Statistikkonsulterna
JOSTAT & MR SAMPLE AB

Predicting migration patterns in Sweden using gravity model and neural networks

John Pavia, Jonny Olofsson and Magnus Pettersson
Statistikkonsulterna, Göteborg, Sweden

Correspondence: magnus.pettersson@statistikkonsulterna.se

1

Scope

1. *Create a model for predicting migration patterns within and between municipalities in Sweden*
2. *Data will be retrieved from public data sources*
3. *We will apply the gravity model using methods within AI/ML and present and visualize the results*
4. *We will make comparisons with "traditional" regression models*

VINNOVA

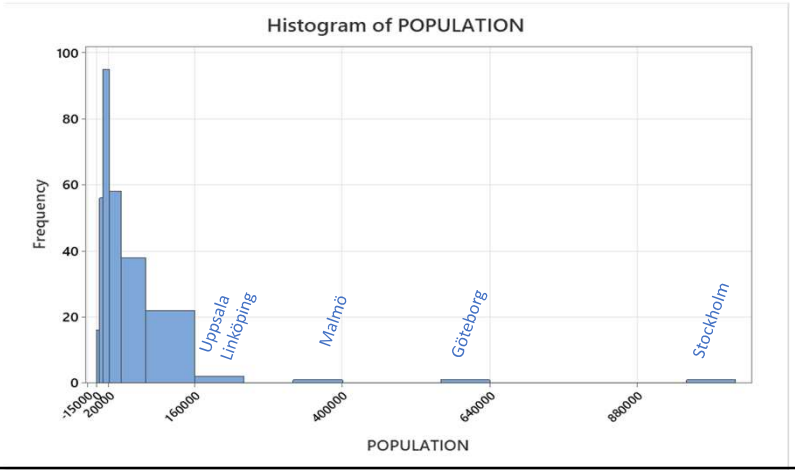
Sponsored by VINNOVA (Sweden's innovation agency) through their funding "Start your AI journey – SME Enterprises".

2

Sweden divided by…

Sweden is divided in 290 municipalities (”Kommun”): NUTS 5 -> LAU2

- Bjurholm (pop 2,408) - Stockholm (pop 975,819)
- 30% live in the 10 biggest, while 30% live in the 215 smallest



3

Intra-national migration

- Planning municipality services
- Promote progress and municipality brand
- Understanding the processes of migration

Attract certain groups	Population prognoses	Local brand
Place of origin	Expectations	Urbanisation processes
Promote people to move to us	Change in net migration	Important reasons for moving

4

Data used in the study

Statistiska centralbyrån
Statistics Sweden

→ Migration data between municipalities, annual 2010-21 (Statistics Sweden). Part of Swedish official statistics.

LANTMÄTERIET

→ Open GeoData, borders and location
Lantmäteriet (Swedish mapping authority)

Google Maps

→ Distances (road, public transport)
Google Map

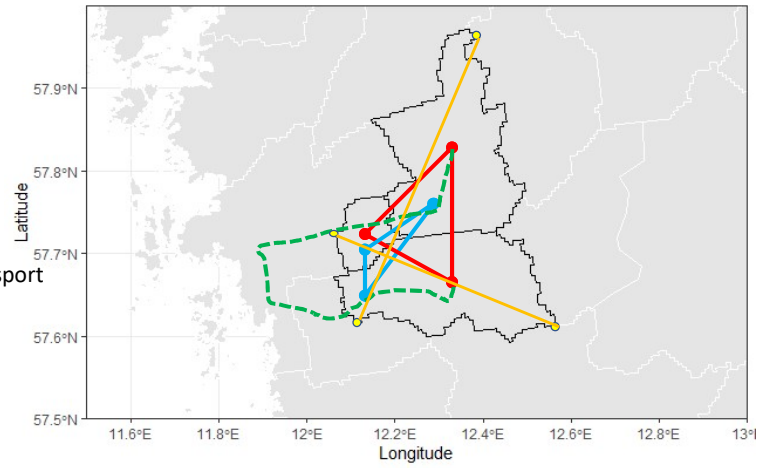
rka

→ Open Source demographic data
KOLADA, Rådet för främjande av kommunala analyser (RKA)

5

Distance measures

- Geographical midpoints
- Population centre of gravity
- Shortest way with public transport
- Points furthest apart



