Institute for Food and Resource Economics

University of Bonn

Discussion Paper 2019:1

Identifying effects of farm subsidies on structural change using neural networks

Hugo Storm, Thomas Heckelei, Kathy Baylis, Klaus Mittenzwei

Editor: Thomas Heckelei

Institute for Food and Resource Economics

University of Bonn

Nußallee 21

53115 Bonn, Germany

Phone: +49-228-732332

Fax: +49-228-734693

E-mail: thomas.heckelei@ilr.uni-bonn.de

Identifying effects of farm subsidies on structural change using neural networks

Hugo Storm, Thomas Heckelei, Kathy Baylis, Klaus Mittenzwei

Abstract: Farm subsidies are commonly motivated by their promise to help keep families in agriculture and reduce farm structural change. Many of these subsidies are designed to be targeted to smaller farms, and include production caps or more generous funding for smaller levels of activity. Agricultural economists have long studied how such subsidies affect production choices, and resulting farm structure. Traditional econometric models are typically restricted to detecting average effects of subsidies on certain farm types or regions and cannot easily incorporate complex subsidy design or the multi-output, heterogeneous nature of many farming activities. Programming approaches may help address the broad scope of agricultural production but have less empirical measures for behavioral and technological parameters. This paper uses a recurrent neural network and detailed panel data to estimate the effect of subsidies on the structure of Norwegian farming. Specifically, we use the model to determine how the varying marginal subsidies have affected the distribution of Norwegian farms and their range of agricultural activities. We use the predictive capacity of this flexible, multi-output machine learning model to identify the effects of agricultural subsidies on farm activity and structure, as well as their detailed distributional effects.

Keywords: farm growth, farm subsidies, structural change, machine learning, recurrent neural network

JEL classification: Q12, Q18, C23, C45

1 Introduction

The agricultural sector in developed countries has long operated in a highly subsidized environment, with farm programs covering a wide range of agricultural activities. These subsidies are frequently justified with the goal of reducing farm consolidation to preserve vibrant rural communities. One overarching question is how these subsidies affect the structure of agriculture (Key and Roberts 2009; Breen, Hennessy and Thorne 2005; Ahearn, Yee and Korb 2005). Because of the complexity of both the subsidies and farming itself, it is often difficult to quantitatively assess this effect and determine how potential program changes might affect farm behaviour and outcomes. In this paper, we develop a novel approach using neural networks to evaluate the effect of the Norwegian agricultural support scheme on farm structure empirically. Norway serves as an excellent case study because its program is particularly complex, large and covers a wide range of agricultural activities, rendering traditional analysis particularly difficult.

The Norwegian policy scheme explicitly aims to equalize income opportunities across the farm and non-farm sector as well as to equalize income opportunities across different farm types and regions. To this end, its policy scheme is highly complex, farm-specific and affects all farm production activities. It has also traditionally been directed at encouraging smaller farms and includes higher subsidy rates for activities below certain levels. The complexity makes it challenging to quantitatively assess potential changes in the subsidy scheme on farmer decisions and resulting farm structure. However, knowledge of these consequences would help both in understanding whether such a detailed scheme meets its goal of encouraging smaller farms and is a valuable input for the annual negotiations between the government and the farmers' organizations. In this paper we develop and illustrate a novel approach to assess the effect of the subsidy on farm structure. Specifically, we aim to identify what effects production caps or more generous funding for smaller farmers have, and how changes in this respect influence farmers decision.

On one hand, while the complexity of the Norwegian system is a challenge for policy analysis, on the other, it provides a unique opportunity to identify the effects of the policy with numerous prior policy changes varying specific payments for individual farms. In other countries, for example EU member states operating under the CAP, empirically identifying and quantifying the effects of subsidies on farmers' decision making is challenging due to a lack of exogenous variation in payments rates. CAP-payments per hectare are either the same across an entire region such that we lack sufficient variation across farms to identify their effects, or vary with past activity levels creating substantial problems of endogeneity. In Norway, payment rates per hectare or head differ by size, region and are interrelated among different subsidy types. Hence, the complexity of the system introduces a considerable amount of exogenous variation that can be exploited to identify the subsidy effects.

In the Norwegian system, payment rates per activity are designed to differ across farms depending on specific farm characteristics. 1) Payments vary by activity level such that the subsidies for the first unit are larger than for the last. 2) There is a total cap on payments for some types of subsidies such that farms already receiving the maximum amount do not receive extra payments for the next unit. 3) Some subsidies have a fixed deduction, which implies that farms do not receive subsidies until their level of activity is large enough to surpass the fixed amount. 4) A certain amount of livestock is required to claim pasture payments, implying that the payment rates for livestock can differ across farms due to their specific livestock-pasture ratio. A farm that does not have sufficient livestock to claim all pasture payments has a higher marginal subsidy rate for an additional unit of livestock compared to a farm that can already claim all pasture payments. 5) Due to the negotiations, new payment schemes may be introduced. That is, eligibility criteria and subsidy rates of existing schemes may change on an annual basis introducing further variation across years and regions. Taken together, these characteristics of the policy imply that a specific change in the subsidy scheme results in heterogeneous payout effects within the farm population. This heterogeneity provides an opportunity to identify the effect subsidy payments have on farmers' decisions. However, the effects of subsidies need to be distinguished from other factors. For example in Norway, subsidies are combined with border controls causing domestic producer prices to be higher than world market prices. The distribution of total agricultural support between border protection and budgetary support depends on political considerations about market developments and general economic conditions outside the agricultural sector.

We use machine learning, specifically a recurrent neural network (RNN) (Schmidhuber 2015; Goodfellow et al. 2016), to ask how these subsidies have affected farm structure. Machine learning tools provide several advantages over more traditional econometric techniques. First, a RNN is easily capable of handling multiple dimensions of the dependent variable that characterize the farm (also called the 'target variable' in machine learning terms). This ability is a crucial improvement over previous studies that only consider farm structure over one dimension, e.g. the number of dairy cows (Zimmermann and Heckelei 2012; Huettel and Jongeneel 2011), farm area (Hüttel, Margarian and von Schlippenbach 2011) or simple binary farm survival/exit models (Breustedt and Glauben 2007; Storm, Mittenzwei and Heckelei 2015; Saint-Cyr et al. forthcoming). Unlike earlier work, we can simultaneously consider multiple dimensions of the farm (e.g. cropped area for cereals, pasture, number of sheep, cows). Capturing multiple dimensions of farm activity not only allows for a fuller prediction of farm structural change, it also allows us to detect the influence of policy on a much more detailed level. It thus becomes possible to identify effects that would likely be lost due to aggregation; for example when working with aggregates for total land or revenue as a single measure of farm size. It also inherently captures farm exits which occur when all production activity is zero. Second, RNN can identify complex interactions between the different types of subsidies and production activities on a farm. For example the decision to extend dairy production might not only depend on the specific subsidies for dairy, but also on subsides and activity levels for all other production activities. Third, the relationships between farm subsidies and farmers' decisions are expected to be highly nonlinear, which RNNs are able to capture. Fourth, we might expect structural changes to have long-run, complex dynamics. RNNs are well suited to detect nonlinear long-term time dependencies, and are substantially more flexible in capturing these dynamics compared to classical autoregressive models.

These advantages are highly useful for our application. However, they come at the cost that the parameters of neural networks cannot be directly interpreted. To derive conclusions about the effects of subsidies, we therefore rely on simulations where we compare the evolution of farm structural change under different policy scenarios. These scenarios are indented to derive how past changes in the policy regime have influenced farm development (i.e., ex-post policy analysis) but can also be used to evaluate the effect of future policy changes (i.e., ex-ante policy analysis). More specifically, we define a baseline and a policy scenario where one or multiple subsidy rates or eligibility criteria are changed. We then derive predictions for the baseline and scenario using the estimated RNN. Finally, we compare the difference in the two predictions which can be interpreted as the causal effect of the change in policy scenario. Importantly, the data set provides information on almost[1] all farms in Norway allowing use to draw conclusions about the full distribution of Norwegian farms. Additionally, information about the exact geographic location of each farm is available, allowing us to determine the geographic distribution of the policy effects.

2 Data and descriptive statistics

We use a Norwegian panel dataset that covers almost all farms in Norway from 1999 to 2015. The data include over 70,000 farms active at different points in time and provide detailed information about production activities of each farm. The

[1] The dataset does not include farms considered to be hobby farms.

definition of the variable codes is provided in table 1, descriptive statistics for the dependent variables and explanatory variables are provided in table 2.

Table 1: Variables considered as dependent variables.

Variable Code	Name	Unit
BAER	Berries	daa (1/10 ha)
CERE	Cereals	daa (1/10 ha)
FODD	Fodder	daa (1/10 ha)
FRUK	Fruits	daa (1/10 ha)
GEIT	Goats	head
GRON	Vegetables Outside	daa (1/10 ha)
SAU	Sheep	head
STOR	Other Cattle	head
USAU	Sheep on outlying fields	head
VSAU	Sheep that is kept inside during winter	head
x120	Dairy cows	head
x121	Suckler cows for special meat production	head
x136	Lamb (i.e., sheep under 1 year) that is kept inside during winter	head
x140	Female goat over 1 year /Milkgoat	head
x142	Suckler goats for special meat production	head
x155	Sows for breeding with minimum one litter	head
x157	Slaughter pigs	head
x160	Laying hens at counting date /Laying hens over 20 weeks	head
x210	Fodder on arable land	daa (1/10 ha)
x211	Fodder (pasture) on arable non-fenced land	daa (1/10 ha)
x212	Fodder on non-arable fenced land	daa (1/10 ha)
x230	Potatoes	daa (1/10 ha)
x410	Dairy Cows on outlying fields	head
x420	Other livestock on outlying fields /Young cattle	head
x440	Sheep and Lamb under one year on outlying fields	head
x521	Sale of feed, High	kg
x522	Sale of feed, silage	kg

Note: The named codes are summary codes of several PT codes.

