Adversarial Method of Moments

BSE Summer Forum
Advances in Econometrics

Ignacio Cigliutti Elena Manresa

New York University

June 13th, 2022

Introduction

GMM/SMM

- Many economic models are defined by moment conditions

$$\mathbb{E}\left[g(x,\theta_0)\right]=0$$

- Two main reasons
  1. Transparent/Derived from theoretical models
  2. Do not rely on strong distributional assumptions

- GMM/SMM methods are a natural estimation strategy

- Bias arise when number of moments is moderate or when estimating higher orders of the data (Altonji & Segal, 1996))

This paper

- Introduce Adversarial Method of Moments (AMM)
  - Based on adversarial estimation (GAN)
  - For models defined by moment conditions/matching moments
- AMM features
  1. Efficiency: Incorporates optimal weighting as OW-GMM/SMM
  2. Smaller finite sample bias
- **Tight connection between adversarial estimation and GEL estimators**

Related Literature

GMM Hansen (1982), Hansen & Singleton (1982), Arellano & Bond (1991), Hansen, Heaton & Yaron (1996)

GAN Goodfellow et al (2014), Kaji, Manresa & Pouliot (2020)

GEL Owen (1988), Qin & Lawless (1994), Imbens (1997), Newey & Smith (2004), Evdokimov, Kitamura & Otsu (2014)

ML Bennett & Kallus (2021), Lewis, Syrgankis et al (2020).

A Preview – Arellano Bond

- Dynamic panel data model

$$Y_{it} = \rho Y_{it-1} + \alpha_i + \epsilon_{it}$$

- Difference out FE

$$\Delta Y_{it} = \rho \Delta Y_{it-1} + \Delta \epsilon_{it}$$

  and use lagged levels/differences as instruments

$$g_L(Y_i, t, \theta) = Y_i^{t-2} \cdot (\Delta Y_{it} - \theta \Delta Y_{it-1})$$
$$g_D(Y_i, t, \theta) = \Delta Y_i^{t-2} \cdot (\Delta Y_{it} - \theta \Delta Y_{it-1})$$

- Have $K \simeq (T-1)(T-2)/2$ moment conditions

Arellano Bond – Bias

		$T = 5$	$T = 10$	$T = 15$	$T = 20$
$\rho = 0.7$	GMM 1–step	-0.061	-0.130	-0.176	-0.217
	GMM 2–step	-0.030	-0.048	-0.082	-0.134
	GMM D	-0.069	-0.133	-0.180	-0.227
	GMM CUE	-0.003	0.000	0.000	0.000
	AMM	0.000	0.001	0.001	-0.002
$\rho = 0.9$	GMM 1–step	-0.317	-0.338	-0.377	-0.419
	GMM 2–step	-0.247	-0.200	-0.242	-0.310
	GMM D	-0.319	-0.387	-0.432	-0.478
	GMM CUE	-0.168	-0.012	-0.001	0.000
	AMM	-0.108	-0.007	0.000	0.001

Table: Bias based on 500 simulations, $N = 500$.

Arellano Bond – SD

		$T = 5$	$T = 10$	$T = 15$	$T = 20$
$\rho = 0.7$	GMM 1–step	0.170	0.089	0.080	0.078
	GMM 2–step	0.119	0.048	0.044	0.051
	GMM D	0.132	0.071	0.063	0.062
	GMM CUE	0.134	0.039	0.024	0.021
	AMM	0.130	0.038	0.023	0.053
$\rho = 0.9$	GMM 1–step	0.393	0.239	0.188	0.167
	GMM 2–step	0.285	0.168	0.140	0.138
	GMM D	0.278	0.168	0.134	0.118
	GMM CUE	0.386	0.099	0.048	0.037
	AMM	0.258	0.081	0.043	0.032

Table: SD based on 500 simulations, $N = 500$.

