Learning and Subjective Expectation Formation: A Recurrent Neural Network Approach

Chenyu (Sev) Hou

Vancouver School of Economics
University of British Columbia

November 6, 2020

Expectation Formation in Macroeconomic Context

Ask yourself a question:

How do you think the unemployment rate is going to change in the coming quarter?

- Personal Experience (Top 3 information source for 36.5% households);
- Media and News (Top 3 information source for 49.2% households);
- Social Connections (Top 3 information source for 52.3% households);

Agents use various sources of information to form expectation.

Expectation Formation in Macroeconomic Context

Ask yourself a question:

How do you think the unemployment rate is going to change in the coming quarter?

- Personal Experience (Top 3 information source for 36.5% households);
- Media and News (Top 3 information source for 49.2% households);
- Social Connections (Top 3 information source for 52.3% households);

Agents use various sources of information to form expectation.

Questions

But how?

- **What's the functional form of agent's expectation formation model?**
- **How do signals on past and future states about macroeconomy affect household's expectation?**

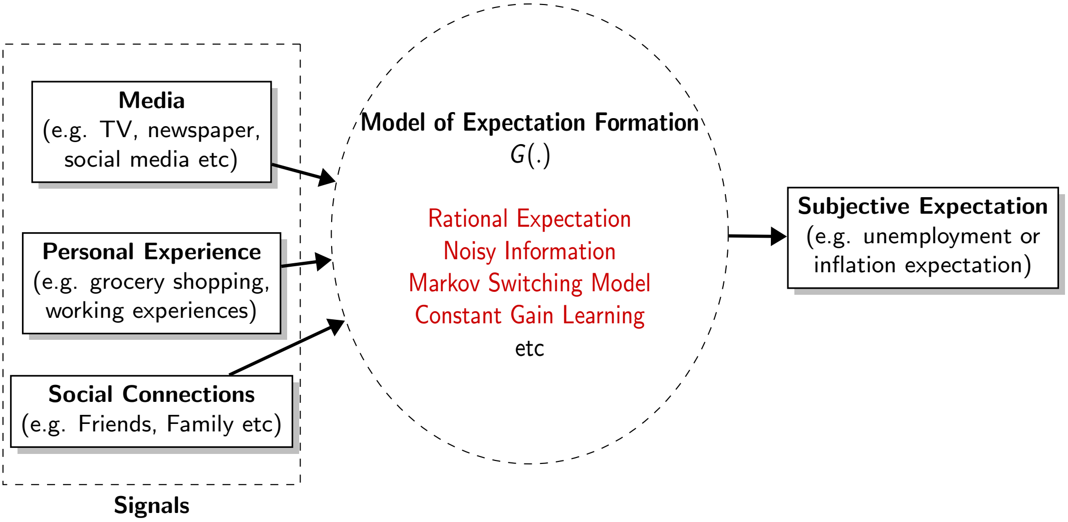
What's New in This Paper?

1. New Method:
   - **Generic Learning Framework: nests most of macroeconomic expectation formation models.**
   - **Flexible non-parametric method: Recurrent Neural Network (RNN).**
   - **DML approach for inference.**
2. New Empirical Findings:
   - Non-linear and asymmetric expectation formation;
   - Attention-shift along Business Cycle;
   - Cause of Attention-shift: signals on unemployment and GDP growth.
3. (Not so new) Model for Explanation:
   - Rational Inattention with Endogenous Value of Information

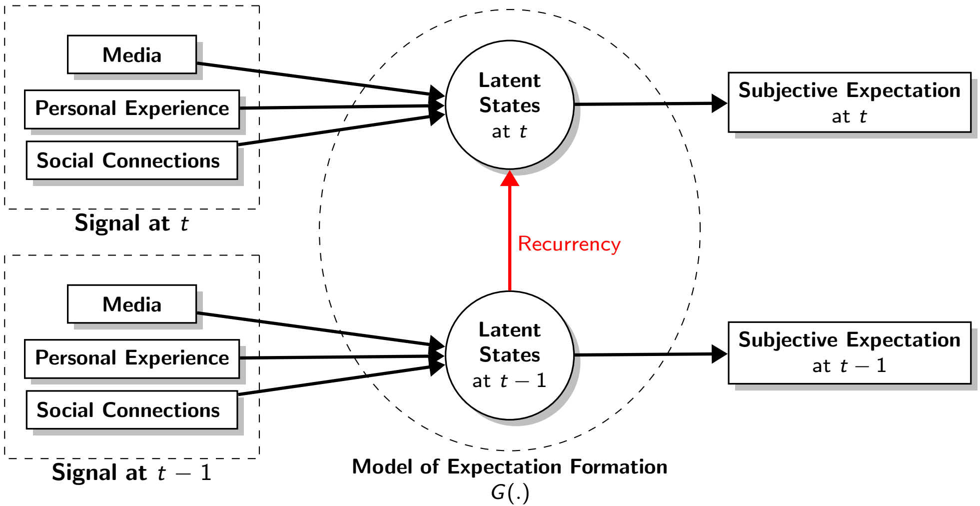
What's New in This Paper?