Table 2: Descriptive statistics dependent variables

	q1	q50	q99	mean	std	#nonzeros
x120	0	0	34	4	8	239,446
x121	0	0	20	1	4	91,602
x136	0	0	249	19	53	257,882
x140	0	0	0	1	8	8,207
x142	0	0	0	0	1	11,124
x155	0	0	30	1	7	32,823
x157	0	0	195	6	42	52,517
x160	0	0	1	0	1	44,678
x210	0	23	425	67	99	616,023
x211	0	0	50	3	11	213,558
x212	0	0	222	20	47	410,334
x230	0	0	60	2	18	75,443
x410	0	0	22	1	5	116,149
x420	0	0	31	2	7	163,026
x440	0	0	7	1	11	13,929
x521	0	0	17000	617	7573	40,460
x522	0	0	300300	11918	68593	112,444
BAER	0	0	3	0	6	18,851
CERE	0	0	607	45	131	245,887
FODD	0	37	554	92	131	623,817
FRUK	0	0	8	0	4	18,333
GEIT	0	0	6	0	3	20,832
GRON	0	0	11	1	16	19,240
SAU	0	0	172	15	37	262,371
STOR	0	0	80	8	18	334,885
USAU	0	0	370	28	78	227,267
VSAU	0	0	152	13	32	259,334

3 Theoretical Framework and identification

In this section we explore in which way we expect changes in the subsidy scheme to affect farmers' production mix. We derive three main policy hypotheses that provide the bases for the construction and selection of our policy variables evaluated in the empirical application. To make the discussion more accessible and to provide further insights about the working to the Norwegian subsidy scheme we illustrate the discussion using a specific policy case.

Specifically, we consider a particular subsidy, production activity and time period, namely the animal payments for sheep between 2014 and 2015. The animal payment is a coupled subsidy paid per head for different animal production activities. The payment rates are provided for the two years in table 1. Payments are differentiated by the number of sheep; for example, in 2014, payments are higher for the first 50 sheep than they are for each sheep between 51-100 sheep, and no payments are made for sheep above 300. These changes in marginal subsidy at different levels of output are known as 'kinks' in the public finance literature (Saez 2009; Chetty 2012). The differentiation according to the level of activity is an example of the size discriminatory component of the subsidy scheme that we aim to analyse. Additionally, there is a total cap on the amount of animal payments any farm can receive. The scenario we explore is the change in per head subsidy from 2014 to 2015, where the number of size categories was reduced, resulting in an increases in marginal subsidy for some operations and a decrease for others.

Table 1 Subsidy rates for the animal payments for sheep

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

3.1 *Hypothesis I: Farmers respond to a change in the marginal subsidy*

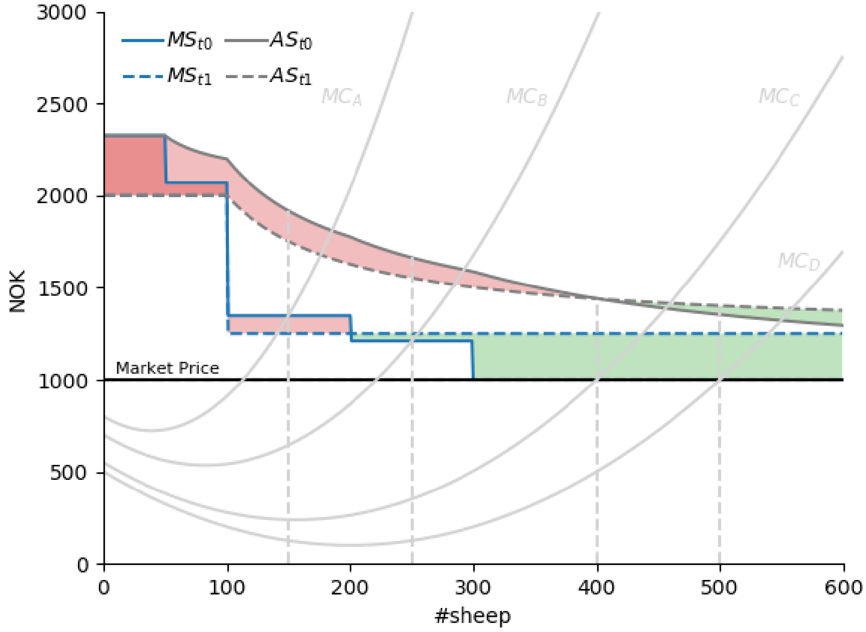
Standard economic theory suggests that profit maximizing farmers choose their level of production activities to set marginal costs equal to marginal revenue, where marginal revenue consists of the market return for the particular activity plus the subsidy. Hence, changes in "marginal subsidies" change marginal revenue and are expected to result in changes in activity levels. In this stylized setting we assume that marginal costs include costs that can be associated with a single production activity but also opportunity costs of pursuing one specific production activity instead of another or instead of renting out the required land or using the required farm labor for off-farm work.

We define "marginal subsidies" as the change in total subsidies for the next unit of production activity. We denote subsidies as S_{kt}, with k and t indexing the production activity and year, respectively. Particularly, we hypothesize that farmers operate in a state where marginal revenue equals marginal cost and that they adjust to changes in marginal subsidy levels, MS_{kt}, over time, expanding production activity if the change in marginal subsidy is positive, $MS_{kt} - MS_{kt-1} > 0$, while reducing it when it is negative, $MS_{kt} - MS_{kt-1} < 0$.

Figure 1 graphically illustrates the effects on marginal subsidies of the policy change discussed in table 1 (the solid blue line represents the marginal subsidy levels in 2014, and the dashed blue line represents the new marginal subsidy levels in 2015). This figure is intended to illustrate that the policy adjustment results in quite different changes in marginal subsidies depending on the number of sheep on the farm. For farms below 200 sheep marginal subsidies decrease between 2014 and 2015 while for farm with >200 sheep marginal subsidies increase.

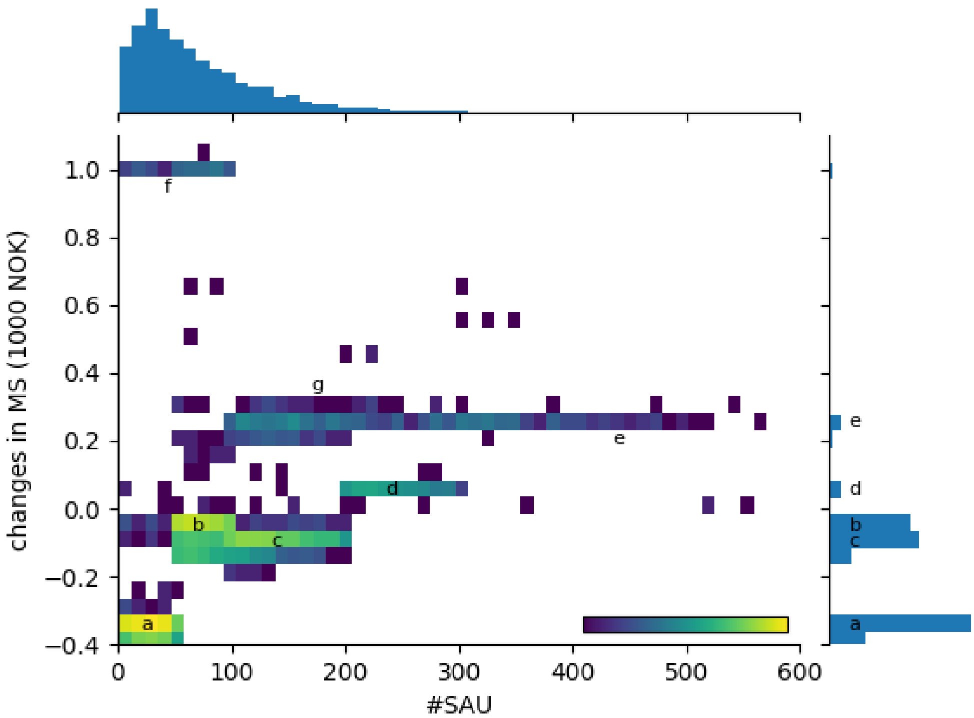
While figure 1 illustrates the conceptual effects of this one change in the policy scheme, Figure 2 illustrates the effects of the entire policy adjustment (not just the animal payments) between 2014 and 2015 on the change in marginal subsidies for sheep for the actual distribution of farms in the population. While other aspects of the policy scheme are changes in addition, it is possible to identify one-to-one relationships between figure 1 and 2. For figure 2, marginal subsides are derived by calculating total subsidies for the policy in time *t* with the observed production activities and total subsidies for the policy in *t* with the same production activities for activity *k* which is increased by one unit. The different in total subsidies provides us with the marginal subsidy MS_{kt}. The same calculation is conducted for the policy scheme in *t-1* in order to obtain S_{kt-1}. With both results we can then calculate the change in marginal subsidies, $\Delta MS_{kt} = MS_{kt} - MS_{kt-1}$.

Figure 1: Illustration of the policy changes for livestock payments for sheep between 2014 (t0) and 2015 (t1). The rates are also provided in table 1. MS_t represents marginal subsidies for sheep while AS_t represent average subsidies. Both are shifted by a hypothesized market price (or market return) set equal to 1000NOK. The shaded areas indicate decreases (red) or increases (green) in MS_t or AS_t between t0=2014 and t1=2015. The grey lines represent marginal costs (MC) for exemplifying farms in a simplified way, ignoring the multi output nature of the farm. Profit maximizing farms choose their activity level such that marginal cost are equal to Price plus MS_{t0}.



It is possible to identify a clear relation between figure 1 and 2. For example cluster a for farms in figure 2 correspond to farms with <50 sheep for which rates are reduced from 1326 to 1000 (resulting in a change in marginal subsidy equal to -326). Cluster b correspond to farms with 50<sheep<100 sheep where rates are reduced from 1070 to 1000. Cluster c correspond to farms with 100<sheep<200 sheep where rates are reduced from 347 to 250. The changes in marginal subsidies for clusters a, b and c are represented by the red areas between the blue lines in figure 1. Cluster d corresponds to farms with between 200 and 300 sheep, where rates where increase from 210 to 250 and cluster e correspond to farm with more than 300 sheep, where rates where increase from 0 to 250. These changes in marginal subsidy are represented by the green shaded areas between the blue lines in figure 1.