Estimation

Generative Adversarial Networks (GAN)

- AMM is inspired in GAN (Goodfellow et al 2016)
- GAN involves 2 models
    - A generative model which creates synthetic observations
    - A discriminative model which takes as inputs real and simulated data, and tries to predict the provenance of each observation
    - The estimator is defined as the value of θ for which the discriminator cannot tell apart real from simulated data.
- AMM is a GAN estimator where the discriminator is a logistic regression.
- **Can use adversarial estimation even when models are defined uniquely by moment restrictions!**

GAN Framework

- Models D and G play a zero–sum game
  - D: Distinguish real from synth
  - G, parametrized by $F(\theta)$: Make both samples as similar as possible
- Loss function

$$\min_{\theta \in \Theta} \max_{D \in \mathcal{D}} \mathcal{L} = \mathbb{E}_{X \sim P_{data}} \left[\log \left(1 - D(X) \right) \right] + \mathbb{E}_{Z \sim F(\theta)} \left[\log D(Z) \right]$$

- Why choose D to be logistic? It ensures...
  1. Inner maximization to be concave
  2. D will use only selected moments to distinguish between samples

Adversarial estimation with moment conditions

- How to 'generate' a synthetic sample from the model?
- Note that for any θ:
$$g(x_i, \theta) = \mathbb{E}\left[g(x, \theta)\right] + \varepsilon_i,$$
where $\mathbb{E}\left[\varepsilon_i\right] = 0$.
- For $\theta = \theta_0$, since $\mathbb{E}\left[g(x_i, \theta_0)\right] = 0$, we have:
$$g(x_i, \theta_0) = \varepsilon_i$$
- We use draws from a mean-zero random variable as 'synthetic' data, and compare them to $\{g(x_i, \theta)\}_{i=1}^N$

AMM Framework

- Logistic function: $\Lambda(x) = (1 + e^{-x})^{-1}$
- Objective function

$$\min_{\theta \in \Theta} \max_{\lambda \in \mathbb{R}^{k+1}} \left\{ n^{-1} \sum_{i=1}^{n} \log \left(1 - \Lambda\left(\lambda' g\left(x_i, \theta\right)\right)\right) + m^{-1} \sum_{j=1}^{m} \log \Lambda\left(\lambda' \epsilon_i\right) \right\}$$

- FOC (inner maximization)

$$n^{-1} \sum_{i=1}^{n} \left(1 - \Lambda\left(\widehat{\lambda}' g\left(x_i, \theta\right)\right)\right) g\left(x_i, \theta\right) = m^{-1} \sum_{j=1}^{m} \Lambda\left(\widehat{\lambda}' \epsilon_i\right) \epsilon_i$$

AMM Framework

- FOC (inner maximization)

$$n^{-1}\sum_{i=1}^{n}\Lambda\left(\widehat{\lambda}' g\left(x_i,\theta\right)\right) g\left(x_i,\theta\right) = m^{-1}\sum_{j=1}^{m}\left(1-\Lambda\left(\widehat{\lambda}'\epsilon_i\right)\right)\epsilon_i$$

- At θ_0 we have $\mathbb{E}\left[g\left(x_i,\theta_0\right)\right]=0$, thus...
    - $\lambda\left(\theta_0\right)=0$ is a solution
    - Concavity of $\mathcal{L}$ w.r.t. λ ensures uniqueness

- $\left[\lambda\left(\theta_0\right)=0 \Rightarrow \widehat{\Lambda}=1-\widehat{\Lambda}=1/2\right]$ so inner FOC yields

$$\frac{1}{n}\sum_{i=1}^{n} g\left(x_i,\theta_0\right)=\frac{1}{m}\sum_{i=1}^{m}\epsilon_i \simeq 0$$