1. New Method:
    - **Generic Learning Framework: nests most of macroeconomic expectation formation models.**
    - **Flexible non-parametric method: Recurrent Neural Network (RNN).**
    - **DML approach for inference.**
2. New Empirical Findings:
    - **Non-linear and asymmetric expectation formation;**
    - **Attention-shift along Business Cycle;**
    - **Cause of Attention-shift: signals on unemployment and GDP growth.**
3. (Not so new) Model for Explanation:
    - **Rational Inattention with Endogenous Value of Information**

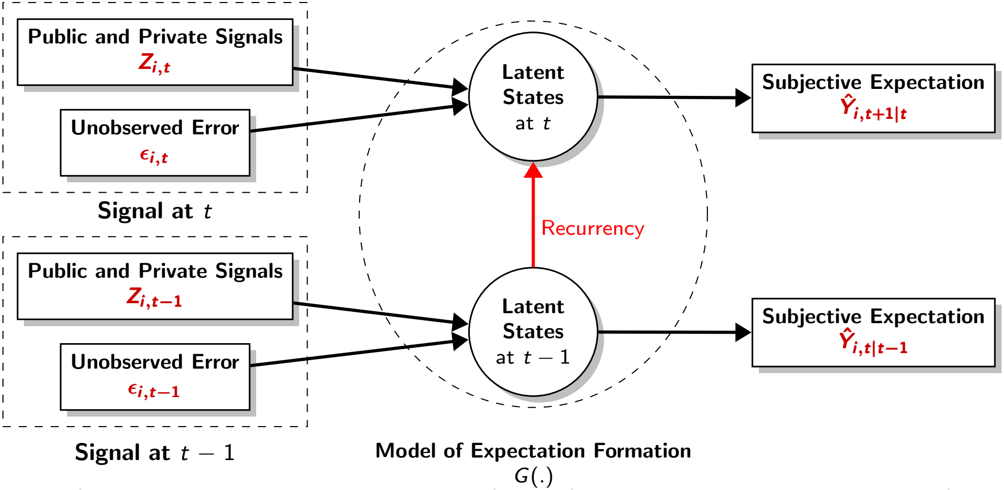
Expectation Formation Models in Macroeconomics



Expectation Formation Models in Macroeconomics: Dynamic Structure

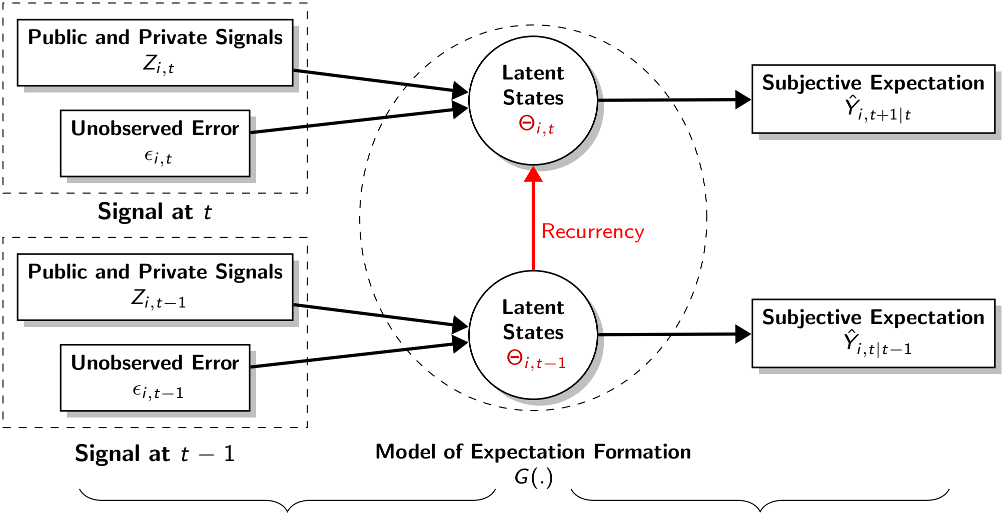


Agent's Problem: Dynamic Structure Formalized by Generic Learning Framework



Updating: $\Theta_{i,t} = H(\Theta_{i,t-1}, Z_{i,t}, \epsilon_{i,t})$ **Forecasting:** $\hat{Y}_{i,t+1|t} = F(\Theta_{i,t})$

Agent's Problem: Dynamic Structure Formalized by Generic Learning Framework



Updating: $\Theta_{i,t} = H(\Theta_{i,t-1}, Z_{i,t}, \epsilon_{i,t})$ **Forecasting:** $\hat{Y}_{i,t+1|t} = F(\Theta_{i,t})$

Flexibility:
Generic Learning Framework nests many Learning Models

$$\begin{aligned} \hat{Y}_{i,t+1|t} &:= G(Z_{i,t}, \epsilon_{i,t} \ldots) \\ &= F(\Theta_{i,t}) \\ &= F(H(\Theta_{i,t-1}, Z_{i,t}, \epsilon_{i,t})) \end{aligned} \tag{1}$$

Model	Θ	$F(.)$ and $H(.)$
Noisy Information Model (Linear Kalman Filter)	"now-cast"	linear functions implied by linear State Space Model
Constant Gain Learning	learned weighting matrix and learned parameters	non-linear functions implied by recursive least squares
Markov Switching Model	posterior beliefs about Markovian State	non-linear functions implied by Bayesian Rule

Econometrician's Information and Goal

Information set:

- Observe: set of signals, $Z_{i,t}$
- Do not observe: $\Theta_{i,t}$, $\epsilon_{i,t}$, dimensionality of $\Theta_{i,t}$, functional form of $F(.)$ and $H(.)$.