Figure 2: Effects of the 2014 (t0) and 2015 (t) changes in the subsidy scheme for sheep. The figure shows the resulting changes in marginal subsidies (in 1000 NOK) depending on the number of sheep (#SAU). The distribution of farms by the number of sheep is given by the blue histogram at the top, and the distribution of the change in marginal subsidies is given by the blue histogram along the right side of the figure. Inside the plot, the colors indicate the frequency of farms for a particular combination of the change in marginal subsidies and the number of sheep.



Note: The color scale represents the density of farm on a log scale with yellow indicating a higher density of farms.

Beyond this, there is additional variation in the change in marginal subsidies in figure 2 that are not represented in figure 1. This variation is caused by the increase in the total cap of the animal payments from 280.000 to 560.000 NOK from 2014 to 2015 (see table 1). This means that farms that already receive 280.000NOK in 2014 (but less than 560.000 NOK) due to other livestock see in increase in marginal subsidy equal to the 1000 NOK for <100 sheep (cluster f) and an increase in marginal subsidy equal to 250 NOK for >100 sheep (cluster g). The remaining variation around these clusters arises due to 1) the payments for replacement of labor which are also limited by an upper bound and 2) by the fact that a certain amount of livestock is required to claim pasture payments. The figure also illustrates that by far the most farms cluster in *a*, *b*, and *c*.

3.2 *Hypothesis II: Average subsidy*

Beside marginal subsidies, we also hypothesize that farmer choice of activity level may be influenced by changes in the "average subsidy": the amount of subsidies that can be attributed to a particular activity (e.g. sheep) divided by the total number of units (sheep). Above, we hypothesize that marginal revenue (including marginal subsidies) determine the precise level of activity. Here we argue that average returns (including average subsidies) are also relevant for determining farmer's decision to keep or end a certain activity. Farms are expected to exit a certain activity once average returns (including average subsidies) are below average costs. Hence we hypothesize that increases/decreases in marginal subsidies increases/decreases the likelihood of abolishing an activity altogether. Again, we understand *average costs* to include all costs that can be attributed to a certain activity and also opportunity costs of forgone alternative activities including renting out land and off-farm income.

In figure 1 average subsidies (black lines) are shown for the payments for replacement of labor for different levels of activity. For illustrative purposes we added four different marginal cost curves in figure 1. We assume that farms choose production levels such that marginal costs are equal to marginal revenues. One observation from figure 1 is that changes in average subsidies might differ from changes in marginal subsidies. Four cases can be distinguished:

1. Farm A: Marginal and average subsidies go down
2. Farm B: Marginal subsidies go up, average subsidies go down
3. Farm C: Marginal subsidies go up, average subsidies do not change
4. Farm D: Marginal and average subsidies go up

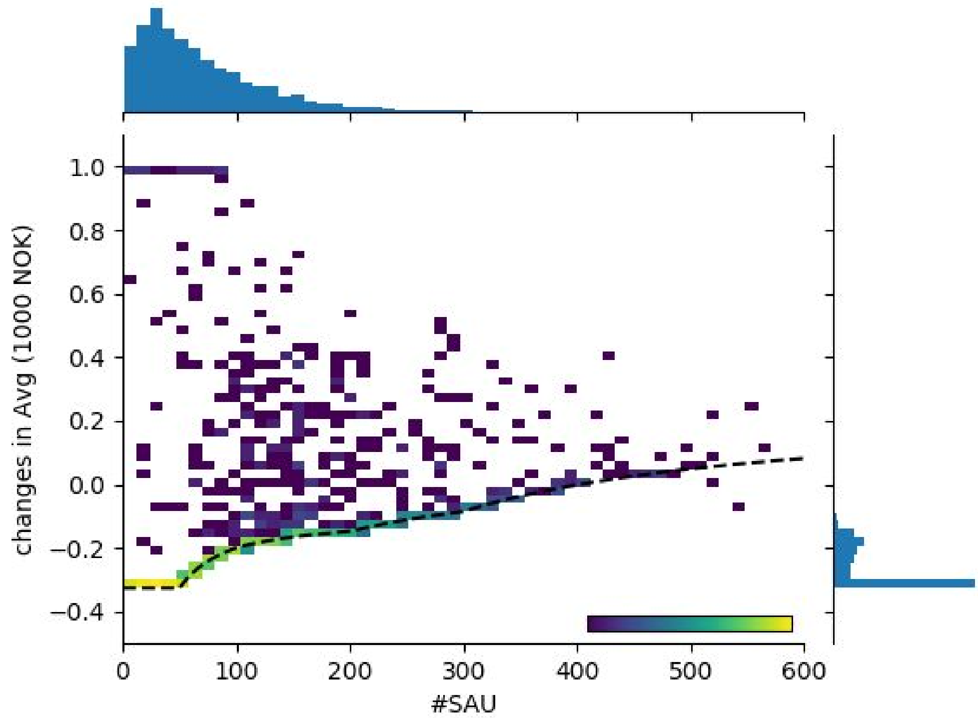
Average subsidies reflect the importance of individual production activities. They show how much the individual unit of an activity (e.g a single sheep/cow/hectare wheat) contributes to the overall amount of subsidy received. For example, imagine a farm that has enough dairy cows to receive the upper bound for the animal payments and the payments for replacement of labor with the dairy cows alone. The marginal subsides and the average subsidy for sheep would both be zero. A second farm might be at the upper bound for animal payments and the payments for replacement of labor when dairy cows and sheep are taken together. Marginal subsidies for sheep would also be zero while average subsidies for sheep are positive. Hence for the second farm, sheep are relatively more important as they contribute more to the overall subsidy income and the farm is less likely to reduce the number of sheep than the first farm when average subsidies decrease.

As with marginal payments, we explore the implications of the policy change for average payments (presented in table 1 and figure 1) considering the entity farm population (figure 3). Additionally we take into account all other policy changes that happened between 2014 and 2015 to illustrate their potential interaction. To calculate the average subsidy empirically we:

1) calculate actual subsidies
2) calculate subsidies with the activity removed (i.e. sheep=0)
3) calculate the difference between 1) and 2) to get the amount of subsidies that can be attributed to on particular activity (sheep)
4) Divide the difference by the number of sheep to get the average subsidy per head
5) Calculate the change in average subsidies by performing steps 1-4 form the subsidy scheme in *t* and *t-1* and calculate the difference between the two.

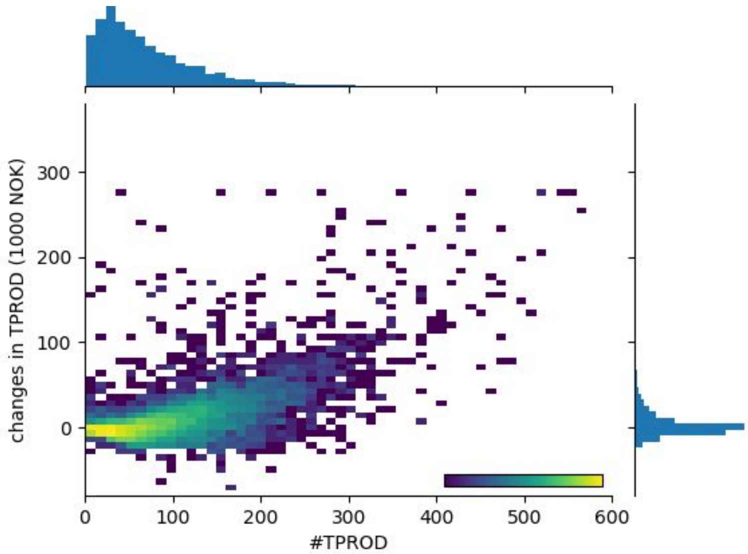
Figure 3 shows the empirical distribution of the change in average subsides for the policy change between 2014 and 2015. Again, we see a clear relation to figure 1. In figure 1 the change in average subsidies is the area between the black solid and dotted line, reflecting the change in average subsidy between 2014 and 2015. This change is also reflected in figure 3 (dotted line) and most of the farm clusters around it. The remaining variation arises from the cap of the animal payments, the payments for the replacement of labor which are also limited by an upper bound and by the fact that a certain amount of livestock is required to claim pasture payments.

Figure 3: Effects of the 2014 (t0) and 2015 (t) changes in the subsidy scheme for sheep. The figure shows the resulting changes in average subsidies (in 1000 NOK) depending on the number of sheep (#SAU). The distribution of farms by the number of sheep is given by the blue histogram at the top, and the distribution of the change in average subsidies is given by the blue histogram along the right side of the figure. Inside the plot, the colors indicate the frequency of farms for a particular combination of the change in average subsidies and the number of sheep.



Note: Observations without any sheep are excluded from the figure. The color scale represents the density of farm on a log scale with yellow indicating a higher density of farms.

Figure 4: Effects of the 2014 (t0) and 2015 (t) changes in the subsidy scheme. The figure shows the resulting changes in total subsidies (in 1000 NOK) depending on the total amount of subsidies (in 1000 NOK). The colors indicate the frequency of farms for a particular combination of the change in average subsidies and the number of sheep.



Note: The color scale represents the density of farm on a log scale with yellow indicating a higher density of farms.

3.3 Hypothesis III: Overall subsidy amount

Finally, we hypothesize that an increase in the overall subsidy amount received might have positive effects on farm activities due to increase of farm liquidity. Increasing farm liquidity might lower marginal cost (for example by creating the opportunity for new investments or reducing their costs). Therefore, we consider the change in total subsidies from t0 to t1 under the condition that the farm would not have changed its production activity levels. Figure 4, illustrates the distributional effects of the policy change in 2014/15 for the entire farm population.

4 Methodology

We apply a recurrent neural network (RNN) approach to estimate how changes in subsidy level and composition affect farm structure. Specifically, we build a RNN using Long Short Term Memory (LSTM) cells (Hochreiter and Schmidhuber 1997; Goodfellow et al. 2016, Chapter 10). The crucial feature of an RNN is that past information is carried across time using a *cell state* vector. This cell state vector as well as the new incoming information from the explanatory variables in each time step is processed in the LSTM cells using different *gates*. Intuitively, these gates, further explained below, determine which past and new information is forgotten or maintained. A crucial feature is that the model learns by itself in which way forgetting and maintaining this information should take place.

