

ROYAL INSTITUTE OF TECHNOLOGY

MASTER THESIS APPLIED MATHEMATICS

**Risks and scenarios in the Swedish
income-based pension system**

Author:
Simon von Mentzer

Supervisors:
Timo Koski KTH
Danne Mikula
Pensionsmyndigheten

October 31, 2015

Abstract

In this master thesis the risks and scenarios in the Swedish income-based pension system are investigated. To investigate the risks one has chosen to look at a vector autoregressive (VAR) model for three variables (AP-fund returns, average wage returns and inflation). Bootstrap is used to simulate the VAR model. When the simulated values are received they are put back in equations that describes real average wage return, real return from the AP-funds, average wage and income index. Lastly the pension balance is calculated with the simulated data.

Scenarios are created by changing one variable at the time in the VAR model. Then it is investigated how different scenarios affect the indexation and pension balance.

The result show a cross correlation structure between average wage return and inflation in the VAR model, but AP-fund returns can simply be modelled as an exogenous white noise random variable. In the scenario when average wage return is altered, one can see the largest changes in indexation and pension balance.

Keywords: Pension, VAR, Bootstrap, Scenario.

Sammanfattning

I det här examensarbetet ("Risker och scenarion i Sveriges inkomstgrundande allmänna pensionssystem") undersöks risker och scenarier i inkomstpensionssystemet. För att kunna undersöka riskerna har en vector autoregressive (VAR) modell valts för tre variabler (AP-fonds avkastning, medelinkomst avkastning och inflation). Bootstrap används för att simulera VAR modellen. När värden från simuleringarna erhållits kan dessa sättas in i ekvationer som beskriver real medelinkomst avkastning, real avkastning från AP-fonderna och inkomst index. Slutligen beräknas pensionsbehållning med simulerad data.

Scenarierna utförs genom att en variabel i taget i VAR modellen störs. Sedan utreds hur denna störning påverkar resterande parametrar som beräknas. Detta görs för olika scenarion.

I VAR modellen finns korrelationer mellan medelinkomst avkastning och inflation, men AP-fonds avkastning kan ses som vitt brus. De scenarier som har störst påverkan på indexeringen är då medelinkomst avkastningen ändras.

Acknowledgements

I would like to thank my professor at KTH, Timo Koski and PhD student Jonas Hallgren for their support during the master thesis. I would also like to thank my supervisor from the Swedish pension agency Danne Mikula.

Contents

1	Introduction	2
1.1	Income-based pension system	2
1.2	Objective	5
2	Background	6
2.1	Time series	6
2.2	Empirical distributions and quantiles	8
2.3	Bootstrapping	8
2.4	Parameter estimation	10
2.5	Macroeconomics	12
2.5.1	CPI and inflation	12
2.5.2	Real and nominal Returns	12
3	Model	14
3.1	VAR and bootstrap	14
3.2	Scenario analysis	17
3.2.1	Scenario drift	17
3.2.2	Scenario probability	17
4	Results scenario analysis	19
4.1	Data	19
4.2	Vector autoregressive model	23
4.3	Uncertainty in coefficient estimation	27
4.4	Scenarios with drift	28
4.4.1	Low average wage returns	28
4.4.2	High average wage returns	31
4.4.3	Low inflation	34
4.4.4	High inflation	37
4.5	Scenarios with probability	40
4.5.1	Low average wage returns	40
4.5.2	High average wage returns	43
4.5.3	Low inflation	45
4.5.4	High inflation	48
5	Results pension balance	50
6	Conclusion	56
7	Discussion	57

Appendices	59
A Algebra	59
B Ordinary least squares	60

Notation

A_i	Coefficient matrix for VAR(1) model where $i \in \{1, 0\}$.
A_i	Coefficient matrix for VAR(1) model where $i \in \{1, 0\}$.
$c_{i,j}$	Coefficients in the VAR(1) model where $i \in \{1, 2, 3\}$ and $j \in \{1, 2, 3\}$.
a_i	Intercept coefficients in the VAR(1) model where $i \in \{1, 2, 3\}$.
Σ	Covariance matrix for the residuals in the VAR(1) model.
AW_t	Average wage at time t .
CPI_t	CPI at time t .
x_t	Average wage logreturns at time t .
x_t^r	Real average wage logreturns at time t .
y_t	AP-funds logreturn at time t .
y_t^r	Real AP-funds logreturn at time t .
z_t	Inflation logreturn t .
ϵ_t^i	Residuals for the Vector autoregressive model $i \in \{1, 2, 3\}$ at time t .
I_t	Income index at time t .
PB_t	Total pension balance balances, year t .
BR_t	Balance ratio year t .
OT_t	Turnover duration year t .
BF_t	Aggregated buffer fund (AP-funds) value year t .
B_t	Balance index year t .
IPR_t	Estimated pension credit earned year t .

1 Introduction

It is of interest for the Swedish pension agency to have a model that describes the risk in the income-based pension system, the risks in this master thesis are referred to how the pension balance is affected in different scenarios.

The Swedish pension system is a pay as you go system that entails risks with the pension balances. To reduce the risks, the Balance ratio is introduced as a "brake" so the pensions are indexed in a slower rate than without the balancing. This is done to make the system financially stable. In the article "Temaartikel: Balanstalet - inkomstpensionens stabila styråra?" it is stated that the Balance ratio is so far the most stable way to reach financial stability [1]. The article further claims that if the pension balance only was to be indexed with respect to the assets' development, this would not be enough to ensure financial stability.

In this master thesis one tries to find a model to the variables: real average wage return, inflation, real return from the AP-funds, average wage and income index. The aim of the modelling is to find a good enough model that can simulate different scenarios and see how changes in one variable affect the other variables. When the mentioned variables above can be simulated, the risk of the pension balance is investigated ,i.e, how does the pension balance change during the different scenarios.

The report describes step by step how a Bootstrap procedure is performed of a VAR(1) time series model that describes relationships (standard deviation and lag structures) of the variables mentioned above. The advantage with bootstrap simulations is that it does not require explicitly the distribution of the data. Bootstrap makes it possible to asses dispersion of complicated functions of random variables [2].

The disposition of the report is as follows. First an introduction to the study is presented, where one describes the income-based pension system, then the objective of the study is presented. In the objectives the purpose and goal of the master thesis is presented. After the objectives the background is presented where all the necessary mathematical theory and some macroeconomics are explained. Then the model and the scenario analysis are presented. Lastly the conclusion is presented where the objectives are adressed followed by discussion about the results.

1.1 Income-based pension system

The purpose of this section is to give the reader an idea of how the different variables introduced in the report are used to index the pension liabilities in the "inkomstpension" and premium pension.

The income-based pension in Sweden consists of the "inkomstpension" and the "premium pension". The income-based pension system is supposed to be a financially stable system. This means that liabilities and assets normally change by the same amount. This is the case for the premium pension, but the "inkomstpension" allows for differences from year to year.

The "inkomstpension" is a pay-as-you go system [3]. This means that one year's income (pension contributions) in the system becomes the same year's liabilities to the retired. The differences in assets and liabilities appear when e.g. the pension contributions are a smaller amount than the pension liabilities or when the pension contributions are larger than the pension liabilities. The buffer fund absorbs these surpluses of the system.

The pension balances can be described by equation (1)

$$PB_t^I = \begin{cases} (PB_{t-1}^I + P_{t-1}) \frac{I_t}{I_{t-1}} \times ACF_t \times IGF_t, & BT_t \geq 1 \\ (PB_{t-1}^I + P_{t-1}) \frac{I_t}{I_{t-1}} \times ACF_t \times IGF_t \times BT_t, & BT_t < 1 \end{cases} \quad (1)$$

The Balance ratio BT_t can be expressed as,

$$BT_{t+2} = \frac{P_t \times \bar{OT}_t + \bar{BF}_t}{PB_t^I + \frac{B_t}{I_{t-1}} \times IPR_t}. \quad (2)$$

In equation (2), B_t is the balance index for year t and IPR_t is the estimated pension credit earned year t .

$$B_t = BT_t \times I_t$$

In equation (2) the \bar{OT}_t is

$$\bar{OT}_t = \text{median}[OT_{t-1}, OT_{t-2}, OT_{t-3}].$$

\bar{BF}_t in equation (2) is calculated as,

$$\bar{BF}_t = \frac{BF_t + BF_{t-1} + BF_{t-2}}{3}.$$

The income index I_t is defined as,

$$I_t = I_{t-1} \cdot \left(\frac{AW_{t-1} \text{CPI}_{t-4}}{AW_{t-4} \text{CPI}_{t-1}} \right)^{\frac{1}{3}} \cdot \frac{\text{CPI}_{t-1}}{\text{CPI}_{t-2}}. \quad (3)$$

In equation (3) the CPI is consumer price index and AW_t is average wage.

In equation (1) I_t is the income index for year t , P_t is the premium paid out to the pensioners year t , ACF_t is the administration cost factor for year t and IGF_t inheritance gain factor for year t .