- Euclidian distance between geographical midpoints
- Euclidian distance between centre of population (cf centre of gravity).
- Euclidian distance between closest border points
- Euclidian distance between furthest border points
- Travelling distance by road in [km]
- Travelling distance expected by public transport

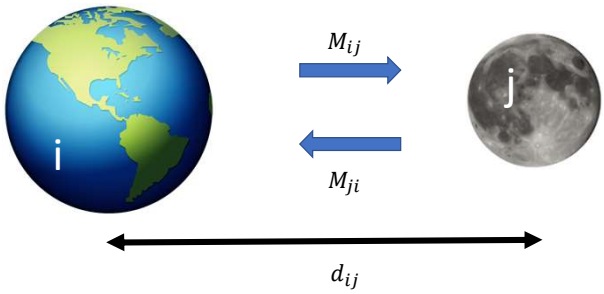
6

Gravity models

Cf to general gravity models: Mass of the two bodies and distance between centre of mass
Replace mass with population and distance with relevant distance measure

$$M_{ij} = \alpha_0 \cdot P_i^{\alpha_1} \cdot P_j^{\alpha_2} \cdot D_{ij}^{\alpha_3} \cdot \eta_{ij}$$

M_{ij} - Migration from *i* to *j*
P_i, P_j - Population in *i* and *j*
d_{ij} - Distance between *i* and *j*
α_0, α_1, α_2, and α_3 are model parameters
η_{ij} is an error term



7

Two approaches

Traditional approaches

Several procedures are applied, e.g.

- Logarithmic transformation, OLS
- Non-linear least squares
- Poisson regression (used for evaluation)

Artificial Neural Network

Architecture of analysis

- Backward propagation
- One hidden layer(3 nodes)

With or without zero migration

8

Traditional model

- Poisson regression model
- Estimation is done on
  - All 290 x 290 municipalities
  - Zeros are included
  - Euclidian distance
  - Aggregated over 10 years

Root Mean Square Error (RMSE) = 21.26

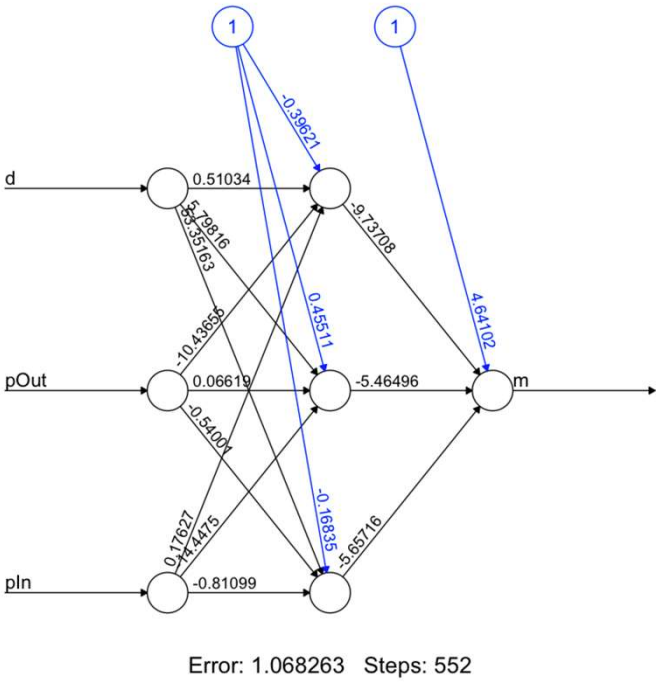
```
Call:
y_ppml ~ dist_log + logpIn + logpOut

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-63.844   -1.106   -0.590   -0.058   47.942

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -10.798216   0.058090  -185.9   <2e-16 ***
dist_log     -1.085836   0.003224  -336.8   <2e-16 ***
logpIn        0.872752   0.003080   283.4   <2e-16 ***
logpOut       0.862846   0.003093   279.0   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for quasipoisson family taken to be 7.866314)

    Null deviance: 2738461  on 83809  degrees of freedom
Residual deviance:  375919  on 83806  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 5
```

