Artificial Neural Networks to solve dynamic programming problems: a bias-corrected Monte Carlo estimator[1]

Julien Pascal

Central Bank of Luxembourg

July 5, 2023

[1]This paper should not be reported as representing the views of the BCL or the Eurosystem. The views expressed are those of the authors and may not be shared by other research staff or policymakers in the BCL or the Eurosystem.

Table of Contents

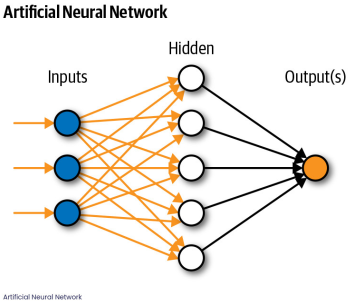
Context: The Unreasonable Effectiveness of Machine Learning

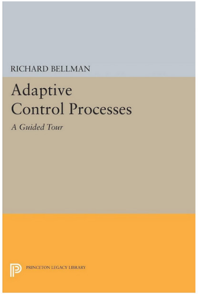
Figure: The success of AlphaGo

Source: AlphaGo (film)

This Paper

Global approach that uses **Artificial Neural Networks** to solve **Economic Models**





$$\mathbf{x} \to \mathcal{N}_\rho(\mathbf{x}) = \sigma_K(\mathbf{W}_K \ldots \sigma_2(\mathbf{W}_2\sigma_1(\mathbf{W}_1\mathbf{x} + \mathbf{b}_1) + \mathbf{b}_2) \ldots + \mathbf{b}_K)$$

Why this paper?

- **High Dimensional Models in Economics**
  - HA models: Aiyagari (1994), Krusell and Smith (1998)
  - HANK models: Kaplan et al. (2018)
  - firm dynamics (Khan and Thomas, 2008), multi-country models (Backus et al., 1992), OLG models (Marchiori and Pierrard, 2015)
- **Why Global Methods?**
  - non-differentiable models
  - linearization may eliminate interesting amplification mechanisms (certainty equivalence)
  - a non-stochastic steady-state may not exist in the first place
- **Why ANNs?**
  - theory: universal function approximation theorems (Hornik et al., 1989), resilient to the curse of dimensionality (Barron, 1993)
  - practice: backpropagation algorithm (Rumelhart et al., 1986), GPUs

Why this paper?

- **High Dimensional Models in Economics**
  - HA models: Aiyagari (1994), Krusell and Smith (1998)
  - HANK models: Kaplan et al. (2018)
  - firm dynamics (Khan and Thomas, 2008), multi-country models (Backus et al., 1992), OLG models (Marchiori and Pierrard, 2015)
- **Why Global Methods?**
  - non-differentiable models
  - linearization may eliminate interesting amplification mechanisms (certainty equivalence)
  - a non-stochastic steady-state may not exist in the first place
- **Why ANNs?**
  - theory: universal function approximation theorems (Hornik et al., 1989), resilient to the curse of dimensionality (Barron, 1993)
  - practice: backpropagation algorithm (Rumelhart et al., 1986), GPUs

Why this paper?

- **High Dimensional Models in Economics**
  - HA models: Aiyagari (1994), Krusell and Smith (1998)
  - HANK models: Kaplan et al. (2018)
  - firm dynamics (Khan and Thomas, 2008), multi-country models (Backus et al., 1992), OLG models (Marchiori and Pierrard, 2015)
- **Why Global Methods?**
  - non-differentiable models
  - linearization may eliminate interesting amplification mechanisms (certainty equivalence)
  - a non-stochastic steady-state may not exist in the first place
- **Why ANNs?**
  - theory: universal function approximation theorems (Hornik et al., 1989), resilient to the curse of dimensionality (Barron, 1993)
  - practice: backpropagation algorithm (Rumelhart et al., 1986), GPUs

Why this paper?

- **High Dimensional Models in Economics**
  - HA models: Aiyagari (1994), Krusell and Smith (1998)
  - HANK models: Kaplan et al. (2018)
  - firm dynamics (Khan and Thomas, 2008), multi-country models (Backus et al., 1992), OLG models (Marchiori and Pierrard, 2015)
- **Why Global Methods?**
  - non-differentiable models
  - linearization may eliminate interesting amplification mechanisms (certainty equivalence)
  - a non-stochastic steady-state may not exist in the first place
- **Why ANNs?**
  - theory: universal function approximation theorems (Hornik et al., 1989), resilient to the curse of dimensionality (Barron, 1993)
  - practice: backpropagation algorithm (Rumelhart et al., 1986), GPUs

Literature Review

1. **Global methods**:
   - Smolyak's sparse grid Krueger and Kubler (2004) and Judd et al. (2014)
   - Adpative sparse grid Brumm and Scheidegger (2017)
   - Gaussian processes Scheidegger and Bilionis (2019)
2. **ANN + Economic Models**:
   - ANN + traditional methods: Fernández-Villaverde et al. (2019)
   - 100% ANN: Azinovic et al. (2022), Maliar et al. (2021)

Literature Review

1. **Global methods:**
   - Smolyak's sparse grid Krueger and Kubler (2004) and Judd et al. (2014)
   - Adpative sparse grid Brumm and Scheidegger (2017)
   - Gaussian processes Scheidegger and Bilionis (2019)
2. **ANN + Economic Models:**
   - ANN + traditional methods: Fernández-Villaverde et al. (2019)
   - 100% ANN: Azinovic et al. (2022), Maliar et al. (2021)

Literature Review

1. **Global methods**:
   - Smolyak's sparse grid Krueger and Kubler (2004) and Judd et al. (2014)
   - Adpative sparse grid Brumm and Scheidegger (2017)
   - Gaussian processes Scheidegger and Bilionis (2019)
2. **ANN + Economic Models**:
   - ANN + traditional methods: Fernández-Villaverde et al. (2019)
   - 100% ANN: Azinovic et al. (2022), Maliar et al. (2021)

Contributions

1. **Contribution 1**: generalize the all-in-one expectation operator of Maliar et al. (2021) with the **bc-MC operator**. all-in-one
2. **Contribution 2**: derive **theoretical properties** for the bc-MC operator
3. **Contribution 3**: **numerical illustrations** and discussion on **time-accuracy trade-offs**

Contributions

1. **Contribution 1**: generalize the all-in-one expectation operator of Maliar et al. (2021) with the **bc-MC operator**. all-in-one
2. **Contribution 2**: derive **theoretical properties** for the bc-MC operator
3. **Contribution 3**: **numerical illustrations** and discussion on **time-accuracy trade-offs**

Contributions

1. **Contribution 1**: generalize the all-in-one expectation operator of Maliar et al. (2021) with the **bc-MC operator**. all-in-one
2. **Contribution 2**: derive **theoretical properties** for the bc-MC operator
3. Contribution 3: **numerical illustrations** and discussion on **time-accuracy trade-offs**

Contributions

1. **Contribution 1**: generalize the all-in-one expectation operator of Maliar et al. (2021) with the **bc-MC operator**. all-in-one
2. **Contribution 2**: derive **theoretical properties** for the bc-MC operator
3. **Contribution 3**: **numerical illustrations** and discussion on **time-accuracy trade-offs**

Table of Contents

General Structure of Economic Models

Example | J equations

- Functional stochastic equation:

$$\mathbb{E}_\varepsilon \Big(f(s, \epsilon) \Big) = 0 \quad \text{for } \forall s \in S \tag{1}$$

Examples: Euler or Bellman equations.

