

# Modeling and Predicting Individual Salaries: A Study of Finland's Unique Dataset

**Lasse Koskinen**

Insurance Supervisory Authority of Finland  
and  
Helsinki School of Economics, Finland

**Tapio Nummi**

University of Tampere, Finland.

**Janne Salonen**

The Finnish Centre for Pensions, Helsinki, Finland.

# OUTLINE

- **Background**
- **Problem:** To model and predict **individual** wages.
- **Data:** A unique Finnish dataset.
- **Model:** A **panel data** models for **subpopulations**.
- **Predictions:** Genuine out-of-sample predictions.
  - Normal growth period and deep recession
- **Concluding remarks.**

# 1. Background (1)

- **Actuarial models** are constructed to aid in the assessment of the financial and economic consequences. This **requires**:
  - understanding the conditions and processes under which past observations were obtained;
  - anticipating changes in those conditions that will affect future
  - evaluating the quality of the available data;
  - bringing judgment to bear on the modeling process, validating the work as it progresses;
  - estimating the uncertainty inherent in the modeling process itself.

## 1. Background (2)

- Different types of models have been proposed for describing **average** salary profiles.
- Moderate average wage is not equivalent to a moderate pension for all individuals.
- **Individual** profiles are rarely modeled.
- Modeling is **often limited** by lack of adequate data; Here a unique Finnish dataset of individuals is exploited.

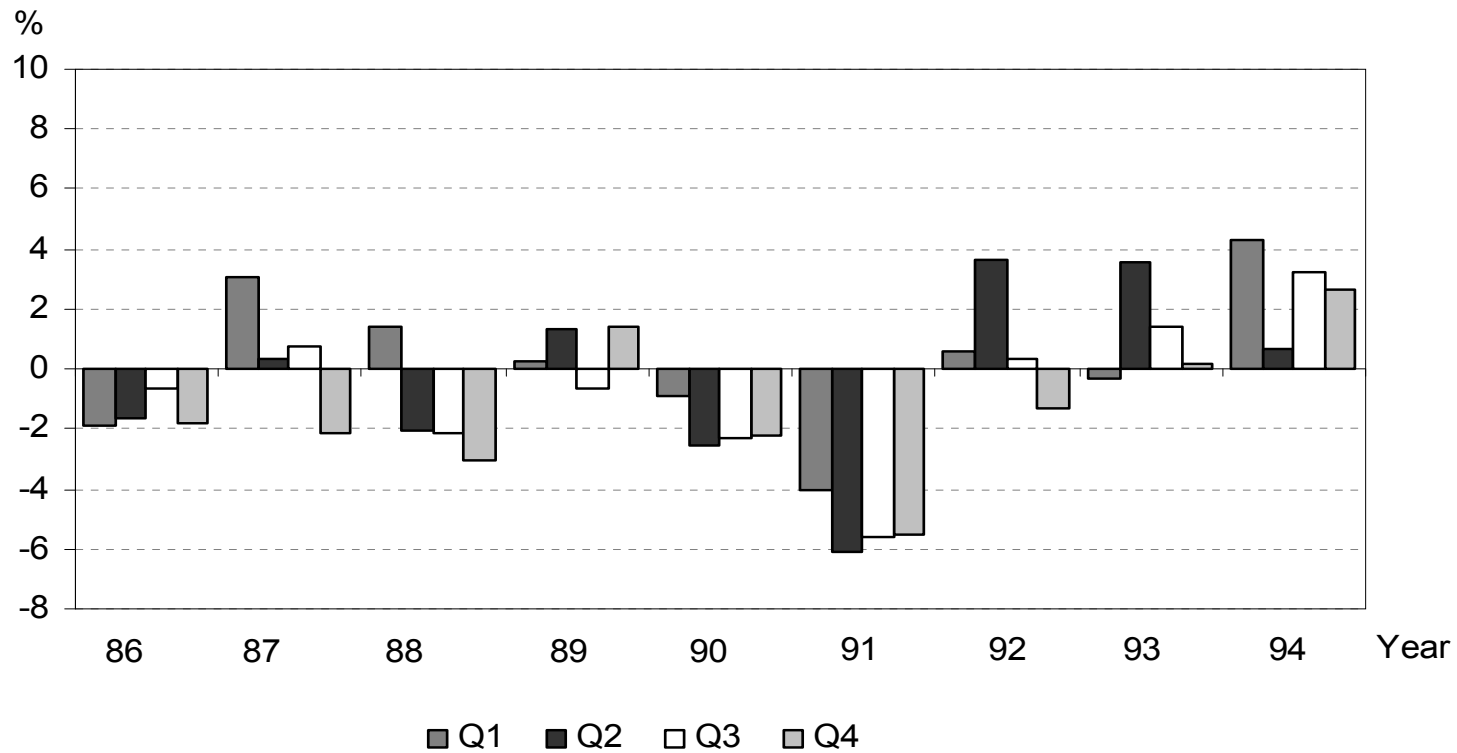
## 2. Problem

- The **general objective** of this study is to develop a model that describes 1) **individual features** of salary development and 2) can be used for **prediction** purposes.
- In this paper a **dataset** of individuals is exploited - all the participants of the Finnish private-sector statutory pension scheme who retired in 1998.
- It is very natural to assume that **genders** are treated as different subpopulations.
- Our approach is to further divide the data according to **income quartiles** in the year 1975. This reflects the effect of certain **socio-economic factors**.

