Potentials and benefits of panel data

Alina Sinisalo and Heikki Mäkinen

MTT Agrifood Research Finland
alina.sinisalo@mtt.fi, heikki.makinen@mtt.fi

Contents

- Introduction
- Benefits of panel data
- Limitations of panel data
- Example for dairy farms
  - Farm level panel data
  - Model specification
  - Model results

Introduction understand changes

- To study long-term effects or development information from several years are needed.
- Panel data allows studying the change by connecting information of an individual with several measurements, rather than only reporting averages in series.
- It is required that subjects are measured repeatedly over time.

Cross-sectional data

subject	year	input	output	age	sex
1	2000	10000	3000	20	1
1	2001	12000	4000	21	1
1	2002	13000	5000	22	1
2	2000	11000	4000	30	2
2	2001	9000	3000	31	2
2	2002	8000	4000	32	2
3	2000	20000	8000	40	1
3	2001	24000	9000	41	1
3	2002	28000	10000	42	1
4	2000	35000	12000	50	2
4	2001	29000	12000	51	2
4	2002	32000	11000	52	2

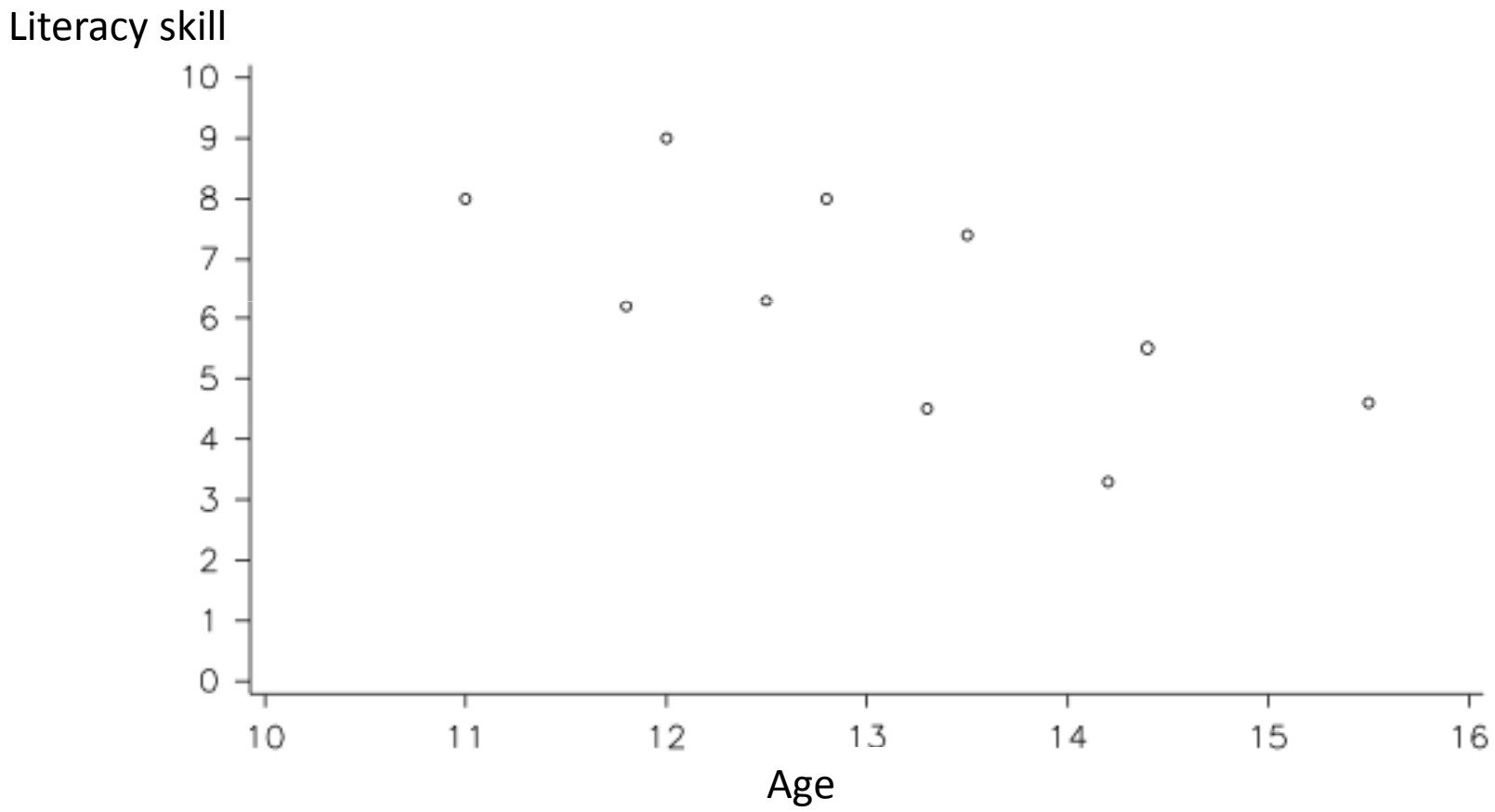
- Cross-sections means several variables from individuals at the same time.
- Inter-individual differences can be measured.
- Individual versus population average can be measured.