9

Neural network model

- A system of connected nodes that tries to replicate a human neuron net

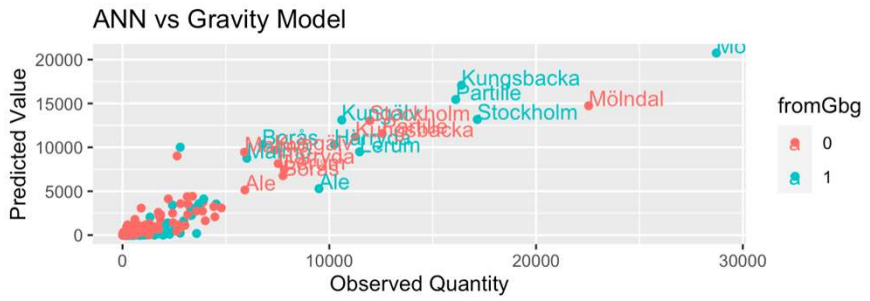
Root Mean Square Error (RMSE) = 36.05



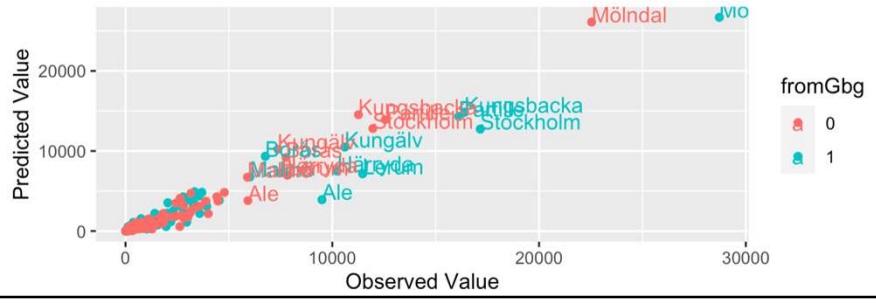
10

Göteborg, metropolitan area

Artificial Neural Network



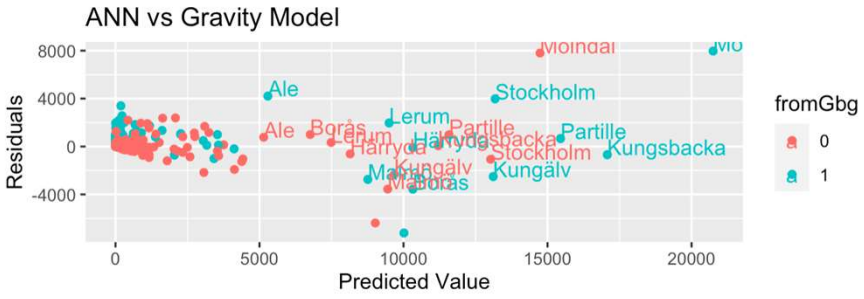
Gravity model



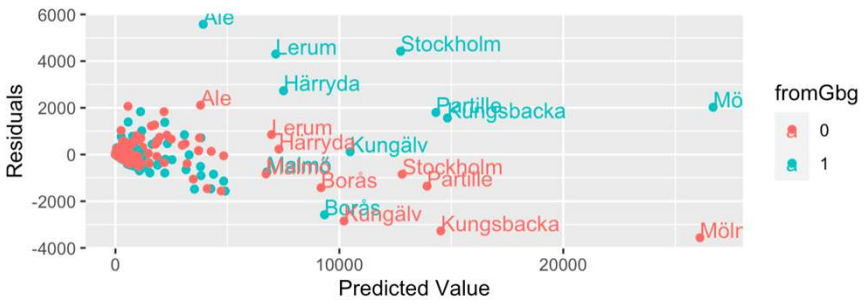