## 3. Data

- The data was collected as a part of the Finnish pension reform package in 2001-2002.
- All people who retired in 1998.
- We focused on the cohorts born between 1933 and 1938 and the years from 1975 to 1994.
- These limitations mean that we have 2986 individuals in the analysis.

## *Annual change (%) of mean wage in each quartile (men).*



### 3. Model (1)

- The model is an **extension of the basic linear model** that allows some model parameters to be drawn from a probability distribution.
- Called **mixed model** since the model parameters contain both fixed and random effects.
- Variables are used:
  - **$Z(ij)$**  age of an individual  $i$  at time  $j$
  - **$d(i)$**  duration of the career of an individual  $i$
  - **$b(j)$**  the change of **GDP** at time  $j$

### 3. Model (2)

- The **linear mixed model** where random parameters  $u_{i0}$  and  $u_{i1}$  are associated with an individual under consideration.

$$y_{ij} = \beta_0 + u_{i0} + (\beta_1 + u_{i1})z_{ij} + \beta_2 z_{ij}^2 + \beta_3 z_{ij}^3 + \beta_4 d_i + \beta_5 b_j + \varepsilon_{ij}.$$

- The fixed parameters  $\beta_0 \beta_1 \dots \beta_5$  are coefficients associated with the entire subdata. Error terms  $\varepsilon_{ij}$  are assumed to be independently and normally distributed.

Here we assume that the joint distribution of  $u_{i0}$  and  $u_{i1}$

- multivariate normal with the expected value zero;
- independent of the random errors.



### 3. Model – potential application

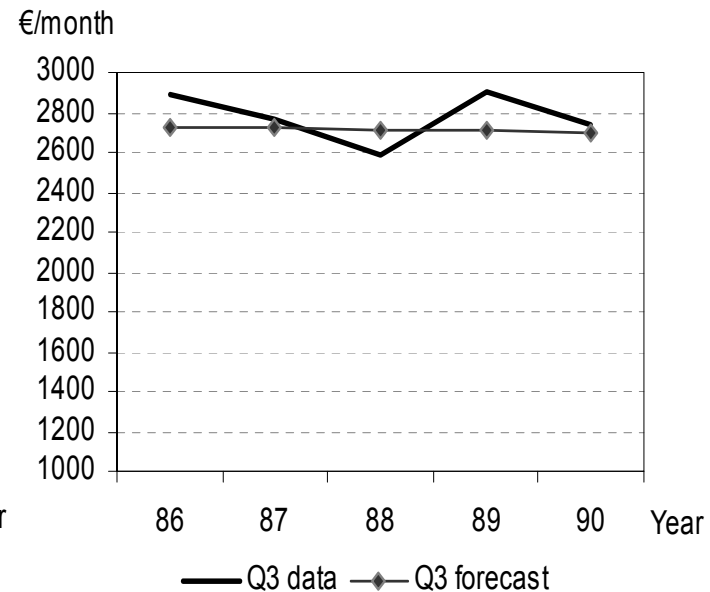
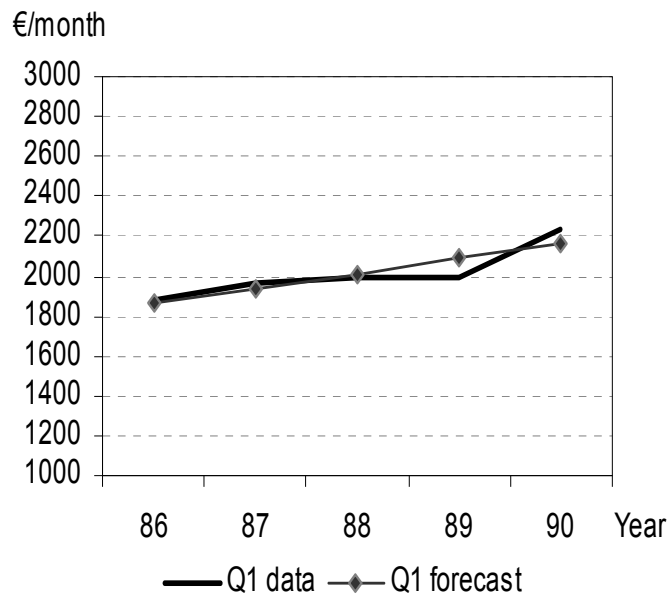
- The **old** system computed the pensionable wage – the base wage for all benefit calculations – for each job separately by **averaging the last 10 years** in each job. This procedure ignores earnings differences among workers in the other years.
- The **new** system bases the **pensionable wage on all earnings** and **does not distinguish among jobs** in different sectors of the economy. The earnings-related pension will be calculated directly as a percentage of the annual earnings.
- A critical factor is to determine what kind of benefits the new system would provide to different employee groups => **Individual subgroup models are needed!**