Goal: Given the information set, approximate Average Structural Function:

$$\begin{aligned}\mathbb{E}[\hat{Y}_{i,t+1|t}|\{Z_{i,\tau}\}_{\tau=0}^{t}] &\equiv \mathbb{E}[G(\{Z_{i,\tau},\epsilon_{i,\tau}\}_{\tau=0}^{t})|\{Z_{i,\tau}\}_{\tau=0}^{t}] \\ &= g(\{Z_{i,\tau}\}_{\tau=0}^{t})\end{aligned} \tag{2}$$

Approximating $g(.)$ with Recurrent Neural Network

Theoretically:

- Exist sufficient statistics $\theta_{i,t}$ for $\Theta_{i,t}$ such that:

$$\begin{aligned} g(\{Z_{i,\tau}\}_{\tau=0}^{t}) &= f(\theta_{i,t}) \\ \theta_{i,t} &= h(\theta_{i,t-1}, Z_{i,t}) \end{aligned} \tag{3}$$

- Recurrent Neural Networks are Universal Approximators for Dynamic System (3) (*Shaffer and Zimmermman 2006*);

$$\hat{g}_{rnn} := \arg\min_{g_w \in \mathcal{G}_{f\circ h}^{RNN}} \sum_{i,t} \frac{1}{2}(\hat{Y}_{i,t+1|t} - g_w(\{Z_{i,\tau}\}_{\tau=0}^{t}))^2$$

Simple RNN

Empirically, RNN recovers correct: KF example

- Functional form of $g(.)$;
- Dynamic structure with latent states θ.

Average Marginal Effect and Inference: Double Machine Learning Method

Average Marginal Effect/Derivative:

$$\beta^j = \mathbb{E}[\frac{\partial g}{\partial z^j_{i,t}}]$$

- Plug-in estimator is biased, inference not available (Chernozhukov et al. 2018);
  - Bias induced by over-fitting and regularization;
  - Slow convergence speed (slower than $\sqrt{n}$);
- (Near Neyman) Orthogonalized moment condition;

$$\mathbb{E}[\psi(W,\beta,\eta)] = \mathbb{E}[\psi^a(W,\eta)\beta + \psi^b(W,\eta)]$$
$$= \mathbb{E}[\beta^j - \frac{\partial g}{\partial z^j_{i,t}} + \frac{\partial ln(P(\{Z_{i,\tau}\}^t_{\tau=0}))}{\partial z^j_{i,t}}(Y_{i,t+1|t} - g(\{Z_{i,\tau}\}^t_{\tau=0}))] = 0$$

  - Less sensitivity to quality of functional estimator;
  - Involve extra nuisance parameter to be estimated (density function);
  - Speed requirement satisfied (Farrell et. al. 2020)

Average Marginal Effect and Inference: Double Machine Learning Method

Average Marginal Effect/Derivative:

$$\beta^j = \mathbb{E}[\frac{\partial g}{\partial z^j_{i,t}}]$$

- Plug-in estimator is biased, inference not available (Chernozhukov et al. 2018);
  - Bias induced by over-fitting and regularization;
  - Slow convergence speed (slower than $\sqrt{n}$);
- (Near Neyman) Orthogonalized moment condition;

$$\mathbb{E}[\psi(W,\beta,\eta)] = \mathbb{E}[\psi^a(W,\eta)\beta + \psi^b(W,\eta)]$$
$$= \mathbb{E}[\beta^j - \frac{\partial g}{\partial z^j_{i,t}} + \frac{\partial ln(P(\{Z_{i,\tau}\}^t_{\tau=0}))}{\partial z^j_{i,t}}(Y_{i,t+1|t} - g(\{Z_{i,\tau}\}^t_{\tau=0}))] = 0$$

  - Less sensitivity to quality of functional estimator;
  - Involve extra nuisance parameter to be estimated (density function);
  - Speed requirement satisfied (Farrell et. al. 2020)

Application

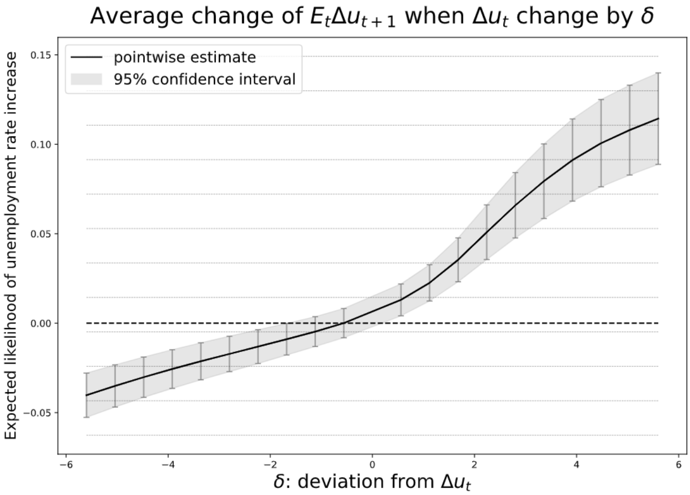
Data Description

- 27 signals:
  - Current signals: realized change of unemployment rate, real GDP growth, inflation etc.
  - Future signals: SPF about change of unemployment rate etc.
  - Local/individual signals;
  - News exposure;
- Expectations: on unemployment, inflation, interest rate and economic condition, from MSC.
- Synthetic panel quarterly 1988q1 to 2019q1.

Estimated ASF: Expectation Formation Model is Non-linear and Asymmetric

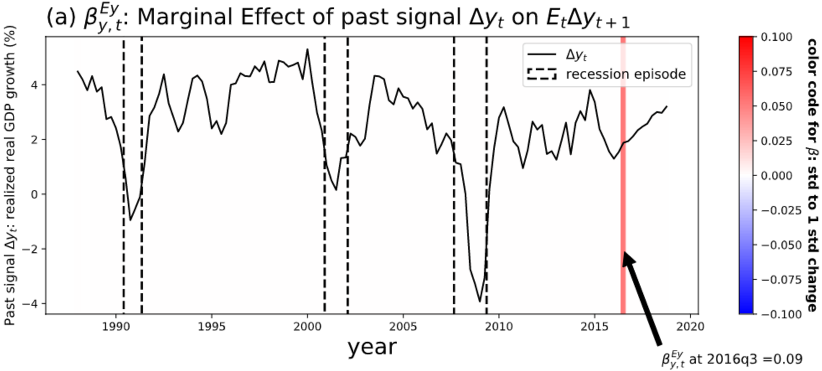
$$\mathbb{E}[\hat{g}_u(\theta_{i,t-1}, Z_{i,t}^{-u}, \Delta u_t + \delta)) - \hat{g}_u(\theta_{i,t-1}, Z_{i,t}^{-u}, \Delta u_t)]$$

- Non-linearity: Slope changes continuously.
- Asymmetry: (Magnitudes of) response to positive and negative signals differ significantly.



