Standard Errors for Calibrated Parameters

Matthew D. Cocci
Amazon

Mikkel Plagborg-Møller
Princeton University

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

May 11, 2022

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?

This paper: Standard errors for calibrated parameters

- "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 $\Longrightarrow$ Valid confidence interval.
- Contribution 2: Moment weighting that minimizes worst-case SE $\Longrightarrow$ Moment selection.
- Further results: Testing, additional information about moment var-cov matrix.

Outline

1. Framework
2. Standard errors
3. Efficient moment weighting
4. Further results
5. Applications
   - Menu cost price setting
   - Heterogeneous agent New Keynesian model
6. 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 $\hat{\mu} \overset{\cdot}{\sim} N(\mu_0, \hat{V})$.
- Moment matching/minimum distance/calibration:
  Newey & McFadden (1994); Hansen & Heckman (1996)

$$\hat{\theta} \equiv \text{argmin}_{\theta \in \Theta} \, (\hat{\mu} - h(\theta))' \hat{W} (\hat{\mu} - h(\theta))$$

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

- Issue (next slide): In applications, often don't know off-diagonal elements of $\hat{V}$.

Var-cov matrix of empirical moments

- 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. Further results
5. Applications
    - Menu cost price setting
    - Heterogeneous agent New Keynesian model
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 parameter of interest $r(\hat{\theta})$ across all possible correlation structures of the moments?
  - E.g., $r(\theta) =$ counterfactual, or $r(\theta) = \theta_j$.
- Delta method (under standard regularity conditions):

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

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

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

Worst-case standard errors (cont.)

- Using the simple result:

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

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

- 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. Further results
5. Applications
   - Menu cost price setting
   - Heterogeneous agent New Keynesian model
6. Summary

Efficient moment weighting/selection

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

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

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

- This is just a median regression. Easy to compute.
- There exists solution z^* such that at least $p - k$ residuals

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

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 $\Longrightarrow$ Moment selection.

Efficient moment weighting/selection: Intuition

- 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 $\Longrightarrow$ 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. Further results
5. Applications
   - Menu cost price setting
   - Heterogeneous agent New Keynesian model
6. Summary

Testing

- Over-identification test ($p > k$). "Checking non-targeted moments."
  - $\tilde{\theta}_0 \equiv \text{argmin}_\theta (\mu_0 - h(\theta))' W (\mu_0 - h(\theta))$.
  - 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 $H_0 : r(\theta_0) = 0_{m \times 1}$.
  - Wald-type test statistic $\hat{\mathscr{T}} \equiv r(\hat{\theta})' \hat{S} r(\hat{\theta})$.
  - Under H_0: $n\hat{\mathscr{T}} \dot{\sim} Z'QZ$, where $Z \sim N(0, I_p)$. Q depends on unknown V.
  - Bound tail probability of $Z'QZ$ to get simple (conservative) critical value.
    Székely & Bakirov (2003)

General knowledge about var-cov matrix

- 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.
- Closed-form formula if block diagonal of $\hat{V}$ is known.
- Optimal weight matrix: Nested concave/convex problems.

$$\min_{W \in \mathcal{S}_p} \max_{V \in \tilde{\mathcal{S}}} \hat{x}(W)' V \hat{x}(W) = \min_{z \in \mathbb{R}^{p-k}} \max_{V \in \tilde{\mathcal{S}}} \{\hat{G}(\hat{G}'\hat{G})^{-1}\hat{\lambda} + \hat{G}^{\perp} z\}' V \{\hat{G}(\hat{G}'\hat{G})^{-1}\hat{\lambda} + \hat{G}^{\perp} z\}$$

Outline

1. Framework
2. Standard errors
3. Efficient moment weighting
4. Further results
5. Applications
   - Menu cost price setting
   - Heterogeneous agent New Keynesian model
6. Summary

Outline

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. $n \approx 38$k.
- $p = 4$ 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 $\hat{E}(|\Delta p|) = 0.145$.

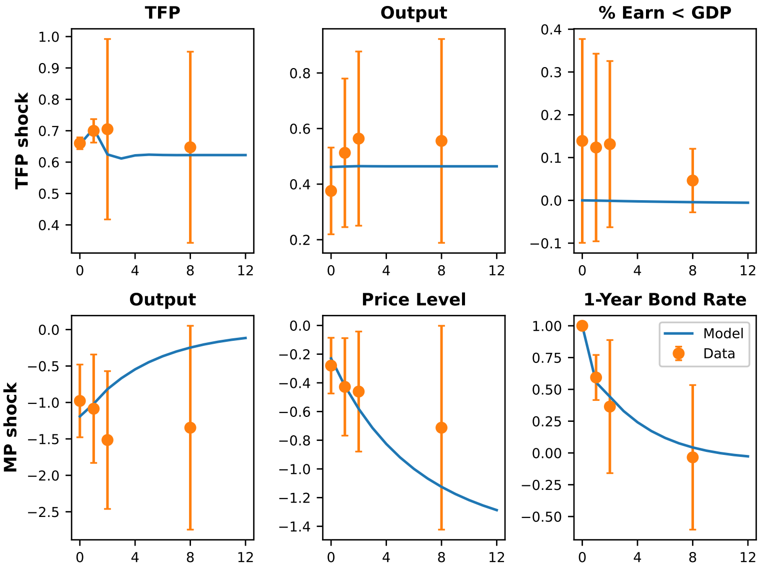
▸ Sim

Outline

1. Framework
2. Standard errors
3. Efficient moment weighting
4. Further results
5. Applications
    - Menu cost price setting
    - Heterogeneous agent New Keynesian model
6. Summary