- Solution is a **parametric decision function** $\underbrace{\mathcal{ANN}(s|\theta)}_{\text{neural network}} = s'$, which **minimizes the loss:**

$$\mathcal{L}(\theta) = \mathbb{E}_s \left[\mathbb{E}_\varepsilon \Big(f(s, \epsilon|\theta) \Big)^2 \right] \tag{2}$$

General Structure of Economic Models

Example · J equations

- Functional stochastic equation:

$$\mathbb{E}_\varepsilon\left(f(s,\epsilon)\right) = 0 \quad \text{for } \forall s \in S \tag{1}$$

Examples: Euler or Bellman equations.

- Solution is a **parametric decision function** $\underbrace{\mathcal{ANN}(s|\theta)}_{\text{neural network}} = s'$, which **minimizes the loss**:

$$\mathcal{L}(\theta) = \mathbb{E}_s\left[\mathbb{E}_\varepsilon\left(f(s,\epsilon|\theta)\right)^2\right] \tag{2}$$

The biased Monte Carlo Estimation

To approximate $\mathcal{L}(\theta)$, replace **population means** by **sample averages** (Monte Carlo integration):

$$\mathcal{L}^B_{M,N}(\theta) = \frac{1}{M}\sum_{m=1}^{M}\Big[\frac{1}{N}\sum_{n=1}^{N} f(s_m,\epsilon_n|\theta)\Big]^2 \tag{3}$$

Bias $(\mathrm{Var}[g(x)] = \mathbb{E}[g(x)^2] - \mathbb{E}[g(x)]^2 \Leftrightarrow \mathbb{E}[g(x)^2] = \mathbb{E}[g(x)]^2 + \mathrm{Var}[g(x)])$:

$$\begin{aligned}\mathbb{E}_\epsilon\Big[\Big(\frac{1}{N}\sum_{n=1}^{N} f(s_m,\epsilon_n|\theta)\Big)^2\Big] &= \Big(\mathbb{E}_\epsilon\Big[\frac{1}{N}\sum_{n=1}^{N} f(s_m,\epsilon_n|\theta)\Big]\Big)^2 + \mathrm{Var}_\epsilon\Big(\frac{1}{N}\sum_{n=1}^{N} f(s_m,\epsilon_n|\theta)\Big)\\ \mathbb{E}_\epsilon\Big[\Big(\frac{1}{N}\sum_{n=1}^{N} f(s_m,\epsilon_n|\theta)\Big)^2\Big] &= \underbrace{\mu^2_{s_m}}_{\text{true value}} + \underbrace{\frac{\sigma^2_{f,s_m}}{N}}_{\text{bias}}\end{aligned} \tag{4}$$

The biased Monte Carlo Estimation

To approximate $\mathcal{L}(\theta)$, replace **population means** by **sample averages** (Monte Carlo integration):

$$\mathcal{L}^B_{M,N}(\theta) = \frac{1}{M}\sum_{m=1}^{M}\Big[\frac{1}{N}\sum_{n=1}^{N} f(s_m, \epsilon_n|\theta)\Big]^2 \tag{3}$$

Bias $(\mathrm{Var}[g(x)] = \mathbb{E}[g(x)^2] - \mathbb{E}[g(x)]^2 \Leftrightarrow \mathbb{E}[g(x)^2] = \mathbb{E}[g(x)]^2 + \mathrm{Var}[g(x)])$:

$$\begin{aligned} \mathbb{E}_\epsilon\Big[\Big(\frac{1}{N}\sum_{n=1}^{N} f(s_m,\epsilon_n|\theta)\Big)^2\Big] &= \Big(\mathbb{E}_\epsilon\Big[\frac{1}{N}\sum_{n=1}^{N} f(s_m,\epsilon_n|\theta)\Big]\Big)^2 + \mathrm{Var}_\epsilon\Big(\frac{1}{N}\sum_{n=1}^{N} f(s_m,\epsilon_n|\theta)\Big) \\ \mathbb{E}_\epsilon\Big[\Big(\frac{1}{N}\sum_{n=1}^{N} f(s_m,\epsilon_n|\theta)\Big)^2\Big] &= \underbrace{\mu^2_{s_m}}_{\text{true value}} + \underbrace{\frac{\sigma^2_{f,s_m}}{N}}_{\text{bias}} \end{aligned} \tag{4}$$

The biased Monte Carlo Estimation

To approximate $\mathcal{L}(\theta)$, replace **population means** by **sample averages** (Monte Carlo integration):

$$\mathcal{L}^B_{M,N}(\theta) = \frac{1}{M}\sum_{m=1}^{M}\Big[\frac{1}{N}\sum_{n=1}^{N} f(s_m, \epsilon_n|\theta)\Big]^2 \tag{3}$$

Bias $(\mathrm{Var}[g(x)] = \mathbb{E}[g(x)^2] - \mathbb{E}[g(x)]^2 \Leftrightarrow \mathbb{E}[g(x)^2] = \mathbb{E}[g(x)]^2 + \mathrm{Var}[g(x)])$:

$$\begin{aligned}\mathbb{E}_\epsilon\Big[\Big(\frac{1}{N}\sum_{n=1}^{N} f(s_m,\epsilon_n|\theta)\Big)^2\Big] &= \Big(\mathbb{E}_\epsilon\Big[\frac{1}{N}\sum_{n=1}^{N} f(s_m,\epsilon_n|\theta)\Big]\Big)^2 + \mathrm{Var}_\epsilon\Big(\frac{1}{N}\sum_{n=1}^{N} f(s_m,\epsilon_n|\theta)\Big) \\ \mathbb{E}_\epsilon\Big[\Big(\frac{1}{N}\sum_{n=1}^{N} f(s_m,\epsilon_n|\theta)\Big)^2\Big] &= \underbrace{\mu^2_{s_m}}_{\text{true value}} + \underbrace{\frac{\sigma^2_{f,s_m}}{N}}_{\text{bias}}\end{aligned} \tag{4}$$