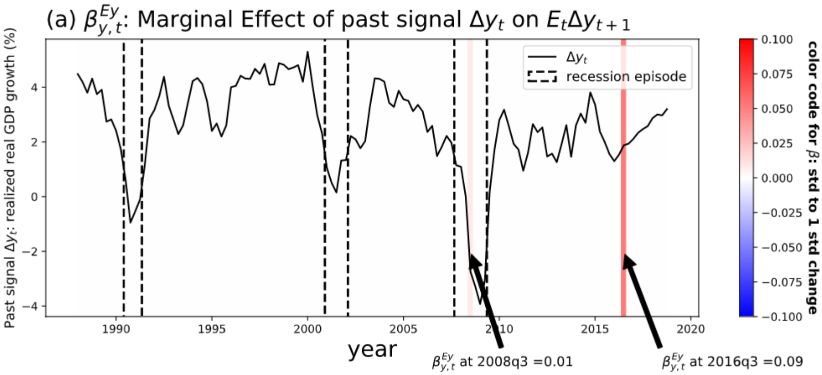
Heat Map: Marginal Effect at Each Quarter

- Each color bar represents magnitude of marginal effect at a time point;



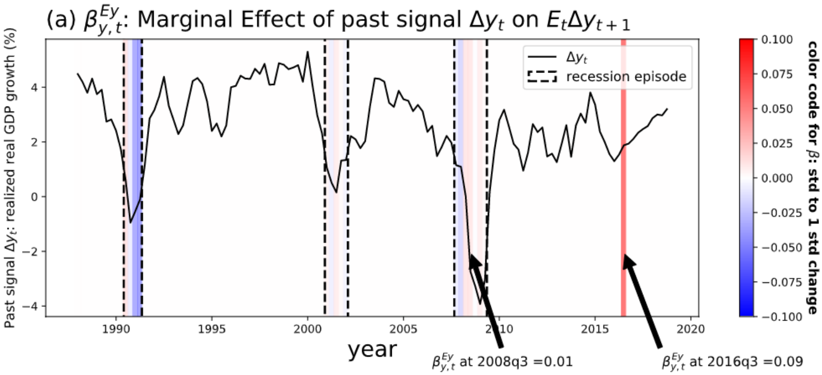
Heat Map: Marginal Effect at Each Quarter

- Color code is slope normalized by standard deviation;



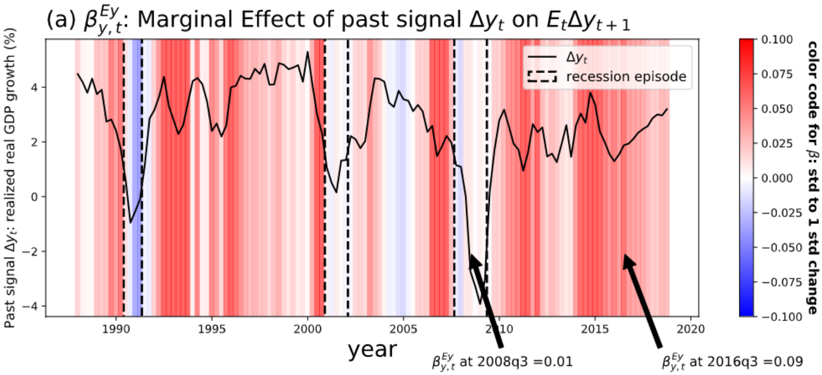
Attention-shift: Lower weight on past signal Δy_t in recession

Marginal effect on past signal Δy_t is smaller in recession.



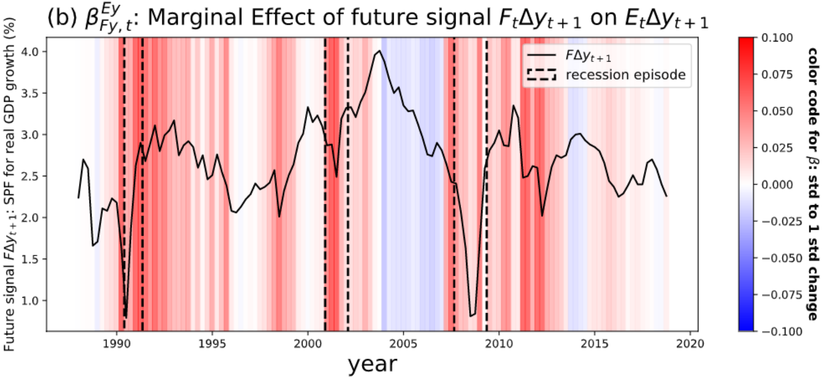
Attention-shift: Lower weight on past signal Δy_t in recession

Marginal effect on past signal Δy_t is much bigger in ordinary period.



Attention-shift: Higher weight on future signal $F_t\Delta y_{t+1}$ in recession

Marginal effect on future signal $F_t\Delta y_{t+1}$ is bigger in recession.