Estimation of HANK model by IRF matching

- One-asset HANK model from Auclert, Bardóczy, Rognlie & Straub (2021), solved via linearization. McKay, Nakamura & Steinsson (2016); Kaplan, Moll & Violante (2018)
- $p = 23$ matched moments: impulse responses for both micro and macro variables to TFP and monetary shocks. Use IRF estimates+SEs from Chang, Chen & Schorfheide (2021) and Miranda-Agrippino & Ricco (2021).
  - IRF matching less efficient than likelihood inference, but robust to specification of other shock processes.
  - We don't need underlying data or extra assumptions required for GMM/bootstrap.
- $k = 7$ estimated parameters: AR(2) shock process parameters, Taylor rule coefficient on inflation, slope of Phillips curve.



Vertical bars: CIs for differences btw. empirical and model-implied moments, centered at the former.

Estimation of HANK model by IRF matching (cont.)

Weight matrix	TR	PC	TFP AR1	TFP AR2	TFP Std	Monetary AR1	Monetary AR2
Diagonal	1.409	0.010	0.076	-0.132	0.007	0.713	0.075
	(4.243)	(0.012)	(0.237)	(0.377)	(0.001)	(0.223)	(0.185)
Efficient	1.583	0.017	0.060	-0.078	0.007	0.723	0.014
	(3.012)	(0.010)	(0.192)	(0.282)	(0.000)	(0.170)	(0.149)

Worst-case SE in parentheses. Parameters: Taylor rule coefficient on inflation ("TR"); slope of Phillips curve ("PC"); first and second autoregressive ("AR1" and "AR2") and standard deviation ("Std") parameters of TFP and monetary disturbance processes.

Outline

1. Framework
2. Standard errors
3. Efficient moment weighting
4. Further results
5. Applications
   - Menu cost price setting
   - Heterogeneous agent New Keynesian model
6. 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 $\Longrightarrow$ Moment selection.
- Further results: Testing, additional info about var-cov matrix.
- Computationally simple (Matlab+Python packages on GitHub).

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 $\Longrightarrow$ Moment selection.
- Further results: Testing, additional info about var-cov matrix.
- Computationally simple (Matlab+Python packages on GitHub).

Thank you!

Appendix

Median regression: Details

- Recall $x(W) = WG(G'WG)^{-1}\lambda$. Lemma:

$$\{x(W)\colon W \in \mathcal{S}_p\} = \{x\colon x \in \mathbb{R}^p,\ G'x = \lambda\} = \left\{G(G'G)^{-1}\lambda + G^{\perp}z\colon z \in \mathbb{R}^{p-k}\right\}.$$

- $G^{\perp}$ is a full-rank $p \times (p-k)$ matrix satisfying $G'G^{\perp} = 0_{k\times(p-k)}$.
- Hence,

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

where

$$\tilde{Y}_j \equiv \sigma_j G_{j\bullet}(G'G)^{-1}\hat{\lambda} \in \mathbb{R}, \quad \tilde{X}_j \equiv -\sigma_j G_{j\bullet}^{\perp\prime} \in \mathbb{R}^{p-k}.$$

Diag General

Over-identification test

$$\tilde{\theta}_0 \equiv \underset{\theta \in \Theta}{\text{argmin}}\ (\mu_0 - h(\theta))' W (\mu_0 - h(\theta))$$

- Standard result:

$$\hat{\mu} - h(\hat{\theta}) - (\mu_0 - h(\tilde{\theta}_0)) \approx (I_p - \hat{G}(\hat{G}'\hat{W}\hat{G})^{-1}\hat{G}'\hat{W})(\hat{\mu} - \mu_0).$$

- 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:

$$n\hat{\mathscr{T}} \xrightarrow{d} Z'QZ, \quad Q \equiv V^{1/2\prime} WG(G'WG)^{-1}\lambda S \lambda'(G'WG)^{-1}G'WV^{1/2}.$$

- 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:

$$\widehat{\text{cv}} \equiv \max_{\tilde{V} \in \mathcal{S}(\text{diag}(V))} \frac{1}{n} \text{trace}\left(\tilde{V} WG(G'WG)^{-1}\lambda S \lambda'(G'WG)^{-1}G'W\right) \times \left(\Phi^{-1}(1-\alpha/2)\right)^2.$$

