SENSITIVITY ANALYSIS USING APPROXIMATE MOMENT CONDITION MODELS

Timothy Armstrong (Yale University)

Michal Kolesár (Princeton University)

Mar 2019

- Economic models often viewed as approximation of reality
  - Yet inference typically proceeds as if approximation exact
- This paper: take this view seriously and consider inference in GMM when moment conditions hold only approximately
  - Includes failure of exclusion restrictions, functional form misspecification, measurement error etc.
  - Sensitivity analysis: do conclusions still hold if model only "approximately" true? How large does misspecification need to be to overturn main result?

- Interested in $h(\theta)$
  - θ_0 is true value of some parameter and, $h\colon \mathbb{R}^{d_\theta} \to \mathbb{R}$ known function
- Moment condition $g(\theta) = E[g(w_i, \theta)]$ locally misspecified

$$g(\theta_0) = c/\sqrt{n}, \qquad c \in \mathcal{C},$$

  with set $\mathcal{C}$ formalizing how moment conditions fail (e.g. $\mathcal{C} = \{c : \|c\| \leq M\}$ for some norm $\|\cdot\|$)
- $\mathcal{C} = \{0\} \implies$ usual (correctly specified) GMM setup

- Valid CIs easy to construct: take GMM estimator ± std error × critical value (cv) that is > 1.96 and takes into account possible bias over C
  - Estimator still asymptotically normal, use of cv only "nonstandard" aspect
- Key insight: because CIs must be widened to take into account bias, optimal weight matrix W different from correctly specified case
  - bias-variance trade-off: less weight on moments with potentially large misspecification according to C
  - Empirically important: in application to Berry, Levinsohn, and Pakes [1995], adjusting W shrinks CI up to 3× or more
- When C convex, cv and optimal W easy to compute by convex programming. Under ℓ_p constraints, obtain weightings analogous to penalized regression (e.g. ridge/LASSO)

- Show our CI (with optimally chosen W) near-optimal at $c = 0$ when $\mathcal{C}$ convex and centrosymmetric ($c \in \mathcal{C} \implies -c \in \mathcal{C}$)
  - Derive limiting experiment for locally misspecified GMM
  - Optimality despite "pessimism": cv based on worst possible misspecification over $\mathcal{C}$
- Two important implications:
  1. Cannot determine magnitude of misspecification from the data; size of $\mathcal{C}$ must be chosen a priori, including M if $\mathcal{C} = \{c : \|c\| \leq M\}$. Instead recommend varying M as a form of sensitivity analysis
  2. Cannot use model selection to determine which moments invalid: valid post-model selections CIs cannot be asymptotically substantively shorter than ours, even if it "turns out" that model correctly specified.

Results for estimation may be different. Leave to future research.

- Liao [2013], Cheng and Liao [2015], DiTraglia [2016]...: use possibly invalid moments for MSE improvements
- Andrews and Guggenberger [2009], DiTraglia [2016], McCloskey [2017] propose valid post-model selection CIs:
  - *This paper:* these CIs cannot substantively improve upon CI based only on valid moments

RELATED LITERATURE: CORRECTLY SPECIFIED GMM

- Correctly specified GMM [Hansen and Singleton, 1982, Hansen, 1982]
    - Unconstrained linear model (Gauss-Markov) + asymptotic linear approximation for nonlinear model = optimal estimator/CI under correct specification
    - Chamberlain [1987] (see also Hansen, 1985): formalize optimality of estimator; van der Vaart [1998, Chapter 25]: one-sided CI optimality
- Locally misspecified GMM
    - Linear model with convex parameter space (Ibragimov & Khasminskii, Donoho, Cai, Low, Armstrong & Kolesár) + asymptotic linear approximation for nonlinear model = optimal estimator/CI under local misspecification
    - *This paper:* formalize notion estimator/CI near-optimal (Le Cam style arguments)

