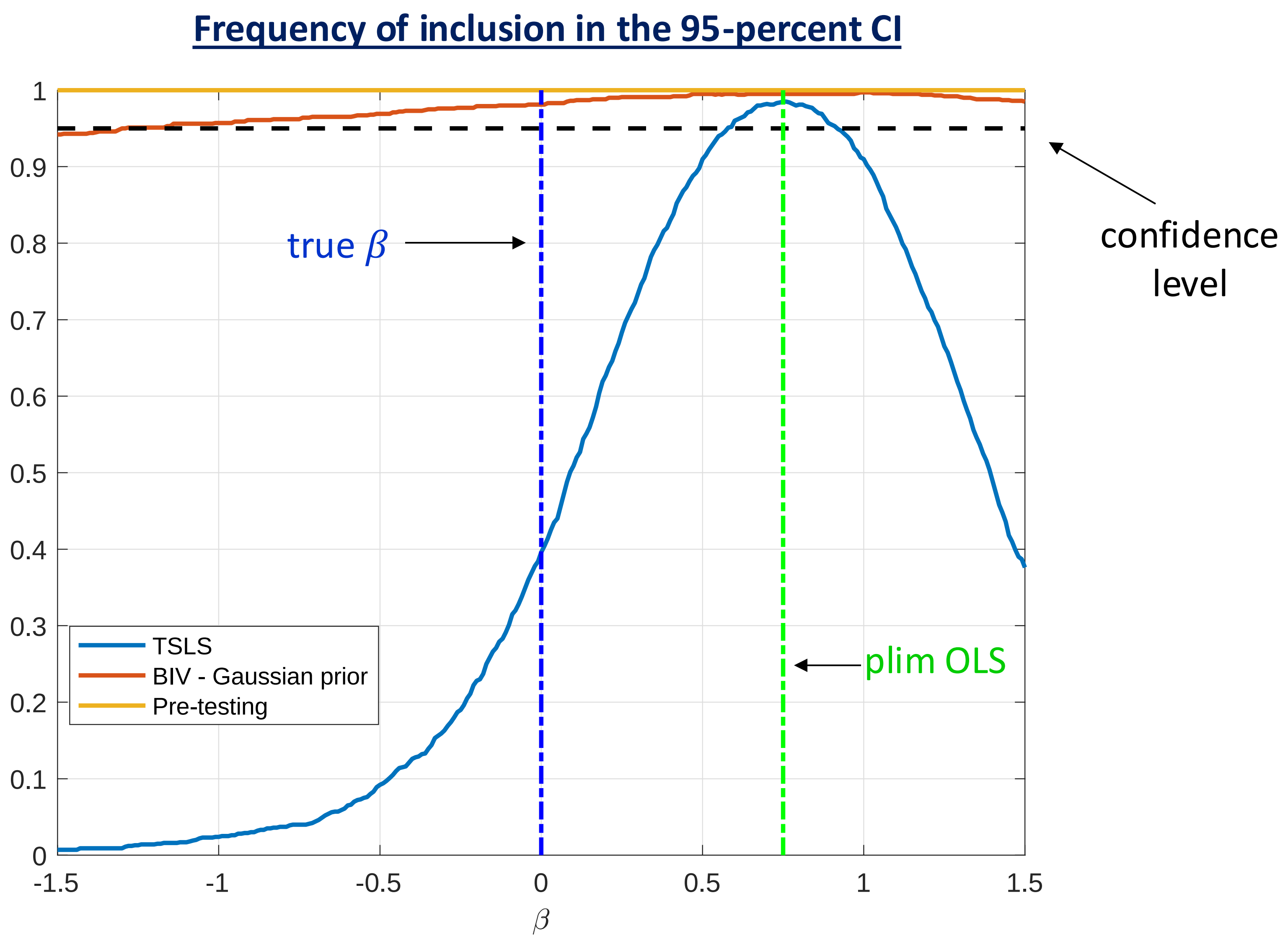
TSLS, BIV and pre-testing with fairly weak instruments



