Discussion: *Estimating Nonlinear Heterogeneous Agents Models with Neural Networks* (Kase, Melosi, Rottner)

Julien Pascal

Central Bank of Luxembourg

28 November 2024

This presentation should not be reported as representing the views of the BCL or the Eurosystem. The views expressed are those of the author and may not be shared by other research staff or policymakers in the BCL or the Eurosystem.

Table of Contents

Review: introduction

- Ambitious goal: **Global solution** and **estimation** of **HANK** + **ZLB** models → technical *tour de force*.
- HANK: Oh and Reis (2012), McKay and Reis (2016), Ravn and Sterk (2017), Challe et al. (2017), Kaplan, Moll, and Violante (2018), Bilbiie (2020), Auclert, Rognlie, and Straub (2024), Bilbiie (2024).
- ZLB: Benhabib, Schmitt-Grohé, and Uribe (2002), Eggertsson et al. (2003), McKay, Nakamura, and Steinsson (2016), Michaillat and Saez (2021).
- Neural Networks: Duffy and McNelis (2001), L. Maliar, S. Maliar, and Winant (2021), Azinovic, Gaegauf, and Scheidegger (2022), Fernández-Villaverde, Hurtado, and Nuno (2023), Folini et al. (2024), V. Duarte, D. Duarte, and Silva (2024), Valaitis and Villa (2024), Pascal (2024).
- Main contribution: the **extended state vector** to solve the "curse of the nested loop" when structural estimation of HANK model. Similar approaches: Norets (2012), Scheidegger and Bilionis (2019).

Review: introduction

- Ambitious goal: **Global solution** and **estimation** of **HANK** + **ZLB** models $\rightarrow$ technical *tour de force*.
- HANK: Oh and Reis (2012), McKay and Reis (2016), Ravn and Sterk (2017), Challe et al. (2017), Kaplan, Moll, and Violante (2018), Bilbiie (2020), Auclert, Rognlie, and Straub (2024), Bilbiie (2024).
- ZLB: Benhabib, Schmitt-Grohé, and Uribe (2002), Eggertsson et al. (2003), McKay, Nakamura, and Steinsson (2016), Michaillat and Saez (2021).
- Neural Networks: Duffy and McNelis (2001), L. Maliar, S. Maliar, and Winant (2021), Azinovic, Gaegauf, and Scheidegger (2022), Fernández-Villaverde, Hurtado, and Nuno (2023), Folini et al. (2024), V. Duarte, D. Duarte, and Silva (2024), Valaitis and Villa (2024), Pascal (2024).
- Main contribution: the **extended state vector** to solve the "curse of the nested loop" when structural estimation of HANK model. Similar approaches: Norets (2012), Scheidegger and Bilionis (2019).

Review: introduction

- Ambitious goal: **Global solution** and **estimation** of **HANK** + **ZLB** models → technical *tour de force*.
- HANK: Oh and Reis (2012), McKay and Reis (2016), Ravn and Sterk (2017), Challe et al. (2017), Kaplan, Moll, and Violante (2018), Bilbiie (2020), Auclert, Rognlie, and Straub (2024), Bilbiie (2024).
- ZLB: Benhabib, Schmitt-Grohé, and Uribe (2002), Eggertsson et al. (2003), McKay, Nakamura, and Steinsson (2016), Michaillat and Saez (2021).
- Neural Networks: Duffy and McNelis (2001), L. Maliar, S. Maliar, and Winant (2021), Azinovic, Gaegauf, and Scheidegger (2022), Fernández-Villaverde, Hurtado, and Nuno (2023), Folini et al. (2024), V. Duarte, D. Duarte, and Silva (2024), Valaitis and Villa (2024), Pascal (2024).
- Main contribution: the **extended state vector** to solve the "curse of the nested loop" when structural estimation of HANK model. Similar approaches: Norets (2012), Scheidegger and Bilionis (2019).