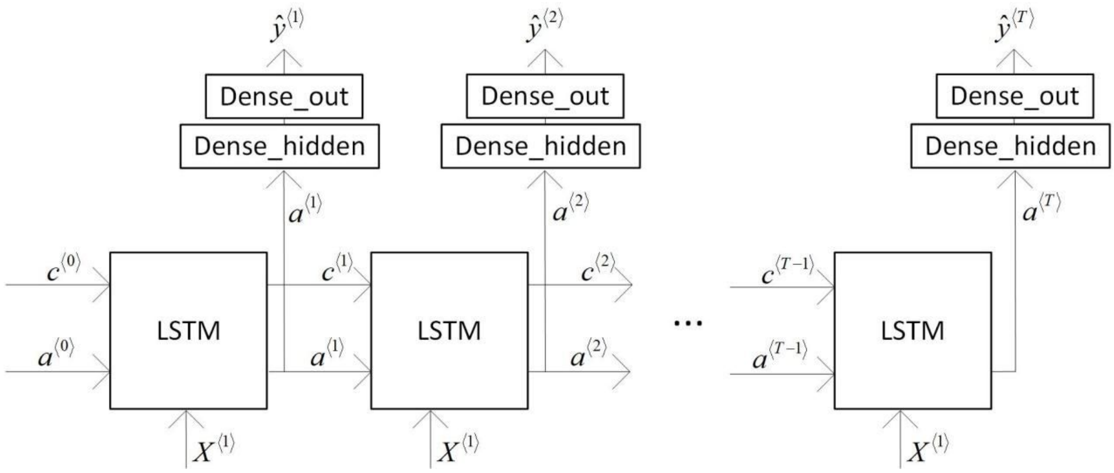
This approach differs from a classical autoregressive (AR) process in two ways. First, in an AR process we need to a priori determine a certain lag structure. In contrast, a RNN decides by itself for how long and which elements of information are maintained. The model can for example decide that certain information is important to maintain over a longer period while other information can be forgotten more quickly. Second, in an AR process, lagged dependent variables enter as explanatory variables, usually in a linear way (i.e. in AR(2) the last two lags of the dependent variable). In a RNN, past information is mapped into a high dimensional cell state vector that can in principle capture more complex past developments, i.e. not just the last two lags but also their interaction or the trajectory between the two. It can also transform the information in arbitrary ways. As a concrete example in our context, if a farm stops dairy production after a couple of years it might not be relevant to store the exact number of cows in each of these prior years; instead it might be sufficient to store the information that the farm had dairy cows and maybe the maximum number of cows, which might serve as information about the potential capacity of the stable. If this information turns out to be relevant, a RNN could learn to encode this information in the cell state vector and remember it over a long time. This approach is substantially more flexible and powerful compared to the restrictive AR process. Crucially, we do not need to specify in which way the information is encoded in the cell state vector a priori; instead the model learns itself how to do this in an optimal way. On the other hand, it should also be clear that this implies that we neither have control over the way information is encoded, if it is maintained for many periods or quickly forgotten, nor is it possible to easily understand which information the models decided to encode in a certain way and maintains.

Technically a RNN is a specific from a neural network (NN). Before turning to the RNN we aim to provide a brief introduction of simple *dense* (or *feed forward*) NNs, which provide the building blocks used in a RNN. For a in depth introduction to NN we refer to Hastie et al. (2009) or Goodfellow et al. (2016). A dense NN maps an input vector x in an output vector y similarly to a regression. Importantly however, x and y refer to one single observation implying that we can have a dependent variable *vector* for one observation (e.g. several farm activities in our application) instead of a scala value as we usually have for regression. The mapping between x and y in a NN is done using different layers in a chain like structure. For example a simple NN with three layer is given by $y = f(x) = f^{(1)}(f^{(2)}(f^{(3)}(x))$, where each layer is given by $f^{(k)}(x) = h^{(x)} = g^{(k)}(W^{(k)\top}h^{k-1} + b^{(k)})$ with $h^{(0)} = x$ and $W^{(k)}$ and $b^{(k)}$ being a matrix and vector, respectively, of unknown coefficients. The link function $g^{(k)}$ needs to be specified, typical choices are the rectified linear unit (relu) or a tanh transformation function. Simple dense NN are well suited for cross sectional data, i.e. mapping x to y. RNN in contrast allows to deal with time series data or panels where we observe $x^{\langle 1\rangle}, x^{\langle 2\rangle}, ..., x^{\langle T\rangle}$ and $y^{\langle 1\rangle}, y^{\langle 2\rangle}, ..., y^{\langle T\rangle}$ and aim to model dynamic relationships over time. For this RNN often use LSTM cells which allow to pass information across time period using a cell state vetore, $c^{\langle t\rangle}$ that aims to capture all past and current information $x^{\langle 1\rangle}, x^{\langle 2\rangle}, ..., x^{\langle t\rangle}$ necessary to predict $y^{\langle t\rangle}$. Formally an LSTM cell can be described as follows. In the first time period, $c^{\langle t\rangle}$, is randomly initialized. In each following time step the cell state $c^{\langle t\rangle}$ is updated based on the incoming information $x^{\langle t\rangle}$. The updating uses several *gates*, defined below. First a new candidate for the cell state $\tilde{c}^{\langle t\rangle}$ is proposed based on

$\tilde{c}^{\langle t\rangle} = \tanh\left(W_c x^{\langle t\rangle} + U_c a^{\langle t-1\rangle} + b_c\right)$. A cell state is updated by $c^{\langle t\rangle} = \Gamma_u * \tilde{c}^{\langle t-1\rangle} + \Gamma_f * c^{\langle t-1\rangle}$ with $\Gamma_f = \operatorname{sigmoid}\left(W_f x^{\langle t\rangle} + U_f a^{\langle t-1\rangle} + b_f\right)$ (forgot gate) and $\Gamma_u = \operatorname{sigmoid}\left(W_u x^{\langle t\rangle} + U_u a^{\langle t-1\rangle} + b_u\right)$ (update gate). The output for the current time step is then calculated as $a^{\langle t\rangle} = \Gamma_o * c^{\langle t\rangle}$ with $\Gamma_o = \operatorname{sigmoid}\left(W_o x^{\langle t\rangle} + U_o a^{\langle t-1\rangle} + b_o\right)$ (output gate). All Ws are trainable parameter matrices with dimension (dim_LSTM, input_dim), all Us are trainable parameter matrices with dimension (LSTM_output_dim, LSTM_output_dim) and all bs are trainable base vectors with dimension (output_dim,1).

In our specific implementation, the output of the LSMT cell $a^{\langle t\rangle}$ is then fed into two *dense* NN layers. The first dense layer (dense_hidden) is specified with a relu activation function, $h^{\langle t\rangle} = \operatorname{relu}\left(W_{d1} a^{\langle t\rangle} + b_{d1}\right)$, where W_{d1} is a (dim_hidden,LSTM_output_dim) trainable parameter matrix all b_{d1} are trainable base vectors with dimension (hidden,1). The final dense output layer is then given by $\hat{y}^{\langle t\rangle} = \operatorname{relu}\left(W_{d2} h^{\langle t\rangle} + b_{d2}\right)$, where W_{d2} is a (dim_out, dim_hidden) trainable parameter matrix all b_{d2} are trainable base vectors with dimension (dim_out,1). We found that using a linear activation in the output layer during an initial training and then switching to a relu activation achieved superior training results. Figure 5 illustrates the model setup graphically.

Figure 5: Model architecture of the Recurrent Neural Network (RNN)



The parameters are estimated by minimizing a weighted square error loss function specified as $L = \sum_{t=1}^{T} w_t \left(\hat{y}^{\langle t \rangle} - y^{\langle t \rangle}\right)' \left(\hat{y}^{\langle t \rangle} - y^{\langle t \rangle}\right)$, where $y^{\langle t \rangle}$ and $\hat{y}^{\langle t \rangle}$ are observed and predicted outcomes, respectively. The weights w_t are defined such that $w_1 = 0$, $w_2 = 0.5$ and $w_3, \ldots, w_T = 1$. The intentions of this weighting scheme is to give lower weights to prediction very early in the sequence where the algorithm still needs to work with random assignment of the initial $c^{\langle 0 \rangle}$ and $a^{\langle 0 \rangle}$ vectors and could not build up a meaningful state and output vectors.

To monitor overfitting we split the dataset by farms into a training and development set. This allows us to monitor the loss (i.e. prediction error) in both set. An increasing divergence in the loss, i.e. if the loss in the training set decreases while it increases in the development set indicates and overfitting of the model. To further reduce the tendency of overfitting, we apply "dropout" (Srivastava et al. 2014) in the dense hidden layer, with a dropout rate equal to 0.1. Intuitively, dropout eliminates randomly certain elements of W_{d1}, at each optimization iteration. The model is implemented in Keras (Chollet and Others 2017) and trained using RMSprop solver (Hinton, Srivastava and Swersky 2014). The learning rate of the RMSprop solver as well as the dropout rate are determined in a random hyperparameter grid search.

We considered a 27 dimensional target (dependent) vector, $y^{\langle t \rangle}$, specifying the various production activities of a farm (see table 1). In the final scenario analysis, these activities are forecasted for each farm and year. The model takes as input 82 features (explanatory variables), $X^{\langle t \rangle}$, they consist of the 27 target variables from the previous year, the age of the farm holder as well as the change in the marginal and average subsidy (as described above) for the considered target variables[2], as well as the change in total subsidy for the different subsidy types. The concrete specification of the model is given in table 3. This specification of the model has 5,611,547 trainable parameters which are tuned during training.

Table 3: Model architecture of the Recurrent Neural Network (RNN)

	Dimensions	**Parameters**	
Input Layer	dim_input=82	---	
LSTM	dim_lstm=1024	4,534,272	*4x Ws with 1024x82=83968* *4x Us with 1024x1024=1048576* *4x bs with 1024x1=1024*
dense_hidden	dim_hidden=1024	1,049,600	*W with 1024x1014=83968* *b with 1024x1=1024*
dense_out	dim_out = 27	27,675	*W with 27x1014=27648* *b with 27x1=27*
	Total	**5,611,547**	

5 Policy analysis

The policy analysis is performed via scenario analysis using the trained RNN. Specifically, we predict outcomes over farm size and composition one-year ahead, given a baseline and a scenario involving policy change. We use data from 1999 to 2015 for estimation and testing, and then use the trained model to make predictions for 2016. In the baseline we assume that the subsidy scheme does not change (i.e.