DML estimates of AME on past v.s. future signal

Expectation:	Signal	$E_t\Delta y_{t+1}$ $\beta_{recession}$ (std)	$E_t\Delta y_{t+1}$ $\beta_{ordinary}$ (std)	$E_t\Delta y_{t+1}$ $\beta_{rec}=\beta_{ord}$ (p-val)	$E_t\Delta u_{t+1}$ $\beta_{recession}$ (std)	$E_t\Delta u_{t+1}$ $\beta_{ordinary}$ (std)	$E_t\Delta u_{t+1}$ $\beta_{rec}=\beta_{ord}$ (p-val)
Past Signal	Δy_t	0.004*	**0.017*** **	<0.01	-0.006***	**-0.01*** **	0.04
		(0.003)	(0.001)		(0.001)	(0.001)	
	Δu_t	-0.006	**-0.021*** **	0.04	0.005	**0.012*** **	0.08
		(0.006)	(0.004)		(0.004)	(0.002)	
Future Signal	$F_t\Delta y_{t+1}$	**0.049*** **	0.016***	<0.01	**-0.022*** **	-0.009***	<0.01
		(0.005)	(0.003)		(0.002)	(0.001)	
	$F_t\Delta u_{t+1}$	**-0.037*** **	0.009**	<0.01	**0.029*** **	0.007***	<0.01
		(0.004)	(0.002)		(0.003)	(0.002)	

* Results are using panel with 12000 observations. HAC standard errors are reported in brackets. *,**,*** stands for significant at 10%, 5% and 1% level.

back more dml results

DML estimates of AME on past v.s. future signal

Expectation:	Signal	$E_t\Delta y_{t+1}$ $\beta_{recession}$ (std)	$E_t\Delta y_{t+1}$ $\beta_{ordinary}$ (std)	$E_t\Delta y_{t+1}$ $\beta_{rec}=\beta_{ord}$ (p-val)	$E_t\Delta u_{t+1}$ $\beta_{recession}$ (std)	$E_t\Delta u_{t+1}$ $\beta_{ordinary}$ (std)	$E_t\Delta u_{t+1}$ $\beta_{rec}=\beta_{ord}$ (p-val)
Past Signal	Δy_t	0.004*	**0.017*****	<0.01	-0.006***	**−0.01*****	0.04
		(0.003)	(0.001)		(0.001)	(0.001)	
	Δu_t	-0.006	**−0.021*****	0.04	0.005	**0.012*****	0.08
		(0.006)	(0.004)		(0.004)	(0.002)	
Future Signal	$F_t\Delta y_{t+1}$	**0.049*****	0.016***	<0.01	**−0.022*****	-0.009***	<0.01
		(0.005)	(0.003)		(0.002)	(0.001)	
	$F_t\Delta u_{t+1}$	**−0.037*****	0.009**	<0.01	**0.029*****	0.007***	<0.01
		(0.004)	(0.002)		(0.003)	(0.002)	

* Results are using panel with 12000 observations. HAC standard errors are reported in brackets. *,**,*** stands for significant at 10%, 5% and 1% level.

back more dml results

Conclusion

1. New Method:
   - Generic Learning Framework.
   - Non-parametric method for estimation: RNN.
   - DML for inference.
2. New empirical findings on expectation formation:
   - Non-linearity and asymmetry. Expectation more sensitive to bad news.
   - Attention-shift. Adaptive learner in ordinary period, forward looking in recession.
3. Model with Rational Inattention:
   - Information becomes more valuable in bad states due to non-linearity in optimal choices.
   - Agents seek for more information about future when economic status worsen.

Thank you!

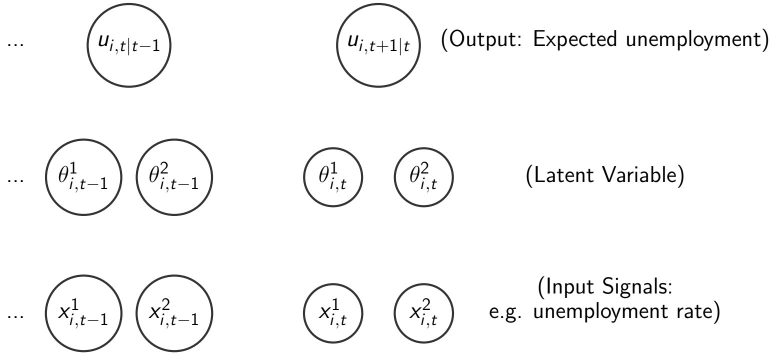
Appendix

Architecture RNN

Table 1: Architecture RNN

Tuned Hyper Parameter	Configuration
Num. of Recurrent Neurons	32
Feed-forward Neurons	20
Dropout on recurrent layer	0.5
Epochs	200
Learning Rate	$1e^{-6}$
Depth	2(4)
Un-tuned Hyper Parameter	**Configuration**
Type of Recurrent Layer	Long-Short Term Memory (LSTM)
Activation Function:	ReLu

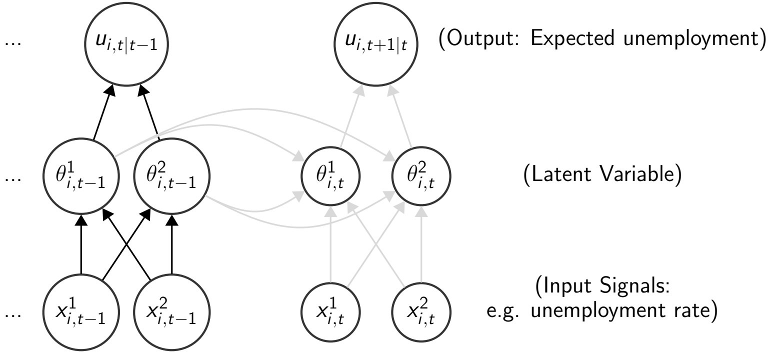
Introduction to RNN: Simple Example



Consider we use this simple RNN to model expected unemployment:

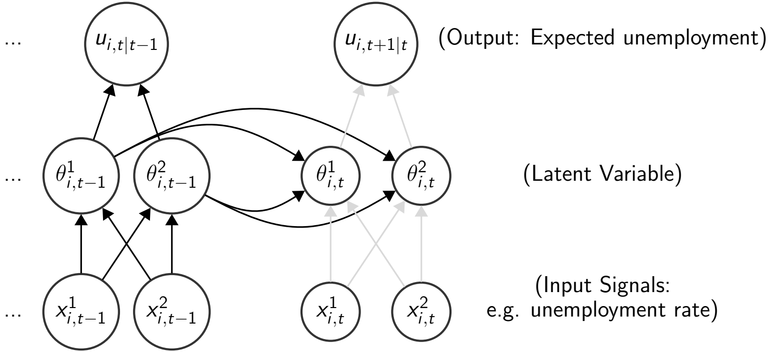
- Observable: two signals $x^1_{i,t}, x^2_{i,t}$, expected unemployment $u_{i,t+1|t}$;
- Unobserved: 2 recurrent hidden neurons: $\theta^1_{i,t}$ and $\theta^2_{i,t}$

Introduction to RNN: Simple Example



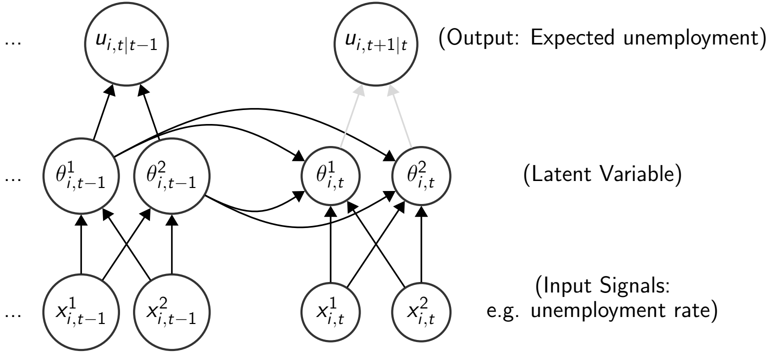
At time $t-1$: information $\{x^1_{i,t-1}, x^2_{i,t-1}\}$ updated into hidden neurons $\{\theta^1_{i,t-1}, \theta^2_{i,t-1}\}$ and used to form $u_{i,t|t-1}$;

Introduction to RNN: Simple Example



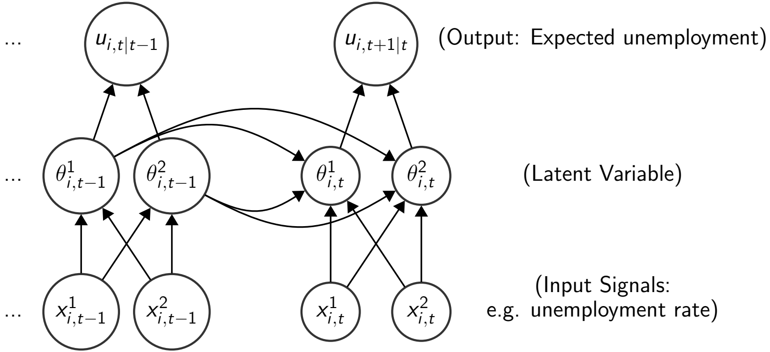
Updating: (1) past values of $\{\theta^1_{i,t-1}, \theta^2_{i,t-1}\}$ are used to update $\{\theta^1_{i,t}, \theta^2_{i,t}\}$

Introduction to RNN: Simple Example



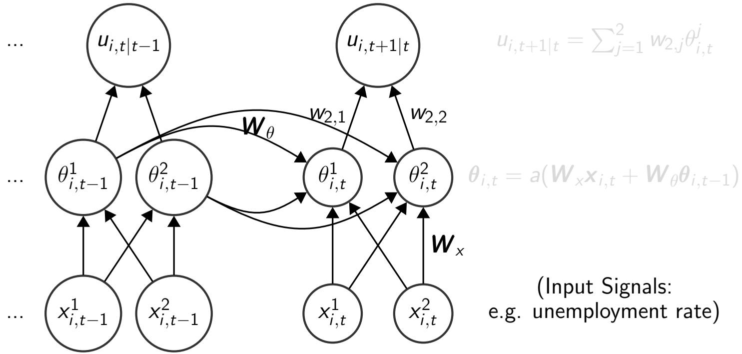
Updating: (2) new signals arrive and are used to update $\{\theta^1_{i,t}, \theta^2_{i,t}\}$

Introduction to RNN: Simple Example



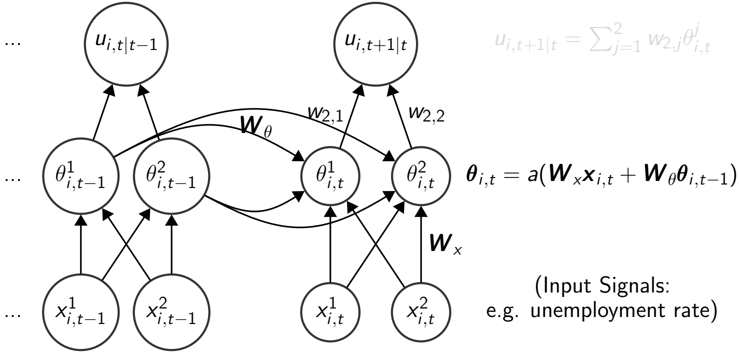
Forecasting: updated $\{\theta^1_{i,t}, \theta^2_{i,t}\}$ are used to form $u_{i,t+1|t}$

Introduction to RNN: Weight Updating



Each branch that connects two neurons has a weight, which is parameter RNN learns; back