RELATED LITERATURE: LOCAL MISSPECIFICATION

- Newey [1985]: use influence fkt weights to calculate local asy bias
- Andrews, Gentzkow, and Shapiro [2017]: one should report these weights, call them "sensitivity"
    - *This paper*: (1) use sensitivity to compute worst-case bias and construct CI; (2) derive optimal sensitivity, and show improvement can be dramatic
- Kitamura, Otsu, and Evdokimov [2013]: derive optimal estimator when misspecification bounded by Hellinger distance
    - *This paper*: under this form of misspecification optimal W same as under correct specification: both Kitamura et al. [2013] estimator and usual GMM are optimal
- Independent work by Bonhomme and Weidner [2018]: inference and optimal estimation when C determined relative to larger class of models.

- True parameter value $\theta_0 \in \mathbb{R}^{d_\theta}$ satisfies

$$g(\theta_0) \equiv Eg(w_i, \theta_0) = c/\sqrt{n}, \quad c \in \mathcal{C} \subset \mathbb{R}^{d_g}$$

- Observe $\{w_i\}_{i=1}^n$. Let $\hat{g}(\theta) = \frac{1}{n}\sum_{i=1}^n g(w_i, \theta)$, $\Gamma = \frac{d}{d\theta'} g(\theta)\big|_{\theta=\theta_0}$.
- Assume standard regularity conditions:

$$\sqrt{n}(\hat{g}(\theta_0) - g(\theta_0)) \xrightarrow{d} \mathcal{N}(0, \Sigma),$$

$$\text{for} \quad \theta_n = \theta_0 + O_P(n^{-1/2}) \colon \hat{g}(\theta_n) - \hat{g}(\theta_0) = \Gamma(\theta_n - \theta_0) + o_P(1/\sqrt{n}),$$

- Interested in $h(\theta_0) \in \mathbb{R}$ with h cont. diff. H $1 \times d_\theta$ derivative matrix at θ_0

- Consider asymptotically linear estimator $\hat{h}$,

$$\sqrt{n}(\hat{h} - h(\theta_0)) = k'\sqrt{n}\hat{g}(\theta_0) + o_P(1) \xrightarrow{d} \mathcal{N}(k'c, k'\Sigma k),$$

  where k is its *sensitivity* (using Andrews et al. [2017] terminology)
- Holds for typical estimators, with $\Sigma = var(g(w_i, \theta_0))$ in iid case.
    - If $\hat{h} = h(\hat{\theta}_W)$ where $\hat{\theta}_W = \text{argmin}_\theta\, \hat{g}(\theta)'W\hat{g}(\theta)$ is GMM estimator with weight matrix W, this holds with [Newey, 1985]

$$k' = -H(\Gamma'W\Gamma)^{-1}\Gamma'W$$

- Let $\hat{k}$ and $\hat{\Sigma}$ be consistent estimates of k and Σ.

- To construct CI, note z-statistic satisfies

$$\frac{\sqrt{n}(\hat{h} - h(\theta_0))}{\sqrt{\hat{k}'\hat{\Sigma}\hat{k}}} \xrightarrow{d} \mathcal{N}\left(\frac{k'c}{\sqrt{k'\Sigma k}}, 1\right)$$

- So mean of z-statistic is asymptotically $|t| \leq \frac{\overline{\text{bias}}_C(k)}{\sqrt{k'\Sigma k}}$, with $\overline{\text{bias}}_C(k) \equiv \sup_{c \in C} |k'c|$.
- Let $\text{cv}_\alpha(t)$ be the $1 - \alpha$ quantile of $|\mathcal{N}(t, 1)|$
- Leads to $100 \cdot (1 - \alpha)\%$ CI

$$\hat{h} \pm \text{cv}_\alpha\left(\frac{\overline{\text{bias}}_C(\hat{k})}{\sqrt{\hat{k}'\hat{\Sigma}\hat{k}}}\right) \cdot \sqrt{\hat{k}'\hat{\Sigma}\hat{k}}/\sqrt{n}$$

- One-sided CI simply $[\hat{h} - \overline{\text{bias}}_C(\hat{k}) - z_{1-\alpha}\sqrt{\hat{k}'\hat{\Sigma}\hat{k}}, \infty)$.