11

Göteborg, metropolitan area

Artificial Neural Network

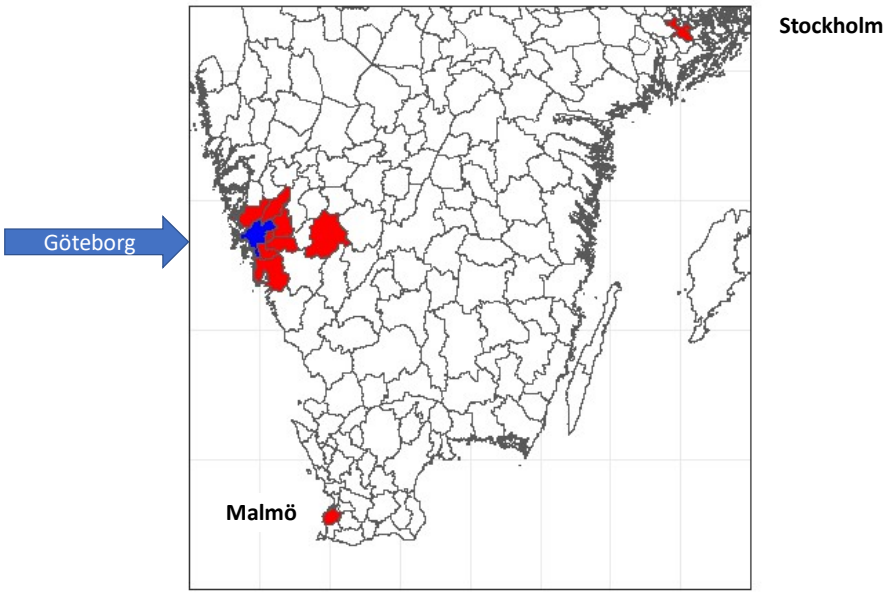


Gravity model



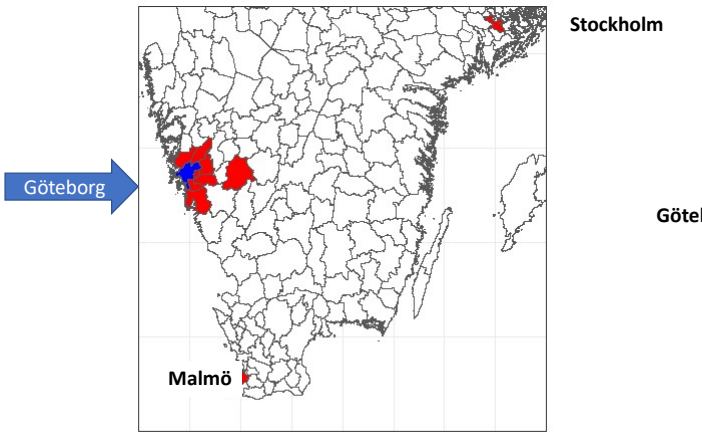
12

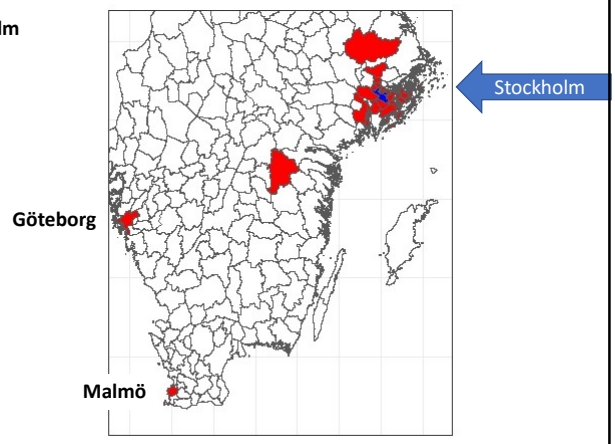
Göteborg, metropolitan area



13

Greater city metropolitan areas





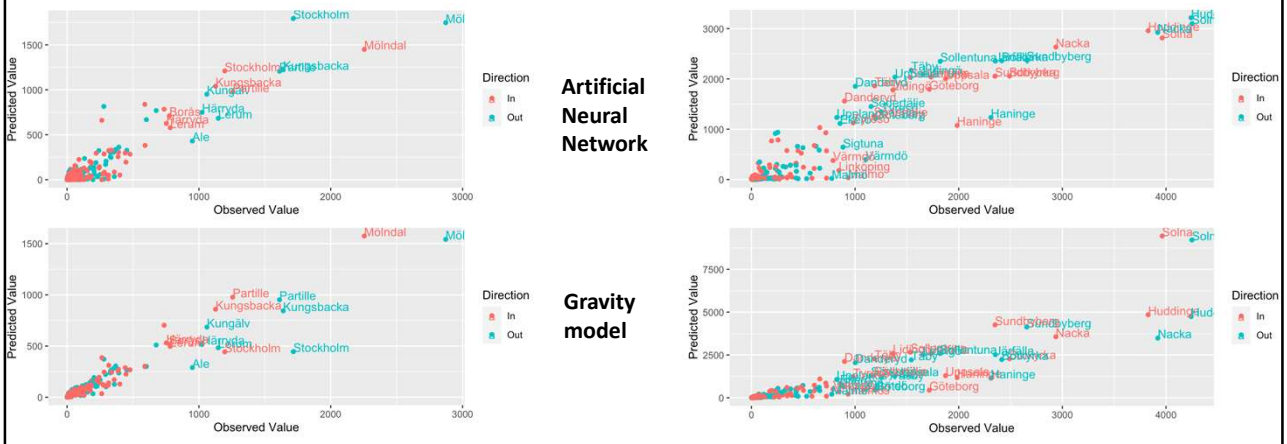
Göteborg
Pop = 583,758 (2)
Dens = 1303 /km^2 (8)

