Standard Errors for Calibrated Parameters

Matthew D. Cocci  Mikkel Plagborg-Møller
Princeton University  Princeton University

Slides: https://scholar.princeton.edu/mikkelpm

June 11, 2021
Data combination for structural inference

• Structural models are often calibrated to match many kinds of “moments”, e.g.:
  • Micro vs. macro.
  • High-frequency vs. low-frequency.
  • Quantiles vs. regression coefficients.
  • Underlying data available vs. only moments available.

• Prominent example: heterogeneous agent macro models.
  Krueger, Mitman & Perri (2016); AKMWW (2017); Kaplan & Violante (2018)

• Key inference challenges:
  1. How do we account for the statistical inter-dependence between the various moments?
  2. How do we exploit the combined data efficiently?
“Calibration”: moment matching (minimum distance) estimation of structural param’s.

SE easy to compute if var-cov matrix of empirical moments is known.

In practice, hard/impossible to estimate *correlations* of moments due to different data sources, methods, etc. But *variances* readily available.

**Contribution 1:** Simple formula for worst-case SE ⇒ Valid confidence interval.

**Contribution 2:** Moment weighting that minimizes worst-case SE ⇒ Moment selection.

**Further results:** Testing, additional information about moment var-cov matrix.
Moment matching estimation

- $\theta_0 \in \Theta \subset \mathbb{R}^k$: structural parameter vector.

- $\mu_0 \equiv h(\theta_0) \in \mathbb{R}^p$: model-implied moment vector.

- $\hat{\mu}$: empirical moment vector. Assume $\hat{\mu} \sim \mathcal{N}(\mu_0, \hat{\hat{\mu}})$.

  
  \[ \hat{\theta} \equiv \arg\min_{\theta \in \Theta} (\hat{\mu} - h(\theta))' \hat{W} (\hat{\mu} - h(\theta)) \]
  
  \[ \sim \mathcal{N} \left( \theta_0, (\hat{G}' \hat{W} \hat{G})^{-1} \hat{G}' \hat{W} \hat{\hat{\mu}} \hat{G} (\hat{G}' \hat{W} \hat{G})^{-1} \right), \quad \hat{G} \equiv \frac{\partial h(\hat{\theta})}{\partial \hat{\theta}}. \]

- Issue (next slide): In applications, often don’t know off-diagonal elements of $\hat{\hat{V}}$. 
• Estimating the \textit{correlations} of different empirical moments is conceptually and practically challenging if they come from... 

1...different data sets (e.g., micro and macro).

2...different estimation methods.

3...previous papers.

• But individual \textit{variances} of the moments are often known/estimable.

• Can we construct SE and CI w/o knowing the correlation structure?

• Note: Joint normality of $\hat{\mu}$ is restrictive. Assume identification.

Hahn, Kuersteiner & Mazzocco (2020a,b)
Outline

1 Framework

2 Standard errors

3 Efficient moment weighting

4 Application

5 Further results

6 Summary
Worst-case standard errors

• Assume we know SE $\hat{\sigma}_j$ of each moment $\hat{\mu}_j$.

• What are the worst-case SE for the scalar $r(\hat{\theta})$ across all possible correlation structures of the moments?

• Delta method (under standard regularity conditions):

$$r(\hat{\theta}) - r(\theta_0) \approx \hat{x}'(\hat{\mu} - \mu_0), \quad \hat{x} \equiv \hat{W}\hat{G}(\hat{G}'\hat{W}\hat{G})^{-1}\hat{\lambda}, \quad \hat{\lambda} \equiv \partial r(\hat{\theta})/\partial \theta.$$  

• So we seek worst-case variance for a particular linear combination of $\hat{\mu}$, given known marginal variances.

• Simple but useful result: By Cauchy-Schwarz,

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y) \leq (\text{Std}(X) + \text{Std}(Y))^2.$$  

$$\leq \text{Std}(X)\text{Std}(Y)$$
Worst-case standard errors (cont.)

- Using the simple result:

\[
WCSE \equiv \max_{\hat{V} \in S(\hat{\theta})} SE(r(\hat{\theta})) = \max_{\hat{V} \in S(\hat{\theta})} SE(\hat{x}' \hat{\mu}) = \sum_{j=1}^{p} |\hat{x}_j| \hat{\sigma}_j,
\]

\[
\hat{x} \equiv \hat{W} \hat{G} (\hat{G}' \hat{W} \hat{G})^{-1} \hat{\lambda}, \quad \hat{G} \equiv \frac{\partial h(\hat{\theta})}{\partial \theta'}, \quad \hat{\lambda} \equiv \frac{\partial r(\hat{\theta})}{\partial \theta}.
\]

- CI with coverage prob. at least 95% for \(r(\theta_0)\):

\[
r(\hat{\theta}) \pm 1.96 \times WCSE.
\]

Exact coverage under worst-case correlation structure: perfect correlation (±1).