[2] For some codes the marginal and average subsidy are by definition the same. In this case they are only considered once (for example x211 and x212).

that it will be the same in 2016 as in 2015). We make specific changes to the subsidy scheme such that we can analyse the resulting adjustments predicted by the model by comparing the baseline prediction with the scenario prediction.

With such a baseline/scenario comparison approach we are able to consider arbitrary policy scenarios at a very detailed level. We can investigate multiple/broad changes affecting many activities at once allowing us, for example, to analyze complete subsidy reform packages proposed by different parties during the yearly negotiations to adjust the subsidy scheme. It is also possible to analyze the effects of a very specific change in the subsidy scheme, for example one that affects only one particular subsidy type, production activity, size class or region. In light of our aim to analyse the specific effects of change to the size discriminatory component of the subsidy scheme, we focus on changes to the animal payments for sheep that are also considered before in section 2. In the base scenario we assume that the subsidy scheme in 2016 is the same as in 2015. The baseline predictions are compared to two different scenarios. In the first scenario "flat rate," we assume that the size-discriminatory component of the subsidy is removed and all farms receive the same payment per sheep, where the rate up to 100 sheep is reduced and the rate for more than 100 is increased (see table 4). In the second scenario, "size discriminatory," we assume that the size discriminatory component of the subsidy is enhanced, such that payment up to 100 sheep are increased while no payments are made for sheep after 100 (table 4).

Table 4 Subsidy rates for animal payments for sheep

	NOK/Head				
Size (head)	2014	2015	2016 *base*	2016 scenario *flat rate*	2016 *scenario size discriminatory*
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	

Figure 6 shows the results for the baseline/scenario comparison for the two different scenarios considered (scenario “flat rate” left column, scenario “size discriminatory” right column). On the x-axis and the histogram on top is the number of sheep in 2015 (farms without any sheep are excluded from the figure). On the y-axis and the vertical histogram on the right is the change in units between the baseline and the scenario of one particular production activity. The heat map in the middle shows the number of observations across these two dimensions.

Figure 6 shows four selected different target variables. The results for all 27 target variables are provided in the appendix. One reassuring result indicated by Figure 6 (but more clearly be appendix A1-A7), is that the change in the policy scheme for sheep seems to largely affect activities directly related to sheep or activities where a cross subsidy effect is intuitive. For example, in Figure 6 we find that the subsidy change in sheep primarily affects sheep (SAU) and the fodder (FODD) activities. For cereals (CERE) we find a weak cross correlation effect and for dairy cows (x120) we find almost no effect.

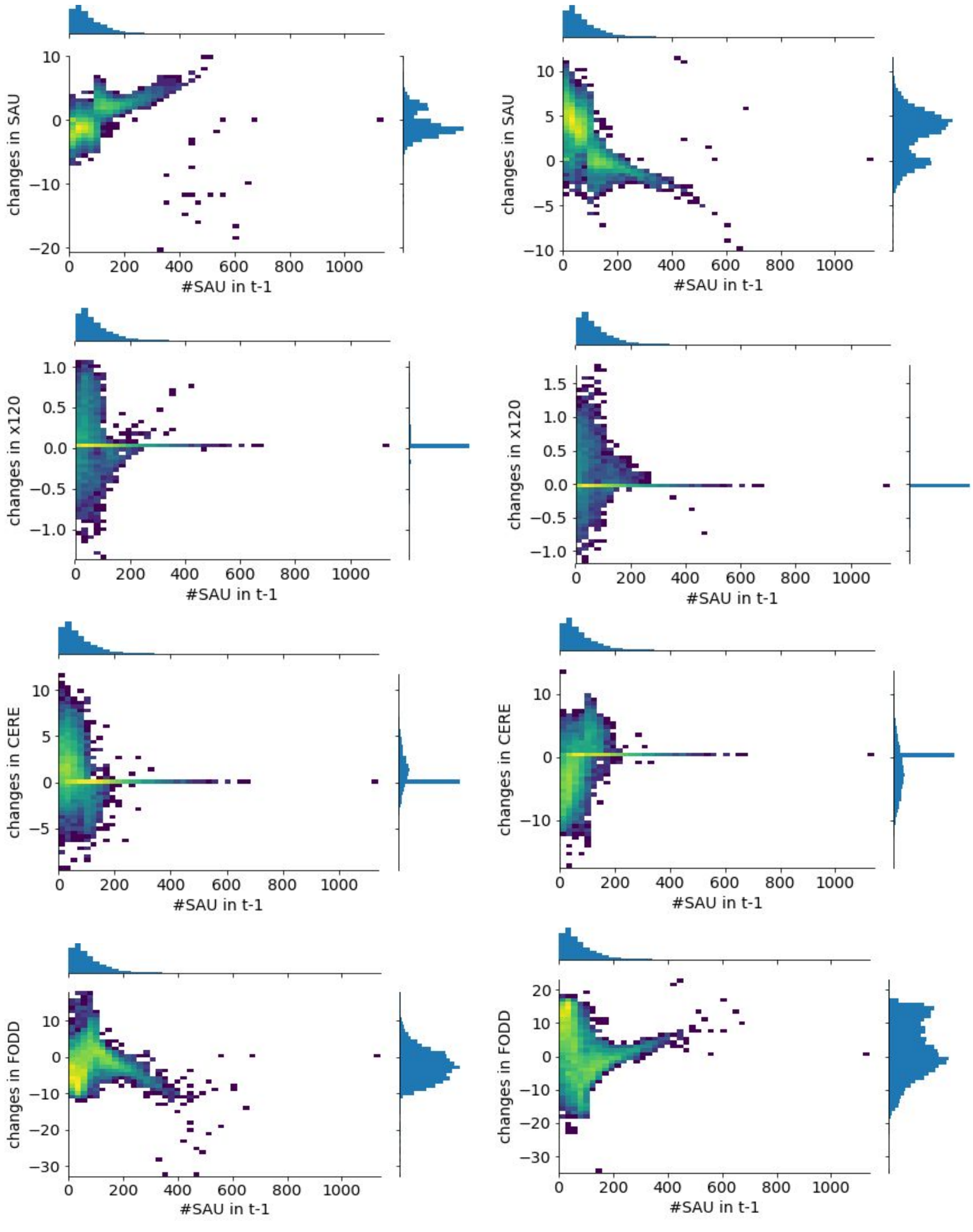
In the first row we see the effects of the policy change in the two scenarios with respect to the number of sheep (SAU). For the scenario “flat rate” we see that compared to the baseline, small farms with less than 100 sheep reduce their number of sheep as expected, because both their average and marginal subsidies decrease. For farms with more than 100 sheep, we see an increase compared to the baseline. The positive effects get stronger the larger the farm size in 2015 (to the right on the axis), which is intuitive since a change in the subsidy rate for size classes of 0-50 and 51-100 affects the average subsidies for larger farms (see figure 3). This corresponds well to the curved tail in the change in number of sheep (first row, left column in figure 6). It does not however, correspond well to the rather drastic change in the effects for farms that have right around 100 sheep. Here, it is more likely that this shift is largely driven by changes in the marginal rate which shifts substantially in the area around 100 sheep. This result underlines that changes in marginal subsidies rates are indeed important for farmers' decision, and reflects the bunching that occurs around kink points seen in the public finance literature (Saez 2009; Chetty 2012). For the scenario "size discriminatory" (right column) we find more or less the picture that we would expect. Farms with fewer than 100 sheep profit most from the policy change and we find a positive change in the number of sheep, while farms with more than 100 sheep lose, and we find a reduction in the number of sheep.

The second row displays results of the policy change with respect to the number of dairy cows. The result indicates that the changes in the subsidy rate for sheep do not have a substantial influence on the number of dairy cows. The vertical histogram on the right illustrates that there are few farms that are predicted to change the number of their cows between the baseline and either scenario. For cereals (third row), we find some small cross effects where changes in the subsidy scheme for sheep lead to changes in the cereals production. In the scenario "flat rate," we found that some small farms increase their cereal production area while in "size discriminatory" scenario we observe the opposite. Large farms in terms of sheep show almost no reaction to the policy change in both scenarios.

For fodder production (third row) we find a larger and more diverse effect and when comparing scenario "flat rate" and "size discriminatory," we find almost a reversed effect. The diversity of the effects might be explained by different initial levels of fodder production that require different changes when the number of sheep is changed and/or it could be driven by interactions with other production activities that might influence the change in marginal and average subsidies with respect to fodder. For example, a farm that cannot claim the fodder payment because they do not have sufficient livestock has little incentive to increase fodder production when extending sheep production, compared to a farm that already has sufficient livestock to claim fodder payments.

Figure 6: Comparison between baseline and scenario for two different types of scenarios. (SAU=sheep, x120=milk cows, CERE= cereals, FODD= fodder)

Scenario flat rate *Scenario size discriminatory*



Note: The color scale represents the density of farm on a log scale with yellow indicating a higher density of farms.

To put the findings into perspective, it should be emphasized that we do not provide any information to the model that the changes in the marginal or average subsidies for sheep are in any way linked to the specific sheep code. The model itself has determined that the change in the subsidy scheme with respect to sheep is more important for the activities related to sheep production and less relevant for variables related to other activities. Appendix A1-A7 shows that we find no effect for most of the other activities except for example codes x136 (lambs kept inside during the winter), USAU (sheep on outlying fields) and VSAU (sheep kept inside during winter), which also represent production activities with respect to sheep. It should also be mentioned that we find a similar pattern when changing subsidy rates for dairy production. In that case, the sheep codes are largely unaffected while we find a clear response for the codes related to dairy production.