## 4. Predictions

- When assessing **the solvency** of a scheme pension experts are mainly interested in predicting **average** wages.
- Instead, in system development, individual variation in wages is essential - **a high average wage does not guarantee an adequate pension for all members of the group.**
  - Hence predicting individual salary growth is very important for planning purposes and risk assessment.

## 4. Predictions (2)

*Examples of individual wage predictions and actual values (men).*



## 4. Predictions (3)

- The wage quartile was reflected in the model specification in a number of ways.
- Next we consider group level predictions.
  - The middle quartiles (Q2 and Q3) are well predicted. The first and fourth quartiles are rather more challenging to predict. This holds for both men and women.
  - The deep recession is certainly a factor affecting wage risk especially for Q1s.

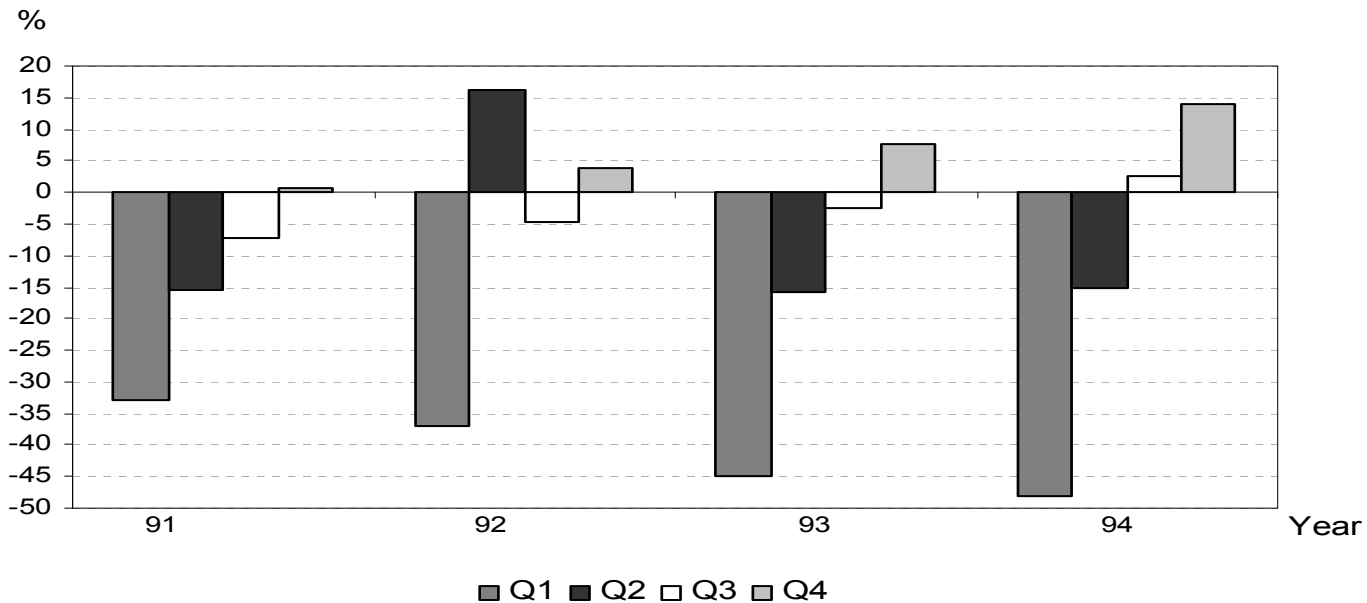
## 4. Predictions (4)

- The estimation and forecasting periods are
  - Estimation period: 1975 – 1985 (Normal economic growth);
  - Forecasting period I: 1986-1990 (Normal economic growth);
  - Forecasting period II: 1991-1994 (Deep recession).
- First predictions were needed for the exogenous variable GDP => Holt-Winters predictions for GDP were made



## 4. Predictions (6)

Absolute prediction error as percentage of mean wage 1991-1994 (men).



## 5. Conclusions (1)

The model specifications and prediction results allow for the following general conclusions:

- The wage formation seems to be essentially different in different wage quartiles. **Better forecasts** may be obtained by using **quarter-specific models**.
- **Individual variation** within a wage quartile is **large** and an important risk factor.
- The workers in the **lowest quarter** have difficulty in maintaining their wages in periods of depression. In this study the link with wages in other groups is much weaker.