Introduction to RNN: Weight Updating

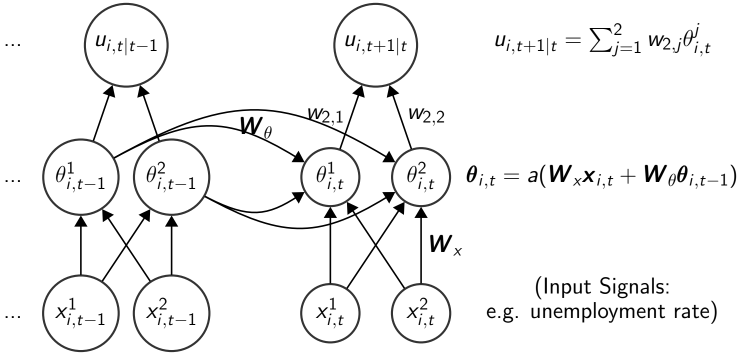


at time t, RNN compute:

(1) $\boldsymbol{\theta}_{i,t} = a(\boldsymbol{W}_x \boldsymbol{x}_{i,t} + \boldsymbol{W}_\theta \boldsymbol{\theta}_{i,t-1})$

(2) $u_{i,t+1|t} = \sum_{j=1}^{2} w_{2,j} l_{i,t}^{j}$ back

Introduction to RNN: Weight Updating

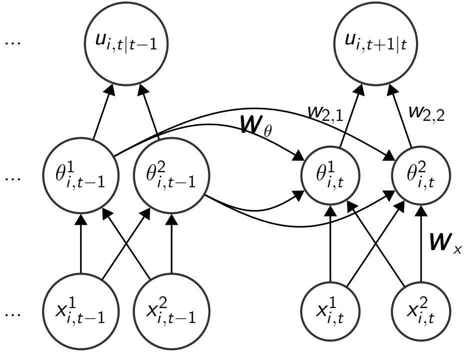


at time t, RNN compute:

(1) $\boldsymbol{\theta}_{i,t} = a(\boldsymbol{W}_x\boldsymbol{x}_{i,t} + \boldsymbol{W}_{\theta}\boldsymbol{\theta}_{i,t-1})$

(2) $u_{i,t+1|t} = \sum_{j=1}^{2} w_{2,j}\theta_{i,t}^{j}$ back

Introduction to RNN: Weight Updating



$u_{i,t+1|t} = \sum_{j=1}^{2} w_{2,j}\theta^j_{i,t}$

$\boldsymbol{\theta}_{i,t} = a(\boldsymbol{W}_x \boldsymbol{x}_{i,t} + \boldsymbol{W}_\theta \boldsymbol{\theta}_{i,t-1})$

(Input Signals:
e.g. unemployment rate)

All weights $w_{2,j}$, $\boldsymbol{W}_x$ and $\boldsymbol{W_\theta}$ are chosen by Gradient Descent; back

Monte Carlo Example: Noisy Information Model with Linear Kalman Filter

The Gaussian Linear State Space Model agent believes in (Perceived Law of Motion);

$$\begin{bmatrix} \pi_t \\ L_t \end{bmatrix} \equiv X_t = \boldsymbol{A} X_{t-1} + \epsilon_t$$

Observe noisy signal:

$$\begin{bmatrix} \pi_{i,t} \\ s_{i,t} \end{bmatrix} \equiv O_{i,t} = \boldsymbol{G} X_t + \nu_{i,t}$$

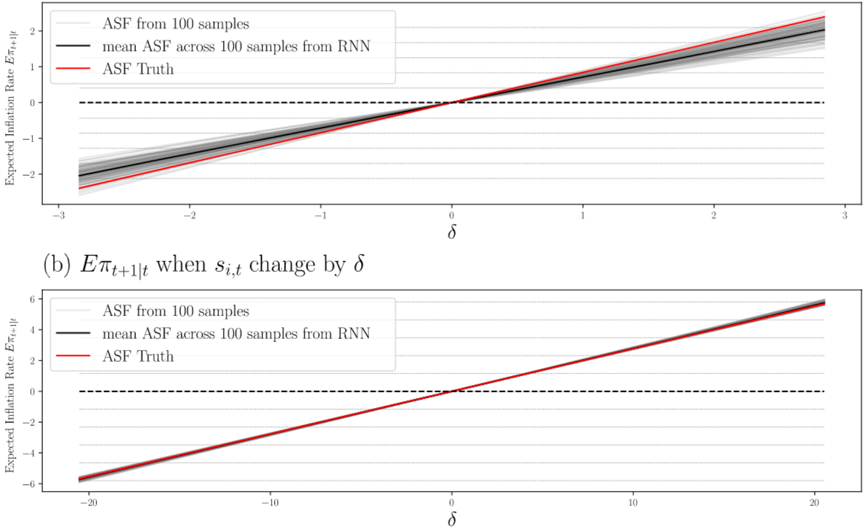
Use Kalman Filter to form forecast:

$$\begin{bmatrix} \pi_{i,t+1|t} \\ L_{i,t+1|t} \end{bmatrix} \equiv X_{i,t+1|t} = \boldsymbol{A}(X_{i,t|t-1} + \boldsymbol{K}(O_{i,t} - \boldsymbol{G} X_{i,t|t-1}))$$