At retirement the pension for an individual's first year is calculated as

$$\text{pension}_t = \frac{\text{PB}_t^I}{\text{AD}}$$

Here AD is the annuity divisor.

$$\text{pension}_t = \begin{cases} \text{pension}_{t-1} \cdot \frac{I_t}{I_{t-1} \cdot 1.016}, & \text{BT}_t \geq 1 \\ \text{pension}_{t-1} \cdot \frac{I_t}{I_{t-1} \cdot 1.016} \text{BT}_t, & \text{BT}_t < 1 \end{cases} \quad (4)$$

The pension balances depend on the income index. This index is determined from the average change in income in Sweden in combination with the Balance ratio in years [3]. The "inkomstpension" is affected by a number of different economic and demographic factors. Factors that effect the income pension are employment, changes in the stock and bond markets [3].

It is reasonable to think that the pensions should reflect the wages of the people that work today. Therefore they are indexed as mentioned above with the income index, but if the Balance ratio is smaller than one, the indexation is changed and the income index becomes $I_t \cdot \text{BT}_t$ called the balance index, see equation (1). Here BT_t is the Balance ratio for year t .

The premium pension is another part of the income-based pension system that is invested in funds. The pension balance for this can be described as

$$\text{PB}_t^p = \text{PB}_s^p \cdot \text{IGF}_t - \text{AC}_t, \quad s < t, \quad (5)$$

where PB_t^p is the pension balance for the premium pension year t , AC_t is a administration cost for year t . The pension for year t is calculated in the same way as the inkomst pension.

Next year, all included smoothing disappear, including the geometric smoothing of the income index. The balance ratio will be attenuated with factor $\frac{1}{3}$. There will also be some changes in how the turnover duration is calculated.

1.2 Objective

The objectives of this master thesis are to;

- Create a model that captures variance and lag structure of average wage returns, inflation and returns from the AP-funds.
- Simulate different scenarios from this model and see how these scenarios affect inflation, average wage return, average wage, real average wage return, AP-funds return, real AP-funds return and income index.
- Simulate the pension balance with the result from previous simulations of inflation, average wage return, average wage, real average wage return, AP-funds return, real AP-funds return and income index.

To meet the objectives a VAR (Vector autoregressive model) has been used to model the variables average wage return, return from the AP-funds and inflation. The VAR model is estimated with real data and simulated with bootstrap. From average wage return, return from the AP-funds and inflation it is possible to calculate real average wage return, real return from the AP-funds, average wage and income index.

When a model has been found it is tested in four different scenarios. The first scenario is when a constant μ is added to the average wage return. The second scenario is when a constant $-\mu$ is added to the average wage return. This creates two scenarios, which describes when average wage returns become smaller than expected or larger than expected. The same procedure is performed on inflation and hence the third and fourth scenarios are obtained. Another method for create scenarios is to assign a higher probability to draw positive residuals from the residuals and lower probability to draw negative residuals, this becomes a high scenario and the opposite conditions on the probability becomes the low scenario.

2 Background

In this section relevant background is presented. First a mathematical background is presented containing time series analysis, empirical distributions and bootstrapping. Next the parameter estimation for the VAR model is described and last some basic macroeconomics is presented.

2.1 Time series

In this section multivariate time series concepts are presented. First the definition of times series is explained, then stationarity conditions, mean and covariance functions are defined.

A multivariate time series $z_t = [z_{1t}, \dots, z_{kt}]'$ is random vector consisting of k random variables, where the index t is time. This means that there exists a probability space for which these random variables are defined on and the observations (data) are realizations of them [4].

A k -dimensional time series is said to be weakly stationary (in this report stationary and weakly stationary are used as synonyms) if

$$\begin{aligned} E[z_t] &= \boldsymbol{\mu} \\ \text{cov}(z_t) &= E[(z_t - \boldsymbol{\mu})(z_t - \boldsymbol{\mu})'] = \Sigma_z \end{aligned}$$

are constant $k \times 1$ vector and a constant positive definite $k \times k$ matrix respectively [4].

The time series z_t is said to follow a VAR (vector autoregressive) model of order p if

$$z_t = \phi_0 + \sum_{i=1}^p \phi_i z_{t-i} + \epsilon_t \quad (6)$$

In equation (21) ϕ_0 is a $k \times 1$ dimensional vector, ϕ_i is a $k \times k$ dimensional constant matrix and a_t is a time series with independent and identically distributed (iid) random vectors with mean zero and covariance matrix Σ_a [4].

To check, if the VAR process is stationary, one can calculate the eigenvalues for the coefficient matrices. For a VAR(1) model this means solving the determinant equation:

$$|\lambda \mathbf{I}_k - \phi_1| = 0$$

Here one checks whether all λ are smaller than 1. If this is the case the VAR(1) model is said to be stationary.

The *cross covariance* matrix are defined as

$$\Gamma_l = \text{cov}(z_t, z_{t-l}) = E[(z_t - \boldsymbol{\mu})(z_{t-l} - \boldsymbol{\mu})']$$

$$\Gamma_l = \begin{bmatrix} E[\tilde{z}_{1,t}\tilde{z}_{1,t-l}] & E[\tilde{z}_{1,t}\tilde{z}_{2,t-l}] & \dots & E[\tilde{z}_{1,t}\tilde{z}_{k,t-l}] \\ \vdots & \vdots & & \vdots \\ E[\tilde{z}_{k,t}\tilde{z}_{1,t-l}] & E[\tilde{z}_{k,t}\tilde{z}_{2,t-l}] & \dots & E[\tilde{z}_{k,t}\tilde{z}_{k,t-l}] \end{bmatrix}$$

Here $\tilde{z}_t \equiv z_t - \boldsymbol{\mu}$.

From this it is natural to define the cross correlation matrix as:

$$\boldsymbol{\rho}_l = \mathbf{C}^{-1}\boldsymbol{\Gamma}_l\mathbf{C}^{-1}$$

Here $\mathbf{C} = \text{dig}\{\sigma_1, \dots, \sigma_k\}$ is the diagonal matrix with standard deviations for each time series on its diagonals.

The *Ljung Box test* is a statistical test for dynamic dependence in data i.e. it tests if there are correlations between the time lags or between different time series. The null hypothesis and the alternative hypothesis can be written as below.

$$H_0: \boldsymbol{\rho}_1 = \dots = \boldsymbol{\rho}_m = 0$$

$$H_1: \boldsymbol{\rho}_i \neq 0 \text{ for some } i \in \{1, \dots, m\}.$$

One calculates a test statistic as in [5] or [4].

$$\hat{Q}_l = T^2 \sum_{i=1}^l \frac{1}{T-l} \text{tr}(\hat{\Gamma}_i \hat{\Gamma}_0^{-1} \hat{\Gamma}_i \hat{\Gamma}_0^{-1}).$$

\hat{Q}_l can be written in the form

$$\hat{Q}_l = T^2 \sum_{i=1}^l \frac{1}{T-l} \hat{\mathbf{b}}_i \hat{\boldsymbol{\rho}}_i^{-1} \otimes \hat{\boldsymbol{\rho}}_i^{-1} \hat{\mathbf{b}}_i.$$

Here $\hat{\mathbf{b}}$ is $\text{vec}(\hat{\boldsymbol{\rho}}_i')$. Here vec is the notation for vectorization of a matrix, see Appendix A. The computed statistic is compared against χ^2 distribution to see if the null hypothesis can be rejected.

2.2 Empirical distributions and quantiles

Consider samples $\mathbf{x}_1, \dots, \mathbf{x}_n$ of the iid d -dimensional random variables $\mathbf{X}_1, \dots, \mathbf{X}_n$ with the unknown distribution function $F(x) = P(\mathbf{X} < \mathbf{x})$. Here $\mathbf{X} < \mathbf{x}$ if and only if $X_j < x_j$ for $j \in \{1, \dots, d\}$. One can approximate the distribution function by assigning probability weights $\frac{1}{n}$ to each \mathbf{x}_k [6]. The empirical distribution F_n is defined by

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I(x_i < x)$$

The I is the indicator function that takes the values 1 and 0 if corresponding input condition are true or false. The empirical distribution shall be interpreted as an outcome of the random variable.

$$F_{n,X}(X) = \frac{1}{n} \sum_{i=1}^n I(X_i < X).$$

Empirical quantiles are quantiles of the empirical distribution and defined as below

$$F_n^{-1}(p) = \min\{x : F_n(x) \geq p\}.$$

It turns out that the empirical quantile can be obtained by ordering x in decreasing order and then picking the x number $[n(1-p)]+1$.

$$F_n^{-1}(p) = X_{[n(1-p)]+1,n}.$$

2.3 Bootstrapping

In this section the bootstrap procedure is presented. First some explanation what bootstrap is, then how bootstrap is applied on time series models [2].

Bootstrapping is a method to create fictive data points (resampling) from data samples, often with simulating methods. A common use of bootstrap is to investigate different estimates, e.g. mean values and standard deviations. The procedure is rather simple, data is drawn with replacement from a sample to generate a new series of data [2].