AMM Computation

Algorithm

1. Fix random draw $\boldsymbol{\varepsilon} = \{\varepsilon_i\}_{i=1}^m$
2. Initialize with $\theta = \theta^{(0)}$
3. At each step s...
   1. Compute $\mathbf{g}(\theta) = \{g(x_i, \theta)\}_{i=1}^n$
   2. Run logistic regression using $(\mathbf{g}(\theta), \boldsymbol{\varepsilon})$ to obtain predicted probabilities $\widehat{\Lambda}^{(s)}$
   3. Use $\widehat{\Lambda}^{(s)}$ to compute numerical gradient $\nabla Q(\theta)$
   4. Update $\theta^{(s+1)} = \theta^{(s)} - \eta \nabla Q(\theta^{(s)})$
   5. Repeat till convergence

AMM Example – OLS

- OLS moment condition

$$\mathbb{E}\left[x_i(y_i - \beta' x_i)\right] = 0$$

- Yields the following *dataset*

$$(\mathbf{X}(\theta)\,|\mathbf{d}) = \left[\begin{array}{cc|c} 1 & x_1\left(y_1 - \beta' x_1\right) & 1 \\ 1 & x_2\left(y_2 - \beta' x_2\right) & 1 \\ \vdots & \vdots & \vdots \\ 1 & x_n\left(y_n - \beta' x_n\right) & 1 \\ \hline 1 & \nu\varepsilon_1 & 0 \\ 1 & \nu\varepsilon_2 & 0 \\ \vdots & \vdots & \vdots \\ 1 & \nu\varepsilon_m & 0 \end{array}\right]$$

Statistical Properties

Main Results

1. Asymptotic equivalence between AMM and optimally–weighted GMM/SMM
2. $\text{Bias}(\widehat{\theta}_{AMM}) \leq \text{Bias}(\widehat{\theta}_{GMM})$

How do we do it

- Link between AMM and GEL (Newey & Smith (2004))
- Derive finite sample bias from stochastic expansion
- Results draw from smoothness of the Logit and g

Generalized Empirical Likelihood

- Let $\rho(v)$ be a function of a scalar v that is concave on its domain, an open interval $\mathcal{V}$ containing zero,

$$\hat{\mathcal{B}}_n(\theta) = \left\{\lambda : \lambda' g_i(\theta) \in \mathcal{V}, i = 1, \ldots, n\right\}$$

- θ_0 is unique value such that $\mathbb{E}[g(x, \theta_0)] = 0$
- $\hat{\theta}_{\text{GEL}}$ is the solution to saddle point problem

$$\min_{\theta \in \Theta} \sup_{\lambda \in \hat{\mathcal{B}}_n(\theta)} \sum_{i=1}^{n} \rho\left(\lambda' g_i(\theta)\right)$$

- EL, ET and CUE are special cases. In particular, **EL estimator is a special case with:** $\rho(v) = \log(1 - v)$ and $\mathcal{V} = (-\infty, 1)$

AMM as GEL estimator

- $\widehat{\theta}_{GEL}$ is the solution to saddle point problem

$$\min_{\theta\in\Theta} \sup_{\lambda\in\hat{\mathcal{B}}_n(\theta)} \sum_{i=1}^{n} \rho\left(\lambda' g_i\left(\theta\right)\right)$$

- Choose $\rho(v) = \log\left(1 - \Lambda(v)\right)$ and degenerate $\{\epsilon_i\}_{i=1}^{m} = \mathbf{0}$

$$\widehat{\theta}_{AMM} = \arg\min_{\theta\in\Theta} \max_{\lambda\in\mathbf{R}^{k+1}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \log\left(1 - \Lambda\left(\lambda' g_i\left(\theta\right)\right)\right) \right\} - 1/2$$

AMM Asymptotics

Assumptions

- In general, parametrize $\epsilon \sim \widetilde{F}\left(0, \nu^2 I\right)$
- Define the following

$$\begin{aligned} \Omega &= \mathbb{E}\left[g\left(x, \theta_0\right) g\left(x, \theta_0\right)'\right] \\ G &= \mathbb{E}\left[\partial g\left(x, \theta\right) / \partial \theta|_{\theta_0}\right] \\ \Omega_\kappa &= \Omega + \kappa I \\ \Sigma_\kappa &= \left(G' \Omega_\kappa^{-1} G\right)^{-1} \end{aligned}$$