- WCSE at most \(\sqrt{p}\) times larger than SE that assume independence.
Efficient moment weighting/selection

- Which moment weight matrix $\hat{W}$ minimizes the WCSE for $r(\hat{\theta})$?

$$\min_{\hat{W} \in S} \sum_{j=1}^{p} \vert \hat{x}_j(\hat{W}) \vert \hat{\sigma}_j = \min_{z \in \mathbb{R}^{p-k}} \sum_{j=1}^{p} \vert \tilde{Y}_j - \tilde{X}'_j z \vert,$$

for certain artificial data $\tilde{Y}_j$ and $\tilde{X}_j$, $j = 1, \ldots, p$.

- This is just a **median regression**. Easy to compute.

- There exists solution $z^*$ such that at least $p - k$ residuals

$$\hat{e}_j^* = \tilde{Y}_j - \tilde{X}'_j z^*, \quad j = 1, \ldots, p,$$

are zero. Koenker & Basset (1978)

- Efficient $\hat{W}$: zero weight on $\hat{\mu}_j$ for which $\hat{e}_j^* = 0$. Depends on $r(\cdot)$.

- Efficient estimator $\hat{\theta}_{\text{eff}}$ uses only $k$ moments $\implies$ Moment selection.
Outline

1. Framework
2. Standard errors
3. Efficient moment weighting
4. Application
5. Further results
6. Summary
Application: Menu cost price setting for multiproduct firms

- Alvarez & Lippi (2014): Continuous-time model of optimal pricing for multiproduct firm subject to menu cost. Once menu cost is paid, all prices may be adjusted.

- Shape of price change distribution depends on (i) number of products, (ii) volatility of frictionless optimal prices, and (iii) scaled menu cost.

- Data: Price changes of beer from a single supermarket branch (Dominick’s). 499 UPCs, on average 76 weekly obs. per UPC. \( n \approx 38k \).

- Estimation moments: freq. of price change and \( E(|\Delta p|^j), j \in \{1, 2, 4\} \).

- Compare full-info (use estimated moment correlations) with limited-info (pretend we get moments + SE from other paper).
Application: Menu cost price setting for multiproduct firms (cont.)

|                   | Just-ID: $E(|\Delta p|)$ not targeted | All moments |
|-------------------|----------------------------------------|-------------|
|                   | #prod   | Vol  | MC   | Over-ID | #prod   | Vol  | MC   |
| Full-info         | 3.012   | 0.090| 0.291| 0.0019  | 3.255   | 0.089| 0.305|
|                   | (0.046) | (0.001)| (0.003)| (0.0001)| (0.052) | (0.001) | (0.003) |
| Limited-info      | 3.012   | 0.090| 0.291| 0.0019  | 2.786   | 0.090| 0.278|
|                   | (0.235) | (0.001)| (0.016)| (0.0022)| (0.148) | (0.001) | (0.011) |

SE in parentheses. Over-ID: error in matching $E(|\Delta p|)$. 
1 Framework
2 Standard errors
3 Efficient moment weighting
4 Application
5 Further results
6 Summary
Further results

- Tests: over-identification ("non-targeted moments"), parameter restrictions.

- Known block diagonal of $V$.
  - Compute variance of part of linear combination $\hat{x}'\hat{\mu}$ corresponding to each known block of $V$.
  - Then reduces to previous case.

- Other info about elements of $V$.
  - General problem: Given linear combination $\hat{x}$,
    $$\max / \min \ \hat{x}'V\hat{x} \quad \text{s.t.} \quad (i) \text{ known elements of } V, \quad (ii) \ V \text{ symm. pos. semidef.}$$
  - Easily computable using (convex) semidefinite programming.
1 Framework
2 Standard errors
3 Efficient moment weighting
4 Application
5 Further results
6 Summary
• In practical moment matching (minimum distance) estimation, the correlation structure of the empirical moments can be difficult/impossible to estimate.

• We construct worst-case SE and valid CI given marginal variances of moments.

• Efficient moment weighting $\implies$ Moment selection.

• Extensions: Testing, additional info about var-cov matrix.

• Computationally simple.
Summary

• In practical moment matching (minimum distance) estimation, the correlation structure of the empirical moments can be difficult/impossible to estimate.

• We construct worst-case SE and valid CI given marginal variances of moments.

• Efficient moment weighting $\rightarrow$ Moment selection.