HOW TO CONSTRUCT ESTIMATORS WITH GIVEN SENSITIVITY k

- Let $\hat{\theta}_{\text{initial}}$ be some initial $\sqrt{n}$-consistent estimator
- Then one-step estimator

$$\hat{h} = h(\hat{\theta}_{\text{initial}}) + \hat{k}'\hat{g}(\hat{\theta}_{\text{initial}}) \quad \text{where } \hat{k} \xrightarrow{p} k.$$

has sensitivty k, so long as $H = -k'\Gamma$, since by Taylor expansion:

$$\begin{aligned}\sqrt{n}\left(\hat{h} - h(\theta_0)\right) &= H\sqrt{n}(\hat{\theta}_{\text{initial}} - \theta_0) + \hat{k}'\sqrt{n}\hat{g}(\hat{\theta}_{\text{initial}}) + o_P(1) \\ &= (H + \hat{k}'\Gamma)\sqrt{n}(\hat{\theta}_{\text{initial}} - \theta_0) + \hat{k}'\sqrt{n}\hat{g}(\theta_0) + o_P(1)\end{aligned}$$

- Class of one-step estimators asymptotically equivalent to class of GMM estimators: optimal k determines optimal weight matrix W

- Asymptotic width of CI is

$$2\,\mathrm{cv}_\alpha\left(\overline{\mathrm{bias}}_C(k)/\sqrt{k'\Sigma k}\right)\cdot\sqrt{k'\Sigma k}/\sqrt{n}$$

- Depends on k only through $\overline{\mathrm{bias}}_C(k)$ (worst-case bias) and $k'\Sigma k$ (variance), increasing in both. Thus, to compute optimal sensitivity, enough to
    1. Minimize variance subject to bound $\overline{B}$ on bias, and subject to $H = -k'\Gamma$
    2. Vary $\overline{B}$ to find optimal k
- Generalizes to other criteria: maximum MSE ($[\overline{\mathrm{bias}}_C(k)^2 + k'\Sigma k]/\sqrt{n}$), absolute error loss… Optimal k depends on criterion

- Bias-variance tradeoff problem

$$\min_{k} k'\Sigma k \quad \text{s.t.} \quad H = -k'\Gamma, \quad \sup_{c\in C} |k'c| \leq \overline{B}$$

- Problem particularly tractable when C convex: can cast as convex optimization problem
    - See paper for details; closed-form solutions available in special cases

To compute (feasible) optimal sensitivity:

1. Obtain initial estimates $\hat{\theta}_{\text{initial}}$ and $\hat{H}$, $\hat{\Gamma}$ and $\hat{\Sigma}$. For each $\overline{B}$, solve for $\hat{k}_{\overline{B}}$ (optimal k with $\hat{H}$, $\hat{\Gamma}$ and $\hat{\Sigma}$ plugged in)
2. Choose $\overline{B}$ to minimize length of resulting CI, or else maximum MSE $\overline{B}^2 + \hat{k}'_{\overline{B}}\hat{\Sigma}\hat{k}_{\overline{B}}$.
3. Report $\hat{h} = h(\hat{\theta}_{\text{initial}}) + \hat{k}'_{\overline{B}}\hat{g}(\hat{\theta}_{\text{initial}})$ as estimate and $\hat{h} \pm \text{cv}_{\alpha}\left(\overline{B}/\sqrt{\hat{k}'_{\overline{B}}\hat{\Sigma}\hat{k}_{\overline{B}}}\right) \cdot \sqrt{\hat{k}'_{\overline{B}}\hat{\Sigma}\hat{k}_{\overline{B}}}/\sqrt{n}$ as CI
    - Alternatively, use GMM with weight matrix implied by $\hat{k}_{\overline{B}}$