Consider x_1, \dots, x_n to be a sample from iid random variables X_1, \dots, X_n with the distribution function F . Let us further consider the numerical estimate $\hat{\theta}(x_1, \dots, x_n)$ and the similar stochastic variable $\hat{\theta}(X_1, \dots, X_n)$, which is used to estimate the true value θ . To generate new samples of $\hat{\theta}(X_1, \dots, X_n)$ one needs to find an estimate for the distribution F i.e. \hat{F}_n . The samples simulated from \hat{F}_n are denoted $\hat{\theta}^* = \hat{\theta}(X_1^*, \dots, X_n^*)$.

In the basic bootstrap hypothesis it is assumed that the copy \hat{F}_n is a good approximation of the true distribution F [2]. For a wide class of functions S it is assumed that

$$S(\hat{F}_n) \approx S(F).$$

This assumption implies that the distribution of $\hat{\theta}(X_1, \dots, X_n) - \theta$ is well approximated by the distribution of $\hat{\theta}(X_1^*, \dots, X_n^*) - \hat{\theta}(x_1, \dots, x_n)$.

There are different kinds of bootstrap methods. Two of them are non-parametric and parametric. In the non-parametric bootstrap the estimate of F is done with an empirical distribution. In the parametric bootstrap the estimation of F is done for a certain parametric family. The parameters for the parametric-family can be estimated, which makes it possible to simulate from this parametric model [2].

When bootstrapping time series data a similar parametric and non-parametric modelling can be used as described above [2]. The first method is one that has a parametric model as base e.g. one find that an AR(1) model is suitable for the data. Subtracting this model from the data points gives the residual. To get new estimates of the parameters in the AR(1) model one can draw with replacement the residuals and create a set of new time series. For each time series one gets a estimate of the parameters in the AR(1) model. A disadvantage with this model is that it is heavily dependent on which time series model that has been chosen.

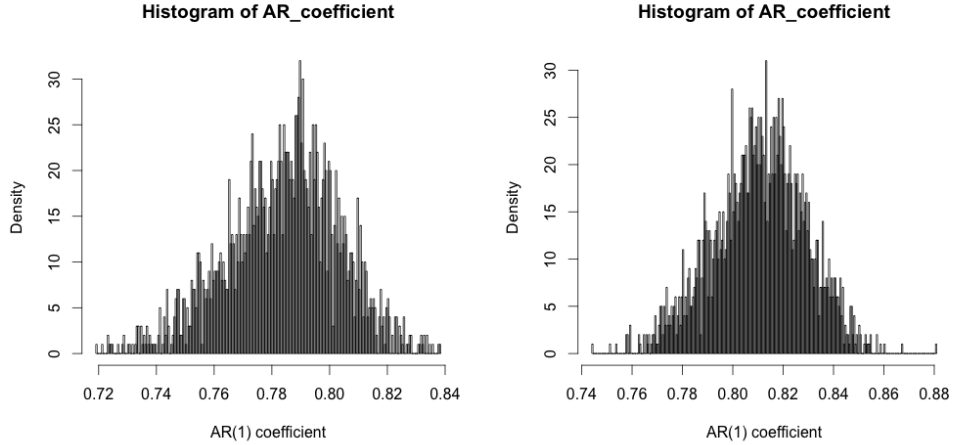


Figure 1: *Left:* Estimated lag coefficients from bootstrapped simulation, where empirical mean is 0.78 *Right:* Estimated lag coefficients from bootstrapped simulation, where empirical mean is 0.815

Figure 1 illustrates estimates of lag coefficient in an AR(1) process. The estimates are done with parametric bootstrap for time series models. The data comes from two simulations of an AR(1) model with lag coefficient 0.8. It can be seen that the method is model dependent.

The second method is a non-parametric method. One divides the time series in overlapping time blocks of a certain length. Then one draws from these blocks with replacement. This procedure is repeated. The idea with this method is that the short time dependence is preserved. There is a big arbitrariness in the choice of block length. There are advantages to choose a large block size to preserve the long time dependence, but then one underestimates the variability, but if one chooses a too small block length one implicitly assumes a rapid decaying dependence in the time series, but get a good variability in the simulated time series.

2.4 Parameter estimation

To estimate the parameters in the VAR model OLS is used in this report. The model is

$$x_t = a_1 + c_{11}x_{t-1} + c_{12}y_{t-1} + c_{13}z_{t-1} + \epsilon_t^1$$

$$y_t = a_2 + c_{21}x_{t-1} + c_{22}y_{t-1} + c_{23}z_{t-1} + \epsilon_t^2$$

$$z_t = a_2 + c_{31}x_{t-1} + c_{32}y_{t-1} + c_{33}z_{t-1} + \epsilon_t^3.$$

The equations above can be written in matrix notation,

$$\begin{bmatrix} x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & y_1 & z_1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n-1} & y_{n-1} & z_{n-1} \end{bmatrix} \begin{bmatrix} a_1 \\ c_{11} \\ c_{12} \\ c_{13} \end{bmatrix} + \begin{bmatrix} \epsilon_2^1 \\ \vdots \\ \epsilon_n^1 \end{bmatrix}$$

$$\begin{bmatrix} y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & y_1 & z_1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n-1} & y_{n-1} & z_{n-1} \end{bmatrix} \begin{bmatrix} a_2 \\ c_{21} \\ c_{22} \\ c_{23} \end{bmatrix} + \begin{bmatrix} \epsilon_2^2 \\ \vdots \\ \epsilon_n^2 \end{bmatrix}$$

$$\begin{bmatrix} z_2 \\ \vdots \\ z_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & y_1 & z_1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n-1} & y_{n-1} & z_{n-1} \end{bmatrix} \begin{bmatrix} a_3 \\ c_{31} \\ c_{32} \\ c_{33} \end{bmatrix} + \begin{bmatrix} \epsilon_2^3 \\ \vdots \\ \epsilon_n^3 \end{bmatrix}.$$

The vectors in the above equations can be set to \mathbf{x}_t , \mathbf{y}_t , \mathbf{z}_t respectively. The matrices in the right-hand side of the above equations can be set to \mathbf{X} and the coefficient vectors for each regression is called \mathbf{c}^x , \mathbf{c}^y and \mathbf{c}^z . Then the coefficients can be estimated as

$$\hat{\mathbf{c}}^x = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{x}_t$$

$$\hat{\mathbf{c}}^y = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_t$$

$$\hat{\mathbf{c}}^z = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{z}_t.$$

2.5 Macroeconomics

In this section some macro economic concepts are introduced. These are the meaning of real, nominal, CPI, inflation, pay as you go system and funded system.

2.5.1 CPI and inflation

Consumer price index (CPI) is the most common measure of price development and is used for measuring the inflation. The purpose of CPI is to show how consumer prices on average change in the domestic consumption [7]. The CPI is used to calculate the inflation. The inflation is calculated as log return, which is described in the next section.

2.5.2 Real and nominal Returns

The word nominal is an indication that something is measured in money e.g. nominal domestic gross product (nominal GDP) [8]. The term real indicates that something is measured in the amount of goods and services that can be purchased with the income [8].

The real return for e.g. wages in Sweden gives more information about the wage growth than the nominal return for wages. Consider the case where goods have increased very much in price from one year to the next. If the wages increase in the same rate, a worker cannot buy more goods for his money even though he has a larger amount of money. The real return of wages takes the inflation into account and can be interpreted as how much more a worker can buy for his wage from one year to the next.

When modelling e.g. historical share prices it can be helpful to consider returns of share prices. The returns can usually be considered weakly dependent and close to identical distributed [6]. Equation (7) is used to calculate real returns [8].

$$1 + \text{nominal interest rate} = (1 + \text{real interest rate})(1 + \text{inflation rate}). \quad (7)$$

If log returns are used one can rewrite equation (7) as

$$\begin{aligned} r^N &= \ln(1 + \text{nominal interest rate}) \\ r^R &= \ln(1 + \text{real interest rate}) \\ r^I &= \ln(1 + \text{inflation rate}). \end{aligned}$$

$$\ln(1 + \text{nominal interest rate}) = \ln(1 + \text{real interest rate}) + \ln(1 + \text{inflation rate}) \quad (8)$$

This yields:

$$r^R = r^N - r^I. \tag{9}$$

The Swedish “inkomstpension” can be described as a pay as you go system which means that this years income becomes this years payments to the pensioners [8]. An alternative to the pay as you go system is a fully funded system where one saves money and receives that money when it is time for the first pension payment [8].

3 Model

To fulfill the objectives of the thesis, a VAR model is fitted to the variables, average wage return x_t , returns from the AP-funds y_t and inflation z_t . When a model is found it is simulated using bootstrap and new values are received. The values are put back in the dependent variables real average wage return x_t^r , average wage AW_t , real returns from the AP-funds y_t^r and income index I_t . The standard deviation is calculated in each time step for the dependent variables.