Panel data

subject	year	input	output	age	sex
1	2000	10000	3000	20	1
1	2001	12000	4000	21	1
1	2002	13000	5000	22	1
2	2000	11000	4000	30	2
2	2001	9000	3000	31	2
2	2002	8000	4000	32	2
3	2000	20000	8000	40	1
3	2001	24000	9000	41	1
3	2002	28000	10000	42	1
4	2000	35000	12000	50	2
4	2001	29000	12000	51	2
4	2002	32000	11000	52	2

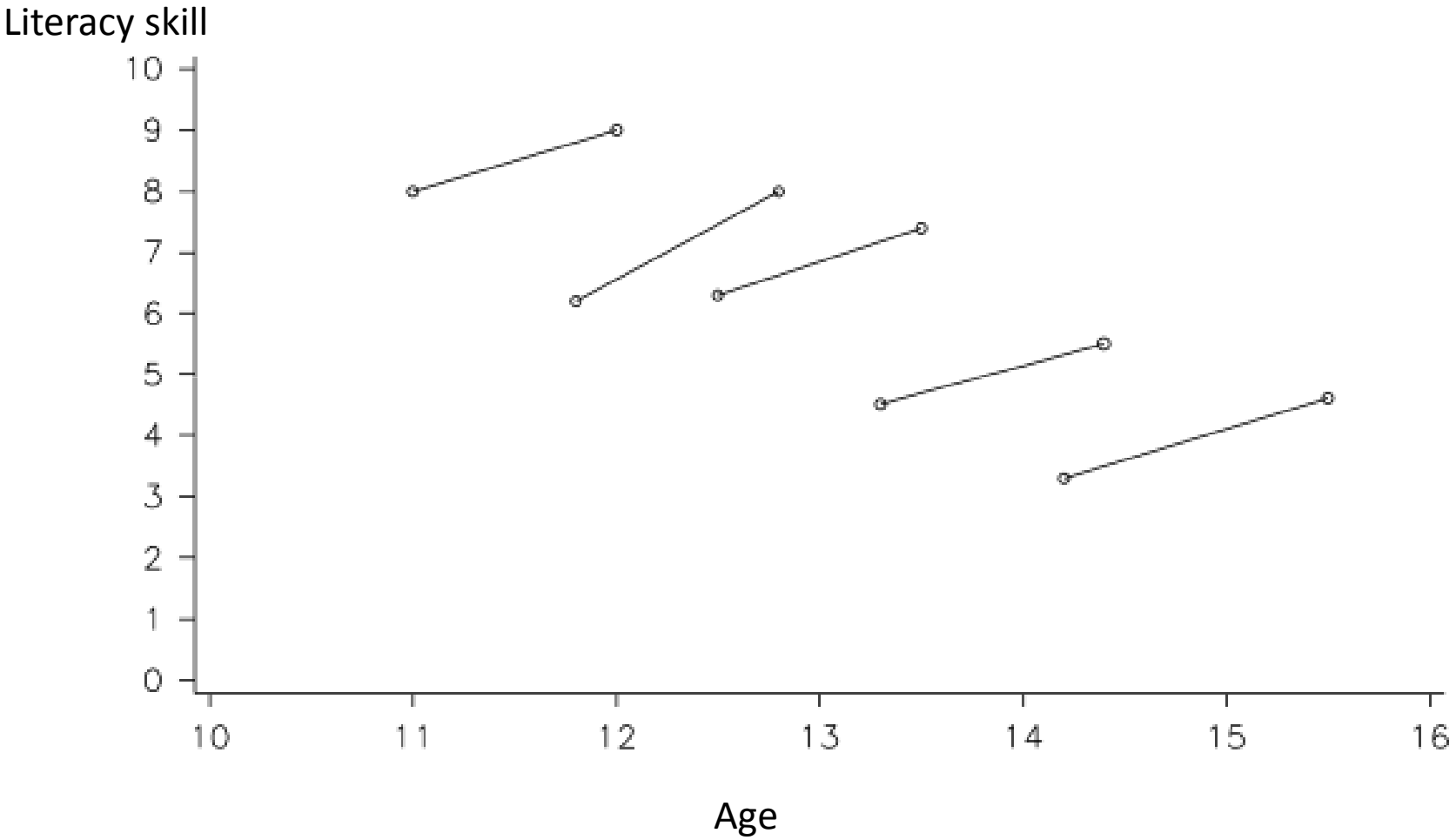
- The same cross-sections are collected every year.
- Inter-individual and intra-individual differences can be measured.
- It is required to collect the same variables each time.