TSLS, BIV and pre-testing with very weak instruments



TSLS, BIV and pre-testing with irrelevant instruments

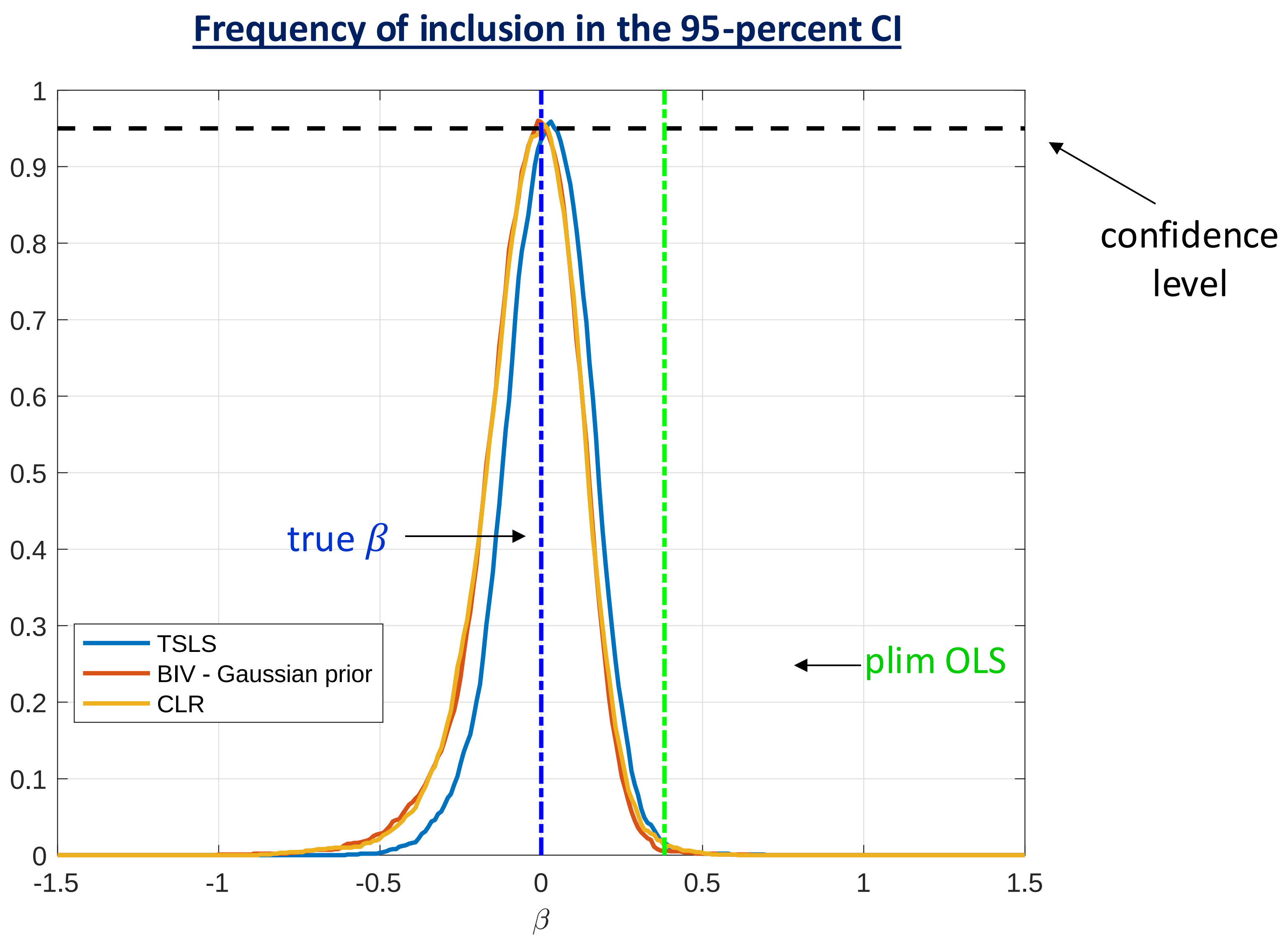


Next few slides

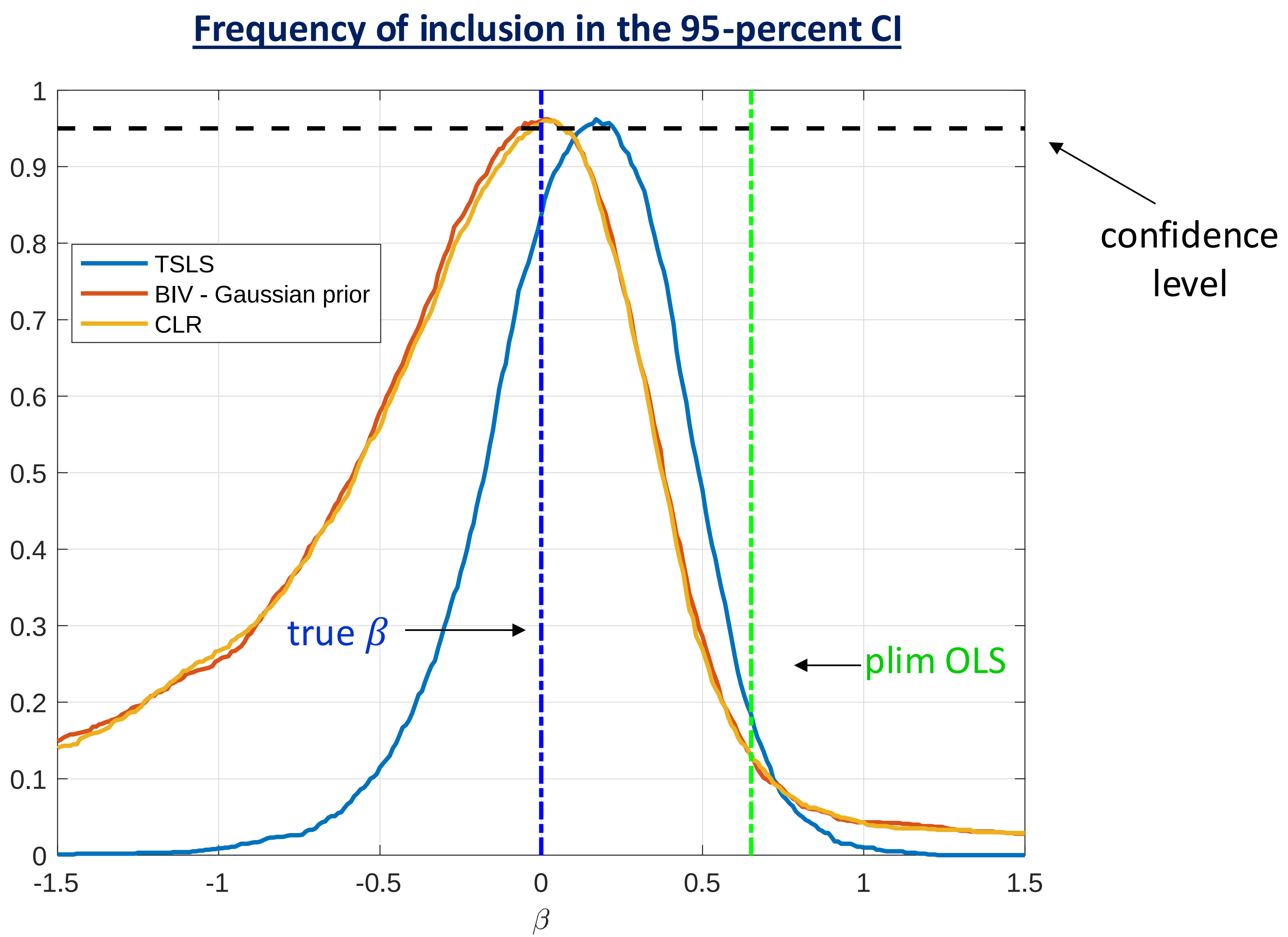
- Simulation evidence
  - Instruments:
    - strong
    - fairly weak
    - very weak
    - irrelevant
  - Compare methods:
    - TSLS
    - BIV—Gaussian prior

↑ Working as intended!