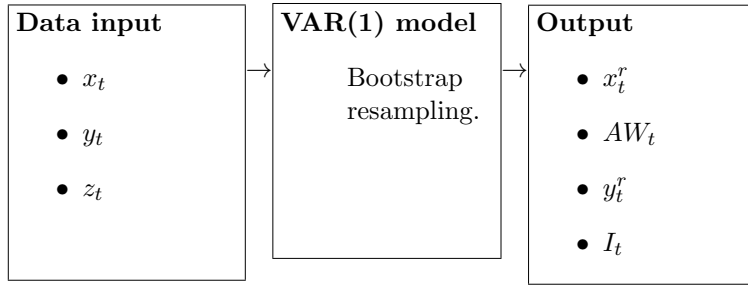


Figure 2: Flow chart illustrates the input and output of the model. Here x_t is average wage return, y_t is AP-funds return, z_t inflation, x_t^r is the real average wage return, AW_t is average wage, y_t^r is real return from the AP-funds and I_t is the income index.

3.1 VAR and bootstrap

The model procedure is a bootstrap scheme that is similar to the scheme found in [9]. The procedure includes 4 steps that are described below.

- (1) First one estimate the VAR(1) model from x_t , y_t and z_t . Here $t \in \{1, \dots, n\}$.

$$\mathbf{X}_t = \hat{A}_0 + \hat{A}_1 \mathbf{X}_{t-1} + \hat{\epsilon}_t.$$

Where,

$$\mathbf{X}_t = \begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix}, \quad \hat{A}_0 = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}, \quad \hat{A}_1 = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix}, \quad \hat{\epsilon}_t = \begin{bmatrix} \hat{\epsilon}_t^1 \\ \hat{\epsilon}_t^2 \\ \hat{\epsilon}_t^3 \end{bmatrix}$$

The estimation is performed with the VAR function in R, which calculates the constants in the VAR model with ordinary least squares (OLS). The residuals are calculated with the following equation.

$$\hat{\epsilon}_t = \mathbf{X}_t - (\hat{A}_0 + \hat{A}_1 \mathbf{X}_{t-1})$$

If a process $\hat{\epsilon}_t$ is uncorrelated in each time step, then every $\hat{\epsilon}_t$ has covariance matrix Σ and $\hat{\epsilon}_t$ is referred to as white noise [5].

- (2) The next step in the procedure is to draw with replacement from the residuals $\hat{\epsilon}_t$. This is done by simulating a uniform distribution $U(1, n - 1)$ and letting $u \in \{1, \dots, n - 1\}$ be the outcomes of the distribution (there exists $n - 1$ residuals from a VAR(1) model if $t \in \{1, \dots, n\}$). Hence for each draw the residual $\hat{\epsilon}_u^*$ is obtained. Note that by the definition of uniform distribution each residual is drawn with the same probability.
- (3) When the residuals have been drawn new points can be calculated. Note that t_1 is our start value and hence it is always the same but for $t \in \{2, \dots, n\}$ a new residual is drawn and a new point is calculated.

$$\mathbf{X}_t^* = \hat{A}_0 + \hat{A}_1 \mathbf{X}_{t-1}^* + \hat{\epsilon}_u^*. \quad (10)$$

When this is done for all t new \hat{A}_0 \hat{A}_1 can be estimated.

- (4) The VAR model is now complete and it is possible to simulate values for x_t , y_t and z_t . The crucial part is to put the simulated values back in our directly dependent variables CPI_t , AW_t and I_t . Then the variance is calculated on these variables.

To check how good our coefficient estimates are, confidence intervals are calculated. This is done with a chosen confidence level $q = 0.05$ on the empirical quantiles of each element in the coefficient matrices \hat{A}_0 and \hat{A}_1 .

The confidence intervals are calculated using equation (11) from [6]

$$I_{a,q} = (\hat{a}_{obs} + R_{[N(1+q)/2]+1,N}^*, \hat{a}_{obs} + R_{[N(1-q)/2]+1,N}^*). \quad (11)$$

Here a is the true values of our coefficients, \hat{a}_{obs} is the estimated value from x_t , y_t and z_t , q is the confidence level and $R_j^* = \hat{a}_{obs} - a_j^*$. The R^* vector is sorted in decreasing order i.e. $R_{1,N}^* \geq R_{2,N}^* \geq R_{3,N}^* \geq \dots R_{N,N}^*$. Then one choose the element $[N(1 + q)/2] + 1$ and $[N(1 - q)/2] + 1$ in the R^* vector. Here the $[]$ bracket denote the floor function.

Next, the variables real return from the AP-funds, real average wage return, average wage and income index are calculated from the simulated values in the equation (10). Let real returns from the AP-funds be denoted by y_t^r then

$$y_{i,t}^r = y_{i,t} - z_{i,t}. \quad (12)$$

Here $i \in \{1, \dots, m\}$ is an index for each simulation, there are m simulations. Similarly the real average wage return x_t^r is calculated.

$$x_{i,t}^r = x_{i,t} - z_{i,t}. \quad (13)$$

The average wage $AW_{i,t}^*$ is calculated as

$$AW_{i,t}^* = AW_{i,t-1}^* \cdot e^{x_{i,t}^r}. \quad (14)$$

Here the start value is $AW_{t=1967}$ used for simulation of the years 1967-2013.

Finally the income index $I_{i,t}^*$ is calculated from $CPI_{i,t}^*$ values and average wages

$$CPI_{i,t}^* = CPI_{i,t-1}^* \cdot e^{z_{i,t}^r}. \quad (15)$$

Here $CPI_{t=1967}$ is the start value used for simulation of the years 1967-2013.

$$I_{i,t}^* = I_{i,t-1}^* \cdot \left(\frac{AW_{i,t-1}^*}{AW_{i,t-4}^*} \frac{CPI_{i,t-4}^*}{CPI_{i,t-1}^*} \right)^{\frac{1}{3}} \cdot \frac{CPI_{i,t-1}^*}{CPI_{i,t-2}^*} \quad (16)$$

Next the standard deviation is calculated for real return from the AP-funds, real average wage return, average wage and income index. The calculations are the same for every parameter therefore it is only shown for $y_{i,t}^r$

$$\sigma_t = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_{i,t}^r - \mu_t)^2}$$

where μ_t is

$$\mu_t = \frac{1}{n} \sum_{i=1}^n y_{i,t}^r.$$

3.2 Scenario analysis

3.2.1 Scenario drift

To investigate different scenarios, a drift is added to the variables average wage return and inflation.

Consider the data set $\mathbf{X} = [x_t, y_t, z_t]$ here \mathbf{X} is a $3 \times n$ matrix with the data as columns where $t \in \{1, \dots, n\}$. In the first scenario, a drift μ is added to the variable x_t . Hence the data set can be written as $\mathbf{X} = [\mu \cdot I_{n \times 1} + x_t, y_t, z_t]$, where, $I_{n \times 1}$ is a row vector of ones. This procedure is done to create a low and high scenario for average wage where $\mu = -0.07$ for the low scenario and $\mu = 0.07$ for the high scenario.

This procedure is repeated for inflation and hence $\mathbf{X} = [x_t, y_t, \mu \cdot I_{n \times 1} + z_t]$ is generated for the same values of μ . This creates high and low values for inflation.

The variable y_t (AP-funds return) is not investigated as the other two variables since it can be described as white noise.

When this procedure has been performed the steps under section VAR bootstrap are repeated to see how the coefficient estimates vary when the altered data is used to estimate the new model.

Two scenarios are investigate for the variables average wage return and inflation in the VAR model, to see how these scenarios affect the output of equations (13), (14), (16), (15), i.e. how does the coefficients in the VAR model change according to the changes in input data and how income index, real average wage return and average wage change during the scenarios.

3.2.2 Scenario probability

Another method to obtain high and low scenarios is to sort the residuals in a negative and a positive part and use equation (17).

$$\mathbf{X}_t = \begin{cases} A_0 + A\mathbf{X}_{t-1} + \epsilon_t^+, & p \leq p_1 \\ A_0 + A\mathbf{X}_{t-1} + \epsilon_t^-, & p > p_1 \end{cases} \quad (17)$$

In equation (17) ϵ_t^+ are the positive residuals and ϵ_t^- negative residuals. p are random samples from the uniform distribution $U(0, 1)$ and $p_1 \in [0, 1]$ is a constant by design. If $p_1 = 0.3$ then it is more likely that one draw negative residuals than positive residuals see

(18), this creates the low scenario. The high scenario is created by simply switch the inequalities see equation (19)

$$\mathbf{X}_t = \begin{cases} A_0 + A\mathbf{X}_{t-1} + \epsilon_t^+, & p \leq 0.3 \\ A_0 + A\mathbf{X}_{t-1} + \epsilon_t^-, & p > 0.3 \end{cases} \quad (18)$$

$$\mathbf{X}_t = \begin{cases} A_0 + A\mathbf{X}_{t-1} + \epsilon_t^+, & p > 0.3 \\ A_0 + A\mathbf{X}_{t-1} + \epsilon_t^-, & p \leq 0.3 \end{cases} \quad (19)$$

4 Results scenario analysis

Below are the results of the modelling presented. For the VAR model to be of any use in the scenario analysis one need to find at least one non-diagonal coefficient in the matrix of coefficients to be non zero. If all the non-diagonal coefficients are zero then the processes can be modelled as three univariate time series i.e. a change in one parameter does not affect the other parameters.

