Standard Errors for Calibrated Parameters

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

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

June 30, 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 $\Rightarrow$ Valid confidence interval.

• Contribution 2: Moment weighting that minimizes worst-case SE $\Rightarrow$ Moment selection.

• Further results: Testing, additional information about moment var-cov matrix.
Outline

1. Framework
2. Standard errors
3. Efficient moment weighting
4. Testing
5. Application
6. Extensions: General knowledge of covariance matrix
7. Summary
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 $\hat{\mu} \sim N(\mu_0, \hat{\Sigma})$.

- Moment matching/minimum distance/calibration:
  Newey & McFadden (1994); Hansen & Heckman (1996)

  $$\hat{\theta} \equiv \arg\min_{\theta \in \Theta} (\hat{\mu} - h(\theta))^\prime \hat{W} (\hat{\mu} - h(\theta))$$

  $$\sim N\left(\theta_0, (\hat{G}' \hat{W} \hat{G})^{-1} \hat{G}' \hat{W} \hat{\Sigma} \hat{W} \hat{G}(\hat{G}' \hat{W} \hat{G})^{-1}\right), \quad \hat{G} \equiv \frac{\partial h(\hat{\theta})}{\partial \theta'}.$$ 

- Issue (next slide): In applications, often don’t know off-diagonal elements of $\hat{\Sigma}$. 
Estimating the *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 *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. Testing
5. Application
6. Extensions: General knowledge of covariance matrix
7. 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 parameter of interest $r(\hat{\theta})$ across all possible correlation structures of the moments?
  
  - E.g., $r(\theta) = \text{counterfactual}$, or $r(\theta) = \theta_j$.

- Delta method (under standard regularity conditions):
  
  $$r(\hat{\theta}) - r(\theta_0) \approx \hat{x}'(\hat{\mu} - \mu_0), \quad \hat{x} \equiv \hat{W}\hat{G}(\hat{G}'\hat{W}\hat{G})^{-1}\hat{\lambda}, \quad \hat{\lambda} \equiv \partial r(\hat{\theta})/\partial \theta.$$

- What is worst-case variance of $\hat{x}'\hat{\mu}$, given known marginal variances of $\hat{\mu}_j$?

- Simple but useful result:
  
  $$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \leq (\text{Std}(X) + \text{Std}(Y))^2 \leq \text{Std}(X)\text{Std}(Y).$$
Worst-case standard errors (cont.)

- Using the simple result:

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

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

- Easy to compute, given \( h(\cdot) \) and \( r(\cdot) \).

- 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.
Outline

1. Framework
2. Standard errors
3. Efficient moment weighting
4. Testing
5. Application
6. Extensions: General knowledge of covariance matrix
7. Summary
Efficient moment weighting/selection

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

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

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

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

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.
Efficient moment weighting/selection (cont.)

- Financial portfolio analogy: How do we form the lowest-variance portfolio subject to achieving a given expected return?

- Diversification argument suggests that we should use all available assets.

- But suppose we do not know the cross-correlations of the assets. How to guard against high variance in the worst case?

- Worst case: Perfect correlation $\implies$ No diversification motive.

- Robust solution: Buy only the single asset with highest Sharpe ratio ("signal-to-noise").

- If we had $k$ constraints (not just expected return target), we would need $k$ assets.
  - Minimum distance: Asy. estimator unbiasedness yields $k$ constraints.
Outline

1 Framework
2 Standard errors
3 Efficient moment weighting
4 Testing
5 Application
6 Extensions: General knowledge of covariance matrix
7 Summary
Testing

• Over-identification test \((p > k)\). “Checking non-targeted moments.”
  
  \[ \tilde{\theta}_0 \equiv \arg \min_{\theta} (\mu_0 - h(\theta))^\prime W (\mu_0 - h(\theta)). \]

• 95% 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 \(H_0: r(\theta_0) = 0_{m \times 1}\).
  
  \[ \hat{T} \equiv nr(\hat{\theta})^\prime \hat{S} r(\hat{\theta}). \]

• Under \(H_0\): \[ \hat{T} \sim Z'QZ, \text{ where } Z \sim N(0, I_p). \] \(Q\) depends on unknown \(V\).

• Bound tail probability of \(Z'QZ\) to get simple (conservative) critical value. 
  \cite{SzekelyBakirov2003}
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 $k = 3$ param’s: (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. $n \approx 38k$.