Monte Carlo Example:
ASF from Noisy Information Model

(a) $E\pi_{t+1|t}$ when π_t change by δ

(b) $E\pi_{t+1|t}$ when $s_{i,t}$ change by δ



Monte Carlo Example: Marginal Effect

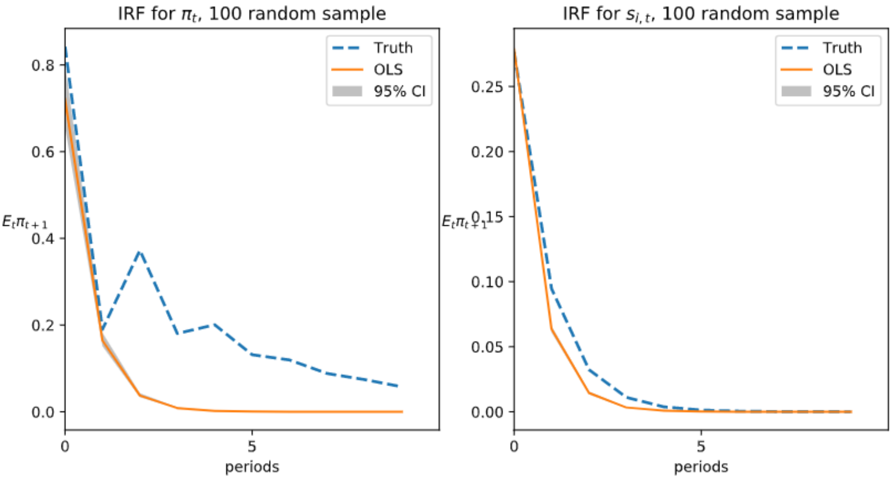
Table 2: Performance of RNN v.s. OLS

	MSE	π_t	$s_{i,t}$
(1) RNN	2.91	0.82	0.276
	(0.054)	(0.037)	(0.003)
(2) OLS mis-specified	3.296	0.720	0.279
	(0.023)	(0.033)	(0.001)
(3) OLS correct	2.835	0.841	0.277
	(0.014)	(0.005)	(0.001)
Truth		0.842	0.277

* The first column is mean squared error on the whole sample, the second column is estimated marginal effect on signal π_t and third column is estimated marginal effect on signal $s_{i,t}$. In brackets I report the standard deviation of the statistics using 100 simulated random samples.

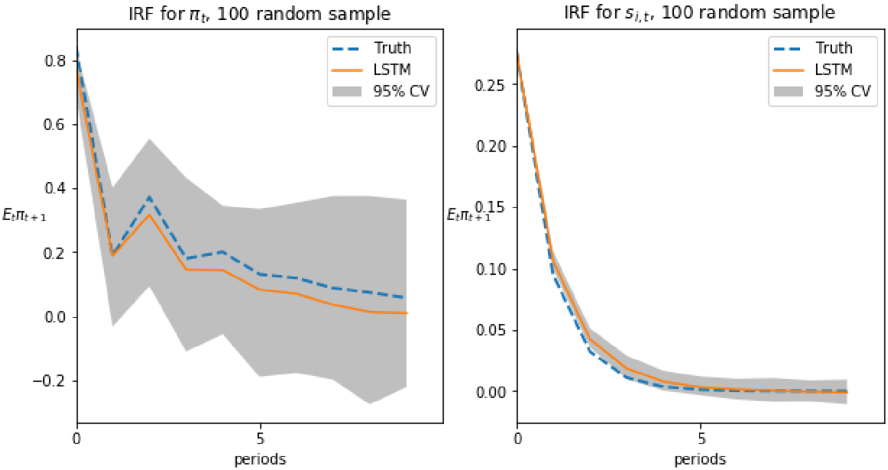
Monte Carlo Example:
Noisy Information Model with Linear Kalman Filter

IRF from mis-specified OLS (missing L_t) back

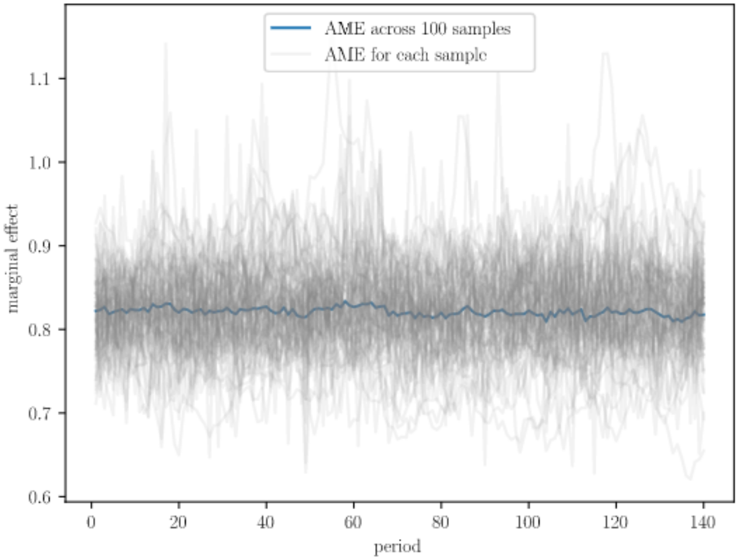


Monte Carlo Example: Noisy Information Model with Linear Kalman Filter

IRF from RNN (LSTM) back



Monte Carlo Example:
AME from Noisy Information Model



RNN Architecture

Table 3: Architecture RNN

Tuned Hyper Parameter	Configuration
Num. of Recurrent Neurons	32
Feed-forward Neurons	20
Dropout on recurrent layer	0.5
Epochs	200
Learning Rate	$1e^{-6}$
Depth	2(4)
Un-tuned Hyper Parameter	**Configuration**
Type of Recurrent Layer	Long-Short Term Memory (LSTM)
Activation Function:	ReLu
Total parameters:	8,424

* Tuned hyper parameters are picked using 6-Fold cross-validation across individuals. This satisfies the requirement for fast enough convergence of estimated Average Structural Function so that functional estimators from this Neural Network can be used to obtain inference on DML estimators.