Literacy development over age



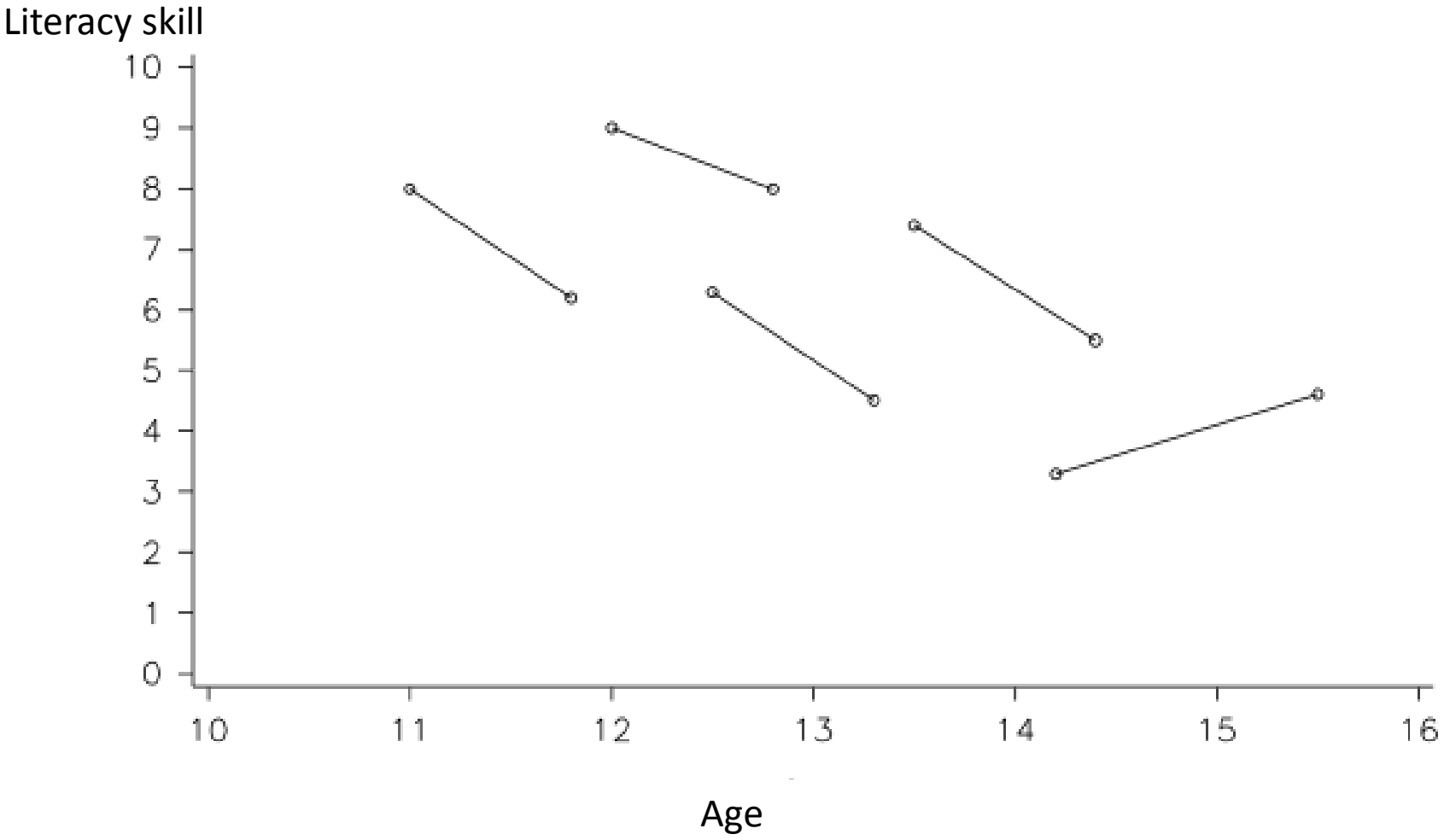
Rahiala, M. (2009) Lineaaristen sekamallien käyttö paneeliaineistojen analysoinnissa.