The biased-corrected Monte Carlo estimator

The **minimum variance unbiased estimator** (MVUE) of μ^2 is $\hat{\mu}^2 - \frac{S_n^2}{N}$ (Das (1975)):

$$\mathcal{L}^U_{M,N}(\theta) = \frac{1}{M}\sum_{m=1}^{M}\left\{\left[\frac{1}{N}\sum_{n=1}^{N} f(s_m, \epsilon_n|\theta)\right]^2 - \underbrace{\frac{S^2_{m,n}}{N}}_{\text{remove the bias with s. variance}}\right\} \tag{5}$$

Proposition

1. The biased-corrected Monte Carlo estimator (5) can be expressed as:

$$\mathcal{L}^{U}_{M,N}(\theta) = \frac{2}{MN(N-1)} \sum_{m=1}^{M} \sum_{1 \leq i < j}^{N} f(s_m, \epsilon^i_m|\theta) f(s_m, \epsilon^j_m|\theta) \tag{6}$$

where ϵ^i and ϵ^j are i.i.d shocks with the same distribution as ϵ (N series of i.i.d shocks).

2. In the special case with $N = 2$:

$$\mathcal{L}^{U}_{M,2}(\theta) = \frac{1}{M} \sum_{m=1}^{M} f(s_m, \epsilon^1_m|\theta) f(s_m, \epsilon^2_m|\theta)$$

This is the all-in-one operator of Maliar et al. (2021) all-in-one

Proposition

1. The biased-corrected Monte Carlo estimator (5) can be expressed as:

$$\mathcal{L}^{U}_{M,N}(\theta) = \frac{2}{MN(N-1)} \sum_{m=1}^{M} \sum_{1 \leq i < j}^{N} f(s_m, \epsilon^i_m|\theta) f(s_m, \epsilon^j_m|\theta) \tag{6}$$

where ϵ^i and ϵ^j are i.i.d shocks with the same distribution as ϵ (N series of i.i.d shocks).

2. In the special case with $N = 2$:

$$\mathcal{L}^{U}_{M,2}(\theta) = \frac{1}{M} \sum_{m=1}^{M} f(s_m, \epsilon^1_m|\theta) f(s_m, \epsilon^2_m|\theta)$$

This is the all-in-one operator of Maliar et al. (2021) all-in-one

Proposition

1. The biased-corrected Monte Carlo estimator (5) can be expressed as:

$$\mathcal{L}^{U}_{M,N}(\theta) = \frac{2}{MN(N-1)} \sum_{m=1}^{M} \sum_{1 \leq i < j}^{N} f(s_m, \epsilon^i_m | \theta) f(s_m, \epsilon^j_m | \theta) \tag{6}$$

where ϵ^i and ϵ^j are i.i.d shocks with the same distribution as ϵ (N series of i.i.d shocks).

2. In the special case with $N = 2$:

$$\mathcal{L}^{U}_{M,2}(\theta) = \frac{1}{M} \sum_{m=1}^{M} f(s_m, \epsilon^1_m | \theta) f(s_m, \epsilon^2_m | \theta)$$

This is the all-in-one operator of Maliar et al. (2021) all-in-one

Proposition (choice of hyperparameters M and N) Proposition 4

1. Let $e_{M,N}(f|\theta)$ denote the **integration error**:

$$e_{M,N}(f|\theta) \equiv \mathbb{E}_s\left[\mathbb{E}_\varepsilon\left(f(s,\epsilon|\theta)\right)^2\right] - \underbrace{\mathcal{L}^U_{M,N}(\theta)}_{\text{stochastic}}$$

2. The **mean squared integration error** is equal to:

$$\mathbb{E}\left[e_{M,N}(f|\theta)^2\right] = \underbrace{\text{Var}(\mathcal{L}^U_{M,N}(\theta))}_{\text{calculable}} \tag{7}$$

- **Procedure**: grid for N (and M), select N to minimize $\text{Var}(\mathcal{L}^U_{M,N}(\theta))$

Proposition (choice of hyperparameters M and N) Proposition 4

1. Let $e_{M,N}(f|\theta)$ denote the **integration error**:

$$e_{M,N}(f|\theta) \equiv \mathbb{E}_s\left[\mathbb{E}_\varepsilon\left(f(s,\epsilon|\theta)\right)^2\right] - \underbrace{\mathcal{L}^U_{M,N}(\theta)}_{\text{stochastic}}$$

2. The **mean squared integration error** is equal to:

$$\mathbb{E}\left[e_{M,N}(f|\theta)^2\right] = \underbrace{\text{Var}(\mathcal{L}^U_{M,N}(\theta))}_{\text{calculable}} \tag{7}$$

- **Procedure**: grid for N (and M), select N to minimize $\text{Var}(\mathcal{L}^U_{M,N}(\theta))$

Proposition (choice of hyperparameters M and N) Proposition 4

1. Let $e_{M,N}(f|\theta)$ denote the **integration error**:

$$e_{M,N}(f|\theta) \equiv \mathbb{E}_s\left[\mathbb{E}_\varepsilon\left(f(s,\epsilon|\theta)\right)^2\right] - \underbrace{\mathcal{L}^U_{M,N}(\theta)}_{\text{stochastic}}$$

2. The **mean squared integration error** is equal to:

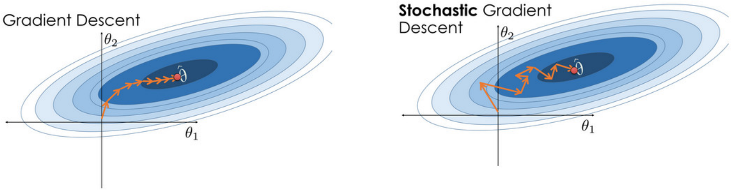
$$\mathbb{E}\left[e_{M,N}(f|\theta)^2\right] = \underbrace{\mathrm{Var}(\mathcal{L}^U_{M,N}(\theta))}_{\text{calculable}} \quad (7)$$

- **Procedure**: grid for N (and M), select N to minimize $\mathrm{Var}(\mathcal{L}^U_{M,N}(\theta))$

Training by stochastic gradient descent

$$\theta_{i+1} = \theta_i - \gamma \nabla_\theta \mathcal{L}^U_{M,N}(\theta_i) \quad (8)$$

Figure: GD and SGD



Source

The **smaller the variance** of the stochastic gradient, the **faster the training** (Katharopoulos and Fleuret (2018)).