First the data used for modeling is presented, then the VAR model followed by results form scenarios analysis and lastly everything is put together to analyse the effect on the pension balance.

4.1 Data

In this section data for the modelling is presented. The raw data has been collected from "Statistiska centralbyrån" (SCB) [10] and [7] and the Pension agency [11] and [12]. The data consists of average wages, CPI and returns from the AP-funds. By using the raw data income index is calculated with (16) and pension balance is calculated with (1). The time period for which the data were selected is 1967-2013. The reason for choosing this time period was that for certain years before 1967 the average wage was replaced with median wage that seemed like a poor substitute for the average.

The CPI data used has a reference year, which is the year 1914, and therefore prices are compared to that year with the reference value 100.

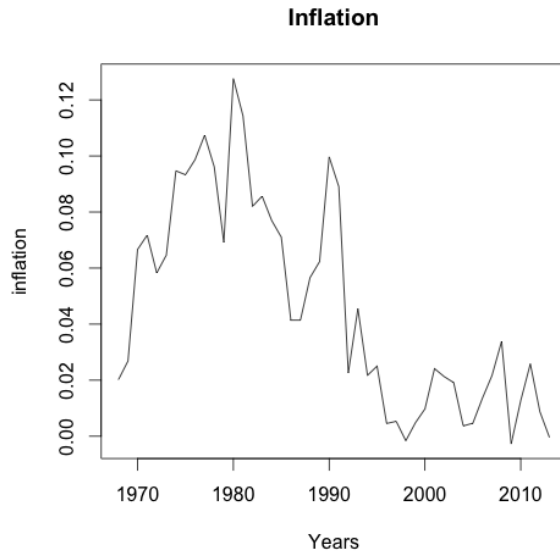


Figure 3: Inflation calculated using CPI from SCB.

Figure 3 shows a time series for inflation which is calculated as log returns of CPI.

The average wage is defined as total earned income for residents in Sweden from the first of januari to the 31 december the same year. All income classes are included in the data even persons who have a zero income. The pension agency does not use average wage when they calculate the parameters mention above, instead they use PGI (pensionable income). When the average of PGI is calculated it excludes persons who have a zero income. Hence the PGI is greater than average income. All calculations in this report are done with average income from SCB. Before 1991 capital income was included in the calculations this is not included after 1991.

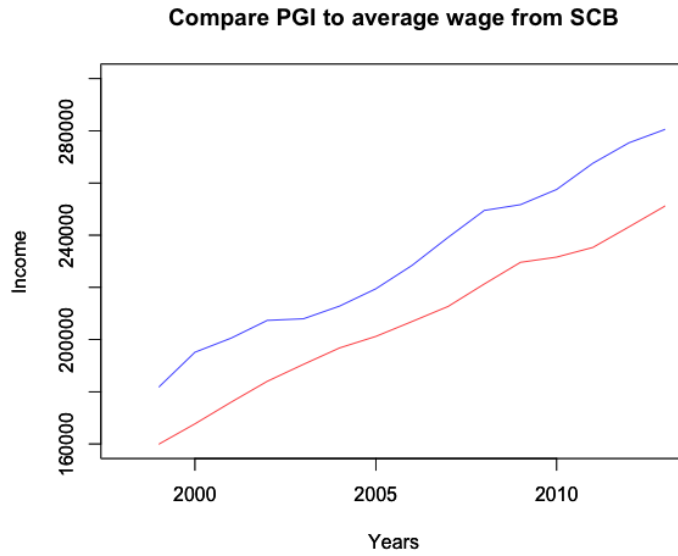


Figure 4: The red line is the average wage from SCB and the blue line is average PGI from the pension agency.

Note that the average wage is smaller than the pensionable income, which can be seen in Figure 4.

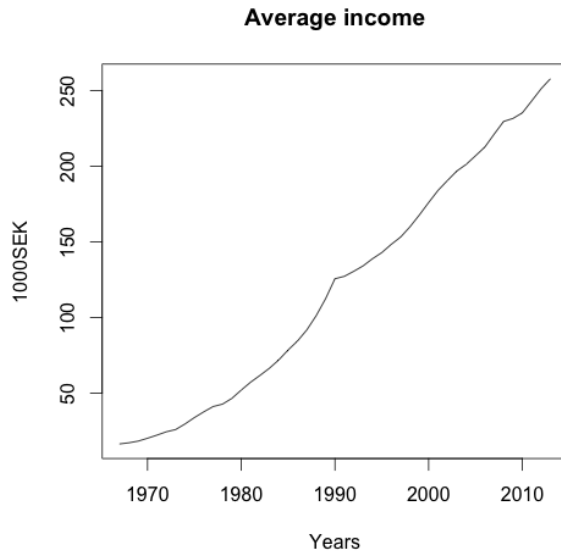


Figure 5: Average income from SCB (years 1967-2013).

Figure 5 shows the average income in Sweden for the years 1967-2013. The data has been collected from SCB as described above. The average wage for year t will be denoted as AW_t .

Returns from AP-funds in units of percent are received from the pension agency, calculated as the net yield divided by fund capital and half of "flow" (i.e. fees and other income are subtracted with expenses) [11]. To make this data consistent with the average wage returns and inflation, log returns are calculated. Finally the real AP-fund returns are obtained with the following equation

$$y_t^r = y_t - z_t. \quad (20)$$

Here y_t^r is the real return from the AP-funds, y_t is the return from the AP-funds and z_t is the inflation. Real returns from the AP-funds for the considered observation period is shown in Figure 6.

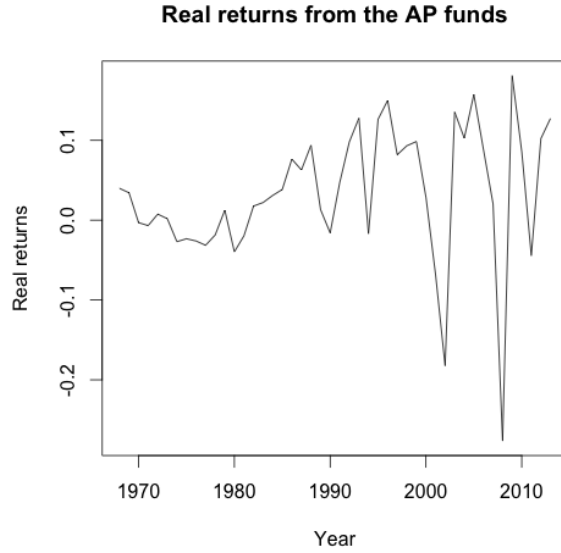


Figure 6: The real return from the AP funds for the years 1967-2013.

4.2 Vector autoregressive model

When one fits the VAR(1) model with OLS to the data set x_t (average wage return), y_t (AP-funds return) and z_t (inflation) the model below is obtained.

$$\mathbf{X}_t = \begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix} = \begin{bmatrix} 0.01 \\ 0.07 \\ 0.0 \end{bmatrix} + \begin{bmatrix} 0.74 & 0.05 & 0.00 \\ -0.25 & 0.02 & 0.42 \\ 0.50 & 0.03 & 0.50 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t^1 \\ \epsilon_t^2 \\ \epsilon_t^3 \end{bmatrix} \quad (21)$$

The following cross correlation were obtained from the data sets.

$$\boldsymbol{\rho}_0 = \begin{bmatrix} 1.00 & -0.02 & 0.76 \\ -0.02 & 1.00 & 0.01 \\ 0.76 & 0.01 & 1.00 \end{bmatrix} \quad (22)$$

$$\boldsymbol{\rho}_1 = \begin{bmatrix} 0.71 & 0.11 & 0.53 \\ 0.04 & 0.03 & 0.11 \\ 0.79 & 0.07 & 0.80 \end{bmatrix} \quad (23)$$

l	$Q(l)$	df	p-value
1.00	4.13	9.00	0.90
2.00	14.72	18.00	0.68
3.00	20.87	27.00	0.79
4.00	22.28	36.00	0.96
5.00	36.76	45.00	0.80
6.00	45.09	54.00	0.80
7.00	47.94	63.00	0.92
8.00	52.26	72.00	0.96
9.00	64.01	81.00	0.92
10.00	76.81	90.00	0.84

Table 1: The table shows results of testing the residuals from the fitted VAR(1) model with a Ljung-Box test.

It can be seen in Table 1 that all p-values are high which indicate that the null hypothesis that residuals are not white noise cannot be rejected.