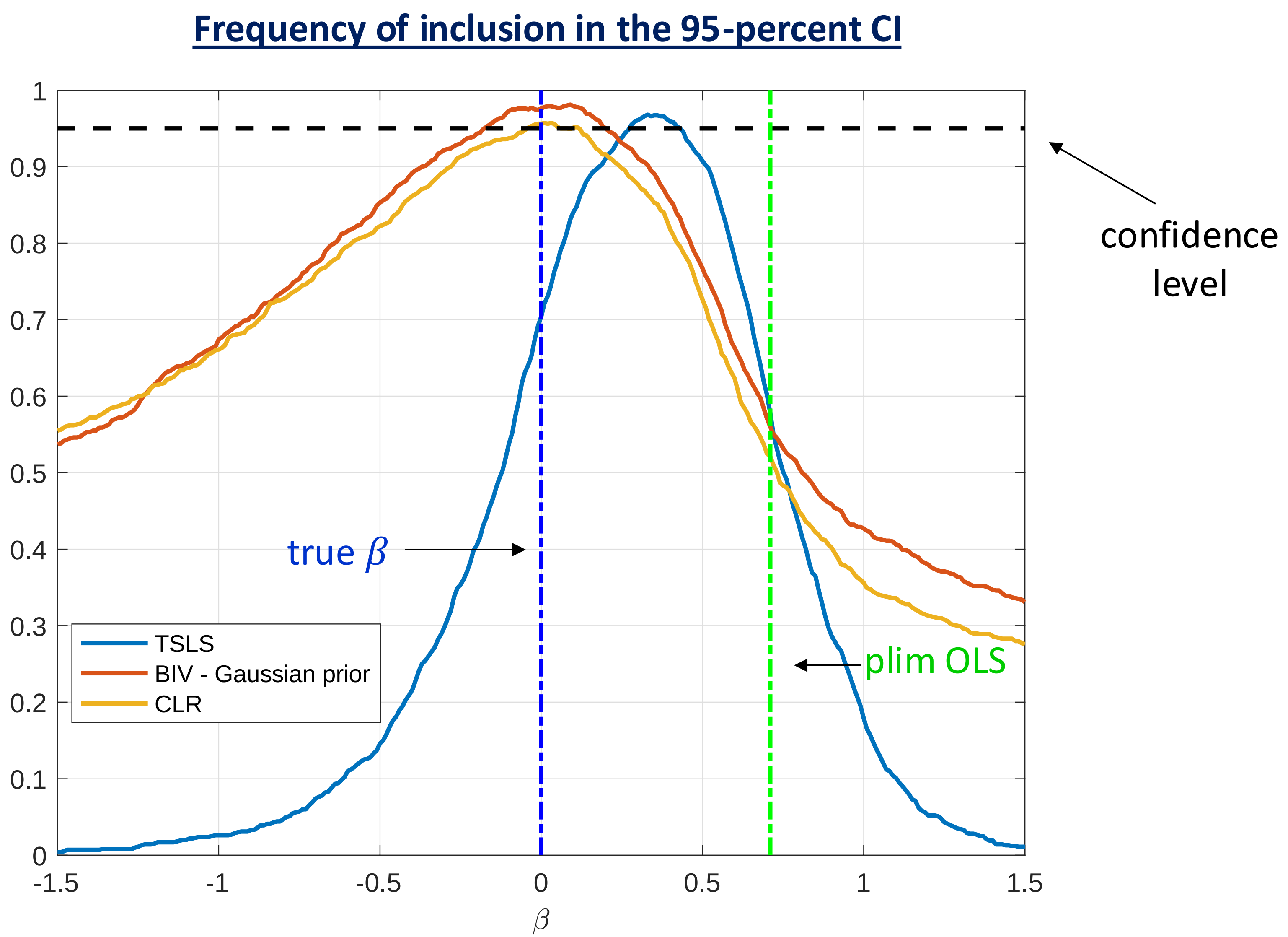
TSLS, BIV and CLR with strong instruments