Takeaways

The biased-corrected Monte Carlo estimator:

$$\mathcal{L}^{U}_{M,N}(\theta) = \frac{2}{MN(N-1)} \sum_{m=1}^{M} \sum_{1 \leq i < j}^{N} f(s_m, \epsilon^i_m | \theta) f(s_m, \epsilon^j_m | \theta)$$

- Model with a **lot of uncertainty**: set N high (use many different series of independent shocks) Neogrowth model
- Model with a **lot of non-linearities**: set M high (use many draws in the state space) Model with a borrowing constraint
- See proposition 4 in the paper Proposition 4

Table of Contents

Model with a borrowing constraint Case $l = 2$

$$\max_{\{c_t\}_{t=0}^{\infty}} \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t u(c_t) \exp(\delta_t)\right] \tag{9}$$

- constraint: $0 \leq c_t \leq w_t$
- $w_{t+1} = (w_t - c_t)\bar{r}\exp(r_{t+1}) + \exp(y_{t+1})$, $y_t = \exp(\sum_{i=1}^{l} p_{i,t})$
- $\beta \in (0,1)$, $\bar{r} \in (0, \frac{1}{\beta})$, $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$

AR(1) processes:

$$\begin{aligned} p_{i,t+1} &= \rho_{i,p} p_{i,t} + \sigma_{i,p}\varepsilon^p_{i,t+1}, \quad \forall i \in 1,2,...,l \\ r_{t+1} &= \rho_r r_t + \sigma_r \varepsilon^r_{t+1} \\ \delta_{t+1} &= \rho_\delta \delta_t + \sigma_\delta \varepsilon^\delta_{t+1} \end{aligned} \tag{10}$$

state $s = (w, r, \delta, p_1, ..., p_l)$ with $d_s \equiv 3 + l$ elements, shock $\varepsilon = (\varepsilon^r, \varepsilon^\delta, \varepsilon^p_1, ..., \varepsilon^p_l)$
$d_\varepsilon \equiv 2 + l$ elements.

Model with a borrowing constraint

Case $l = 2$

$$\max_{\{c_t\}_{t=0}^{\infty}} \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t u(c_t)\exp(\delta_t)\right] \tag{9}$$

- constraint: $0 \leq c_t \leq w_t$
- $w_{t+1} = (w_t - c_t)\bar{r}\exp(r_{t+1}) + \exp(y_{t+1})$, $y_t = \exp(\sum_{i=1}^{l} p_{i,t})$
- $\beta \in (0, 1)$, $\bar{r} \in (0, \frac{1}{\beta})$, $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$

AR(1) processes:

$$\begin{aligned} p_{i,t+1} &= \rho_{i,p} p_{i,t} + \sigma_{i,p}\varepsilon^p_{i,t+1}, \quad \forall i \in 1, 2, ..., l \\ r_{t+1} &= \rho_r r_t + \sigma_r \varepsilon^r_{t+1} \\ \delta_{t+1} &= \rho_\delta \delta_t + \sigma_\delta \varepsilon^\delta_{t+1} \end{aligned} \tag{10}$$

state $s = (w, r, \delta, p_1, ..., p_l)$ with $d_s \equiv 3 + l$ elements, shock $\varepsilon = (\varepsilon^r, \varepsilon^\delta, \varepsilon^p_1, ..., \varepsilon^p_l)$ $d_\varepsilon \equiv 2 + l$ elements.

Model with a borrowing constraint (Case $l = 2$)

$$\max_{\{c_t\}_{t=0}^{\infty}} \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t u(c_t)\exp(\delta_t)\right] \tag{9}$$

- constraint: $0 \leq c_t \leq w_t$
- $w_{t+1} = (w_t - c_t)\bar{r}\exp(r_{t+1}) + \exp(y_{t+1})$, $y_t = \exp(\sum_{i=1}^{l} p_{i,t})$
- $\beta \in (0,1)$, $\bar{r} \in (0, \frac{1}{\beta})$, $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$

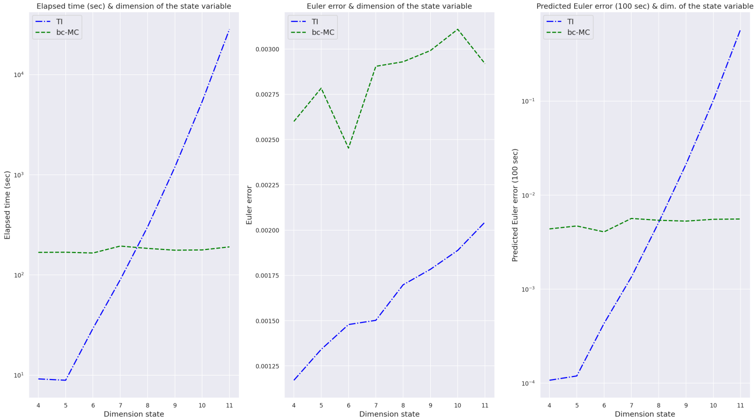
AR(1) processes:

$$\begin{aligned} p_{i,t+1} &= \rho_{i,p} p_{i,t} + \sigma_{i,p}\varepsilon^{p}_{i,t+1}, \quad \forall i \in 1, 2, ..., l \\ r_{t+1} &= \rho_r r_t + \sigma_r \varepsilon^r_{t+1} \\ \delta_{t+1} &= \rho_\delta \delta_t + \sigma_\delta \varepsilon^\delta_{t+1} \end{aligned} \tag{10}$$

state $s = (w, r, \delta, p_1, ..., p_l)$ with $d_s \equiv 3 + l$ elements, shock $\varepsilon = (\varepsilon^r, \varepsilon^\delta, \varepsilon^p_1, ..., \varepsilon^p_l)$ $d_\varepsilon \equiv 2 + l$ elements.