Stockholm
Pop = 975,819 (1)
Dens = 5,213.82 /km^2 (2)

14

Greater city metropolitan areas

GÖTEBORG **STOCKHOLM**

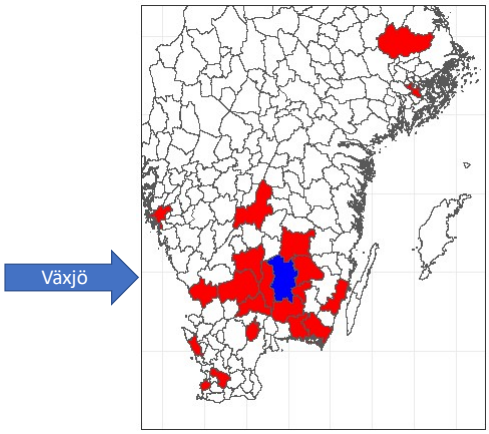


Artificial Neural Network

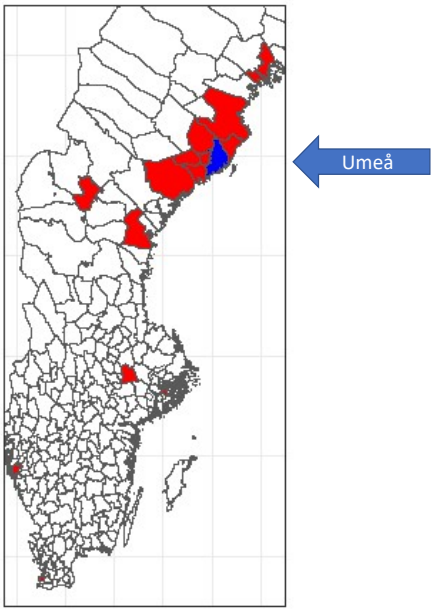
Gravity model

15

Regional centres



Växjö



Umeå

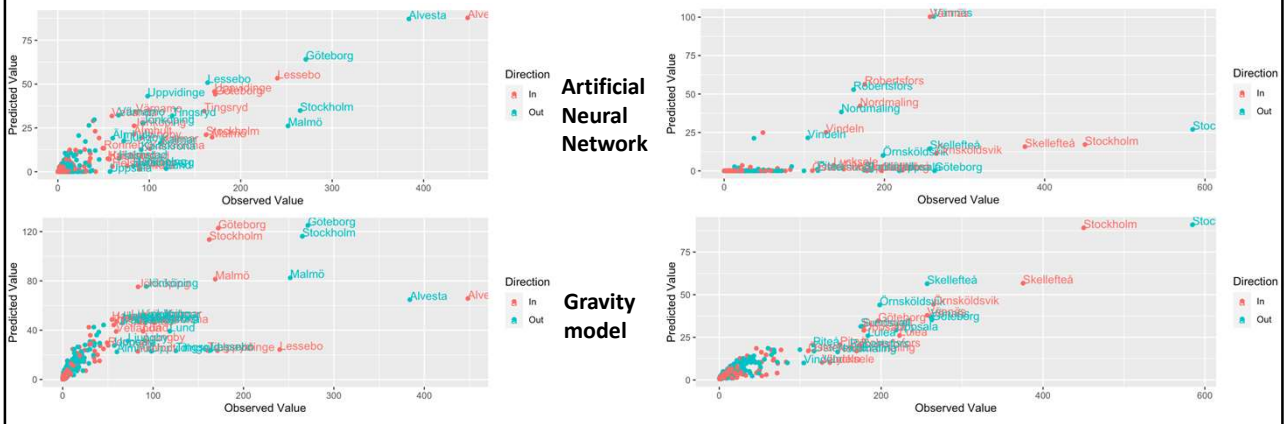
<u>Växjö</u>
Pop = 94,869 (22)
Dens = 59.96 /km^2 (92)

<u>Umeå</u>
Pop = 129,683 (11)
Dens = 55.97 /km^2 (93)