Review: introduction

- Ambitious goal: **Global solution** and **estimation** of **HANK** + **ZLB** models → technical *tour de force*.
- HANK: Oh and Reis (2012), McKay and Reis (2016), Ravn and Sterk (2017), Challe et al. (2017), Kaplan, Moll, and Violante (2018), Bilbiie (2020), Auclert, Rognlie, and Straub (2024), Bilbiie (2024).
- ZLB: Benhabib, Schmitt-Grohé, and Uribe (2002), Eggertsson et al. (2003), McKay, Nakamura, and Steinsson (2016), Michaillat and Saez (2021).
- Neural Networks: Duffy and McNelis (2001), L. Maliar, S. Maliar, and Winant (2021), Azinovic, Gaegauf, and Scheidegger (2022), Fernández-Villaverde, Hurtado, and Nuno (2023), Folini et al. (2024), V. Duarte, D. Duarte, and Silva (2024), Valaitis and Villa (2024), Pascal (2024).
- Main contribution: the **extended state vector** to solve the "curse of the nested loop" when structural estimation of HANK model. Similar approaches: Norets (2012), Scheidegger and Bilionis (2019).

Review: introduction

- Ambitious goal: **Global solution** and **estimation** of **HANK** + **ZLB** models → technical *tour de force*.
- HANK: Oh and Reis (2012), McKay and Reis (2016), Ravn and Sterk (2017), Challe et al. (2017), Kaplan, Moll, and Violante (2018), Bilbiie (2020), Auclert, Rognlie, and Straub (2024), Bilbiie (2024).
- ZLB: Benhabib, Schmitt-Grohé, and Uribe (2002), Eggertsson et al. (2003), McKay, Nakamura, and Steinsson (2016), Michaillat and Saez (2021).
- Neural Networks: Duffy and McNelis (2001), L. Maliar, S. Maliar, and Winant (2021), Azinovic, Gaegauf, and Scheidegger (2022), Fernández-Villaverde, Hurtado, and Nuno (2023), Folini et al. (2024), V. Duarte, D. Duarte, and Silva (2024), Valaitis and Villa (2024), Pascal (2024).
- Main contribution: the **extended state vector** to solve the “curse of the nested loop” when structural estimation of HANK model. Similar approaches: Norets (2012), Scheidegger and Bilionis (2019).

Review: the "curse of the nested loop"

Estimation: standard approach

$$p^* = \arg\max_{p \in \Omega} \text{Likelihood}(p, \text{data})$$

with policies $\approx NN(s|\theta)$.

```
#I. Minimization ("outside loop")
p_star = maximize(likelihood)
#II. Likelihood evaluation ("inside loop")
def likelihood(p, data):
    """Return the value of likelihood"""
    #1. Solution step: solve the model, conditional on parameter p (and data)
    # Stochastic gradient descent to find NN parameter theta*
    for i in range(I):
        theta += - l*Gradient_Loss(theta, p)
    #2. Evaluation step: simulate the model conditional on theta*, calculate the likelihood
    return likelihood_value
```

Algorithm: Pseudo-code for the standard estimation approach

Review: the "curse of the nested loop"

Estimation: **extended state vector** approach

Calculate policies $\approx NN(s, p|\theta)$.

```
# Solve model (SGD), using (s,p) as the state vector:
for i in range(I):
    # random draws of parameter vector
    p = random.rand()
    # SGD step:
    theta += - l*Gradient_Loss(theta, p)
```

Algorithm: Pseudo-code for the extended state vector approach

Output θ^*: policy functions for **all parameter values**. Estimation is then "easy":

$$p^* = \arg\max_{p\in\Omega} \text{Likelihood}(p, NN(s,p|\theta^*), \text{data})$$

Extra step. Approximate the likelihood with the "NN particle filter":

$$NN_L(p, data|\theta_L) \approx \text{Likelihood}(p, NN(s,p|\theta^*), \text{data})$$

Review: the "curse of the nested loop"