Literacy development over age



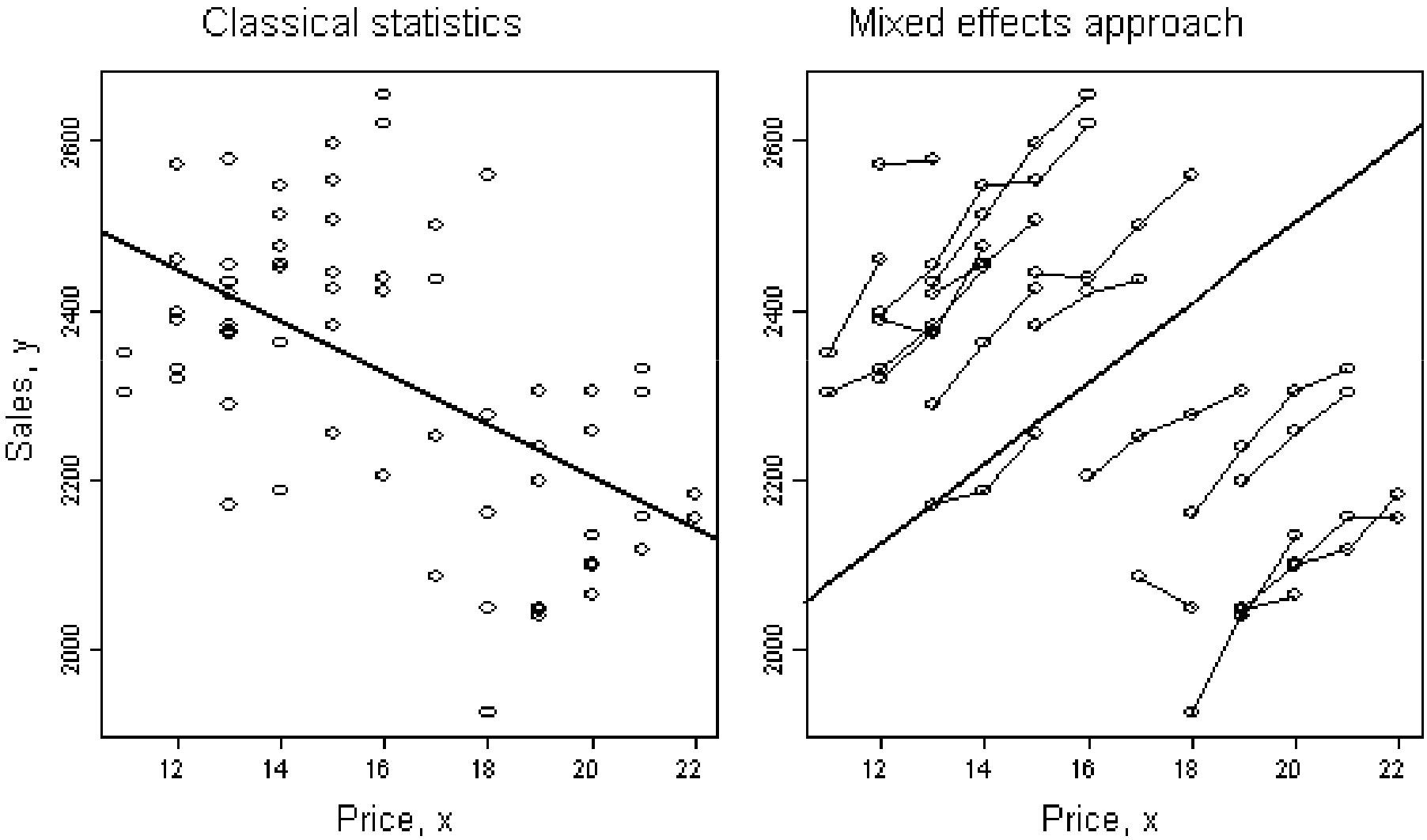
Rahiala, M. (2009) Lineaaristen sekamallien käyttö paneeliaineistojen analysoinnissa.