bc-MC and Time Iteration: time-accuracy trade-off

Figure: bc-MC vs TI: time and accuracy



bc-MC and Time Iteration: time-accuracy trade-off

Figure: bc-MC vs TI: time and accuracy large scale model

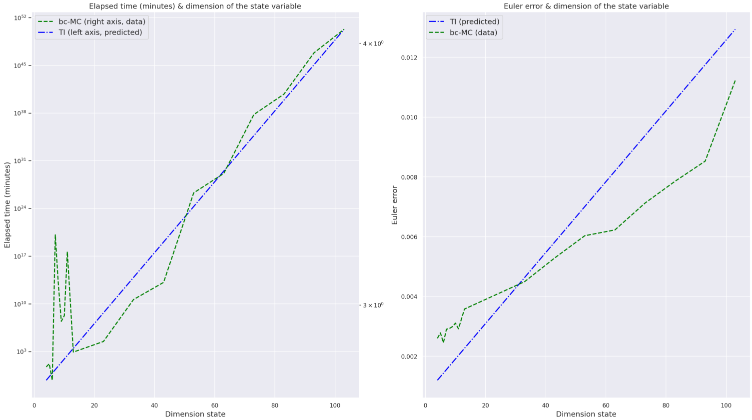


Table of Contents

Conclusion

1. Generalize the all-in-one expectation operator of Maliar et al. (2021) with the **bc-MC operator**.
2. Derive **theoretical properties** for the bc-MC operator
3. **Numerical illustrations** and discussion on **time-accuracy trade-offs**

Conclusion

1. Generalize the all-in-one expectation operator of Maliar et al. (2021) with the **bc-MC operator**.
2. Derive **theoretical properties** for the bc-MC operator
3. **Numerical illustrations** and discussion on **time-accuracy trade-offs**

Conclusion

1. Generalize the all-in-one expectation operator of Maliar et al. (2021) with the **bc-MC operator**.
2. Derive **theoretical properties** for the bc-MC operator
3. **Numerical illustrations** and discussion on **time-accuracy trade-offs**

Conclusion

1. Generalize the all-in-one expectation operator of Maliar et al. (2021) with the **bc-MC operator**.
2. Derive **theoretical properties** for the bc-MC operator
3. **Numerical illustrations** and discussion on **time-accuracy trade-offs**

Thank You

References I

Aiyagari, S. R. (1994). Uninsured idiosyncratic risk and aggregate saving. *The Quarterly Journal of Economics*, 109(3):659–684.

Azinovic, M., Gaegauf, L., and Scheidegger, S. (2022). Deep equilibrium nets. *International Economic Review*.

Backus, D. K., Kehoe, P. J., and Kydland, F. E. (1992). International real business cycles. *Journal of political Economy*, 100(4):745–775.

Barron, A. R. (1993). Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information theory*, 39(3):930–945.

Brumm, J. and Scheidegger, S. (2017). Using adaptive sparse grids to solve high-dimensional dynamic models. *Econometrica*, 85(5):1575–1612.

Das, B. (1975). Estimation of $\mu 2$ in normal population. *Calcutta Statistical Association Bulletin*, 24(1-4):135–140.

Fernández-Villaverde, J., Hurtado, S., and Nuno, G. (2019). Financial frictions and the wealth distribution. Technical report, National Bureau of Economic Research.

References II

Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366.

Judd, K. L., Maliar, L., Maliar, S., and Valero, R. (2014). Smolyak method for solving dynamic economic models: Lagrange interpolation, anisotropic grid and adaptive domain. *Journal of Economic Dynamics and Control*, 44:92–123.

Kaplan, G., Moll, B., and Violante, G. L. (2018). Monetary policy according to hank. *American Economic Review*, 108(3):697–743.

Katharopoulos, A. and Fleuret, F. (2018). Not all samples are created equal: Deep learning with importance sampling. In *International conference on machine learning*, pages 2525–2534. PMLR.

Khan, A. and Thomas, J. K. (2008). Idiosyncratic shocks and the role of nonconvexities in plant and aggregate investment dynamics. *Econometrica*, 76(2):395–436.

Krueger, D. and Kubler, F. (2004). Computing equilibrium in olg models with stochastic production. *Journal of Economic Dynamics and Control*, 28(7):1411–1436.

References III

Krusell, P. and Smith, Jr, A. A. (1998). Income and wealth heterogeneity in the macroeconomy. *Journal of political Economy*, 106(5):867–896.

Maliar, L., Maliar, S., and Winant, P. (2021). Deep learning for solving dynamic economic models. *Journal of Monetary Economics*, 122:76–101.

Marchiori, L. and Pierrard, O. (2015). *LOLA 3.0: Luxembourg OverLapping generation model for policy Analysis: Introduction of a financial sector in LOLA*. Banque centrale du Luxembourg.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *nature*, 323(6088):533–536.

Scheidegger, S. and Bilionis, I. (2019). Machine learning for high-dimensional dynamic stochastic economies. *Journal of Computational Science*, 33:68–82.

All-in-one Maliar et al. (2021) Contributions Proposition 2

Key idea (AIO):

$$\Big(E_\varepsilon[f(\varepsilon)]\Big)^2 = E_{\varepsilon_1}[f(\varepsilon_1)]E_{\varepsilon_2}[f(\varepsilon_2)]$$

But also (bc-MC):

$$\Big(E_\varepsilon[f(\varepsilon)]\Big)^2 = \frac{1}{3}\Big(E_{\varepsilon_1}[f(\varepsilon_1)]E_{\varepsilon_2}[f(\varepsilon_2)] + E_{\varepsilon_1}[f(\varepsilon_1)]E_{\varepsilon_3}[f(\varepsilon_3)] + E_{\varepsilon_2}[f(\varepsilon_2)]E_{\varepsilon_3}[f(\varepsilon_3)]\Big)$$

Or

$$\Big(E_\varepsilon[f(\varepsilon)]\Big)^2 = \frac{1}{6}\Big(E_{\varepsilon_1}[f(\varepsilon_1)]E_{\varepsilon_2}[f(\varepsilon_2)] + E_{\varepsilon_1}[f(\varepsilon_1)]E_{\varepsilon_3}[f(\varepsilon_3)] + E_{\varepsilon_1}[f(\varepsilon_1)]E_{\varepsilon_4}[f(\varepsilon_4)] + \ldots\Big)$$

etc.

All-in-one Maliar et al. (2021) Contributions Proposition 2

Key idea (AIO):

$$\left(E_\varepsilon[f(\varepsilon)]\right)^2 = E_{\varepsilon_1}[f(\varepsilon_1)]E_{\varepsilon_2}[f(\varepsilon_2)]$$

But also (bc-MC):

$$\left(E_\varepsilon[f(\varepsilon)]\right)^2 = \frac{1}{3}\Big(E_{\varepsilon_1}[f(\varepsilon_1)]E_{\varepsilon_2}[f(\varepsilon_2)] + E_{\varepsilon_1}[f(\varepsilon_1)]E_{\varepsilon_3}[f(\varepsilon_3)] + E_{\varepsilon_2}[f(\varepsilon_2)]E_{\varepsilon_3}[f(\varepsilon_3)]\Big)$$

Or

$$\left(E_\varepsilon[f(\varepsilon)]\right)^2 = \frac{1}{6}\Big(E_{\varepsilon_1}[f(\varepsilon_1)]E_{\varepsilon_2}[f(\varepsilon_2)] + E_{\varepsilon_1}[f(\varepsilon_1)]E_{\varepsilon_3}[f(\varepsilon_3)] + E_{\varepsilon_1}[f(\varepsilon_1)]E_{\varepsilon_4}[f(\varepsilon_4)] + ...\Big)$$

etc.

