Machine Learning in Agricultural, Environmental and Applied Economics

Hugo Storm, Kathy Baylis, Thomas Heckelei

Outline Part I ML from an applied econometrics perspective

- Approach to avoid overfitting
- Supervised approaches
  - Shrinkage methods
  - Tree based methods
  - Neural networks
- Unsupervised approaches
- Model complexity versus interpretability

Outline Part II What ML can add to our toolbox

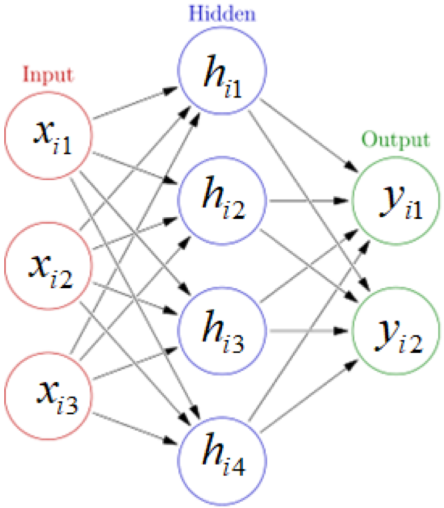
- Restrictive functional forms
- Extract information from unstructured data
- Large number of explanatory variables
- Causal inference and Identification
- Limitations of simulation models for policy analysis

For today

- Shameless self promotion (✓)
- Quick overview of Neural Networks (NN)
- How can they help?
  - Restrictive functional forms
  - Extract information from unstructured data
  - Deal with a large number of explanatory variables
  - Causal inference and Identification
- Example: NN for policy simulation

Neural networks (NN)

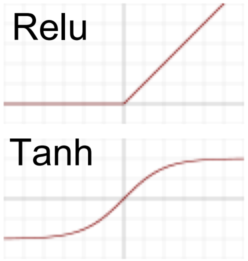
- Can learn highly nonlinear multi-input/output relations
- Highly flexible/nonlinear
- Deep NN have many layers => representation learning
- Less automatic to train than trees but can handle more complex data structures (e.g. Convolutional NN (CNN) for images or Recurrent NN (RNN) for time series/panel data)



$$\underset{(4\times1)}{h_i} = \sigma^{[1]}\left(\underset{(4\times3)}{W^{[1]}}\ \underset{(3\times1)}{x_i} + \underset{(4\times1)}{b^{[1]}}\right)$$

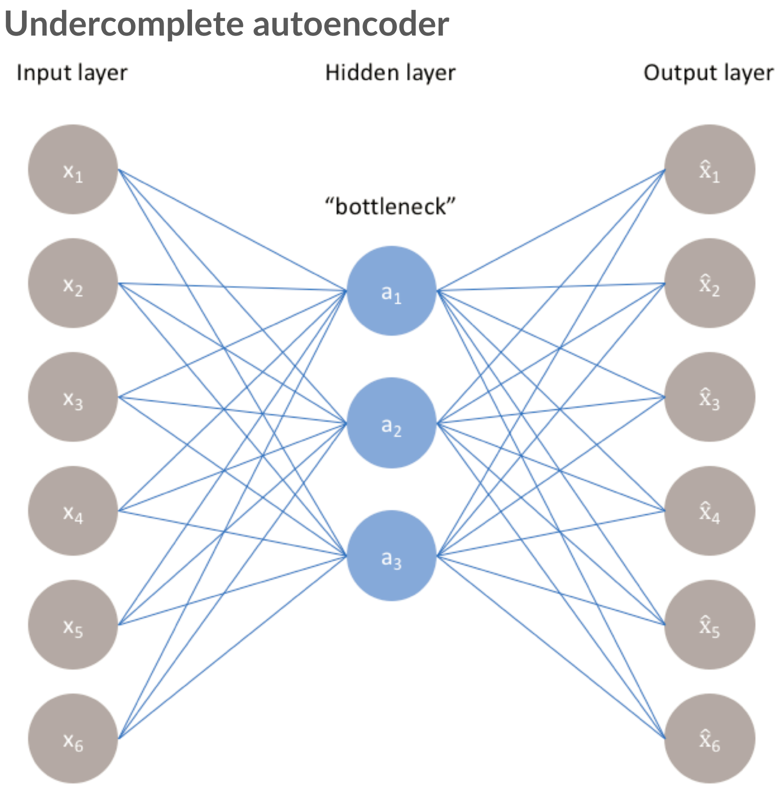
$$\underset{(2\times1)}{y_i} = \sigma^{[2]}\left(\underset{(2\times4)}{W^{[2]}}\ \underset{(4\times1)}{h_i} + \underset{(2\times1)}{b^{[2]}}\right)$$