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

• Compare:
  1. Full-info: Use estimated moment correlations.
  2. 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 \( \hat{E}(|\Delta p|) = 0.145 \).
Outline

1 Framework
2 Standard errors
3 Efficient moment weighting
4 Testing
5 Application
6 Extensions: General knowledge of covariance matrix
7 Summary
General knowledge of covariance matrix

- 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 $\hat{x}$,
    $$\max / \min \ \hat{x}'V\hat{x} \quad \text{s.t. (i) known elements of } V,$$
    $$\quad \text{(ii) } V \text{ symm. pos. semidef.}$$
  - Easily computable using (convex) semidefinite programming.
Outline

1. Framework
2. Standard errors
3. Efficient moment weighting
4. Testing
5. Application
6. Extensions: General knowledge of covariance matrix
7. Summary
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.

• Further results: 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 $\implies$ Moment selection.

- Further results: 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

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

- \(\hat{G}'\) 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

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

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

- 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 $H_0$:
  \[
  \hat{\mathcal{F}} \overset{d}{\to} Z'QZ, \quad Q \equiv V^{1/2}WG(G'WG)^{-1}\lambda S\lambda'(G'WG)^{-1}G'WV^{1/2}.
  \]

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

- Then, for any $\alpha \leq 0.215$, we have under $H_0$:
  \[
P(\hat{\mathcal{F}} \leq \hat{cv}) \geq P \left( \hat{\mathcal{F}} \leq \text{trace}(Q) \times (\Phi^{-1}(1 - \alpha/2))^2 \right)
  \rightarrow P \left( Z'QZ \leq \text{trace}(Q) \times (\Phi^{-1}(1 - \alpha/2))^2 \right)
  \leq P \left( Z_1^2 \leq (\Phi^{-1}(1 - \alpha/2))^2 \right) = 1 - \alpha.
  \]
Alvarez & Lippi (2014) model

- Firm chooses stopping times $\tau_j$ and price changes $\Delta p_i(\tau_j)$ to minimize

$$E \left[ \sum_{j=1}^{\infty} e^{-r\tau_j} \psi + B \int_0^{\infty} e^{-rt} \left( \sum_{i=1}^{n} p_i(t)^2 \right) dt \right] \mid p(0) = p.$$ 

- Price gap evolution:

$$p_i(t) = \sigma \mathcal{W}_i(t) + \sum_{j: \tau_j < t} \Delta p_i(\tau_j), \quad t \geq 0, \ i = 1, \ldots, n.$$ 

- Following A&L, consider limit $r \to 0$.

- Parameters to be estimated: number $n$ of products, volatility $\sigma$, scaled menu cost $\sqrt{\psi/B}$.
Simulation study

- Simulate from A&L model with just-ID parameter estimates. Same sample size $n \approx 37k$.

<table>
<thead>
<tr>
<th></th>
<th>Just-identified specification</th>
<th>Efficient specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidence interval coverage rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-info</td>
<td>94.5%</td>
<td>95.0%</td>
</tr>
<tr>
<td>Independence</td>
<td>100.0%</td>
<td>95.0%</td>
</tr>
<tr>
<td>Worst case</td>
<td>100.0%</td>
<td>99.4%</td>
</tr>
<tr>
<td>Confidence interval average length</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-info</td>
<td>0.179</td>
<td>0.002</td>
</tr>
<tr>
<td>Independence</td>
<td>0.627</td>
<td>0.002</td>
</tr>
<tr>
<td>Worst case</td>
<td>0.878</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>Just-identified specification</td>
<td>Efficient specification</td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td><strong>RMSE relative to true parameter values</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-info</td>
<td>1.53%</td>
<td>0.59%</td>
</tr>
<tr>
<td>Independence</td>
<td>1.53%</td>
<td>0.59%</td>
</tr>
<tr>
<td>Worst case</td>
<td>1.53%</td>
<td>0.59%</td>
</tr>
<tr>
<td><strong>Rejection rate of over-identification test</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-info</td>
<td>5.01%</td>
<td></td>
</tr>
<tr>
<td>Independence</td>
<td>0.00%</td>
<td></td>
</tr>
<tr>
<td>Worst case</td>
<td>0.00%</td>
<td></td>
</tr>
<tr>
<td><strong>Rejection rate of joint test of true parameter values</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-info</td>
<td>4.79%</td>
<td></td>
</tr>
<tr>
<td>Independence</td>
<td>7.54%</td>
<td></td>
</tr>
<tr>
<td>Worst case</td>
<td>2.47%</td>
<td></td>
</tr>
</tbody>
</table>