J stochastic functional equations Structure

Economic model:

$$\mathbb{E}_{\varepsilon}\left(f_j(s,\epsilon)\right) = 0 \quad \text{for } s \in S \text{ and } j \in 1, ..., J \tag{11}$$

Loss:

$$\mathcal{L}(\theta) = \sum_{j=1}^{J} \vartheta_j \, \mathbb{E}_s\left[\mathbb{E}_{\varepsilon}\left(f_j(s,\epsilon|\theta)\right)^2\right] \tag{12}$$

The biased-corrected Monte Carlo estimator writes:

$$\mathcal{L}^U_{M,N}(\theta) = \sum_{j=1}^{J} \vartheta_j \Big(\frac{1}{M} \sum_{m=1}^{M} \Big\{ \Big[\frac{1}{N} \sum_{n=1}^{N} f_j(s_m, \epsilon_n|\theta) \Big]^2 - \frac{S^2_{j,m,n}}{N} \Big\} \Big) \tag{13}$$

J stochastic functional equations Structure

Economic model:

$$\mathbb{E}_\varepsilon\left(f_j(s,\epsilon)\right) = 0 \quad \text{for } s \in S \text{ and } j \in 1, ..., J \tag{11}$$

Loss:

$$\mathcal{L}(\theta) = \sum_{j=1}^{J} \vartheta_j \, \mathbb{E}_s\left[\mathbb{E}_\varepsilon\left(f_j(s,\epsilon|\theta)\right)^2\right] \tag{12}$$

The biased-corrected Monte Carlo estimator writes:

$$\mathcal{L}^U_{M,N}(\theta) = \sum_{j=1}^{J} \vartheta_j \left(\frac{1}{M}\sum_{m=1}^{M}\left\{\left[\frac{1}{N}\sum_{n=1}^{N} f_j(s_m,\epsilon_n|\theta)\right]^2 - \frac{S^2_{j,m,n}}{N}\right\}\right) \tag{13}$$

Stochastic neogrowth model Structure Takeaways

$$\max_{\{c_t\}_{t=0}^{\infty}} \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t u(c_t)\right] \tag{14}$$

- constraints $0 \leq c_t \leq y_t$
- $y_{t+1} = g(y_t - c_t)\eta_{t+1}$, $\eta_t \equiv \eta(\nu_t) = \exp(\mu + \sigma_\nu \nu_t)$, $\nu \sim \mathcal{N}(0,1)$
- $u(c) = \log(c)$, $g(k) = k^\alpha$, $\beta \in (0,1)$

Euler equation characterizing the model:

$$\mathbb{E}_\nu \left[u'\Big(c(y|\theta)\Big) - \beta u'\Big(c\Big(g\big(y - c(y|\theta)\big)\eta(\nu)\Big|\theta\Big)g'\big(y - c(y|\theta)\big)\eta(\nu)\Big)\right] = 0 \tag{15}$$

Equation (15) is an example of equation (1):

$$f(s,\varepsilon) = u'\Big(c(s|\theta)\Big) - \beta u'\Big(c\Big(g\big(s - c(s|\theta)\big)\eta(\varepsilon)\Big|\theta\Big)g'\big(s - c(s|\theta)\big)\eta(\varepsilon)\Big)$$

Stochastic neogrowth model

Structure Takeaways

$$\max_{\{c_t\}_{t=0}^{\infty}} \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t u(c_t)\right] \tag{14}$$

- constraints $0 \leq c_t \leq y_t$
- $y_{t+1} = g(y_t - c_t)\eta_{t+1}$, $\eta_t \equiv \eta(\nu_t) = \exp(\mu + \sigma_\nu \nu_t)$, $\nu \sim \mathcal{N}(0,1)$
- $u(c) = \log(c)$, $g(k) = k^\alpha$, $\beta \in (0,1)$

Euler equation characterizing the model:

$$\mathbb{E}_\nu\left[u'\Big(c(y|\theta)\Big) - \beta u'\Big(c\Big(g\big(y - c(y|\theta)\big)\eta(\nu)\Big|\theta\Big)g'\big(y - c(y|\theta)\big)\eta(\nu)\Big)\right] = 0 \tag{15}$$

Equation (15) is an example of equation (1):

$$f(s,\varepsilon) = u'\Big(c(s|\theta)\Big) - \beta u'\Big(c\Big(g\big(s - c(s|\theta)\big)\eta(\varepsilon)\Big|\theta\Big)g'\big(s - c(s|\theta)\big)\eta(\varepsilon)\Big)$$

Stochastic neogrowth model

$$\max_{\{c_t\}_{t=0}^{\infty}} \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t u(c_t)\right] \tag{14}$$

- constraints $0 \leq c_t \leq y_t$
- $y_{t+1} = g(y_t - c_t)\eta_{t+1}$, $\eta_t \equiv \eta(\nu_t) = \exp(\mu + \sigma_\nu \nu_t)$, $\nu \sim \mathcal{N}(0,1)$
- $u(c) = \log(c)$, $g(k) = k^\alpha$, $\beta \in (0,1)$

Euler equation characterizing the model:

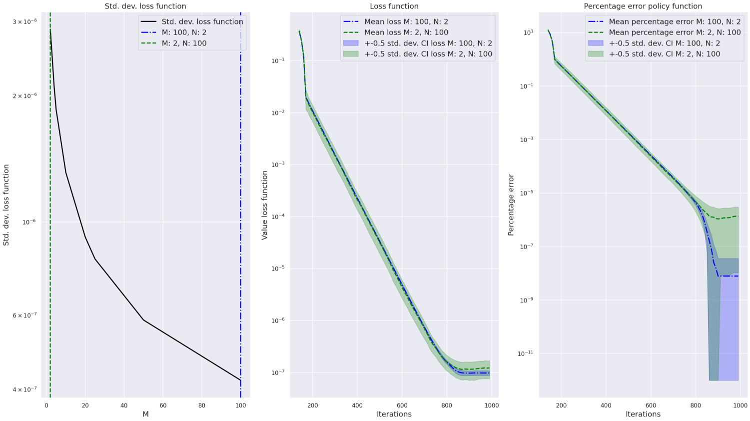
$$\mathbb{E}_\nu\left[u'\Big(c(y|\theta)\Big) - \beta u'\Big(c\Big(g\big(y - c(y|\theta)\big)\eta(\nu)\Big|\theta\Big)g'\big(y - c(y|\theta)\big)\eta(\nu)\Big)\right] = 0 \tag{15}$$