6 Conclusion

In this paper we employ a RNN to analyze the effects of subsidies on farmers behavior in Norway. The approach operates on a very detailed level, providing results at an individual farm and activity level, and is able to evaluate a highly detailed representation of the subsidy scheme. The approach exploits the complexity of the Norwegian subsidy scheme in order to overcome endogeneity issues that usually complicate the identification of policy effects. The results indicate that despite allowing the model to determine the relationship between subsidies and output on its own directly from the data, the RNN is capable of detecting plausible relationships between subsidies and agricultural activities, without making any prior assumptions that relate specific subsidies to production activities. Additionally, it is able to pick up attributes like bunching that would be predicted by economic theory. The approach can be useful for providing detailed policy analysis that can help to assess the consequences of policy reforms discussed during the yearly negotiations to adjust the Norwegian subsidy scheme.

Despite the advantages of the approach and the encouraging results, several limitations or drawbacks of the approach should be pointed out. The complexity of the neural network (with >5.5 million parameters) does not allow a direct interpretation of the parameters, as usually done in classical econometric approaches. Further, as already pointed out in section 4, the model is highly capable of capturing complex relationships in the data. However, it is hardly possible to assess exactly which effects are picked up by the model. For the results, this implies that it is very difficult, if not impossible, to assess why the model makes certain predictions. Two aspects add to this challenge. First, working on an individual farm level results in very heterogeneous effects. In general, this is an important strength of the approach, but it also makes the interpretation of the results more complex. Secondly, the large number of farm activities that are considered in parallel, also a strength of our approach compared to previous studies in the literature, creates an additional challenge to process and inspect results in each of these dimensions. Overall this results in a Black-box character of the model. However, this Black-box character is largely inherent in the desire to work on a detailed level as well as the ability to capture complex (nonlinear/dynamic) behavior and similar problems would occur in complex and flexible econometric approaches.

It should also be reflected that in our context, due to the way we defined our explanatory variables, a direct interpretation of the parameters (or marginal effects) is less relevant. We consider the change in marginal, average and total subsides. Even though it would be possible to derive marginal effects for those variables, they are not really of interest to policy makers[3]. First because policy makers cannot control them directly and secondly, because they do not necessarily have a meaningful ceteris paribus interpretation. For example, a ceteris paribus effect of the change in average subsidy has very limited relevance because it is inherently related to changes in marginal subsidies and total subsidies.

In order to improve the model for policy analysis, further extensions should include additional explanatory variables. Particularly, detailed information about

[3] Similarly as in a simple logistics or probit binary choice regression they would also depend on farm characteristics. Given the flexibility of the RNN the dependencies can be more complex involving dynamics and non-linearity's.

input and output prices as well as detailed geophysical variables are interesting candidates.

7 References

Ahearn, M.C., J. Yee, and P. Korb. 2005. "Effects of Differing Farm Policies on Farm Structure and Dynamics." *American journal of agricultural economics* 87(5):1182–1189.

Breen, J.P., T.C. Hennessy, and F.S. Thorne. 2005. "The effect of decoupling on the decision to produce: An Irish case study." *Food policy* 30(2):129–144.

Breustedt, G., and T. Glauben. 2007. "Driving Forces behind Exiting from Farming in Western Europe." *Journal of Agricultural Economics* 58(1):115–127.

Chetty, R. 2012. "Bounds on Elasticities With Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply." *Econometrica: journal of the Econometric Society* 80(3):969–1018.

Chollet, F., and Others. 2017. "Keras (2015)."

Goodfellow, I., Y. Bengio, A. Courville, and Y. Bengio. 2016. *Deep learning*. MIT press Cambridge.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer Science & Business Media.

Hinton, G., N. Srivastava, and K. Swersky. 2014. "Neural Networks for Machine Learning. A bag of tricks for mini-atch gradient descent. Lecture notes 6b." Available at: https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf.

Hochreiter, S., and J. Schmidhuber. 1997. "Long short-term memory." *Neural computation* 9(8):1735–1780.

Huettel, S., and R. Jongeneel. 2011. "How has the EU milk quota affected patterns of herd-size change?" *European Review of Agricultural Economics*. Available at: https://academic.oup.com/erae/article-abstract/38/4/497/408393.

Hüttel, S., A. Margarian, and V. von Schlippenbach. 2011. "Growth and exit in the agricultural sector." In *European Association of Agricultural Economists 2011* *congress*. pp. 30–08.

Key, N., and M.J. Roberts. 2009. “Nonpecuniary Benefits to Farming: Implications for Supply Response to Decoupled Payments.” *American journal of agricultural economics* 91(1):1–18.

Saez, E. 2009. “Do tax filers bunch at kink points? Evidence, elasticity estimation, and salience effects.” *University of California at Berkeley. Mimeo*. Available at: http://elsa.berkeley.edu/~saez/saez09bunching.pdf.

Saint-Cyr, L.D.F., H. Storm, T. Heckelei, and L. Piet. forthcoming. “Heterogeneous impacts of neighbouring farm size on the decision to exit: evidence from Brittany.” *European Review of Agricultural Economics* jby029. Available at: http://dx.doi.org/10.1093/erae/jby029 [Accessed October 22, 2018].

Schmidhuber, J. 2015. “Deep learning in neural networks: an overview.” *Neural networks: the official journal of the International Neural Network Society* 61:85–117.

Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. 2014. “Dropout: a simple way to prevent neural networks from overfitting.” *Journal of machine learning research: JMLR* 15(1):1929–1958.

Storm, H., K. Mittenzwei, and T. Heckelei. 2015. “Direct Payments, Spatial Competition, and Farm Survival in Norway.” *American journal of agricultural economics* 97(4):1192–1205.

Zimmermann, A., and T. Heckelei. 2012. “Structural Change of European Dairy Farms - A Cross-Regional Analysis: Dairy Farm Structural Changes in Europe.” *Journal of Agricultural Economics* 63(3):576–603.

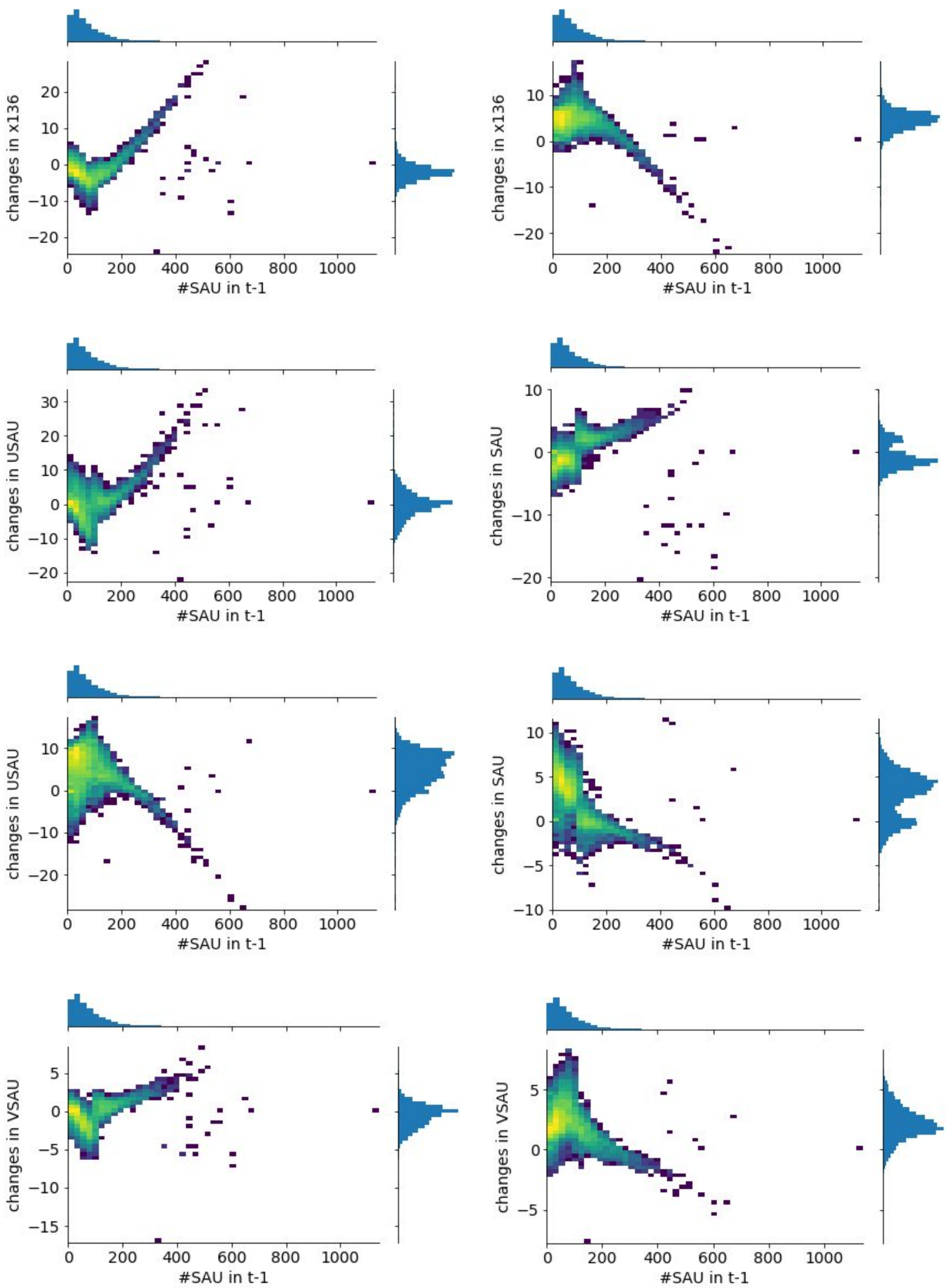
Acknowledgement

The work was supported by the German Research Foundation (DFG) – Deutsche Forschungsgemeinschaft [grant number STO 1087/1-1].