Estimation: **extended state vector** approach

Calculate policies $\approx NN(s, p|\theta)$.

```
# Solve model (SGD), using (s,p) as the state vector:
for i in range(I):
    # random draws of parameter vector
    p = random.rand()
    # SGD step:
    theta += - l*Gradient_Loss(theta, p)
```

Algorithm: Pseudo-code for the extended state vector approach

Output θ^*: policy functions for **all parameter values**. Estimation is then "easy":

$$p^* = \arg\max_{p\in\Omega} \text{Likelihood}(p, NN(s, p|\theta^*), \text{data})$$

Extra step. Approximate the likelihood with the "NN particle filter":

$$NN_L(p, data|\theta_L) \approx \text{Likelihood}(p, NN(s, p|\theta^*), \text{data})$$

Review: the "curse of the nested loop"

Estimation: **extended state vector** approach

Calculate policies $\approx NN(s, p|\theta)$.

```
# Solve model (SGD), using (s,p) as the state vector:
for i in range(I):
    # random draws of parameter vector
    p = random.rand()
    # SGD step:
    theta += - l*Gradient_Loss(theta, p)
```

Algorithm: Pseudo-code for the extended state vector approach

Output θ^*: policy functions for **all parameter values**. Estimation is then "easy":

$$p^* = \arg\max_{p \in \Omega} \text{Likelihood}(p, NN(s, p|\theta^*), \text{data})$$

Extra step. Approximate the likelihood with the "NN particle filter":

$$NN_L(p, data|\theta_L) \approx \text{Likelihood}(p, NN(s, p|\theta^*), \text{data})$$

Major comments: separation between solving and estimating models

Sampling strategy

Uniform sampling when "training" likelihood $NN_L(p, \text{data}|\theta_L)$ (or policies $\approx NN(s, p|\theta)$).

Ideal sampling strategy

1. draw more in the direction of the maximizer p^* ,
2. draw more where functions are "unknown"

Literature: surrogate model optimization

Algorithms that balance the **explore-exploit trade-off**. Sampling the most unknown region *and* sampling in minimizing region:
Expected improvement (EI) criterion (Jones, Schonlau, and Welch, 1998), Stochastic RBF (Rommel G Regis, 2011) Lower confidence-bound (LCB) strategy (Srinivas et al., 2012), Dynamic coordinate search (DYCORS) (Rommel G. Regis and Shoemaker, 2013).

Major comments: separation between solving and estimating models

Sampling strategy

Uniform sampling when "training" likelihood $NN_L(p, \text{data}|\theta_L)$ (or policies $\approx NN(s, p|\theta)$).

Ideal sampling strategy

1. draw more in the direction of the maximizer p^* ,
2. draw more where functions are "unknown"

Literature: surrogate model optimization

Algorithms that balance the **explore-exploit trade-off**. Sampling the most unknown region *and* sampling in minimizing region:
Expected improvement (EI) criterion (Jones, Schonlau, and Welch, 1998), Stochastic RBF (Rommel G Regis, 2011) Lower confidence-bound (LCB) strategy (Srinivas et al., 2012), Dynamic coordinate search (DYCORS) (Rommel G. Regis and Shoemaker, 2013).

Major comments: separation between solving and estimating models

Sampling strategy

Uniform sampling when "training" likelihood $NN_L(p, \text{data}|\theta_L)$ (or policies $\approx NN(s, p|\theta)$).

Ideal sampling strategy

1. draw more in the direction of the maximizer p^* ,
2. draw more where functions are "unknown"

Literature: surrogate model optimization

Algorithms that balance the **explore-exploit trade-off**. Sampling the most unknown region *and* sampling in minimizing region:
Expected improvement (EI) criterion (Jones, Schonlau, and Welch, 1998), Stochastic RBF (Rommel G Regis, 2011) Lower confidence-bound (LCB) strategy (Srinivas et al., 2012), Dynamic coordinate search (DYCORS) (Rommel G. Regis and Shoemaker, 2013).