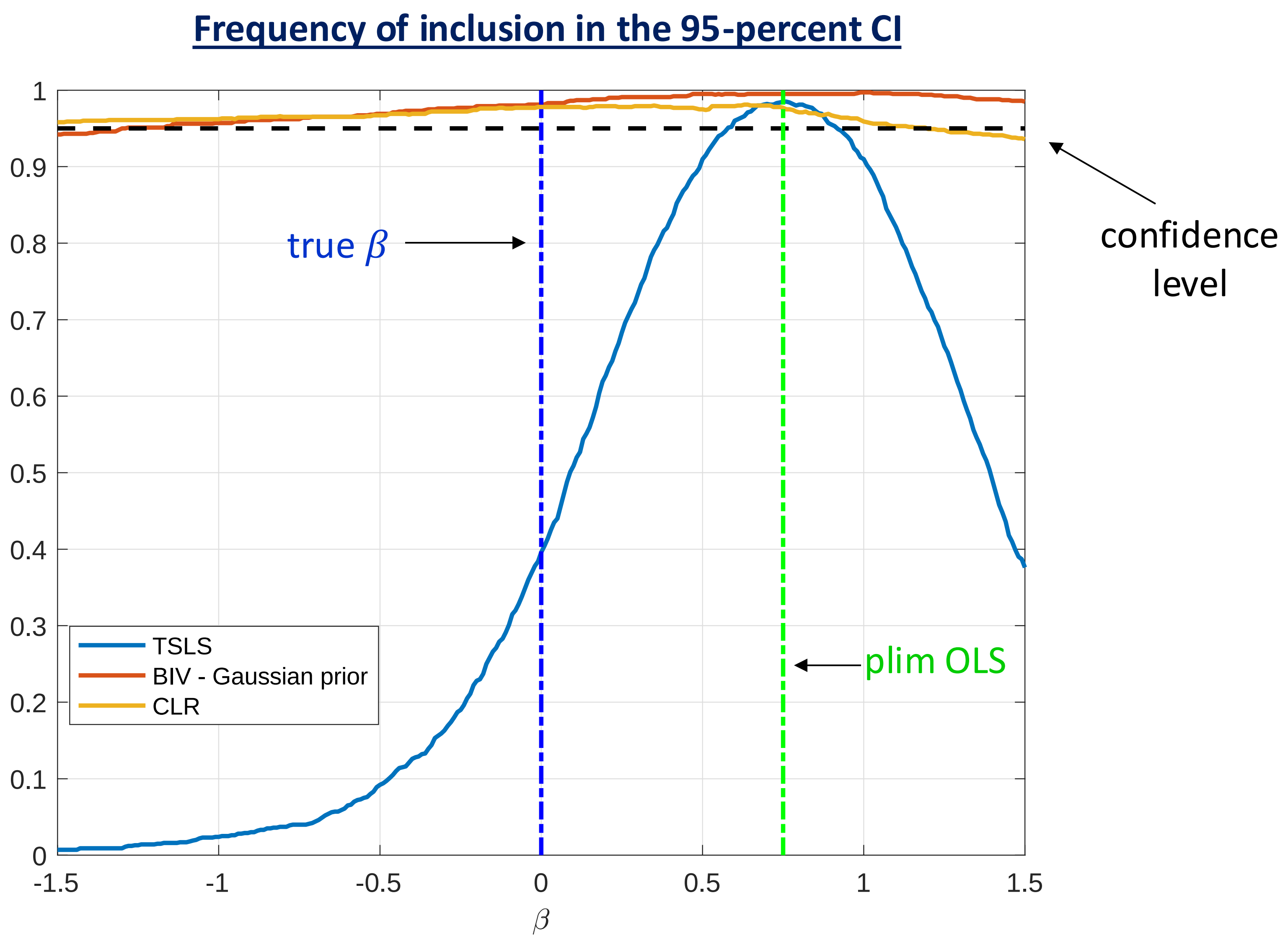
TSLS, BIV and CLR with fairly weak instruments