Activation functions



Unsupervised approaches
Autoencoder

- Aim to learn joined probability of (x) instead of E(y|x)
- Non-linear version of PCA
- Aim: Extract most relevant features in (x)

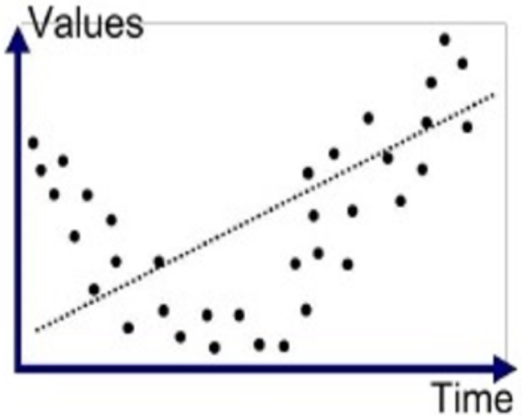


Source: https://www.jeremyjordan.me/autoencoders/

How can NNs help?

Restrictive functional forms

- Many problems are non-linear. Theory provides only weak guidance
- Particularly environmental/physical processes
- Underfitting can lead to bias
- Existing flexible econometric tools struggle with large K and/or N
- NN good for time series (instead of AR)



Source:https://cdn-images-1.medium.com/max/1200/1*_7OPgojau8hkiPUiHoGK_w.png

Extract information from unstructured data

- *Unstructured data* = everything that cannot be processed in a spreadsheet
- Examples: (remote sensing) images, text, cell phone records, weather
- Current approach: extract hand crafted features

ML approches
1) End-to-end learning
2) Unsupervised pre-training
3) Transfer learning
4) Automatic feature creation
5) Text methods

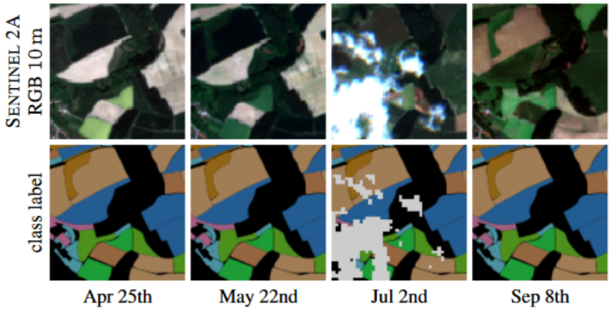
Unstructured data: End-to-end learning

- Raw data as input variables (without preprocessing)
  - \+ No manual feature extraction
  - \+ No information loss
  - \- Requires large labeled datasets

Rußwurm and Körner (2017)

- Crop classification with remote sensing
- >137.000 labelled fields, 19 crops

l ata)

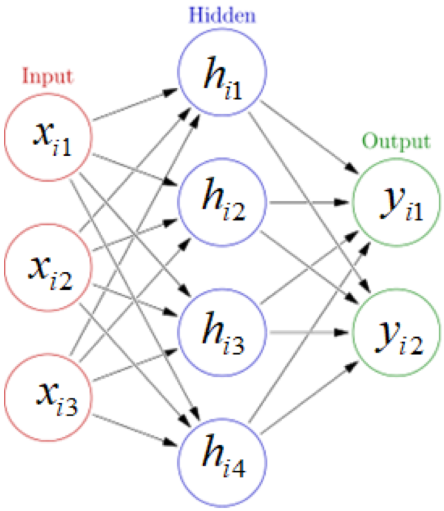


Unstructured data: Unsupervised pre-training

First successful training of deep NN (Hinton et al. 2006)

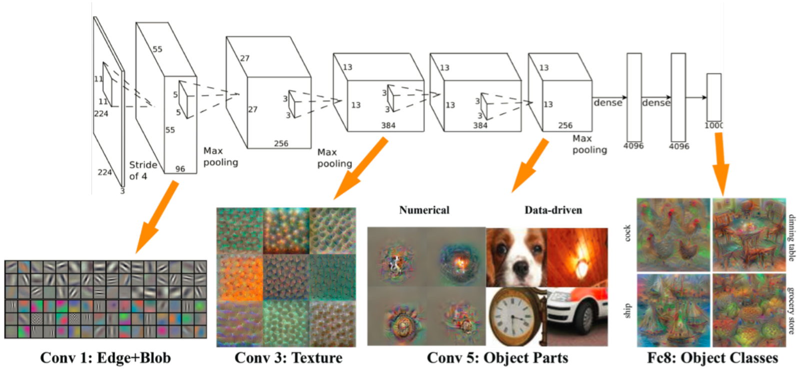
1) Pretraining NN layers as autoencoder using unlabeled data
2) Train last layer using labeled data