The covariance matrix for the residual can be seen in equation (24). The covariance values are small, approximately zero, between the residuals from the VAR model.

$$\Sigma = \begin{bmatrix} 0.001 & 0.000 & 0.000 \\ 0.000 & 0.006 & 0.000 \\ 0.000 & 0.000 & 0.000 \end{bmatrix} \quad (24)$$

Figures, 7-10 are created from simulations of equation 21.

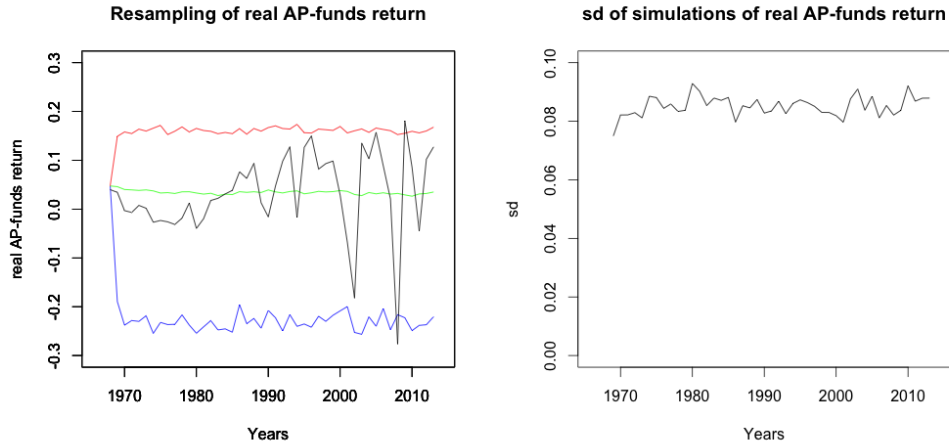


Figure 7: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line). *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real return from the AP-funds.

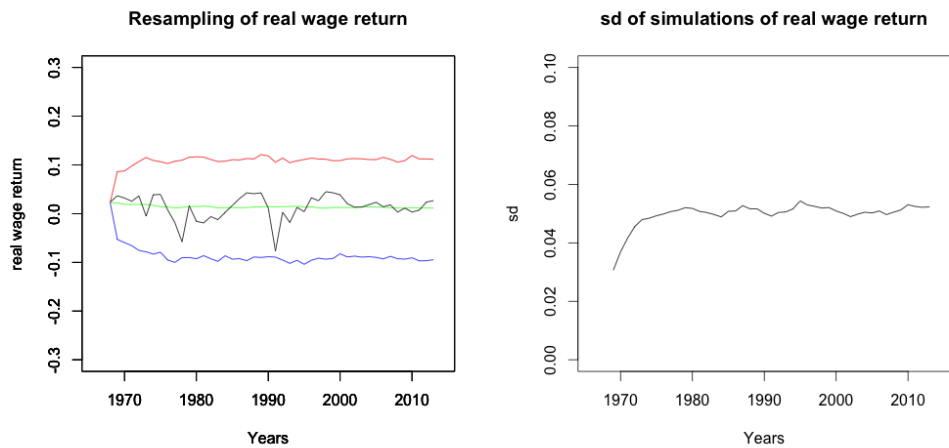


Figure 8: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line). *Right:* Plot shows standard deviation for each time step in the bootstrapped resampling procedure of real average wage returns.

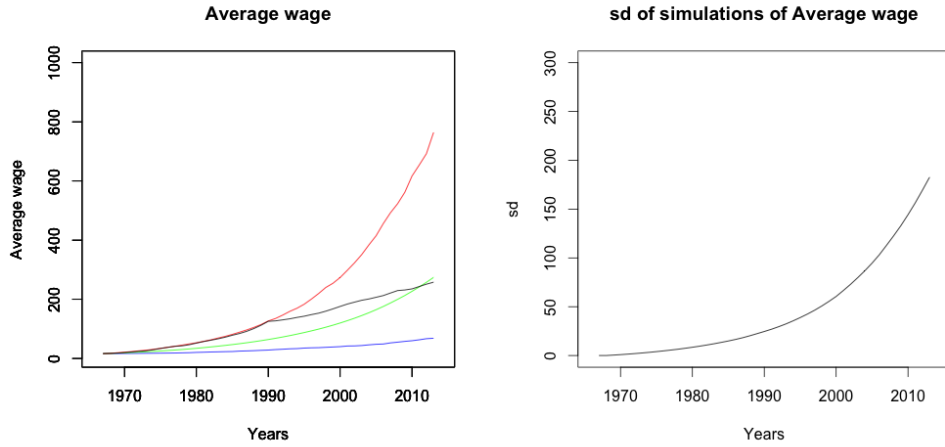


Figure 9: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line). *Right:* Plot shows standard deviation for each time step in the bootstrapped resampling procedure of average wage.

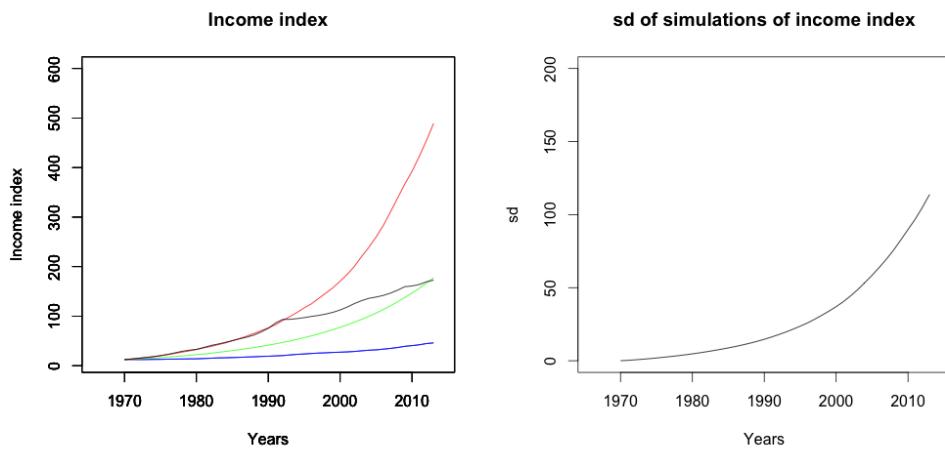


Figure 10: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line). *Right:* Plot shows standard deviation for each time step in the bootstrapped resampling procedure of income index.

4.3 Uncertainty in coefficient estimation

In the previous section "model" it is described in step (4) how the uncertainty in the coefficient estimates are investigated. The result is that returns from the AP-fund can be modelled as a random walk.

Coefficient	I_l	I_u
a_1	-0.02	0.02
a_2	0.00	0.12
a_3	-0.03	0.00
c_{11}	0.49	1.13
c_{21}	-1.36	1.03
c_{31}	0.26	0.76
c_{12}	-0.05	0.17
c_{22}	-0.21	0.34
c_{32}	-0.04	0.11
c_{13}	-0.28	0.32
c_{23}	-0.65	1.57
c_{33}	0.32	0.77

Table 2: The table shows confidence intervals calculated as described under "model" step (4) with average wage return, AP-funds return and inflation. The data are for the years 1967-2013.

From Table 2 it can be seen that on a 95% confidence level the following system of time series can be obtained.

$$\begin{aligned}x_t &= c_{11}x_{t-1} + \epsilon_t^1 \\y_t &= \epsilon_t^2 \\z_t &= c_{31}x_{t-1} + c_{33}z_{t-1} + \epsilon_t^3.\end{aligned}$$

Let us note that the system can be written in triangular form

$$\begin{aligned}z_t &= c_{33}z_{t-1} + c_{31}x_{t-1} + \epsilon_t^3 \\x_t &= c_{11}x_{t-1} + \epsilon_t^1 \\y_t &= \epsilon_t^2.\end{aligned}$$

Which associates the recursive modeling ideas in econometrics by Herman Wold.

4.4 Scenarios with drift

4.4.1 Low average wage returns

In the scenario low average wage return, a drift $\mu = -0.07$ is added to log returns of average wage. The scenario generates equation (25). The following VAR model is contained

$$\mathbf{X}_t = \begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix} = \begin{bmatrix} -0.01 \\ 0.06 \\ 0.03 \end{bmatrix} + \begin{bmatrix} 0.74 & 0.05 & 0.00 \\ -0.25 & 0.02 & 0.42 \\ 0.50 & 0.03 & 0.50 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t^1 \\ \epsilon_t^2 \\ \epsilon_t^3 \end{bmatrix}. \quad (25)$$

Coefficients	I_l	I_u
a_1	-0.03	0.01
a_2	-0.02	0.12
a_3	0.01	0.04
c_{11}	0.50	1.13
c_{21}	-1.49	0.88
c_{31}	0.19	0.73
c_{12}	-0.06	0.17
c_{22}	-0.21	0.32
c_{32}	-0.04	0.12
c_{13}	-0.27	0.32
c_{23}	-0.63	1.66
c_{33}	0.34	0.80

Table 3: Confidence interval for the scenario low wage. A drift of $\mu = -0.07$ is added to the log returns of average wage from the years 1967-2013.