Equation (15) is an example of equation (1):

$$f(s,\varepsilon) = u'\Big(c(s|\theta)\Big) - \beta u'\Big(c\Big(g\big(s - c(s|\theta)\big)\eta(\varepsilon)\Big|\theta\Big)g'\big(s - c(s|\theta)\big)\eta(\varepsilon)\Big)$$

Stochastic neogrowth model

Figure: Low-uncertainty parametrization ($\sigma_\nu = 0.5$)



Stochastic neogrowth model

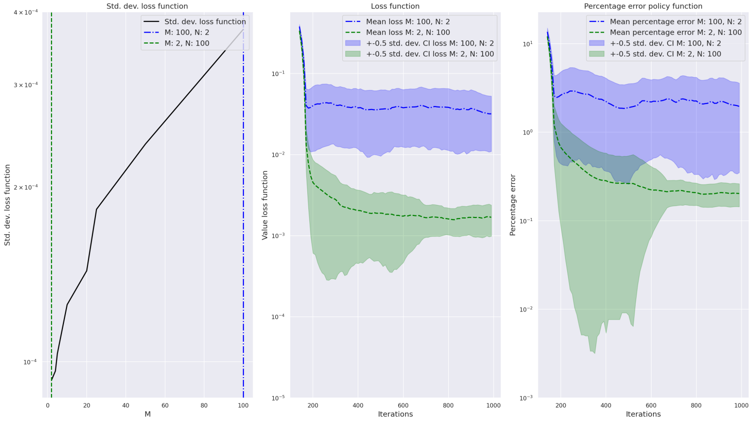


Figure: High-uncertainty parametrization ($\sigma_\nu = 1.5$)

Optimal consumption with a borrowing constraint Takeaways

$$\max_{\{c_t\}_{t=0}^{\infty}} \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t u(c_t)\exp(\delta_t)\right] \tag{16}$$

- constraint: $0 \leq c_t \leq w_t + b$
- $w_{t+1} = (w_t - c_t)\bar{r}\exp(r_{t+1}) + \exp(y_{t+1}),\ y_t = \exp(p_t + q_t)$
- $\beta \in (0,1),\ \bar{r} \in (0, \frac{1}{\beta}),\ u(c) = \frac{c^{1-\gamma}}{1-\gamma}$

The four exogenous variables are assumed to following AR(1) processes:

$$\begin{aligned} p_{t+1} &= \rho_p p_t + \sigma_p \varepsilon^p_{t+1} \\ q_{t+1} &= \rho_q q_t + \sigma_q \varepsilon^q_{t+1} \\ r_{t+1} &= \rho_r r_t + \sigma_r \varepsilon^r_{t+1} \\ \delta_{t+1} &= \rho_\delta \delta_t + \sigma_\delta \varepsilon^\delta_{t+1} \end{aligned} \tag{17}$$

Optimal consumption with a borrowing constraint Takeaways

$$\max_{\{c_t\}_{t=0}^{\infty}} \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t u(c_t)\exp(\delta_t)\right] \tag{16}$$

- constraint: $0 \leq c_t \leq w_t + b$
- $w_{t+1} = (w_t - c_t)\bar{r}\exp(r_{t+1}) + \exp(y_{t+1}),\ y_t = \exp(p_t + q_t)$
- $\beta \in (0,1),\ \bar{r} \in (0, \frac{1}{\beta}),\ u(c) = \frac{c^{1-\gamma}}{1-\gamma}$

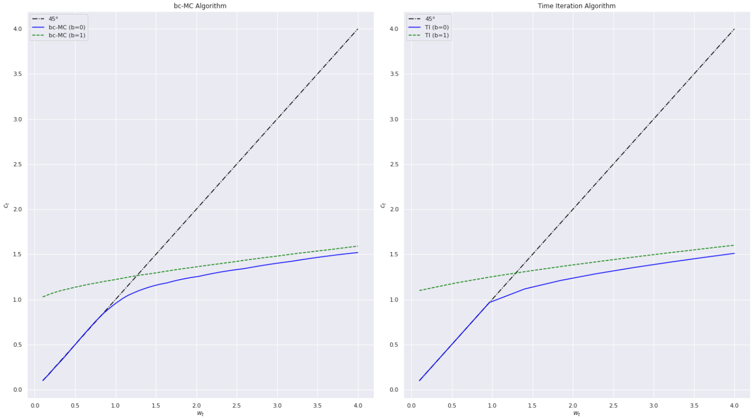
The four exogenous variables are assumed to following AR(1) processes:

$$\begin{aligned}
p_{t+1} &= \rho_p p_t + \sigma_p \varepsilon_{t+1}^p \\
q_{t+1} &= \rho_q q_t + \sigma_q \varepsilon_{t+1}^q \\
r_{t+1} &= \rho_r r_t + \sigma_r \varepsilon_{t+1}^r \\
\delta_{t+1} &= \rho_\delta \delta_t + \sigma_\delta \varepsilon_{t+1}^\delta
\end{aligned} \tag{17}$$

Optimal consumption with a borrowing constraint

Back Large scale model

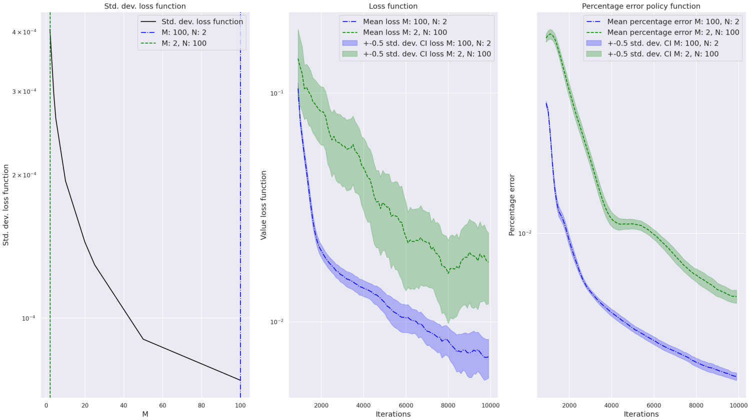
Figure: **bc-MC estimator** (left) and **Time Iteration** (right)