- Standard assumptions on $g(\cdot)$ yield

$$\sqrt{n}\left(\widehat{\theta} - \theta_0\right) \longrightarrow \mathcal{N}\left(0, \Sigma_{\nu^2}\right)$$

Finite Sample Bias

- Main results in Newey & Smith (2004) are bias expressions Bias

$$\text{Bias}(\widehat{\theta}_{GMM}) = B_I + B_G + B_\Omega + B_W$$
$$\text{Bias}(\widehat{\theta}_{GEL}) = B_I + (1 + \rho_3/2)\, B_\Omega$$

- Derive analogous expression for AMM

$$\text{Bias}(\widehat{\theta}_{AMM}) = B_I + B_{\Omega_{\nu^2}}$$

Monte Carlo

Testing AMM performance

- We consider 2 models with many moment conditions
  1. Arellano & Bond (1991): Use of lagged levels/differences as instruments in FE models provide large number of moments
  2. Altonji & Segal (1996): 2–step GMM ill-behaved when estimating 2nd moments

Arellano & Bond

- Consider the dynamic panel data model

$$Y_{it} = \rho Y_{it-1} + \alpha_i + \epsilon_{it}$$

where $\alpha_i \sim \mathcal{N}, \epsilon_{it} \sim t_3$

- Moments involve differencing out FE

$$\Delta Y_{it} = \rho \Delta Y_{it-1} + \Delta \epsilon_{it}$$

and use lagged levels/differences as instruments

$$g_L (Y_i, t, \theta) = Y_i^{t-2} \cdot (\Delta Y_{it} - \theta \Delta Y_{it-1})$$
$$g_D (Y_i, t, \theta) = \Delta Y_i^{t-2} \cdot (\Delta Y_{it} - \theta \Delta Y_{it-1})$$

- Note: We have $K \simeq (T-1)(T-2)/2$ moment conditions

Arellano Bond – Bias

		$T=5$	$T=10$	$T=15$	$T=20$
$\rho=0.7$	GMM 1–step	-0.061	-0.130	-0.176	-0.217
	GMM 2–step	-0.030	-0.048	-0.082	-0.134
	GMM D	-0.069	-0.133	-0.180	-0.227
	GMM CUE	-0.003	0.000	0.000	0.000
	AMM $\nu=0.5$	0.000	0.001	0.001	-0.002
	AMM $\nu=1$	-0.001	0.000	0.002	-0.003
	EL	0.002	0.000	0.000	0.001
$\rho=0.9$	GMM 1–step	-0.317	-0.338	-0.377	-0.419
	GMM 2–step	-0.247	-0.200	-0.242	-0.310
	GMM D	-0.319	-0.387	-0.432	-0.478
	GMM CUE	-0.168	-0.012	-0.001	0.000
	AMM $\nu=0.5$	-0.108	-0.007	0.000	0.001
	AMM $\nu=1$	-0.109	-0.010	-0.003	-0.002
	EL	N/A	0.002	-0.002	0.002

Table: Bias based on 500 simulations, $N=500$. ν denotes dispersion in AMM estimator

Arellano Bond – SD

		$T=5$	$T=10$	$T=15$	$T=20$
$\rho=0.7$	GMM 1–step	0.170	0.089	0.080	0.078
	GMM 2–step	0.119	0.048	0.044	0.051
	GMM D	0.132	0.071	0.063	0.062
	GMM CUE	0.134	0.039	0.024	0.021
	AMM $\nu=0.5$	0.130	0.038	0.023	0.053
	AMM $\nu=1$	0.129	0.038	0.022	0.052
	EL	0.137	0.038	0.023	0.020
$\rho=0.9$	GMM 1–step	0.393	0.239	0.188	0.167
	GMM 2–step	0.285	0.168	0.140	0.138
	GMM D	0.278	0.168	0.134	0.118
	GMM CUE	0.386	0.099	0.048	0.037
	AMM $\nu=0.5$	0.258	0.081	0.043	0.032
	AMM $\nu=1$	0.257	0.080	0.043	0.031
	EL	N/A	0.112	0.047	0.036