TSLS, BIV and CLR with very weak instruments



TSLS, BIV and CLR with irrelevant instruments



Back to empirics: Estimating the return to education

	AK	
TSLS	**.083** **(.009)**	

Back to empirics: Estimating the return to education

	AK	
TSLS	**.083** **(.009)**	
BIV **Gaussian prior**	**0.097** **(0.017)**	

- BIV estimate
  - with true instruments → Similar to AK estimates

Back to empirics: Estimating the return to education

	AK	
TSLS	**.083** **(.009)**	
BIV **Gaussian prior**	**0.097** **(0.017)**	
CLR	**[0.065–0.128]**	

- BIV estimate
  - with true instruments → Similar to AK estimates

Back to empirics: Estimating the return to education

	AK	Fake Zs
TSLS	**.083 (.009)**	**44.8%**
BIV **Gaussian prior**	**0.097 (0.017)**	**0.4%**
CLR	**[0.065–0.128]**	

"significant" estimates across simulations

- BIV estimate
  - with true instruments → Similar to AK estimates
  - with fake instruments → Detects the irrelevance of IV

Back to empirics: Estimating the return to education

	AK	Fake Zs
TSLS	**.083 (.009)**	**44.8%**
BIV Gaussian prior	**0.097 (0.017)**	**0.4%**
CLR	**[0.065–0.128]**	**6.2%**

"significant" estimates across simulations

- BIV estimate
  - with true instruments → Similar to AK estimates
  - with fake instruments → Detects the irrelevance of IV

Summing up

- Traditional IV inference is distorted by an implicit prior
  - It favors instrument strength
  - Unintended consequence: Standard errors might be unrealistically tight

- A simple *agnostic prior* on instrument strength solves the problem
  - Bayesian inference robust to weak instruments

Additional slides

Relation with the literature

- Weak instruments
  - Large frequentist literature
    - Staiger and Stock (1997), Stock and Wright (2000), Moreira (2003), Andrews, Moreira and Stock (2006), Mikusheva (2012), Andrews, Stock and Sun (2019),…
  - ✓ We study the problem of weak instruments from Bayesian perspective