8 Appendix

A1: Comparison between baseline and scenario prediction for two different types of scenarios. (x136=lambs kept inside during the winter, SAU=sheep, USAU=virtual property sheep on outlying fields, VSAU=Virtual property sheep that is kept inside during winter)

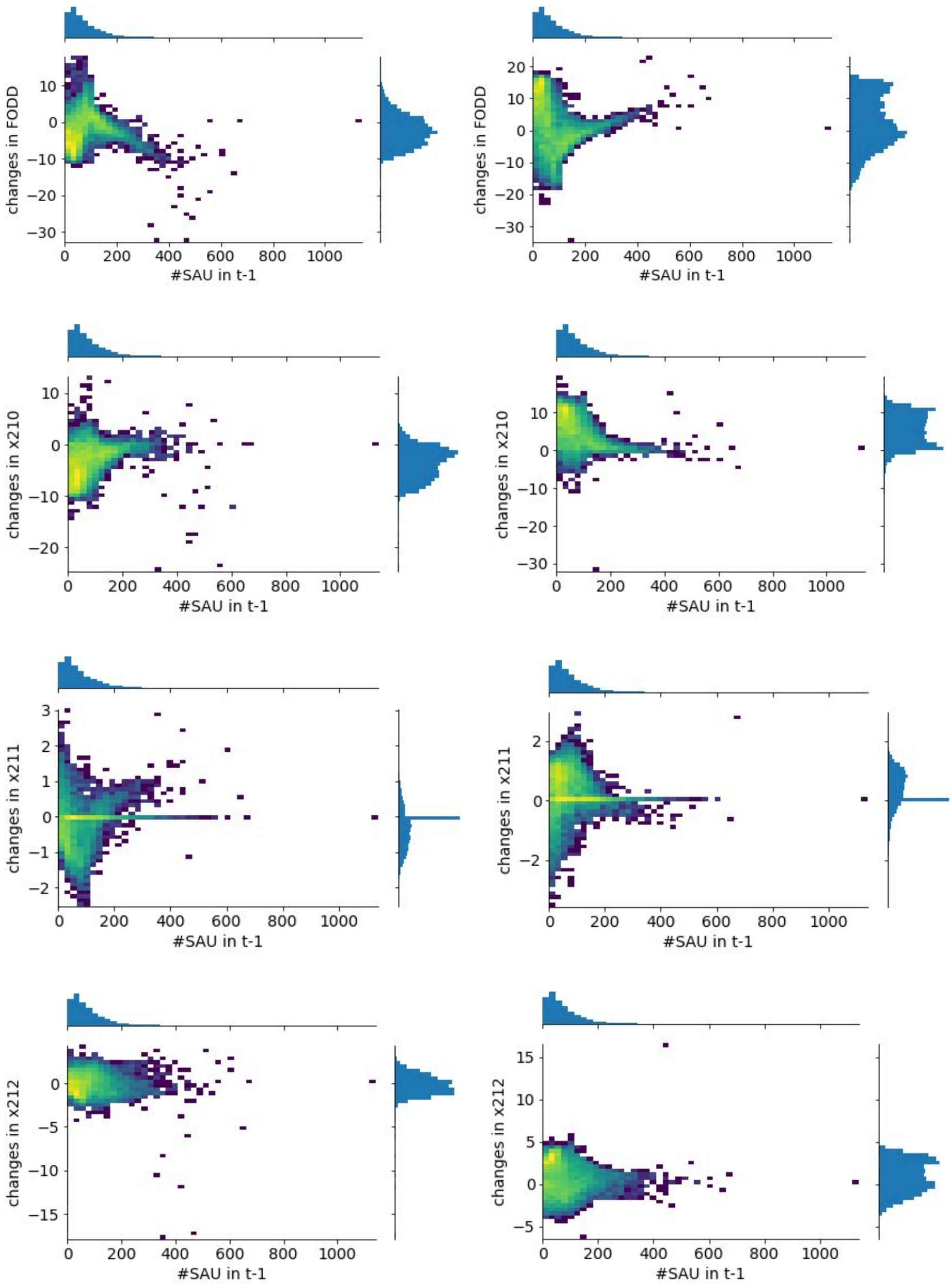
Scenario flat rate *Scenario size discriminatory*



Note: The color scale represents the density of farm on a log scale with yellow indicating a higher density of farms

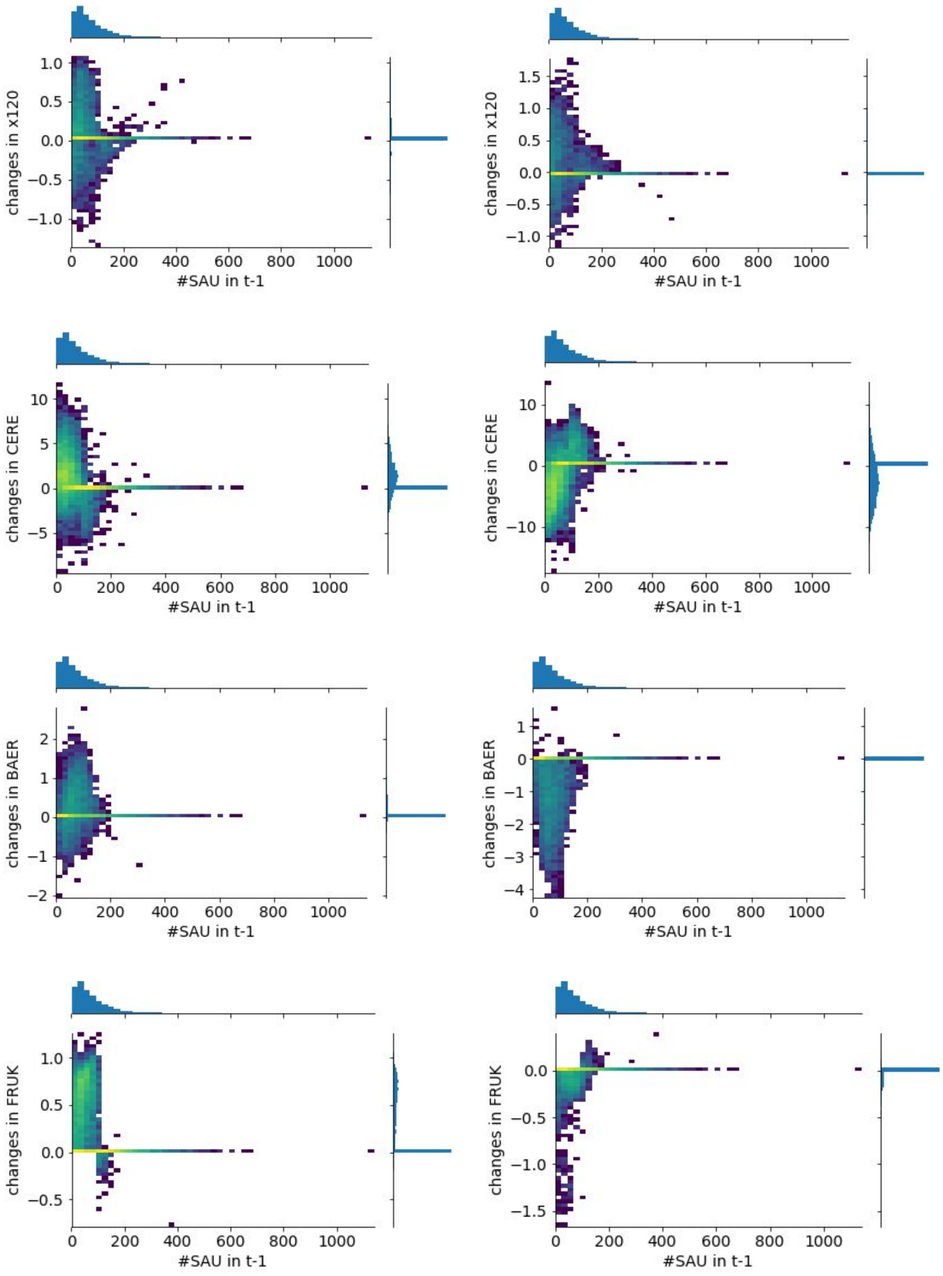
A2: Comparison between baseline and scenario prediction for two different types of scenarios. (FODD=fodder, x210=fodder on arable land, x211=fodder (pasture) on arable non-fenced land, x212=fodder on non-arable fenced land)

Scenario flat rate | *Scenario size discriminatory*



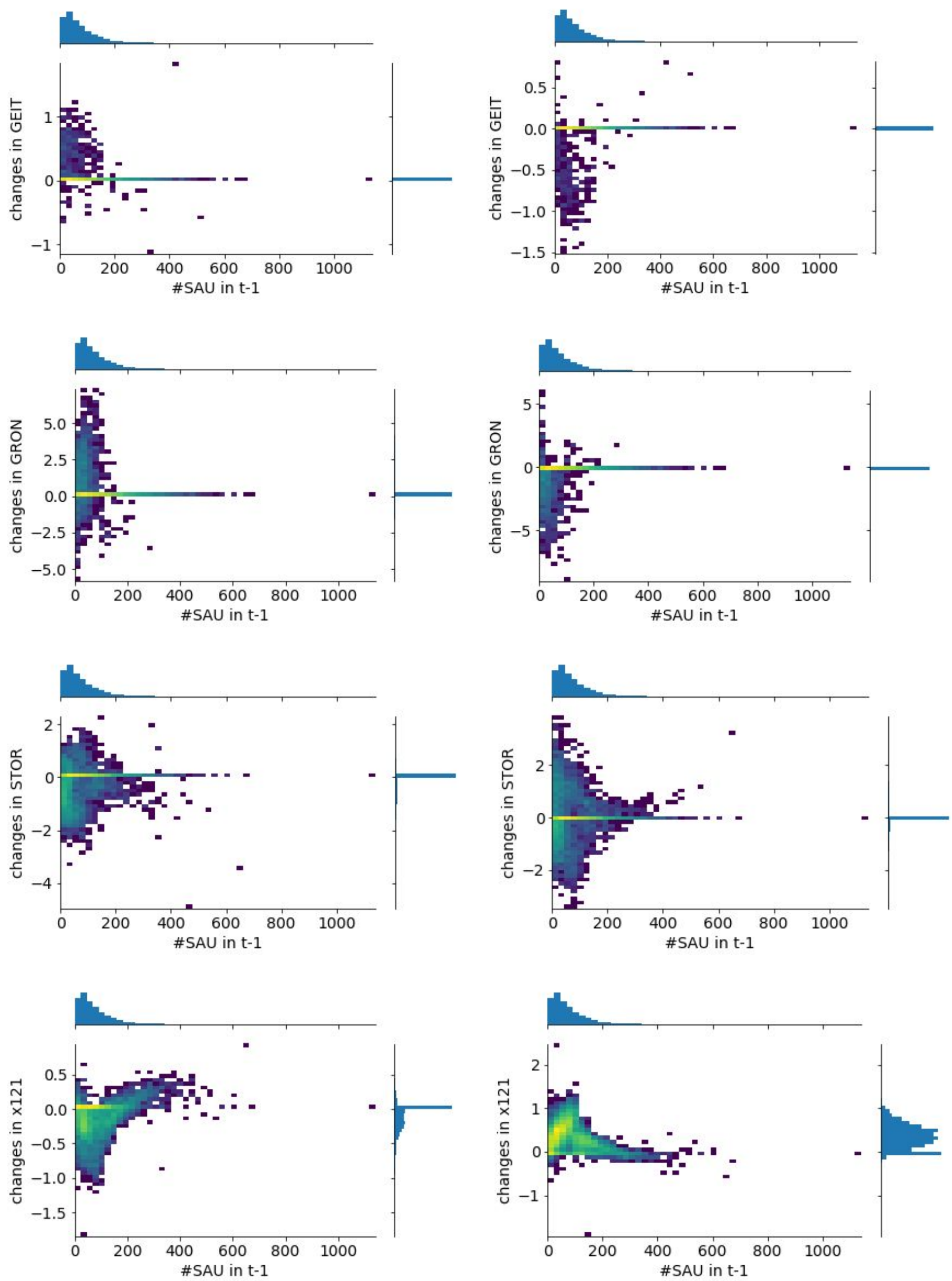
A3: Comparison between baseline and scenario prediction for two different types of scenarios. (x120=milk cows, CERE=cereals, BAER=berries, FRUK=fruits,)