Table: SD based on 500 simulations, $N=500$. ν denotes dispersion in AMM estimator

ν smoothes the loss

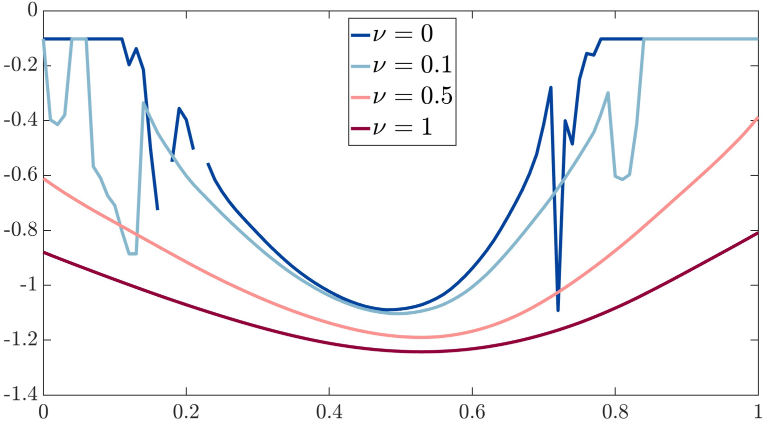

Figure: Realization for $\rho = 0.5$, with $(N, T) = (500, 20)$

Altonji & Segal (1996)

- Panel of individuals $(i = 1, \ldots, n)$ observed across $(t = 1, \ldots, T)$
- Observations $x_{it} \overset{iid}{\sim} F$, $\mathbb{E}[x_{it}] = 0, \mathbb{V}[x_{it}] = \sigma^2$
- Stacked observations $x_i = (x_{i1}, \ldots, x_{iT})$
- Parameter of interest is $\theta = \sigma^2$ and moment of the data to consider are

$$g\left(\sigma^2, x_i\right) = \left(x_{i1}^2 - \sigma^2, \ldots, x_{iT}^2 - \sigma^2\right)$$

- Consider $F = t_\eta, \log \mathcal{N}$
- Note: 1-step GMM is efficient in this case

Bias Student–t

	ν	$T = 1$	$T = 5$	$T = 10$	$T = 15$	$T = 20$	$T = 30$
AMM	0	0.006	-0.011	-0.014	-0.015	-0.016	-0.016
	0.05	0.006	-0.009	0.008	0.004	-0.002	-0.003
	0.1	0.006	-0.009	-0.012	-0.011	-0.013	-0.010
	0.5	0.006	-0.01	-0.012	-0.014	-0.014	-0.014
	1	0.007	-0.007	-0.009	-0.010	-0.011	-0.011
GMM	1–step	-0.003	0.001	0	0	-0.001	0
	2–step	-0.003	-0.038	-0.042	-0.044	-0.046	-0.045
	IT	-0.003	-0.038	-0.042	-0.044	-0.046	-0.045
	CUE	-0.003	-0.039	-0.042	-0.044	-0.046	-0.045
	Diagonal W	-0.003	-0.039	-0.042	-0.044	-0.046	-0.045

Table: Bias based on 500 simulations, sample size $= 500$. Student–t with 3 degrees of freedom. ν denotes the noise coefficient of the AMM estimator