Figure 11-14 are obtained from simulations of equation (25).

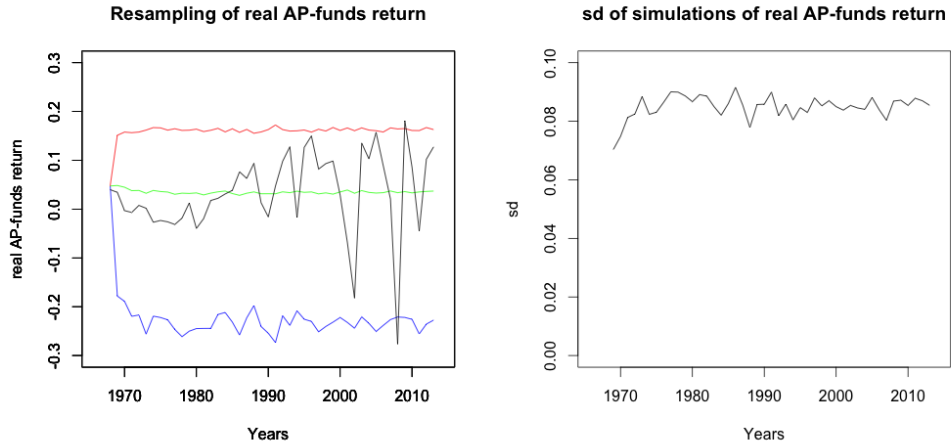


Figure 11: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line). *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average AP-funds return.

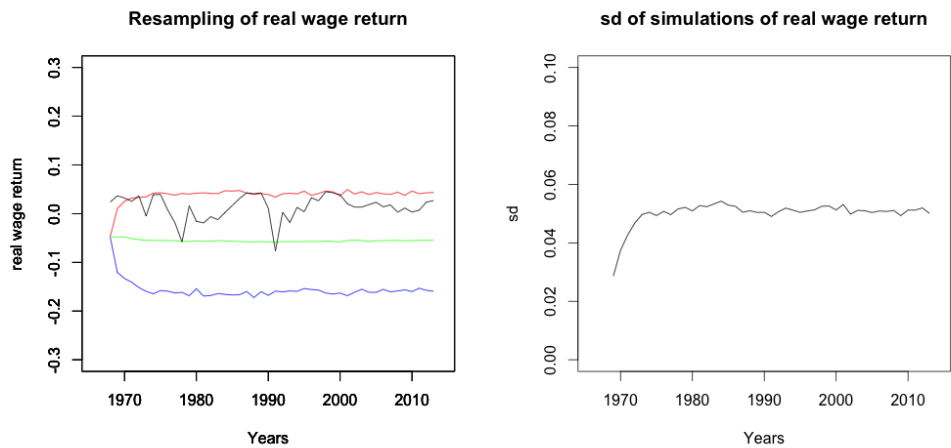


Figure 12: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line). *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average wage return.

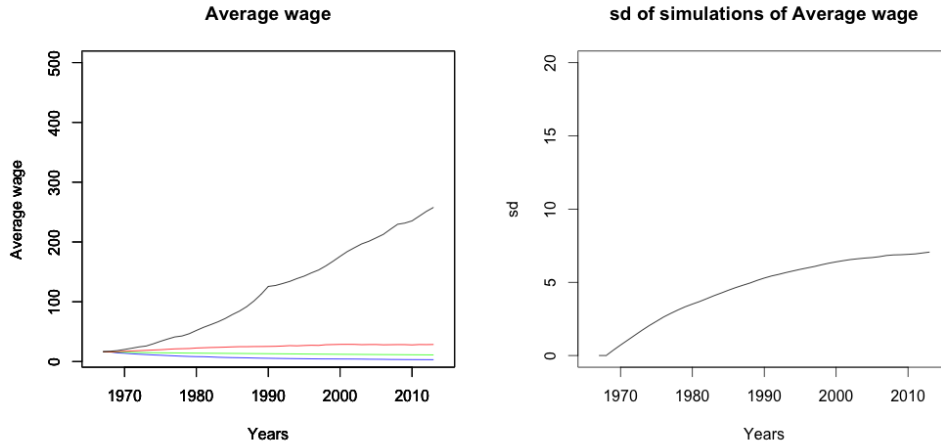


Figure 13: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line), the values are in 1000 SEK. *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average wage return.

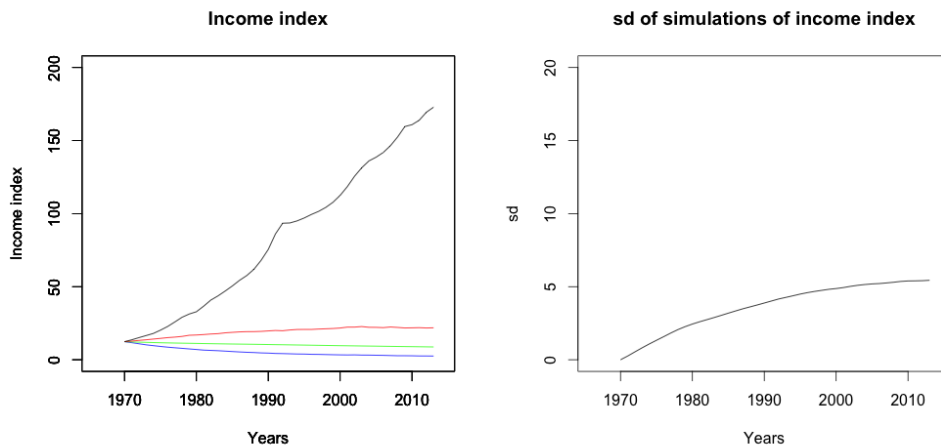


Figure 14: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line). *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average wage return.

4.4.2 High average wage returns

In the scenario High average wage return, a drift $\mu = 0.07$ is added to log returns of average wage.

$$\mathbf{X}_t = \begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix} = \begin{bmatrix} 0.03 \\ 0.09 \\ -0.04 \end{bmatrix} + \begin{bmatrix} 0.74 & 0.05 & 0.00 \\ -0.25 & 0.02 & 0.42 \\ 0.50 & 0.03 & 0.50 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t^1 \\ \epsilon_t^2 \\ \epsilon_t^3 \end{bmatrix} \quad (26)$$

Coefficients	I_l	I_u
a_1	-0.01	0.06
a_2	-0.17	0.13
a_3	-0.07	-0.01
c_{11}	0.45	1.08
c_{21}	-1.06	1.66
c_{31}	0.21	0.71
c_{12}	-0.01	0.15
c_{22}	-0.15	0.69
c_{32}	-0.02	0.10
c_{13}	-0.27	0.30
c_{23}	-0.97	1.67
c_{33}	0.34	0.79

Table 4: Confidence interval for the scenario low wage. A drift of $\mu = 0.07$ is added to the log returns of average wage from the years 1967-2013.

Figures 15-18 are obtained from simulations of equation (26).

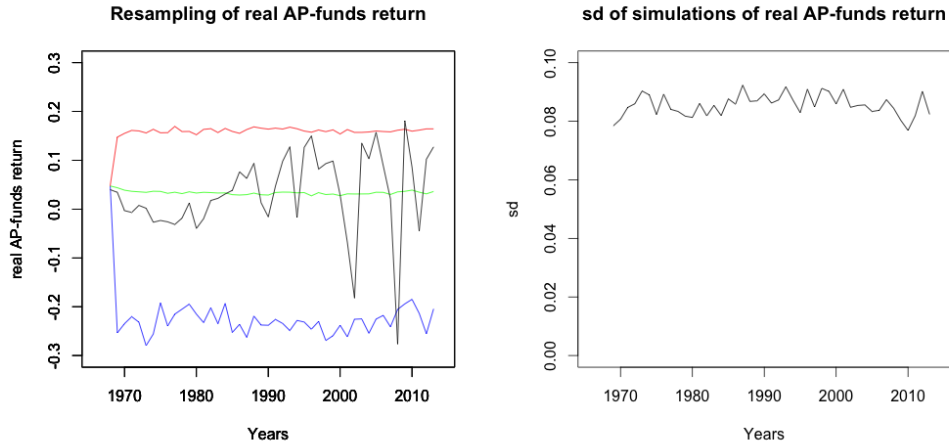


Figure 15: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line). *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average AP-funds return.