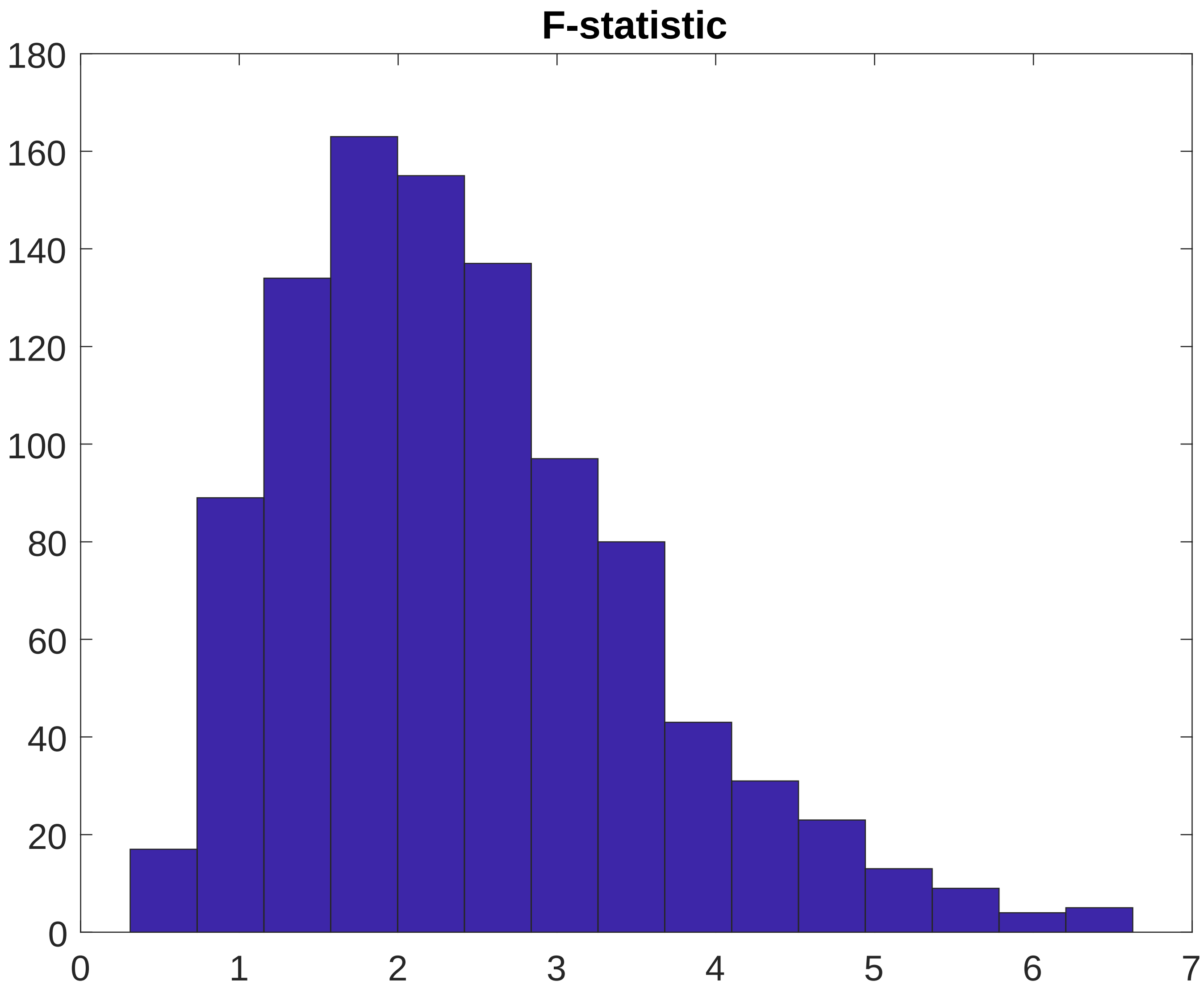
- Bayesian inference
  - Focus on deriving implicit priors that justify standard frequentist results
    - Zellner (1971), Drèze (1976), Maddala (1976), Bawens and Van Djik (1986), Kleibergen (1997), Kleibergen and Van Dijk (1998), Chao and Phillips (1998), Kleibergen and Zivot (2003), Lopez and Polson (2014),…
  - ✓ We study the pathology of the posterior that emerges when instruments are weak, connect it to overfitting in the first stage, and suggest informative priors