RMSE Student–t

	ν	$T=1$	$T=5$	$T=10$	$T=15$	$T=20$	$T=30$
AMM	0	0.059	0.026	0.022	0.021	0.021	0.019
	0.05	0.059	0.048	0.133	0.113	0.084	0.070
	0.1	0.059	0.047	0.047	0.051	0.040	0.048
	0.5	0.059	0.025	0.021	0.020	0.019	0.019
	1	0.059	0.024	0.019	0.017	0.016	0.015
GMM	1–step	0.121	0.054	0.038	0.032	0.028	0.024
	2–step	0.121	0.060	0.053	0.052	0.052	0.049
	IT	0.121	0.060	0.053	0.052	0.052	0.049
	CUE	0.121	0.060	0.053	0.052	0.052	0.050
	Diagonal W	0.121	0.060	0.053	0.052	0.052	0.050

Table: RMSE based on 500 simulations, sample size = 500. Student–t with 3 degrees of freedom. ν denotes the noise coefficient of the AMM estimator

Bias Log–Normal

	ν	$T=1$	$T=5$	$T=10$	$T=15$	$T=20$	$T=30$
AMM	0	-0.006	-0.058	-0.236	-0.324	-0.28	-0.159
	0.05	-0.006	-0.049	-0.097	-0.088	-0.063	0.039
	0.1	-0.006	-0.017	-0.091	-0.110	-0.073	-0.048
	0.5	-0.006	0.003	-0.088	-0.110	-0.117	-0.112
	1	-0.006	-0.052	-0.096	-0.108	-0.111	-0.116
GMM	1–step	-0.003	-0.002	-0.013	-0.006	-0.007	-0.006
	2–step	-0.003	-0.209	-0.236	-0.238	-0.243	-0.247
	IT	-0.003	-0.209	-0.236	-0.238	-0.243	-0.247
	CUE	-0.003	-0.211	-0.240	-0.242	-0.247	-0.253
	Diagonal W	-0.003	-0.211	-0.240	-0.242	-0.247	-0.253

Table: Bias based on 500 simulations, sample size $=500$. Log–Normal distribution. ν denotes the noise coefficient of the AMM estimator

RMSE Log–Normal

	ν	$T=1$	$T=5$	$T=10$	$T=15$	$T=20$	$T=30$
AMM	0	0.169	0.203	0.348	0.415	0.36	0.197
	0.05	0.169	0.222	0.181	0.22	0.267	0.282
	0.1	0.169	0.261	0.186	0.171	0.247	0.263
	0.5	0.169	0.296	0.185	0.154	0.151	0.159
	1	0.169	0.214	0.159	0.128	0.13	0.127
GMM	1–step	0.369	0.198	0.133	0.104	0.099	0.093
	2–step	0.369	0.238	0.250	0.247	0.25	0.252
	IT	0.369	0.238	0.250	0.247	0.25	0.252
	CUE	0.369	0.241	0.254	0.251	0.255	0.258
	Diagonal W	0.369	0.241	0.254	0.251	0.255	0.258

Table: RMSE based on 500 simulations, sample size = 500. Log–Normal distribution. ν denotes the noise coefficient of the AMM estimator

Conclusion

Conclusion

- AMM is based on adversarial estimation
  - Contribution: Adapt adversarial estimation to models defined by moment conditions
  - The discriminator looks for patterns in the moment of g to distinguish true to simulated data
- AMM is a GEL estimator:
  - Use results similar to Newey & Smith (2004)
  - Derive finite sample bias
- Properties of AMM
  - Asymptotically equivalent to SMM/GMM
  - Better finite sample performance in terms of smaller bias
  - Computationally more tractable than other GEL estimators

Thanks!