Misspecification-robust CIs

Efficiency bounds

Applications / Examples

Empirical Application

- Estimator with sensitivity k, s.t. $H = -k'\Gamma$ has asy distribution $\mathcal{N}(k'c, k'\Sigma k)$
- Isomorphic to approximately linear regression model

$$Y = -\Gamma\theta + c + \Sigma^{1/2}\varepsilon, \quad c \in \mathcal{C},\ \theta \in \mathbb{R}^{d_\theta}, \quad \varepsilon \sim \mathcal{N}(0, I),$$

  design matrix $-\Gamma$, c is approximation error, interested in $H\theta$.
- Estimator $k'Y$ has $\text{var}(k'Y) = k'\Sigma k$ and bias $k'c - (k'\Gamma - H)\theta$.
  - Condition $H = -k'\Gamma$, means we restrict attention to estimators with finite worst-case bias, $\sup_{c\in\mathcal{C}}|k'C|$

- Sacks and Ylvisaker [1978] considered a special case of this model with $\Sigma = I$ and $C =$ hyperrectangle. Generalization to our case covered by Donoho [1994], Low [1995]
- Our CI corresponds asymptotically to Donoho's CI in limit experiment,

$$k_{B_*}' Y \pm \mathrm{cv}_\alpha(B_* / \sqrt{k_{B_*}' \Sigma k_{B_*}}) \cdot \sqrt{k_{B_*}' \Sigma k_{B_*}}, \qquad (*)$$

where $B_* = \operatorname{argmin}_B \mathrm{cv}_\alpha(B / \sqrt{k_B' \Sigma k_B}) \cdot \sqrt{k_B' \Sigma k_B}$

- Potential drawback of (*): CI length determined by worst possible misspecification in C: can we do better when $c \approx 0$?
- Efficiency benchmark: among conf. sets with coverage $\geq 1 - \alpha$ for all $\theta \in \mathbb{R}^{d_\theta}$ and $c \in C$, minimize expected length when $\theta = \theta^*$ and $c = 0$.
- Let $\kappa_*(H, \Gamma, \Sigma, C)$ denote ratio of length of this CI to (*). Explicit formula follows from results in Armstrong and Kolesár [2018] when C convex and centrosymmetric (see paper)
  - When $C \subseteq \mathbb{R}^{d_g}$ linear subspace, $\kappa_*(H, \Gamma, \Sigma, C) = \frac{(1-\alpha)z_{1-\alpha} + \phi(z_{1-\alpha})}{z_{1-\alpha/2}} \geq \frac{z_{1-\alpha}}{z_{1-\alpha/2}}$
  - Show in paper that κ_* admits universal lower bound $(z_{1-\alpha}(1-\alpha) - \tilde{z}_\alpha \Phi(\tilde{z}_\alpha) + \phi(z_{1-\alpha}) - \phi(\tilde{z}_\alpha))/z_{1-\alpha/2} = 71.7\%$ at $\alpha = 0.05$, where $\tilde{z}_\alpha = z_{1-\alpha} - z_{1-\alpha/2}$

- Show using Le Cam-style arguments (least favorable submodels, etc.) that bound κ^* indeed translates to locally misspecified GMM
- Require asymptotic coverage uniformly over pairs $(\theta, P) \in \Theta_n \times \mathcal{P}_n$ such that $\sqrt{n}g_P(\theta) \in C$.

Theorem 1

If C is convex and centrosymmetric, and $\mathcal{P}_n$ contains a rich enough set of qmd submodels through correctly specified model, then under regularity conditions, (1) CI (*) is asymptotically valid; and (2) Bound κ_* on ratio of asy lengths holds for any other asy valid CI

To build intuition for this efficiency result, walk through special cases:

1. Correctly specified case, $C = \{0\}$
2. Some moments are valid, some may be invalid, $C = \{(0', \gamma')' : \gamma \in \mathbb{R}^{d_\gamma}\}$
3. General case