+ Can make use of unlabeled data
- Aims to preserve as much variation as possible (not necesssarily the most relevant variation)



Unstructured data: Transfer Learning

- Use models trained in one context as starting point
- E.g. "Representation learning"



Jean et al. (2016) poverty prediction:
Transfer learning

1) VGG trained on ImageNet
2) Predict nightlights intensity classes from daylight images
3) Predict poverty indices

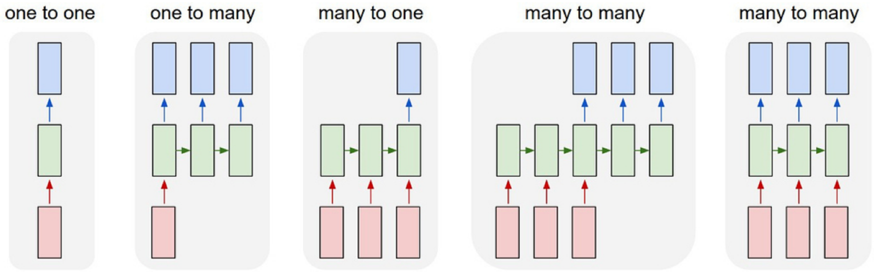
Why does it work? Examples, roofing material and distance to urban areas

Large number of explanatory variables

- In many application K is large and/or high temporal/spatial resolution (scanner/climate date)
- Current approach: Handcrafted collapsing, test driven selection, PCA, Bayesian model selection

ML approches

1) Can better handle large K
2) Unsupervised feature extraction
3) RNN/CNN to deal with high temporal/spatial resolution



Source:https://discuss.pytorch.org/t/example-of-many-to-one-lstm/1728

Causal inference and Identification

Causal inference can be thought of as a prediction problem: What would have happened in absence of the treatment?

(i) **Counterfactual simulation** [exogenous treatment]
(ii) Causal forests [selection on observables, heterogeneous effects]
(iii) Double Machine Learning [selection on observables, average effects]
(iv) **ML Panel Methods** [unobserved time-invariant characteristics]
(v) IV with may instruments [endogenous treatment, linear]
(vi) Deep IV estimation [endogenous treatment, nonlinear]

Example: RNN to evaluate farm policy

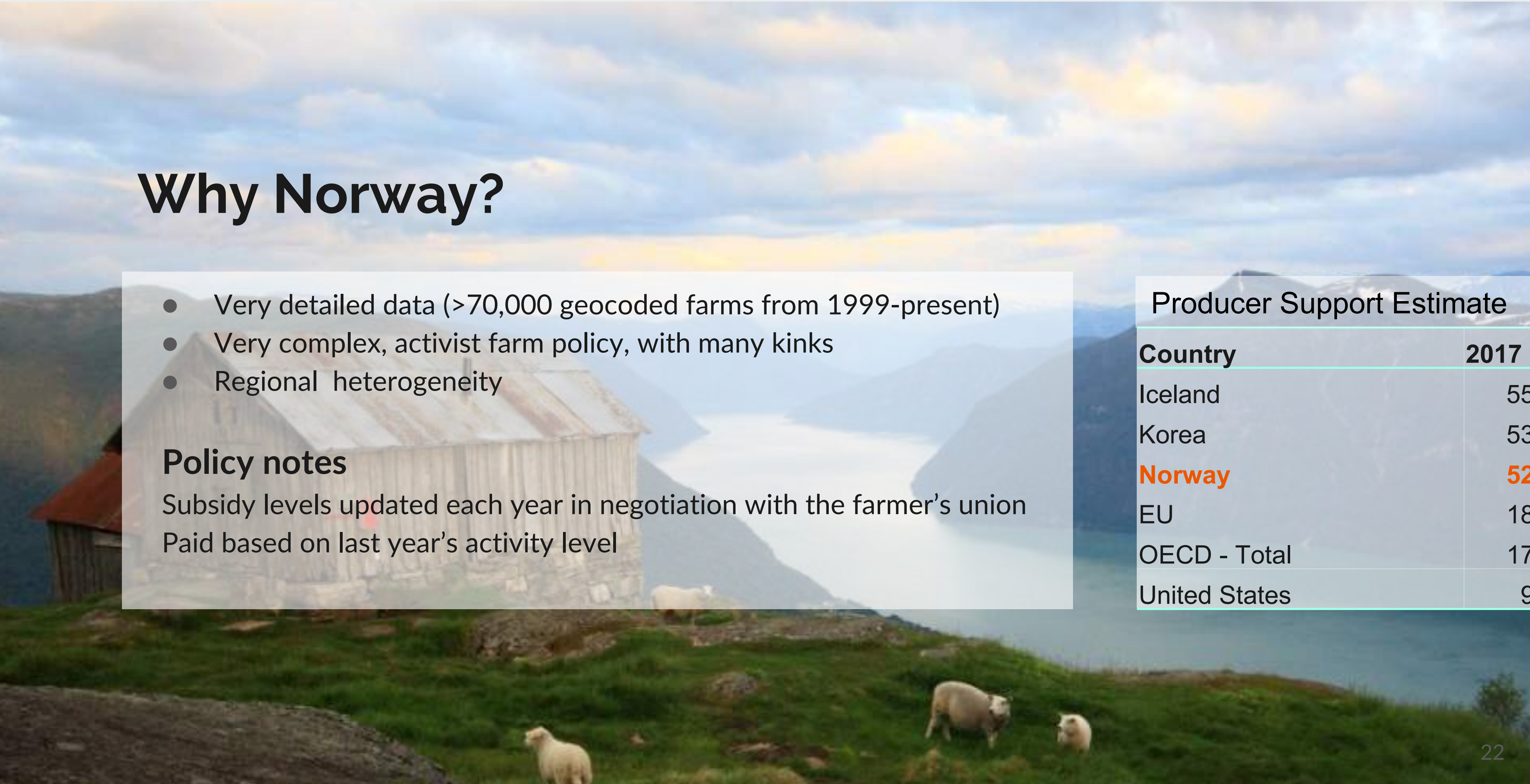
How does farm policy affect farm structure?