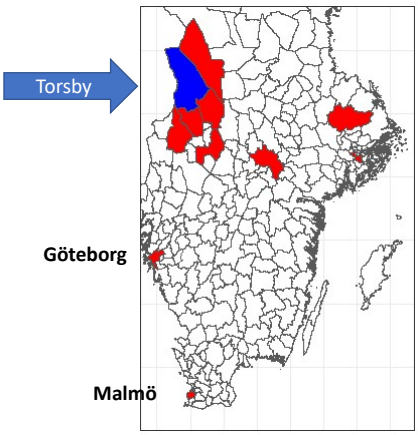
16

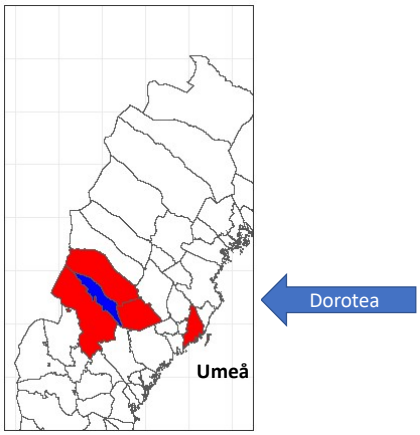
Regional centres

VÄXJÖ **UMEÅ**



Rural municipalities



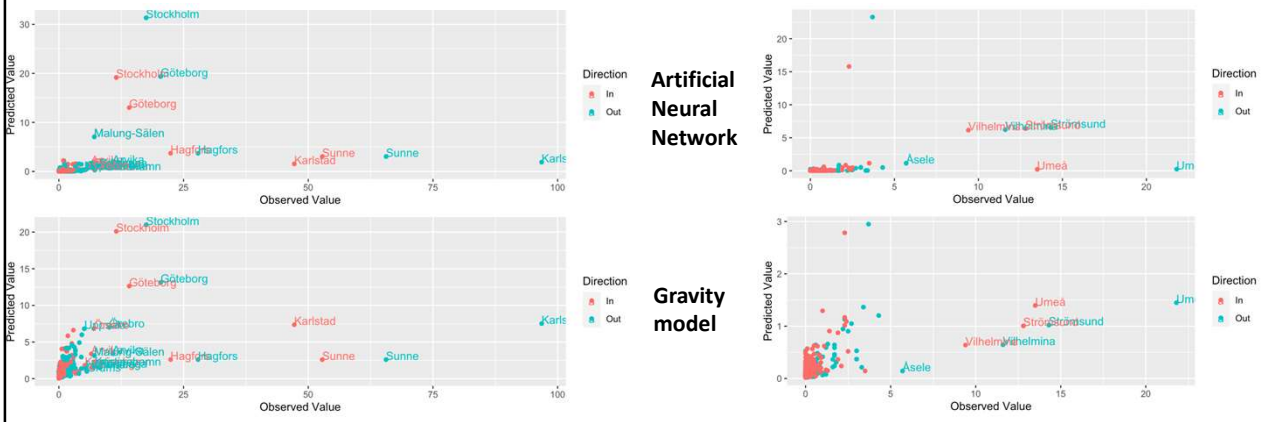


Torsby
Pop = 11,486 (197)
Dens = 2.76 /km² (263)

Dorotea
Pop = 2,488 (288)
Dens = 0.90 /km² (282)

Rural municipalities

TORSBY **DOROTEA**



19

Summary

- Estimations are generally biased (can be adjusted)
- Traditional (gravity) model is more accurate than neural network
- Both models are significant
- Neural Network has more dispersed estiamation
- Greatest problem is two unequal sized municipalites (imbalance)
- Neural Network is inferior for predicting small migrations
- Regional centres have a multimodality
- When more data is acquired ANN and gravity models their performance approach each other

20

Possible additional data

Demographic data

- Population income
- Age structure
- Population education
- Proportion born in Sweden

Surveys of people moving in and moving out

- Motivation for moving
- ”Customer satisfaction” studies of population

Cause specific information – (categorical data)

- University
- Communcation structure (commuting)
- Urbanisation
- Attractiveness, lack and cost for housing
- Migration chains

21

A Great Project Team!

- Magnus Pettersson
  Senior statistician, fil.lic biostatistics, PStat®
  Project leader and coach
- John Pavia
  Statistician, BSc, AI specialist
  Evaluation and programming
- Jonny Olofsson
  Senior statistician, fil lic
  Society analysis research specialist
  Migration analysis
- Ahmet Akdeve
  Statistician BSc, AI MSc student
  R-programming, map construction
- Rebecca Bylund
  Statistician, MSc, R-programmer
  Google maps, distances

22