- If $C = \{0\}$, limiting experiment reduces to $Y = -\Gamma\theta + \Sigma^{1/2}\varepsilon$, $\varepsilon \sim \mathcal{N}(0, I)$.
- Gauss-Markov theorem: best unbiased estimator has sensitivity $k_{LS,0} = -H(\Gamma'\Sigma^{-1}\Gamma)^{-1}\Gamma'\Sigma^{-1}$. *This estimator also minimaxes MSE.*
  - $k_{LS,0}$ is sensitivity of $h(\hat{\theta}_{\Sigma^{-1}})$, GMM with $W = \Sigma^{-1}$. Chamberlain [1987] shows $h(\hat{\theta}_{\Sigma^{-1}})$ is asymptotically minimax for MSE.
  - *Theorem implication:* (linear subspace case) usual two-sided CI based on $h(\hat{\theta}_{\Sigma^{-1}})$ has asy efficiency $\geq z_{1-\alpha}/z_{1-\alpha/2}$ relative to CI that optimizes length at θ^* when indeed $\theta = \theta^*$.
    - New result (to our knowledge)
    - Limiting experiment also implies usual two-sided CI achieves asy minimax expected length

- Suppose $C = \{(0', \gamma')' \colon \gamma \in \mathbb{R}^{d_\gamma}\}$. Then limiting experiment becomes $\binom{Y_1}{Y_2} = -\binom{\Gamma_1\theta}{\delta} + \Sigma^{1/2}\varepsilon$, with δ unrestricted
- In limiting experiment, GLS estimator based on Y_1 best unbiased and minimax, property carries over to GMM model
  - May prefer different estimator under weighted MSE criterion. Liao [2013], Cheng and Liao [2015], and DiTraglia [2016] propose estimators that aim to reduce risk when γ small
- *Theorem implication:* (linear subspace case) CI based on GMM with valid moments only has efficiency $\geq z_{1-\alpha}/z_{1-\alpha/2}$ (and asy minimax)
  - CIs based on shrinkage / model-selection [e.g. Andrews and Guggenberger, 2009, DiTraglia, 2016, McCloskey, 2017] cannot achieve substantive improvement, and must pay for it
  - One-sided version of our theorem $\implies$ one-sided CI 100% efficient

- In order to make use of invalid moments, need to restrict magnitude of misspecification: leads to general case with $\mathcal{C}$ convex
- CI (*) close to optimal for "directing power" at $c = 0$ when $\mathcal{C}$ convex and centrosymmetric.
    - $\kappa_* \geq 71.6\%$, in general, can be higher for particular Γ, Σ and $\mathcal{C}$.
    - When $\mathcal{C} = \{c : \|c\| \leq M\}$ for some norm $\|\cdot\|$, cannot specify a conservative M, and construct a CI that is shorter when it "turns out" model correctly specified

Misspecification-robust CIs

Efficiency bounds

Applications / Examples

Empirical Application

- For general C, k_δ can be computed using convex optimization; In particular cases, closed-form solutions may be available
- Computationally convenient and flexible to consider sets $C = \{B\gamma : \|\gamma\| \le M\}$, where B is $d_g \times d_\gamma$ matrix
    - B may incorporate standardizing moments by their variability; or account for their correlations
    - Can vary M to determine sensitivity of given result
    - When $\|\cdot\| = \ell_\infty$ norm, bounds on each element of γ do not interact.
        - simple interpretation: no single element of γ can be greater than M
    - Optimal sensitivity analogous penalized regression problem: in limiting experiment, $Y = -\Gamma\theta + B\gamma + \Sigma^{1/2}\epsilon$, with $\|\gamma\| \le M$

$\|\cdot\| = \ell_1$ OR ℓ_∞

Use penalized regression analogy to develop a simple algorithm for computing optimal sensitivity (details in paper)

- Similar to the LASSO/LAR algorithm [Efron, Hastie, Johnstone, and Tibshirani, 2004]
- Optimal sensitivities piecewise linear as $\overline{B}$ varies
- Under ℓ_∞ constraints, as M grows, optimal sensitivity successively drops "least informative" moments