Why do we care?

- Much agricultural policy is justified by the 'need to preserve the family farm'
- Most agricultural policies have differentiated payments based on size and/or payment caps

Challenges:

(1) multi-dimensional measure of farm structure
(2) non-linear policy (and policy effects) – over multiple outputs
(3) spatial dependence
(4) dynamics



Why Norway?

- Very detailed data (>70,000 geocoded farms from 1999-present)
- Very complex, activist farm policy, with many kinks
- Regional heterogeneity

Policy notes

Subsidy levels updated each year in negotiation with the farmer's union

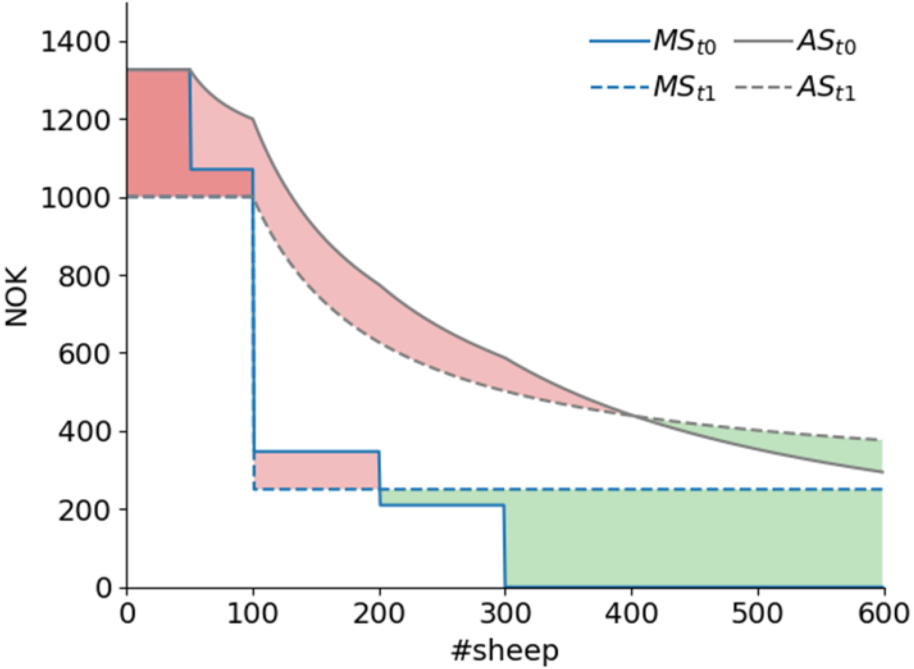
Paid based on last year's activity level