Literacy development over age



Rahiala, M. (2009) Lineaaristen sekamallien käyttö paneeliaineistojen analysoinnissa.

Paradoxical sales



Demidenko, E. (2004) Mixed Models: Theory and Applications.

Benefits of panel data (1)

1. Controlling for individual heterogeneity.

   Panel data suggests that individuals or firms are heterogeneous. Time-series and cross-section studies not controlling this heterogeneity run the risk of obtaining biased results.

2. Panel data give more informative data, more variability, less collinearity among the variables, more degrees of freedom and more efficiency.
3. Panel data are better able to study the dynamics of adjustment.

Baltagi, B. H. (2005) Econometric Analysis of Panel Data.

Benefits of panel data (2)

4. Panel data are better able to identify and measure effects that are simply not detectable in pure cross-section or pure time-series data
5. Panel data models allow constructing and testing more complicated behavioral models than purely cross-section or time-series data. For example, technical efficiency is better studied and modeled with panels.

Baltagi, B. H. (2005) Econometric Analysis of Panel Data.

Limitations of panel data

1. Design and data collection problems
   - Coverage
   - Nonresponse
   - Recall
   - Frequency, spacing
2. Distortions of measurement errors
   - Unclear questions
   - Memory errors
   - Misrecording, interviewer effects
3. Selectivity problems
4. Short time-series dimension
5. Cross-section dependence.

Baltagi, B. H. (2005) Econometric Analysis of Panel Data.


Example for dairy farms

Data farm level panel data 2000-2011

- We studied the development of production costs in dairy farms in 2000—2011 with a linear mixed model taking into account
  - farm-level information (location, economic size, number of cows) and time effect.
- We analyzed interindividual differences in intraindividual changes over time.

Data farm level panel data 2000-2011

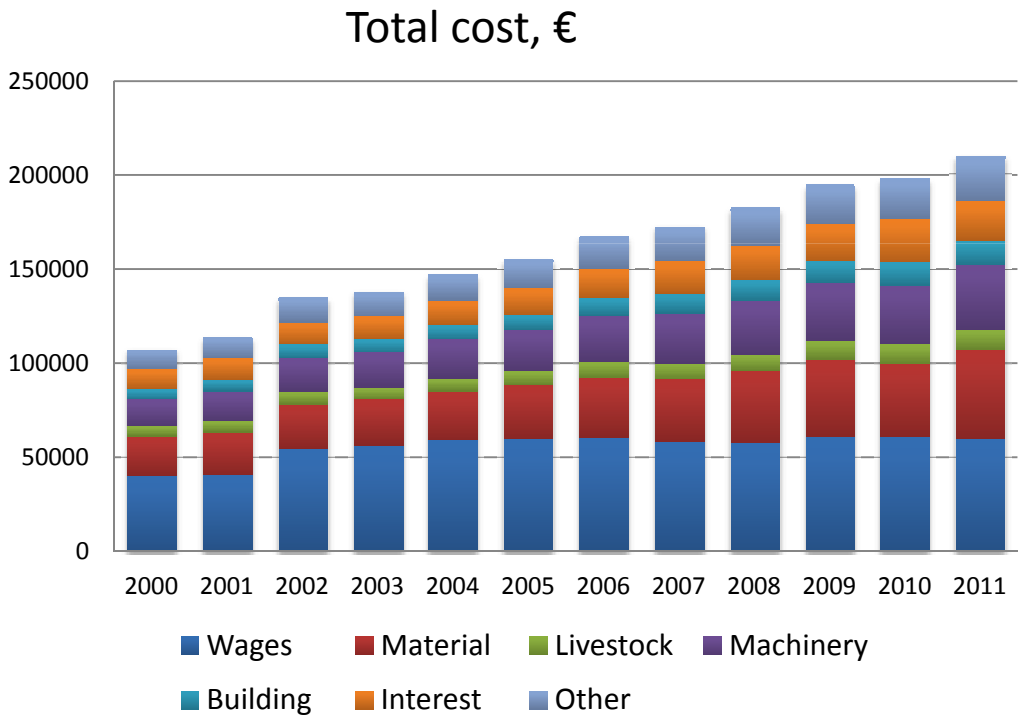
- Dairy farms participating in MTT profitability bookkeeping were studied for the years 2000–2011.
- The data set was formed as panel. Each farm was repeatedly measured in one year intervals.
- There were 4205 observations from 633 different farms and on average 350 different farms every year.
- Data set was unbalanced. This is due to the fact that it is voluntary to participate in MTT bookkeeping activities and, on the other hand, some farms had exited the business.