• Extensions: Testing, additional info about var-cov matrix.

• Computationally simple.

Thank you!
Appendix
Median regression: Details

\[
\min_{\hat{W} \in S} \sum_{j=1}^{p} |\hat{x}_j(\hat{W})| \sigma_j = \min_{z \in \mathbb{R}^{p-k}} \sum_{j=1}^{p} |\tilde{Y}_j - \tilde{X}_j'z|,
\]

where

\[
\tilde{Y}_j \equiv \hat{\sigma}_j \hat{G}_j \cdot (\hat{G}' \hat{G})^{-1} \hat{\lambda} \in \mathbb{R}, \quad \tilde{X}_j \equiv -\hat{\sigma}_j \hat{G}_j' \in \mathbb{R}^{p-k},
\]

- \( \hat{G}^\perp \) is the \( p \times (p - k) \) matrix of eigenvectors of \( I_p - \hat{G}(\hat{G}' \hat{G})^{-1} \hat{G}' \) corresponding to its nonzero eigenvalues.
• Over-identification test ($p > k$). “Checking non-targeted moments.”
  \[ \tilde{\theta}_0 \equiv \arg\min_{\theta} (\mu_0 - h(\theta))' W (\mu_0 - h(\theta)). \]
  \[ 95\% \text{ CI for } (\mu_j - h_j(\tilde{\theta}_0)): (\hat{\mu}_j - h_j(\hat{\theta})) \pm 1.96 \times WCSE_{(\hat{\mu}_j - h_j(\hat{\theta}))}. \]

• Joint test of parameter restrictions $H_0: r(\theta_0) = 0_{m \times 1}$.
  \[ \text{Wald-type test statistic } \hat{T} \equiv nr(\hat{\theta})' \hat{S} r(\hat{\theta}). \]
  \[ \text{Under } H_0: \hat{T} \sim Z' Q Z, \text{ where } Z \sim N(0, I_p). Q \text{ depends on unknown } V. \]
  \[ \text{Bound tail probability of } Z' Q Z \text{ to get simple (conservative) critical value. } \]
  \[ \text{Székely \& Bakirov (2003)} \]
Over-identification test

\[ \tilde{\theta}_0 \equiv \arg\min_{\theta \in \Theta} (\mu_0 - h(\theta))'W(\mu_0 - h(\theta)) \]

- Standard result:
  \[ \hat{\mu} - h(\tilde{\theta}) - (\mu_0 - h(\tilde{\theta}_0)) \approx (I_p - \hat{G}(\hat{G}'\hat{W}\hat{G})^{-1}\hat{G}'\hat{W})(\hat{\mu} - \mu_0). \]

- Computing the WCSE for \((\hat{\mu}_j - h_j(\hat{\theta}))\) just amounts to finding the WCSE for a particular linear combination of \(\hat{\mu}\). Apply earlier result.

- 95% CI for \((\mu_j - h_j(\tilde{\theta}_0))\):
  \[ (\hat{\mu}_j - h_j(\hat{\theta})) \pm 1.96 \times WCSE. \]
Joint test of parameter restrictions

- Under $H_0$:
  \[ \hat{T} \overset{d}{\to} Z'QZ, \quad Q \equiv V^{1/2}WG(G'WG)^{-1}\lambda S\lambda'(G'WG)^{-1}G'WV^{1/2}. \]

- Székely & Bakirov (2003) prove that
  \[ P(Z'QZ \leq \text{trace}(Q) \times \tau) \leq P(Z_1^2 \leq \tau) \]
  for any $p \times p$ sym. pos. semidef. $Q \neq 0$ and any $\tau > 1.5365$.

- Compute using (convex) semidefinite programming:
  \[ \hat{c}v \equiv \max_{\tilde{V} \in S(\text{diag}(V))} \text{trace}\left( \tilde{V}WG(G'WG)^{-1}\lambda S\lambda'(G'WG)^{-1}G'W \right) \times (\Phi^{-1}(1 - \alpha/2))^2. \]

- Then, for any $\alpha \leq 0.215$, we have under $H_0$:
  \[
  P(\hat{T} \leq \hat{c}v) \geq P \left( \hat{T} \leq \text{trace}(Q) \times (\Phi^{-1}(1 - \alpha/2))^2 \right)
  \]
  \[ \rightarrow P \left( Z'QZ \leq \text{trace}(Q) \times (\Phi^{-1}(1 - \alpha/2))^2 \right) \]
  \[ \leq P \left( Z_1^2 \leq (\Phi^{-1}(1 - \alpha/2))^2 \right) = 1 - \alpha. \]