Scenario flat rate *Scenario size discriminatory*



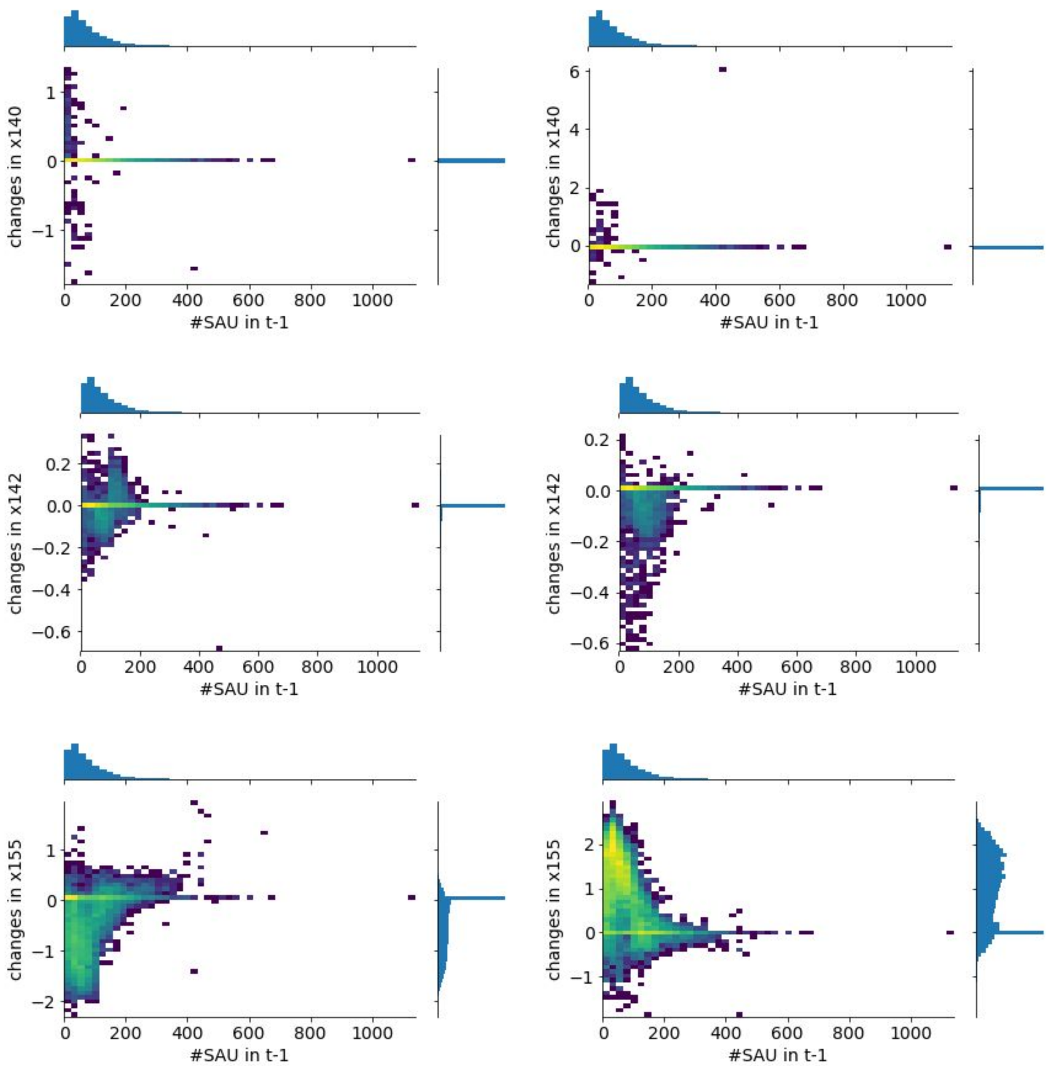
A4: Comparison between baseline and scenario prediction for two different types of scenarios. (GEIT=goats, GRON=manure, STOR=other cattle, x121= suckler cows for special meat production)

Scenario flat rate *Scenario size discriminatory*



A5: Comparison between baseline and scenario prediction for two different types of scenarios. (x140=milkgoats over 1 year, x142=suckler goats for special meat production, x155=sows for breeding with minimum one litter)

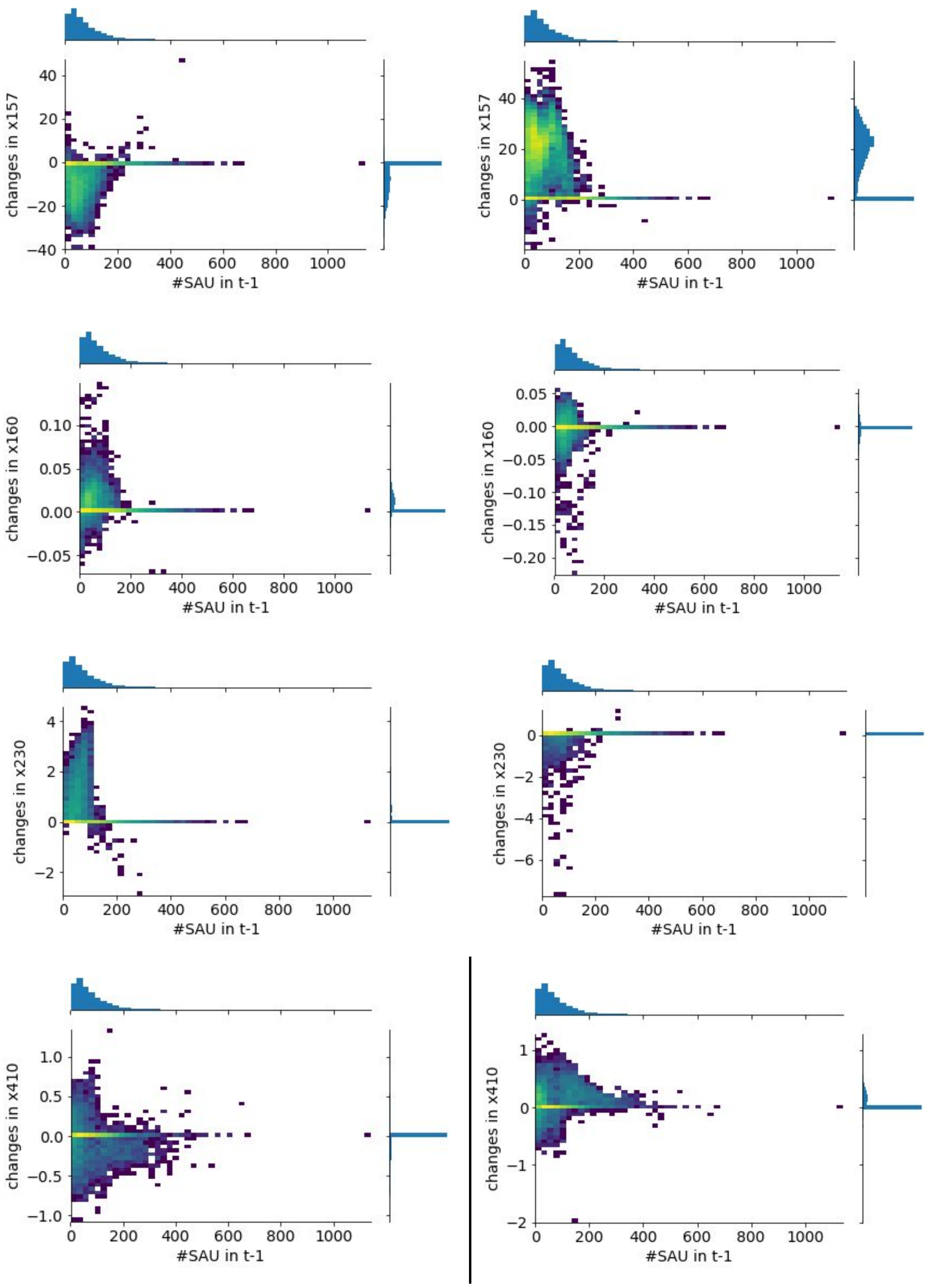
Scenario flat rate *Scenario size discriminatory*



Note: The color scale represents the density of farm on a log scale with yellow indicating a higher density of farms

A6: Comparison between baseline and scenario prediction for two different types of scenarios. (x157=slaughter pigs, x160=laying hens over 20 weeks, x230=potatoes, x410=dairy cows on outlying fields)

Scenario flat rate *Scenario size discriminatory*



A7: Comparison between baseline and scenario prediction for two different types of scenarios (x420=young cattle, x440=goats on outlying fields, x521=sale of feed, high, x522=sale of feed, silage)

Scenario flat rate *Scenario size discriminatory*