$\|\cdot\| = \ell_2$

- Penalized regression analogy leads to ridge regression
- Optimal sensitivity is $k'_\lambda = -H(\Gamma' W_\lambda \Gamma)^{-1}\Gamma' W_\lambda$, where $W_\lambda = (\lambda M^2 BB' + \Sigma)^{-1}$; λ is relative weight on bias
- W_λ = optimal weight under correct specification and variance $\lambda M^2 BB' + \Sigma$
  - Arises under random-effects approach: $\gamma \sim [0, \lambda M^2 I]$, or under Bayes approach with normal prior on γ and diffuse prior on θ
  - RE/Bayes approach leads to different CI: misspecification bias becomes part of additional variability of the moments
  - This RE specification suggested in Conley, Hansen, and Rossi [2012] (they don't consider implications for form of optimal estimator)
  - Conceptual issues when c random but not θ: what's the DGP?

- In general, $W = \Sigma^{-1}$ suboptimal under local misspecification
- Exception: $C = \{\Sigma^{1/2}\gamma : \|\gamma\|_2 \leq M\}$
  - Uncertainty from potential misspecification exactly $\propto$ asymptotic sampling uncertainty
  - Can simply use $h(\hat{\theta}_{\Sigma^{-1}})$, with critical value $\mathrm{cv}_\alpha(M)$, since asy bias bounded by $\sqrt{k'\Sigma^{-1}k}$.
- $\kappa_* \geq z_{1-\alpha}/z_{1-\alpha/2}$ for $\alpha = 0.05$ in this case
- Andrews, Gentzkow, and Shapiro [2018] show this form of C arises whenever misspecification defined in terms of magnitude of any member of Cressie and Read [1984] family of divergences
  - Kitamura et al. [2013] use Hellinger distance: usual GMM estimator, as well as Kitamura et al. [2013] estimator near-optimal for CI construction

- Single equation IV: $y_i = x_i'\theta + \varepsilon_i$
- Suppose $\varepsilon_i = z_{Ii}'\gamma/\sqrt{n} + \eta_i$, where $E[z_i\eta_i] = 0$, and z_{Ii} corresponds to subset of instruments z_i that may be invalid
    - Considered in Hahn and Hausman [2005], Conley et al. [2012], Andrews et al. [2017]…
    - Equivalent to interpreting z_{Ii} as valid control variable, with restriction on magnitude of its coefficient
- Leads to $C = \{B\gamma : \|\gamma\| \leq M\}$ where $B = E[z_i z_{Ii}']$
- Linearity implies that one-step estimator $\hat{h}$ doesn't depend on $\hat{\theta}_{\text{initial}}$: $\hat{h} = \hat{k}'\frac{1}{n}\sum_{i=1}^{n} y_i z_i$
- Under ℓ_2 norm, If all IVs invalid, optimal estimator is usual GMM estimator (TSLS under homoscedastic errors)

- Omitted variables bias in linear regression
- Treatment effect extrapolation [Kowalski, 2016, Brinch, Mogstad, and Wiswall, 2017, Mogstad, Santos, and Torgovitsky, 2017]
- Nonparametric IV with discrete covariates

Misspecification-robust CIs

Efficiency bounds

Applications / Examples

Empirical Application

- Berry et al. [1995, BLP] estimate a supply and demand model of automobile demand using nonlinear IV.
- Andrews et al. [2017, AGS] report sensitivity k that corresponds to GMM, and use this to calculate bias under local misspecification.
- We consider sets C based on perturbations under which AGS calculate bias, fix parameter of interest $h(\theta)$ to average markup as % of price
- How does optimal k differ from GMM sensitivity? How much does using optimal k shrink the CI?