Producer Support Estimate

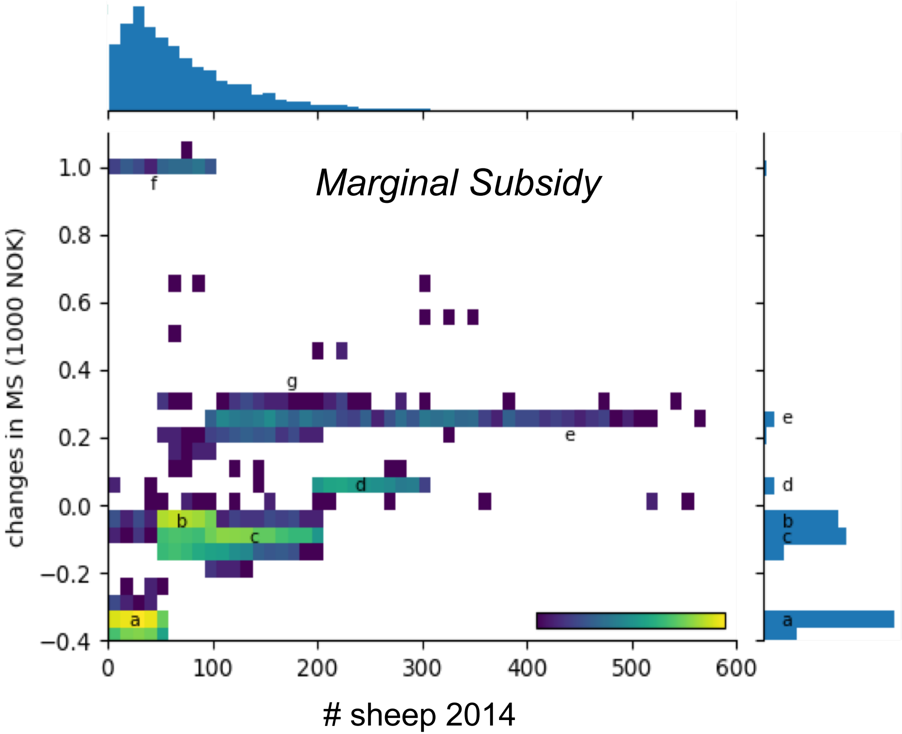
Country	2017
Iceland	55
Korea	53
Norway	**52**
EU	18
OECD - Total	17
United States	9

Policy change

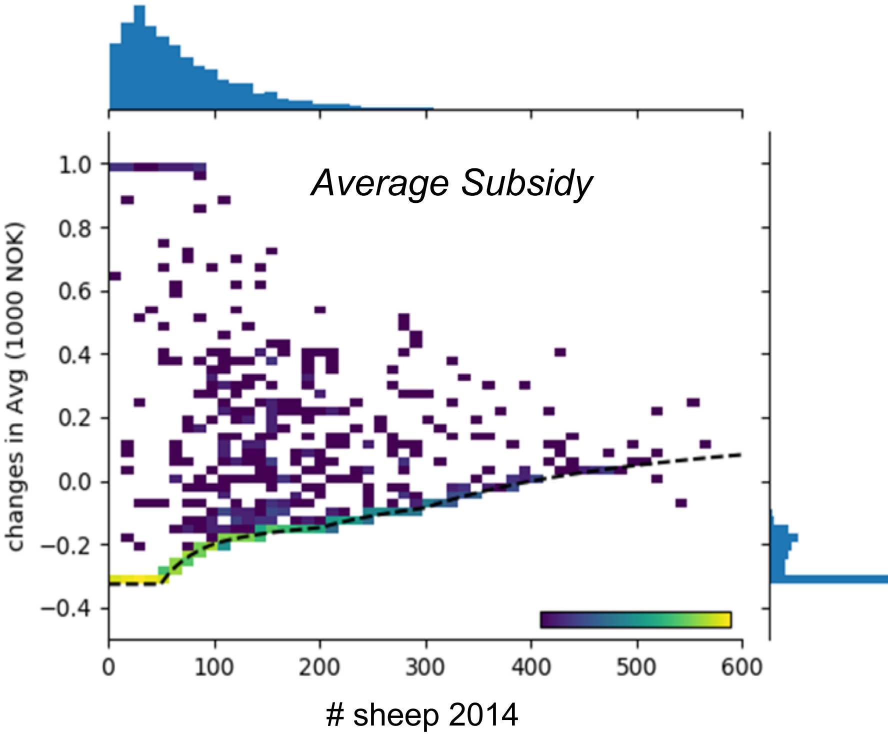
	NOK/Head	
Size (head)	**2014**	**2015**
1-50	1326	1000
51-100	1070	1000
101-200	347	250
201-300	210	250
>300	0	250
Total cap	280k	560k