Major comments: econometric results

- **Linearized RANK $\approx$ Global HANK**: "[...] heterogeneity and non-linearities do not lead to substantial revision to the estimated value of those parameters." (p. 29).
- "The match is somewhat unsatisfactory" (p. 31). Similar to Acharya et al. (2023). RANK $\subseteq$ HANK. **Why** worse fit?
- What about identification? Use of **cross-sectional data** to identify some parameters.

Major comments: econometric results

- **Linearized RANK ≈ Global HANK**: "[...] heterogeneity and non-linearities do not lead to substantial revision to the estimated value of those parameters." (p. 29).
- "The match is somewhat unsatisfactory" (p. 31). Similar to Acharya et al. (2023). RANK $\subseteq$ HANK. **Why** worse fit?
- What about identification? Use of **cross-sectional data** to identify some parameters.

Major comments: econometric results

- **Linearized RANK $\approx$ Global HANK**: "[...] heterogeneity and non-linearities do not lead to substantial revision to the estimated value of those parameters." (p. 29).
- "The match is somewhat unsatisfactory" (p. 31). Similar to Acharya et al. (2023). RANK $\subseteq$ HANK. **Why** worse fit?
- What about identification? Use of **cross-sectional data** to identify some parameters.

Major comments: econometric results

- **Linearized RANK $\approx$ Global HANK**: “[...] heterogeneity and non-linearities do not lead to substantial revision to the estimated value of those parameters.” (p. 29).
- “The match is somewhat unsatisfactory” (p. 31). Similar to Acharya et al. (2023). RANK $\subseteq$ HANK. **Why** worse fit?
- What about identification? Use of **cross-sectional data** to identify some parameters.

Minor comments

Too many NNs?

1. NN for aggregate variable,
2. NN for deterministic steady-state,
3. NN for likelihood.

Compounding approximations. What happens to approximation errors?

- Show Euler equation errors, not just value of the loss.

Monte-Carlo integration

- Monte-Carlo integration (antithetic variates) to approximate expectation w.r.t. next period's innovation + L-2 norm → **bias**, because $\left(\frac{1}{N}\sum_{i=1}^{N} x_i\right)^2$ **biased estimator** of $\mathbb{E}(x)^2$.
- Use "all-in-one" operator (L. Maliar, S. Maliar, and Winant, 2021) or "bias-corrected Monte Carlo" operator (Pascal, 2024).

Minor comments

Too many NNs?

1. NN for aggregate variable,
2. NN for deterministic steady-state,
3. NN for likelihood.

Compounding approximations. What happens to approximation errors?

- Show Euler equation errors, not just value of the loss.

Monte-Carlo integration

- Monte-Carlo integration (antithetic variates) to approximate expectation w.r.t. next period's innovation + L-2 norm → **bias**, because $\left(\frac{1}{N}\sum_{i=1}^{N} x_i\right)^2$ **biased estimator** of $\mathbb{E}(x)^2$.
- Use “all-in-one” operator (L. Maliar, S. Maliar, and Winant, 2021) or “bias-corrected Monte Carlo” operator (Pascal, 2024).

Minor comments

Too many NNs?

1. NN for aggregate variable,
2. NN for deterministic steady-state,
3. NN for likelihood.

Compounding approximations. What happens to approximation errors?

- Show Euler equation errors, not just value of the loss.

Monte-Carlo integration

- Monte-Carlo integration (antithetic variates) to approximate expectation w.r.t. next period's innovation + L-2 norm → **bias**, because $\left(\frac{1}{N}\sum_{i=1}^{N} x_i\right)^2$ **biased estimator** of $\mathbb{E}(x)^2$.
- Use "all-in-one" operator (L. Maliar, S. Maliar, and Winant, 2021) or "bias-corrected Monte Carlo" operator (Pascal, 2024).

Minor comments

Too many NNs?

1. NN for aggregate variable,
2. NN for deterministic steady-state,
3. NN for likelihood.

Compounding approximations. What happens to approximation errors?