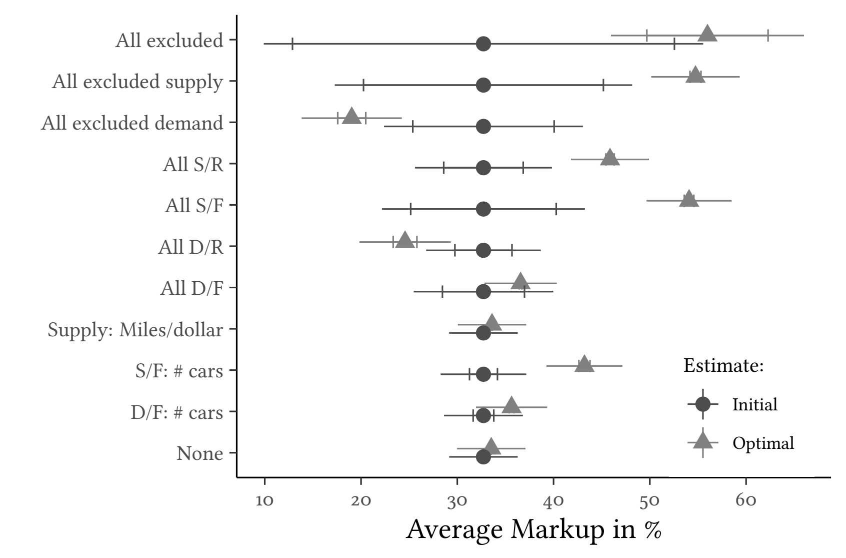
- Can invert market shares to yield unobservable characteristic of car j as a fkt of model parameters, $\xi_i(\theta)$, and use Bertand-Nash eq'm conditions to invert unobservable marginal cost component $\omega_j(\theta)$
- Supply and demand instruments z_j^d and z_j^s yield sample moment condition

$$\hat{g}(\theta) = \frac{1}{n}\sum_{j=1}^{n}\begin{pmatrix} z_j^d \xi_j(\theta) \\ z_j^s \omega_j(\theta) \end{pmatrix},$$

- If z_j^s and z_j^s appear directly in utility/cost function, obtain nonlinear version of misspecified IV model
- Instruments are "BLP instruments": product characteristics/cost components of other cars produced by same firm/produced by rivals.

- As in AGS, scale B so that:
  - Consumer willingness to pay for 1sd increase in the ℓth demand-side instrument is $\gamma_\ell^d\%$ of the average car price
  - Increasing the ℓth supply-side instrument by 1sd decreases MC by $\gamma_\ell^s\%$ of the average car price.
- Given set I of potentially invalid instruments, use scaling $C = \{B_I\gamma : \|\gamma\|_p \leq M|I|^{1/p}\}$, where $\gamma = (\gamma^{d\prime}, \gamma^{s\prime})'$.
  - Scaling by $|I|^{1/p}$ ensures $\gamma = M(1, \dots, 1)'$ always included in the set.
  - Take $M = 1$ as baseline
- Show results when estimator chosen to minimize length of CI

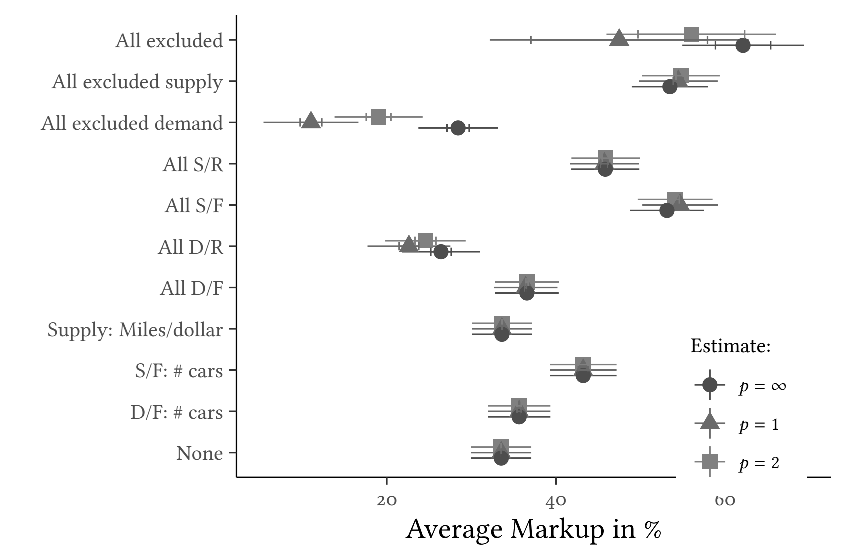
CIS WHEN $p = 2$ UNDER DIFFERENT SETS OF INVALID INSTRUMENTS