Raw data

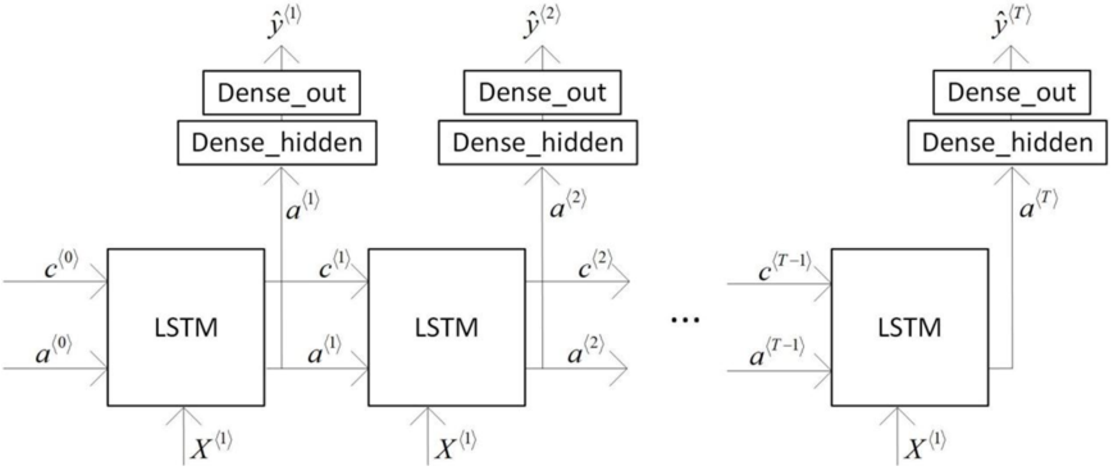


Marginal Subsidy
changes in MS (1000 NOK)
sheep 2014
a
b
c
d
e
f
g



Average Subsidy
changes in Avg (1000 NOK)
sheep 2014

Recurrent NN with Long Short Term Memory (RNN-LSTM)



- LSTM cells pass information across time in a cell-state vector *c*
- Takes new input each period (X) and use *gates* to figure out what information it can keep and what we can forget (*a* = *Gc*)

Intuition

- Cell state encodes past information
- Model learns itself how to encode information
- and which information to keep/forget
- ➢ no need to specify lag structure
- ➢ Lag structure can vary across variables

Example:

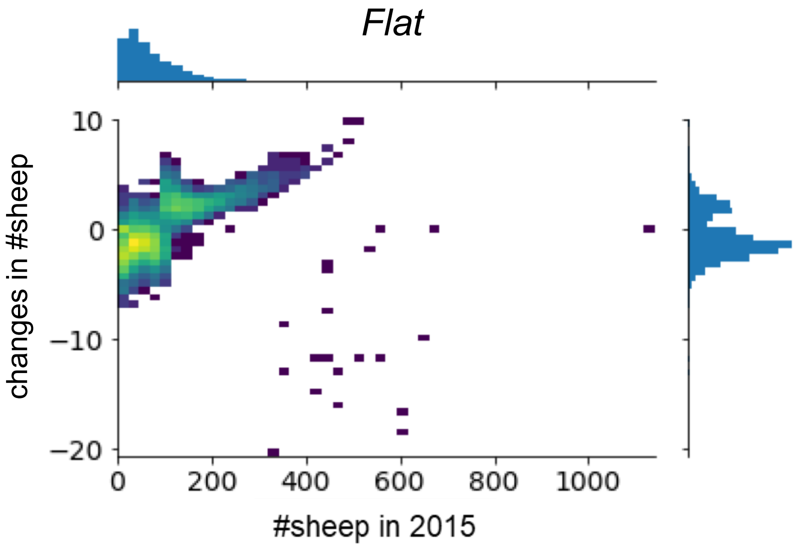
- Farm stops milk production
- Specific number of cows every past year might not be relevant
- Sufficient to encode that it had dairy once (maybe the maximum number of cows)