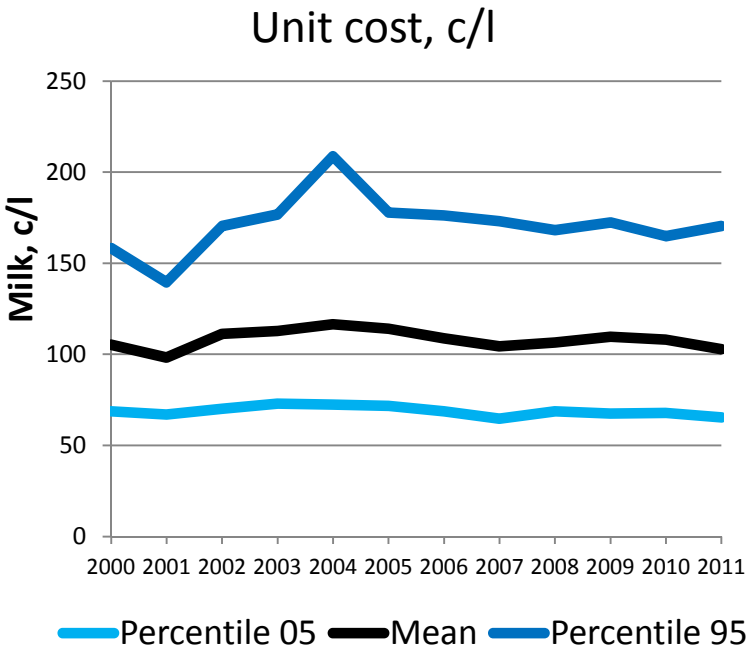
Data farm level panel data 2000-2011

- We studied the **unit production costs** (continuous variable).
- The total production cost is formed as sum of following components:
  - material, livestock, machinery, building, wages and interest costs.
- The production costs were deflated by using Consumer price indices year to 2011 prices (2000=100).
- The farm-level data were weighted with weight factors calculated individually for each farm for every year taking into account
  - the type of operations, economic size and location by support areas.
  - Weights were calibrated taking into account the total arable land in Finland.
- The unit costs of dairy farms were obtained by dividing the total production costs by the amount of produced milk (eurocent per litre).

Data farm level panel data 2000-2011

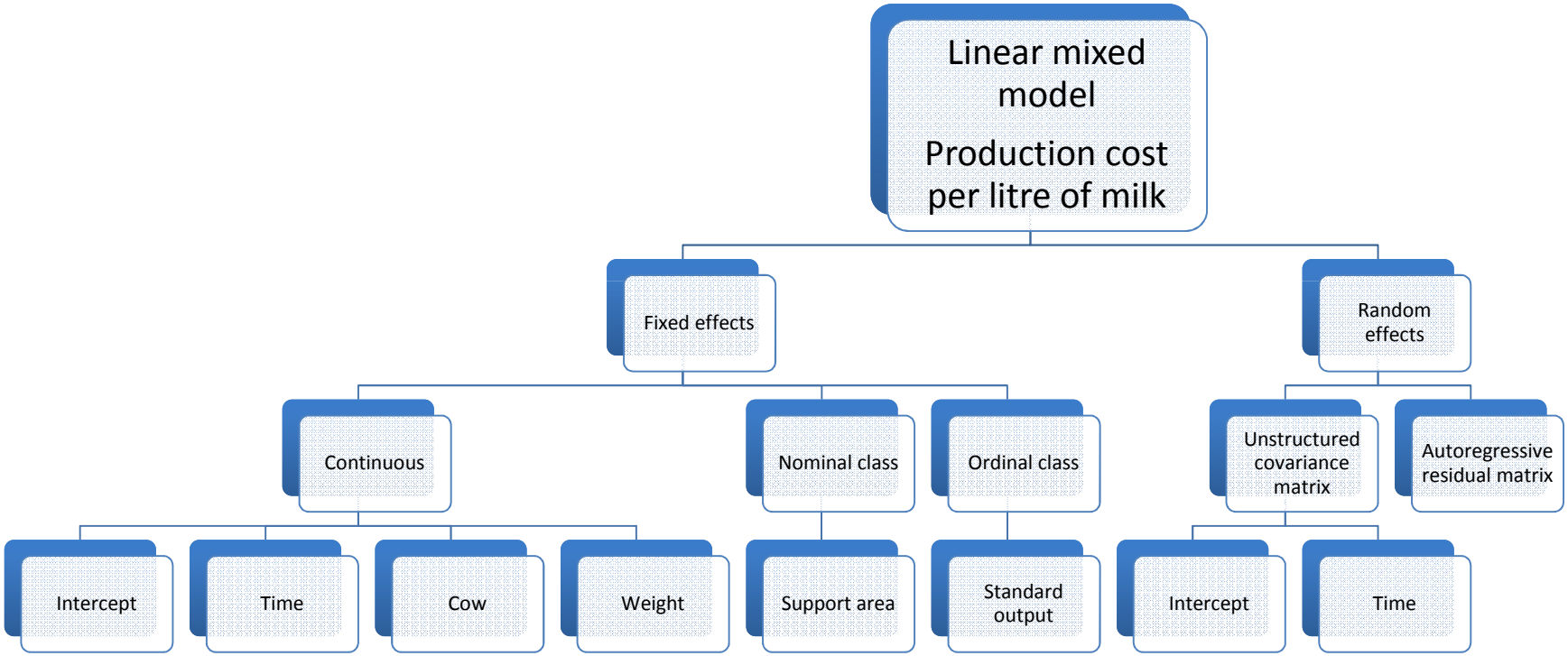
The development of average total and unit production costs in 2000—2011 deflated to 2011 prices and weighted results from MTT bookkeeping farms