Instrument set	$p = 1$	$p = 2$	$p = \infty$
D/F: # cars	9.77	9.77	9.77
S/F: # cars	15.32	15.32	15.32
Supply: Miles/dollar	17.00	17.00	17.00
All D/F	2.38	2.59	2.59
All D/R	4.08	5.22	5.40
All S/F	2.04	2.61	2.62
All S/R	2.47	4.16	6.99
All excluded demand	1.19	1.72	1.88
All excluded supply	1.02	1.64	1.78
All excluded	0.48	1.08	2.54

Notes: The table gives the minimum value of M such that the test of overidentifying restrictions has p-value equal to 0.05. J-statistic: 404.7.

OPTIMAL CIS UNDER $p = 1, 2, \infty$

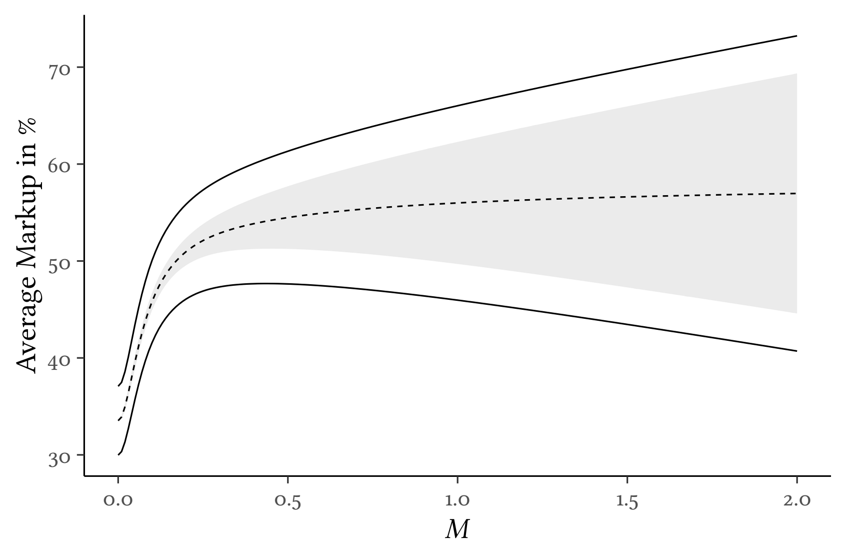


κ_* UNDER ℓ_p MISSPECIFICATION, $M = 1$

Instrument set	$p = 1$	$p = 2$	$p = \infty$
D/F: # cars	85.9	85.9	85.9
S/F: # cars	90.1	90.1	90.1
Supply: Miles/dollar	85.0	85.0	85.0
All D/F	85.4	85.5	85.7
All D/R	94.3	94.8	95.3
All S/F	88.0	88.6	89.1
All S/R	89.5	89.4	89.2
All excluded demand	95.0	95.4	96.4
All excluded supply	89.8	90.3	90.1
All excluded	96.3	97.0	97.5

For comparison: $z_{1-\alpha}/z_{1-\alpha/2} = 83.9\%$

OPTIMAL CIS UNDER $p = 2$



- We derived (near) optimal CIs under local misspecification
  - Use usual GMM estimator, but adjust weight matrix to put more weight on moments with low specification error. Adjusting weight matrix matters a lot in BLP application
  - $\pm$ std error $\times$ critical value that is > 1.96 and takes into account worst-case bias.
- Asymptotic efficiency bounds have implications for pre-tests for correct specification and for model/moment selection