- Show Euler equation errors, not just value of the loss.

Monte-Carlo integration

- Monte-Carlo integration (antithetic variates) to approximate expectation w.r.t. next period's innovation + L-2 norm → **bias**, because $\left(\frac{1}{N}\sum_{i=1}^{N} x_i\right)^2$ **biased estimator** of $\mathbb{E}(x)^2$.
- Use "all-in-one" operator (L. Maliar, S. Maliar, and Winant, 2021) or "bias-corrected Monte Carlo" operator (Pascal, 2024).

Minor comments

Extra noise

- Finite number of agents. Why not use a histogram (Young, 2010)?

Proofs of concept

- POC1: 3-equation NK model. Why log-linearization?
- POC2-3: correctly-centered truncated-Gaussian priors. How much Bayesian updating?
- Comparison with other methods (time-accuracy trade-offs)?

Minor comments

Extra noise

- Finite number of agents. Why not use a histogram (Young, 2010)?

Proofs of concept

- POC1: 3-equation NK model. Why log-linearization?
- POC2-3: correctly-centered truncated-Gaussian priors. How much Bayesian updating?
- Comparison with other methods (time-accuracy trade-offs)?

Minor comments

Extra noise

- Finite number of agents. Why not use a histogram (Young, 2010)?

Proofs of concept

- POC1: 3-equation NK model. Why log-linearization?
- POC2-3: correctly-centered truncated-Gaussian priors. How much Bayesian updating?
- Comparison with other methods (time-accuracy trade-offs)?

Minor comments

Extra noise

- Finite number of agents. Why not use a histogram (Young, 2010)?

Proofs of concept

- POC1: 3-equation NK model. Why log-linearization?
- POC2-3: correctly-centered truncated-Gaussian priors. How much Bayesian updating?
- Comparison with other methods (time-accuracy trade-offs)?

Conclusion

- Technically impressive. Key idea: **pseudo-state vector**, combined with **neural network(s)**.
- **New questions** now answerable.

Conclusion

- Technically impressive. Key idea: **pseudo-state vector**, combined with **neural network(s)**.
- **New questions** now answerable.

References I

- Acharya, Sushant et al. (2023). "Estimating HANK for central banks." In: *FRB of New York Staff Report* 1071.
- Auclert, Adrien, Matthew Rognlie, and Ludwig Straub (2024). "The Intertemporal Keynesian Cross." In: *Journal of Political Economy, forthcoming.* DOI: 10.1086/732531.
- Azinovic, Marlon, Luca Gaegauf, and Simon Scheidegger (2022). "Deep equilibrium nets." In: *International Economic Review.*
- Benhabib, Jess, Stephanie Schmitt-Grohé, and Martin Uribe (2002). "Avoiding liquidity traps." In: *Journal of Political Economy* 110.3, pp. 535–563.
- Bilbiie, Florin O (2020). "The new Keynesian cross." In: *Journal of Monetary Economics* 114, pp. 90–108.
- — (2024). "Monetary policy and heterogeneity: An analytical framework." In: *Review of Economic Studies*, rdae066.
- Challe, Edouard et al. (2017). "Precautionary saving and aggregate demand." In: *Quantitative Economics* 8.2, pp. 435–478.

References II

Duarte, Victor, Diogo Duarte, and Dejanir H Silva (Sept. 2024). "Machine Learning for Continuous-Time Finance." In: *The Review of Financial Studies* 37.11, pp. 3217–3271. ISSN: 0893-9454. DOI: 10.1093/rfs/hhae043. eprint: https://academic.oup.com/rfs/article-pdf/37/11/3217/59762832/hhae043.pdf. URL: https://doi.org/10.1093/rfs/hhae043.

Duffy, John and Paul D McNelis (2001). "Approximating and simulating the stochastic growth model: Parameterized expectations, neural networks, and the genetic algorithm." In: *Journal of Economic Dynamics and Control* 25.9, pp. 1273–1303.