AMM Asymptotics – General Case

Back

- Define the following

$$\Omega = \mathbb{E}\left[g\left(x,\theta_0\right)g\left(x,\theta_0\right)'\right],\ G = \mathbb{E}\left[\partial g\left(x,\theta\right)/\partial\theta\right]|_{\theta_0}$$

$$\Omega_\kappa = \Omega + \kappa I_m,\ \Sigma_\kappa = \left(G'\Omega_\kappa^{-1}G\right)^{-1},\ H_\kappa = \Sigma_\kappa G'\Omega_\kappa^{-1}$$

- Asymptotic Normality

$$\sqrt{n}\left(\widehat{\theta} - \theta_0\right) \longrightarrow \mathcal{N}\left(0, H_{\nu^2}\Omega_{\tau\nu^2}H_{\nu^2}'\right)$$

where $\tau = n/m$

AMM Asymptotics – Matching Moments

SMM

- Define the following

$$\Omega = \mathbb{E}\left[g_i(\theta)\, g_i(\theta)'\right], \; G = \mathbb{E}\left[\frac{\partial g_i(\theta)}{\partial \theta}\right]$$

$$\Sigma = \left(G'\Omega^{-1}G\right)^{-1}$$

- Under regularity conditions, AMM is asymptotically equivalent to SMM

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, (1+\tau)\Sigma\right)$$

where $\tau = \frac{n}{m}$

Assumptions

Back

Assumption (Consistency)

(a) $\theta_0 \in \Theta$ *is the unique solution to* $\mathbb{E}\left[g\left(x,\theta\right)\right] = \mathbb{E}\left[g\left(x^{\theta},\theta\right)\right]$ *(b)* Θ *is compact; (c)* $g\left(x,\theta\right)$ *is continuous at each* $\theta \in \Theta$ *with probability one; (d)* $\mathbb{E}\left[\sup_{\theta\in\Theta}\|g\left(x,\theta\right)\|^{\alpha}\right] < \infty$ *for some* $\alpha > 2$*; (e)* Ω *is nonsingular.*

Assumption (Asymptotic Normality)

(a) $\theta_0 \in int\left(\Theta\right)$*; (b)* $g\left(x,\theta\right)$ *is continuously differentiable in a neighborhood* $\mathcal{N}$ *of* θ_0 *and* $\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\left\|\partial g_i\left(\theta\right)/\partial\theta'\right\|\right] < \infty$*; (c)* $rank\left(G\right) = p$.

Assumptions (cont'd)

Assumption (Stochastic Expansions)

There is $b(x)$ *with* $\mathbb{E}\left[b(x)^6\right] < \infty$ *such that for* $0 \leq j \leq 4$ *and all* $x, \nabla^j g(x, \theta)$ *exists on a neighborhood* $\mathcal{N}$ *of* $\theta_0, \sup_{\theta \in \mathcal{N}} \left\|\nabla^j g(x, \theta)\right\| \leq b(x)$, *and for each* $\theta \in \mathcal{N}, \left\|\nabla^4 g(x, \theta) - \nabla^4 g(x, \theta_0)\right\| \leq b(x) \|\theta - \theta_0\|$

Bias terms

Back

- Terms for GEL estimators

$$B_I = n^{-1} H \left(-a + \mathbb{E}\left[G_i H g_i\right]\right)$$
$$B_G = -n^{-1} \Sigma \mathbb{E}\left[G_i' P g_i\right]$$
$$B_\Omega = n^{-1} H \mathbb{E}\left[g_i g_i' P g_i\right]$$
$$B_W = -n^{-1} H \sum \overline{\Omega}_{\theta_j} \left(H_W - H\right)' e_j$$

- ...and AMM estimators

$$B_{I_{\nu^2}} = n^{-1} H_{\nu^2} \left(-a_{\nu^2} + \mathbb{E}\left[G_i H_{\nu^2} g_i\right]\right)$$
$$B_{\Omega_{\nu^2}} = n^{-1} H_{\nu^2} \mathbb{E}\left[g_i g_i' P_{\nu^2} g_i\right]$$

where $M_\nu = \mathbb{E}\left[\partial m\left(x, \varphi, \nu\right) / \partial \varphi\right],\ A_\nu = \partial m\left(x, \varphi, \nu\right) / \partial \varphi - M_\nu,\ a_\nu = \text{vec}\left(A_\nu\right)$