Optimal consumption with a borrowing constraint

Back

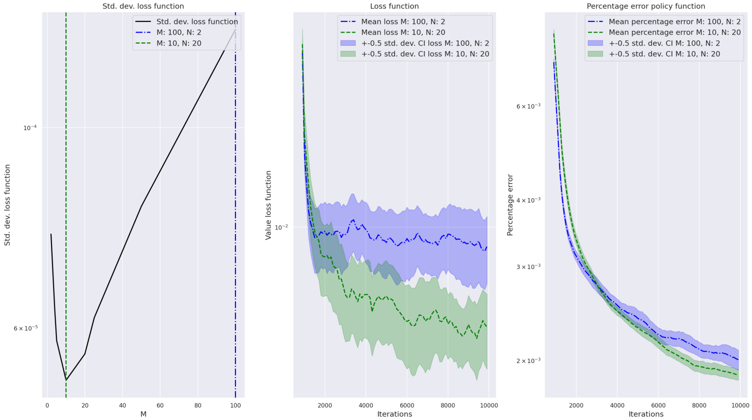
Figure: Model with a borrowing constraint ($b = 0$) solved with the bc-MC estimator



Optimal consumption with a borrowing constraint

Back

Figure: Model with a borrowing constraint ($b = 1$) solved with the bc-MC estimator



Proposition Back

Define $T \equiv \frac{MN}{2}$, $2T$ is the number of function calls $f(.)$ within the loss function

1. $\mathrm{Var}(\mathcal{L}^U_{M,N}(\theta))$ is proportional to $\frac{1}{T}$
2. If $f(s_m, \varepsilon_m|\theta) = f(\varepsilon_m|\theta), \quad \forall s \in S$ ($\approx$ high-variance model):

$$\mathrm{Var}(\mathcal{L}^U_{M,N}(\theta)) = \frac{1}{T(N-1)} \mathrm{Var}\left(f(s_m, \varepsilon^1_m|\theta)\right)^2 + \frac{2}{T} \mathbb{E}\left[f(s_m, \varepsilon^1_m|\theta)\right]^2 \mathrm{Var}\left(f(s_m, \varepsilon^1_m|\theta)\right) \tag{18}$$

3. If $f(s_m, \varepsilon_m|\theta) = f(s_m|\theta), \quad \forall \varepsilon_m \in \mathcal{E}$ ($\approx$ highly non-linear model):

$$\mathrm{Var}(\mathcal{L}^U_{M,N}(\theta)) = \frac{1}{M}\left[\mathrm{Var}\left(f(s_m, \varepsilon^1_m|\theta)^2\right)\right] \tag{19}$$

Proposition Back

Define $T \equiv \frac{MN}{2}$, $2T$ is the number of function calls $f(.)$ within the loss function

1. $\text{Var}(\mathcal{L}^U_{M,N}(\theta))$ is proportional to $\frac{1}{T}$
2. If $f(s_m, \varepsilon_m|\theta) = f(\varepsilon_m|\theta), \quad \forall s \in S$ ($\approx$ high-variance model):

$$\text{Var}(\mathcal{L}^U_{M,N}(\theta)) = \frac{1}{T(N-1)} \text{Var}\left(f(s_m, \varepsilon^1_m|\theta)\right)^2 + \frac{2}{T} \mathbb{E}\left[f(s_m, \varepsilon^1_m|\theta)\right]^2 \text{Var}\left(f(s_m, \varepsilon^1_m|\theta)\right) \tag{18}$$

3. If $f(s_m, \varepsilon_m|\theta) = f(s_m|\theta), \quad \forall \varepsilon_m \in \mathcal{E}$ ($\approx$ highly non-linear model):

$$\text{Var}(\mathcal{L}^U_{M,N}(\theta)) = \frac{1}{M}\left[\text{Var}\left(f(s_m, \varepsilon^1_m|\theta)^2\right)\right] \tag{19}$$

Proposition Back

Define $T \equiv \frac{MN}{2}$, $2T$ is the number of function calls $f(.)$ within the loss function

1. $\mathrm{Var}(\mathcal{L}^U_{M,N}(\theta))$ is proportional to $\frac{1}{T}$
2. If $f(s_m, \varepsilon_m|\theta) = f(\varepsilon_m|\theta), \quad \forall s \in S$ ($\approx$ high-variance model):

$$\mathrm{Var}(\mathcal{L}^U_{M,N}(\theta)) = \frac{1}{T(N-1)} \mathrm{Var}\left(f(s_m, \varepsilon^1_m|\theta)\right)^2 + \frac{2}{T} \mathbb{E}\left[f(s_m, \varepsilon^1_m|\theta)\right]^2 \mathrm{Var}\left(f(s_m, \varepsilon^1_m|\theta)\right) \tag{18}$$

3. If $f(s_m, \varepsilon_m|\theta) = f(s_m|\theta), \quad \forall \varepsilon_m \in \mathcal{E}$ ($\approx$ highly non-linear model):

$$\mathrm{Var}(\mathcal{L}^U_{M,N}(\theta)) = \frac{1}{M}\left[\mathrm{Var}\left(f(s_m, \varepsilon^1_m|\theta)^2\right)\right] \tag{19}$$

Proposition Back

Define $T \equiv \frac{MN}{2}$, $2T$ is the number of function calls $f(.)$ within the loss function

1. $\text{Var}(\mathcal{L}^U_{M,N}(\theta))$ is proportional to $\frac{1}{T}$
2. If $f(s_m, \varepsilon_m|\theta) = f(\varepsilon_m|\theta), \quad \forall s \in S$ ($\approx$ high-variance model):

$$\text{Var}(\mathcal{L}^U_{M,N}(\theta)) = \frac{1}{T(N-1)} \text{Var}\left(f(s_m, \varepsilon^1_m|\theta)\right)^2 + \frac{2}{T}\mathbb{E}\left[f(s_m, \varepsilon^1_m|\theta)\right]^2 \text{Var}\left(f(s_m, \varepsilon^1_m|\theta)\right) \tag{18}$$

3. If $f(s_m, \varepsilon_m|\theta) = f(s_m|\theta), \quad \forall \varepsilon_m \in \mathcal{E}$ ($\approx$ highly non-linear model):

$$\text{Var}(\mathcal{L}^U_{M,N}(\theta)) = \frac{1}{M}\left[\text{Var}\left(f(s_m, \varepsilon^1_m|\theta)^2\right)\right] \tag{19}$$