Eggertsson, Gauti B et al. (2003). "Zero bound on interest rates and optimal monetary policy." In: *Brookings Papers on Economic Activity* 2003.1, pp. 139–233.

Fernández-Villaverde, Jesús, Samuel Hurtado, and Galo Nuno (2023). "Financial frictions and the wealth distribution." In: *Econometrica* 91.3, pp. 869–901.

References III

- Folini, Doris et al. (Jan. 2024). "The Climate in Climate Economics." In: *The Review of Economic Studies*, rdae011. ISSN: 0034-6527. DOI: 10.1093/restud/rdae011. eprint: https://academic.oup.com/restud/advance-article-pdf/doi/10.1093/restud/rdae011/56663801/rdae011.pdf. URL: https://doi.org/10.1093/restud/rdae011.
- Jones, Donald R, Matthias Schonlau, and William J Welch (1998). "Efficient global optimization of expensive black-box functions." In: *Journal of Global optimization* 13, pp. 455–492.
- Kaplan, Greg, Benjamin Moll, and Giovanni L Violante (2018). "Monetary policy according to HANK." In: *American Economic Review* 108.3, pp. 697–743.
- Maliar, Lilia, Serguei Maliar, and Pablo Winant (2021). "Deep learning for solving dynamic economic models.." In: *Journal of Monetary Economics* 122, pp. 76–101.
- McKay, Alisdair, Emi Nakamura, and Jón Steinsson (2016). "The power of forward guidance revisited." In: *American Economic Review* 106.10, pp. 3133–3158.
- McKay, Alisdair and Ricardo Reis (2016). "The role of automatic stabilizers in the US business cycle." In: *Econometrica* 84.1, pp. 141–194.

References IV

Michaillat, Pascal and Emmanuel Saez (2021). "Resolving New Keynesian anomalies with wealth in the utility function." In: *Review of Economics and Statistics* 103.2, pp. 197–215.

Norets, Andriy (2012). "Estimation of dynamic discrete choice models using artificial neural network approximations." In: *Econometric Reviews* 31.1, pp. 84–106.

Oh, Hyunseung and Ricardo Reis (2012). "Targeted transfers and the fiscal response to the great recession." In: *Journal of Monetary Economics* 59, S50–S64.

Pascal, Julien (2024). "Artificial neural networks to solve dynamic programming problems: A bias-corrected Monte Carlo operator." In: *Journal of Economic Dynamics and Control* 162, p. 104853.

Ravn, Morten O and Vincent Sterk (2017). "Job uncertainty and deep recessions." In: *Journal of Monetary Economics* 90, pp. 125–141.

Regis, Rommel G (2011). "Stochastic radial basis function algorithms for large-scale optimization involving expensive black-box objective and constraint functions." In: *Computers & Operations Research* 38.5, pp. 837–853.

References V

- Regis, Rommel G. and Christine A. Shoemaker (2013). "Combining radial basis function surrogates and dynamic coordinate search in high-dimensional expensive black-box optimization." In: *Engineering Optimization* 45.5, pp. 529–555. DOI: 10.1080/0305215X.2012.687731.
- Scheidegger, Simon and Ilias Bilionis (2019). "Machine learning for high-dimensional dynamic stochastic economies." In: *Journal of Computational Science* 33, pp. 68–82.
- Srinivas, Niranjan et al. (2012). "Information-theoretic regret bounds for gaussian process optimization in the bandit setting." In: *IEEE transactions on information theory* 58.5, pp. 3250–3265.
- Valaitis, Vytautas and Alessandro T Villa (2024). "A machine learning projection method for macro-finance models." In: *Quantitative economics* 15.1, pp. 145–173.
- Young, Eric R (2010). "Solving the incomplete markets model with aggregate uncertainty using the Krusell–Smith algorithm and non-stochastic simulations." In: *Journal of Economic Dynamics and Control* 34.1, pp. 36–41.