Total cost has increased over time meaning that prices have grown and dairy farms are larger. However, it seems that the unit cost has remained the same.

Model specification for unit cost



Data used, MTT profitability bookkeeping panel data.

Results of linear mixed model explaining the unit cost

Effect		Estimate	Std. error	Sig.	CI 95% Low	CI 95% Up
Intercept	a_0	135.303	3.120	<0.001	129.182	141.423
time	a_1	1.472	0.176	<0.001	1.125	1.818
cow	a_2	-0.710	0.045	<0.001	-0.797	-0.622
Standard output	a_3					
medium (50000-100000 €)		-20.025	1.869	<0.001	-23.689	-16.361
large (>100000 €)		-22.253	2.390	<0.001	-26.938	-17.567
small (0-50000 €)		0	0			
Support area	a_4					
A		-1.506	4.879	0.758	-11.089	8.077
B		7.385	3.623	0.042	0.268	14.503
C1		-0.921	3.195	0.773	-7.195	5.354
C2		-2.513	2.903	0.387	-8.215	3.189
C2P-C4		0	0			
weight	a_5	-0.011	0.013	0.390	-0.036	0.014
Covariance parameters						
UN (1,1)	σ^2_{b0}	342.126	57.962	<0.001	245.459	476.864
UN (2,1)	$\sigma_{b0,b1}$	5.231	6.890	0.448	-8.274	18.735
UN (2,2)	σ^2_{b1}	2.774	1.205	0.021	1.184	6.499
Residual						
AR1 diagonal	σ^2	435.630	27.519	<0.001	384.898	493.048
AR1 rho	ρ	0.493	0.032	<0.001	0.428	0.554
Observations		4205				
-2 Restricted Log Likelihood		37439				
Akaike's Information Criterion (AIC)		37449				
Schwarz's Bayesian Criterion (BIC)		37481				

Results

- Costs increase year-to-year.
- The unit cost decreased as the number of cows increased.
  - To compensate annual cost increase farms should be expanded with two cows every year.
- Small farms had higher unit cost and annual variation than medium-sized and large farms.
  - Finnish dairy farms have developed fast and the benefits of scale may not have yet been accomplished.
- The farm location by support areas explains only slightly the unit cost.
- Productions costs change at different pace between farms.

Thank you

MTT Agrifood Research Finland, the Finnish Forest Research Institute (Metla), the Finnish Game and Fisheries Research Institute (RKTL) and the statistical services of the Information Centre of the Ministry of Agriculture and Forestry (Tike) are to be merged under a new entity called *Natural Resources Institute Finland (Luke)* as of 1 January 2015.

Alina Sinisalo and Heikki Mäkinen
MTT Agrifood Research Finland
alina.sinisalo@mtt.fi and heikki.makinen@mtt.fi