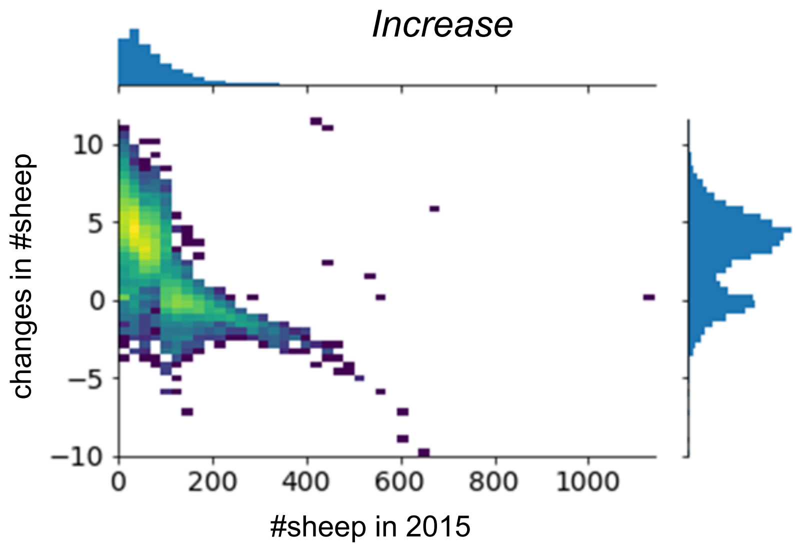
Simulation

Size (head)	2014	2015	2016 base	2016 Flat	2016 Increase
			NOK/Head		
0-50	1326	1000	1000	600	1500
51-100	1070	1000	1000	600	1500
101-200	347	250	250	600	0
201-300	210	250	250	600	0
>300	0	250	250	600	0
Total cap	280k	560k	560k	560k	560k

Changes in sheep

	NOK/Head				
Size (head)	2014	2015	2016 base	2016 Flat	2016 Increase
0-50	1326	1000	1000	600	1500
51-100	1070	1000	1000	600	1500
101-200	347	250	250	600	0
201-300	210	250	250	600	0
>300	0	250	250	600	0
Total cap	280k	560k	560k	560k	560k

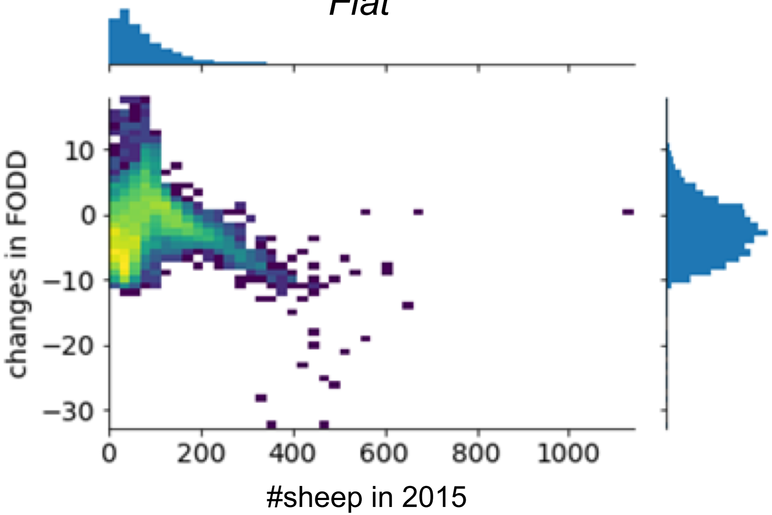




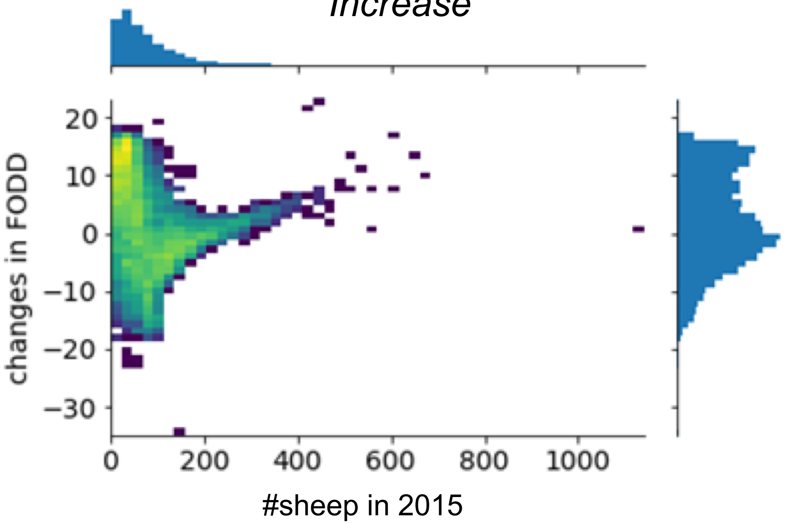
Fodder

	NOK/Head				
Size (head)	2014	2015	2016 **base**	2016 **Flat**	2016 **Increase**
0-50	1326	1000	1000	600	1500
51-100	1070	1000	1000	600	1500
101-200	347	250	250	600	0
201-300	210	250	250	600	0
>300	0	250	250	600	0
Total cap	280k	560k	560k	560k	560k