Relation with the literature

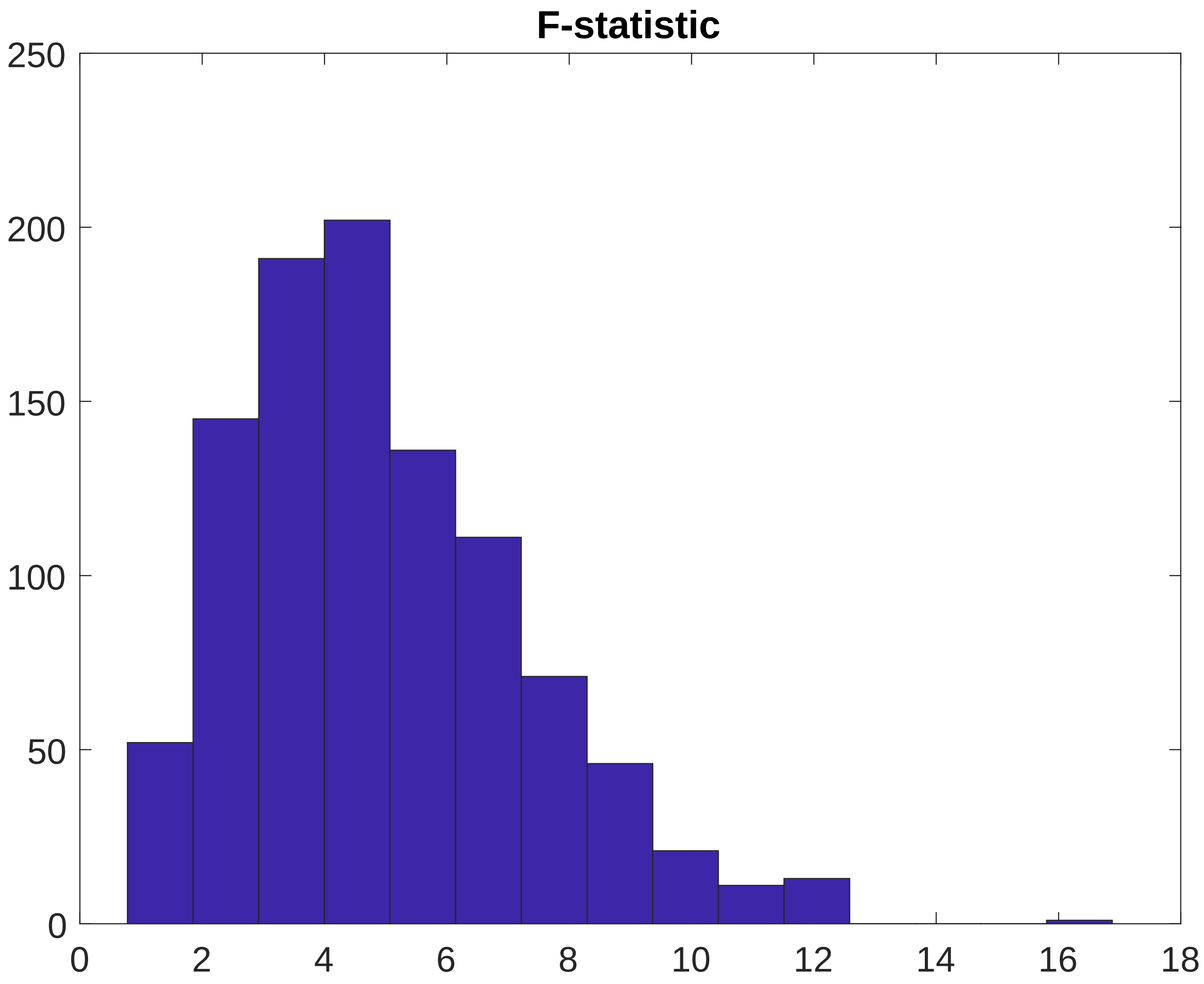
- Shrinkage approaches to the many instruments problem
  - Chamberlain and Imbens (2003): Random effects (Gaussian prior)
  - Carrasco (2012): Tikhonov, PCA, Landweber–Fridman (Gaussian prior)
  - Bai and Ng (2010), Kapetanios and Marcellino (2010), Hahn, Le, and Lopez (2018): PCA (Gaussian prior)
  - Belloni et al. (2012): Lasso (Double exponential prior)
  - Koop, Leon-Gonzalez, and Strachan (2012): BMA (Spike and Slab prior)
  - ✓ We show that these approaches robustly improve inference also when the number of instruments is small

- Robustness
  - Large frequentist literature on inference robust to weak instruments
    - CLR, AR, Wald, LM, …
  - ✓ We find that shrinkage priors give very similar results
    (We are working to establish a theoretical link with CLR, not there yet)

F-stats of simulations with very weak instruments



F-stats of simulations with fairly weak instruments



F-stats of simulations with strong instruments