- Then, for any $\alpha \leq 0.215$, we have under H_0:

$$\begin{aligned} P(\hat{\mathscr{T}} \leq \widehat{\text{cv}}) &\geq P\left(n\hat{\mathscr{T}} \leq \text{trace}(Q) \times (\Phi^{-1}(1-\alpha/2))^2\right) \\ &\to P\left(Z'QZ \leq \text{trace}(Q) \times (\Phi^{-1}(1-\alpha/2))^2\right) \\ &\leq P\left(Z_1^2 \leq (\Phi^{-1}(1-\alpha/2))^2\right) = 1-\alpha. \end{aligned}$$

Alvarez & Lippi (2014) model

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

$$E\left[\underbrace{\sum_{j=1}^{\infty} e^{-r\tau_j}\psi}_{\text{menu cost}} + B\underbrace{\int_0^{\infty} e^{-rt}\left(\sum_{i=1}^{n} p_i(t)^2\right)dt}_{\text{price gaps}} \,\middle|\, p(0) = p\right].$$

- Price gap evolution:

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

- Following A&L, consider limit $r \to 0$.
- Parameters to be estimated: number n of products, volatility σ, scaled menu cost $\sqrt{\psi/B}$.

Simulation study

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

	Just-identified specification			Efficient specification		
	# prod.	Vol.	Menu cost	# prod.	Vol.	Menu cost
	Confidence interval coverage rate					
Full-info	94.5%	95.0%	94.7%	95.0%	95.2%	95.3%
Independence	100.0%	95.0%	100.0%	100.0%	89.1%	100.0%
Worst case	100.0%	99.4%	100.0%	100.0%	99.4%	100.0%
	Confidence interval average length					
Full-info	0.179	0.002	0.010	0.162	0.002	0.009
Independence	0.627	0.002	0.039	0.390	0.002	0.025
Worst case	0.878	0.003	0.059	0.571	0.003	0.041

	Just-identified specification			Efficient specification		
	# prod.	Vol.	Menu cost	# prod.	Vol.	Menu cost
	RMSE relative to true parameter values					
Full-info	1.53%	0.59%	0.86%	1.37%	0.58%	0.76%
Independence	1.53%	0.59%	0.86%	1.72%	0.59%	0.99%
Worst case	1.53%	0.59%	0.86%	1.79%	0.59%	1.03%
	Rejection rate of over-identification test					
Full-info		5.01%				
Independence		0.00%				
Worst case		0.00%				
	Rejection rate of joint test of true parameter values					
Full-info		4.79%				
Independence		7.54%				
Worst case		2.47%				

Auclert et al. (2021) model

- Mass 1 of heterogeneous households with Bellman equation:

$$V_t(e_{it}, a_{i,t-1}) = \max_{c_{it}, n_{it}, a_{it}} \left\{ \frac{c_{it}^{1-\sigma}}{1-\sigma} - \varphi \frac{n_{it}^{1+\nu}}{1+\nu} + \beta \mathbb{E}_t [V_{t+1}(e_{i,t+1}, a_{it})] \right\}$$

$$\text{s.t.} \quad c_{it} + a_{it} = (1+r_t) a_{i,t-1} + w_t e_{it} n_{it} - \tau_t \bar{\tau}(e_{it}) + d_t \bar{d}(e_{it}), \quad a_{it} \geq 0.$$

- Idiosyncratic productivity (discretized): $\log e_{it} = \rho_e \log e_{i,t-1} + \sigma_e \epsilon_{it}$, $\epsilon_{it} \overset{i.i.d.}{\sim} N(0,1)$.
- Competitive final goods firm aggregates continuum of intermediate goods.
  - Intermediate producers: Monop. comp., linear production in labor, quadr. price adj. costs ψ_t.
  - Yields NKPC:

$$\log(1+\pi_t) = \kappa \left(\frac{w_t}{Z_t} - \frac{1}{\mu} \right) + \frac{1}{1+r_{t+1}} \frac{Y_{t+1}}{Y_t} \log(1+\pi_{t+1}).$$

  - Dividends: $d_t = Y_t - w_t N_t - \psi_t$. Incidence rule $\bar{d}(e_{it}) \propto e_{it}$.

Auclert et al. (2021) model (cont.)

- Fiscal policy: $\tau_t = r_t B$. Incidence rule $\bar{\tau}(e_{it}) \propto e_{it}$.
- Monetary policy: $i_t = r_t^* + \phi \pi_t$, where $1 + r_t = (1 + i_{t-1})/(1 + \pi_t)$.
- Exogenous disturbance processes:
  - $\log Z_t - \log Z_{t-1} =$ AR(2) process.
  - $r_t^* =$ AR(2) process.
- Market clearing: $Y_t = \int c_{it}\, di + \psi_t$, $B = \int a_{it}\, di$, $N_t = \int e_{it} n_{it}\, di$.

Auclert et al. (2021) model: Calibrated parameters

Parameter	Value
β	0.982
φ	0.786
σ	2
ν	2
ρ_e	0.966
$\sigma_e/\sqrt{1-\rho_e^2}$	0.5
μ	1.2
B	5.6

Note: $N_t = 1$, $r^{ss} = 0.005$, $Y^{ss} = 1$.