Flat



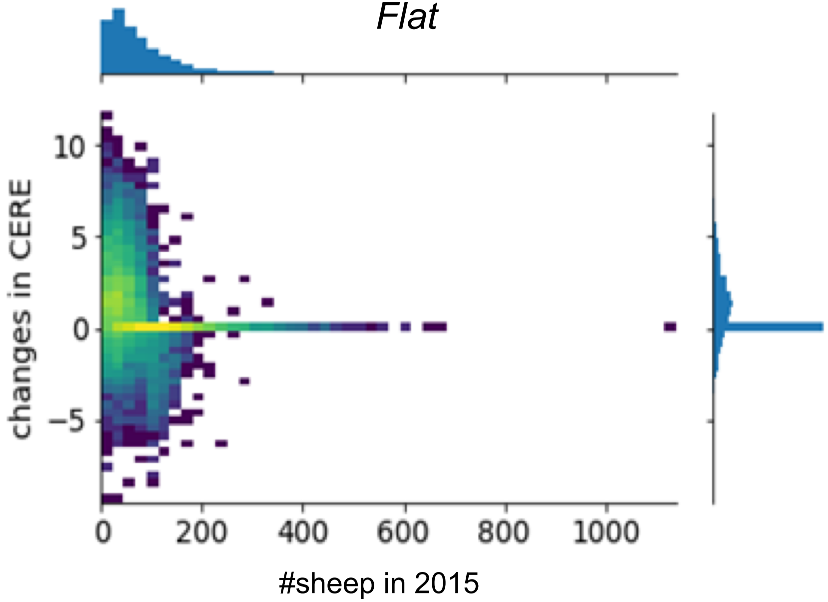
Increase



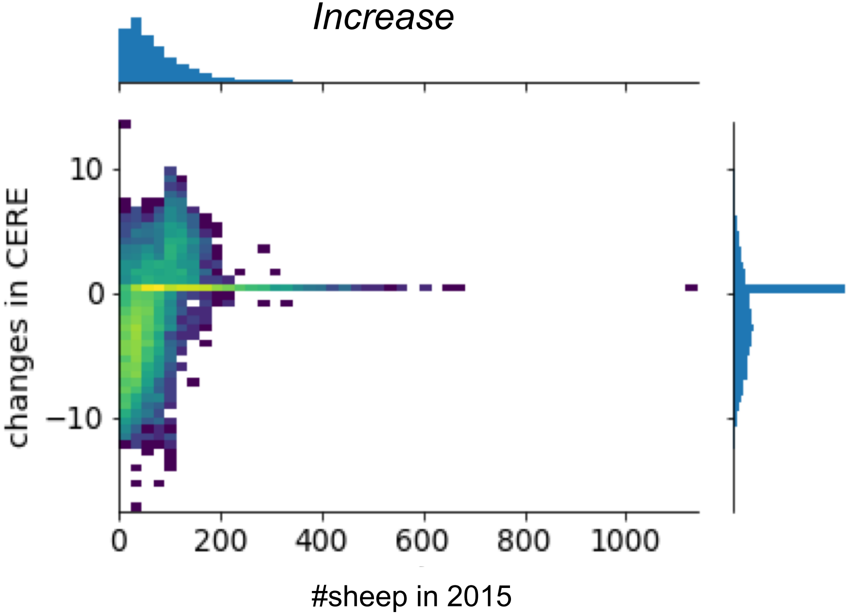
Cereals

NOK/Head					
Size (head)	2014	2015	2016 base	2016 Flat	2016 Increase
0-50	1326	1000	1000	600	1500
51-100	1070	1000	1000	600	1500
101-200	347	250	250	600	0
201-300	210	250	250	600	0
>300	0	250	250	600	0
Total cap	280k	560k	560k	560k	560k

Flat



Increase



Implications

- Without explicit economic information, RNN is able to pick up reasonably direct and indirect responses to changes in subsidies
- Able to model more complex dynamic processes than standard AR models
- And more complex spatial patterns

Lots of things yet to do…

- Explore dynamics
- Embed farm fixed effects
- Compare to standard panel model estimation
- Explore how well CNNs can pick up spatial dependence

ML and Environmental Econ: Coming soon

- Thinking hard about selection of input data
- Introducing structure into ML models
- Statistical properties of ML and uncertainty (probabilistic programming)
- Application to Causal Inference