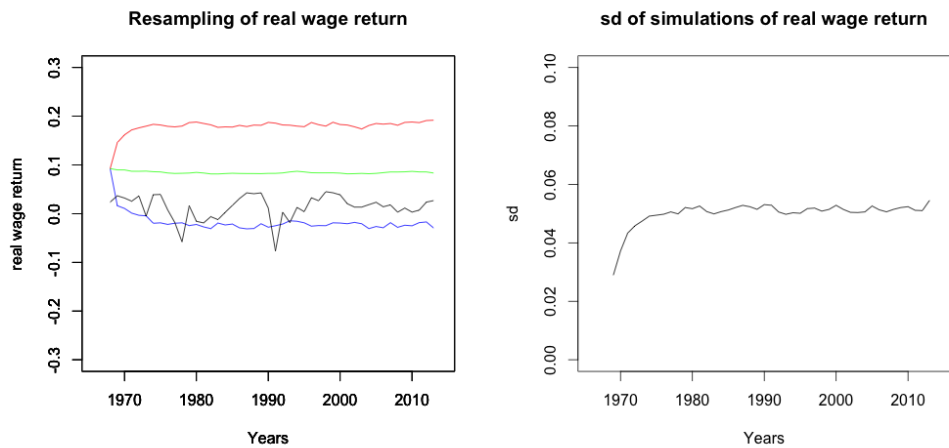


Figure 16: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line). *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average wage returns.

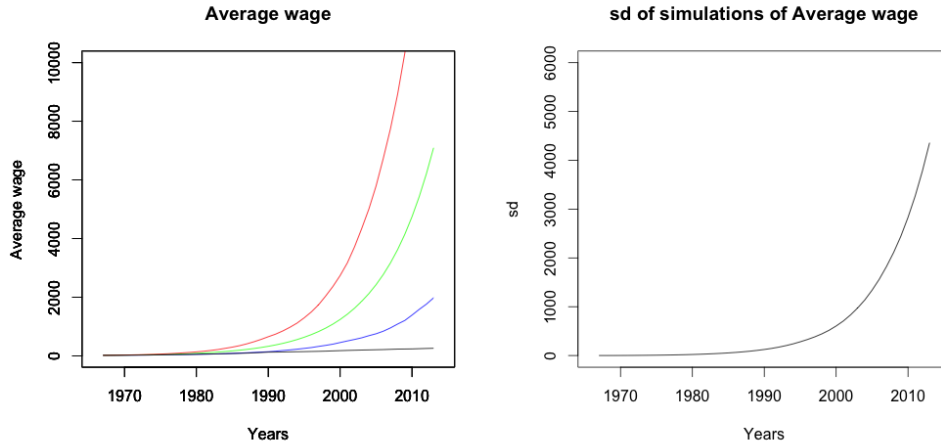


Figure 17: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line), the values are in 1000 SEK. *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average wages.

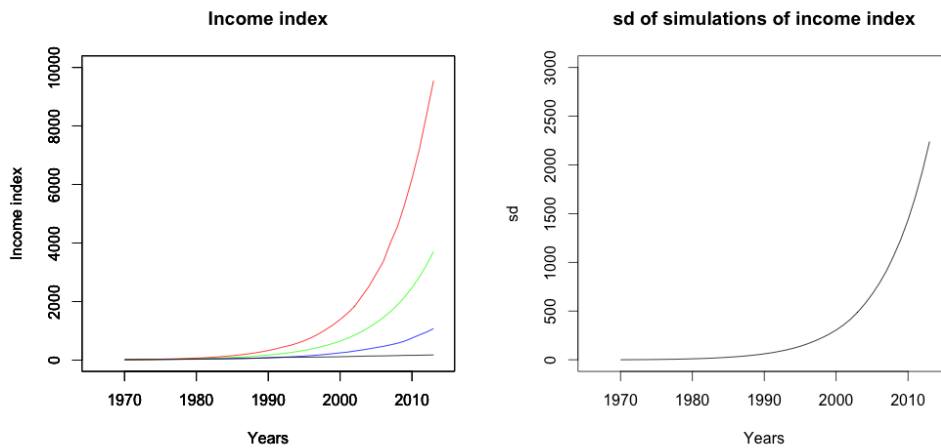


Figure 18: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line). *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average income index

4.4.3 Low inflation

In the scenario low inflation a drift $\mu = -0.07$ is added to log returns of inflation.

$$\mathbf{X}_t = \begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix} = \begin{bmatrix} 0.01 \\ 0.10 \\ -0.04 \end{bmatrix} + \begin{bmatrix} 0.74 & 0.05 & 0.00 \\ -0.25 & 0.02 & 0.42 \\ 0.50 & 0.03 & 0.50 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t^1 \\ \epsilon_t^2 \\ \epsilon_t^3 \end{bmatrix} \quad (27)$$

Coefficients	I_l	I_u
a_1	-0.02	0.04
a_2	-0.07	0.16
a_3	-0.06	-0.02
c_{11}	0.45	1.08
c_{21}	-1.12	1.71
c_{31}	0.20	0.72
c_{12}	-0.01	0.15
c_{22}	0.14	0.70
c_{32}	-0.02	0.10
c_{13}	-0.27	0.30
c_{23}	-0.82	1.82
c_{33}	0.33	0.76

Table 5: Confidence interval for the scenario low wage. A drift of $\mu = -0.07$ is added to the log returns of inflation from the years 1967-2013.

Figures 19-22 are obtained from simulations of equation (27).

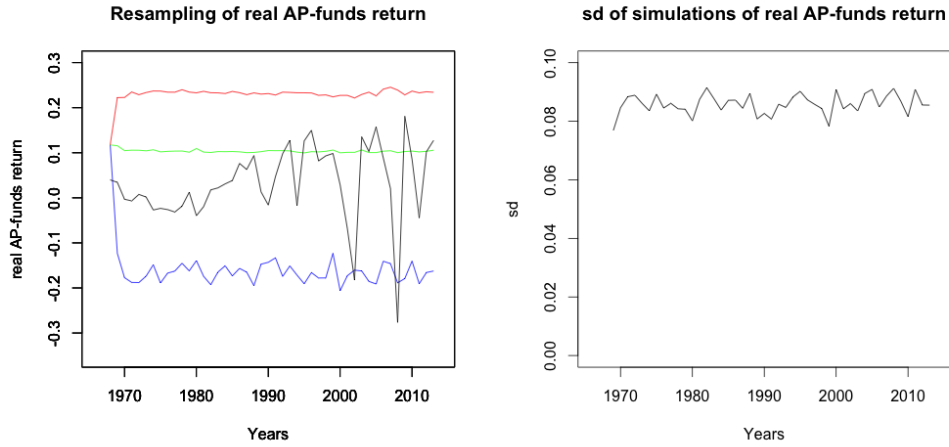


Figure 19: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line). *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average AP-funds return.

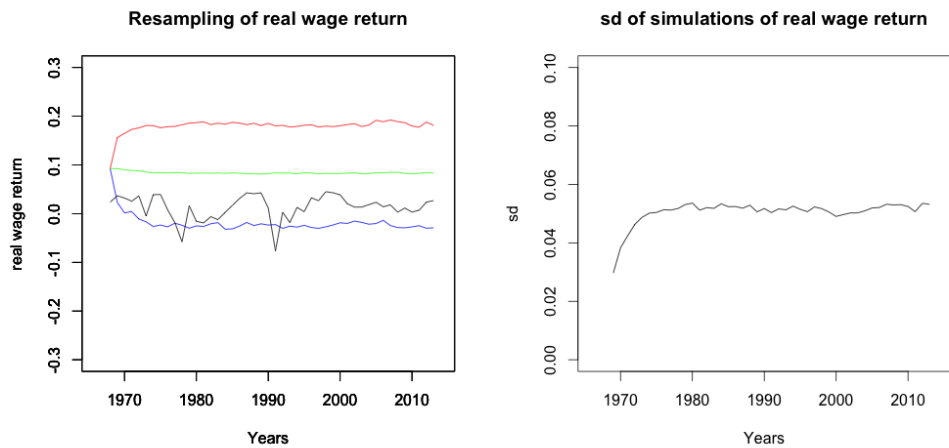


Figure 20: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line). *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average wage return.

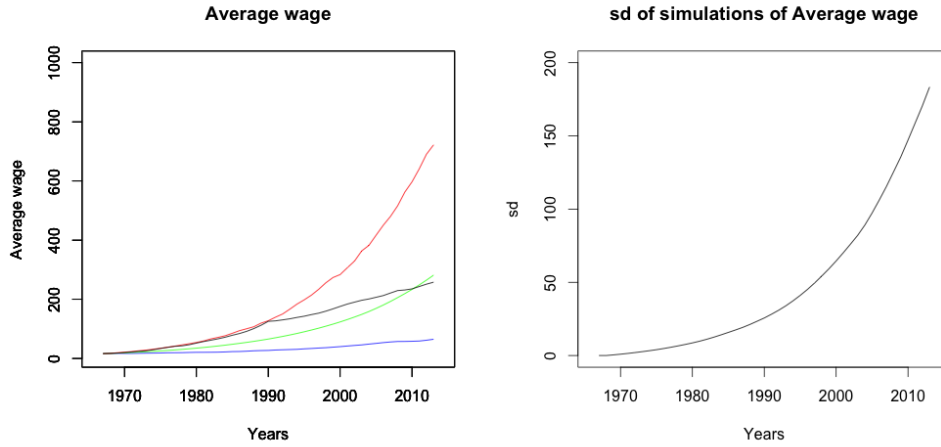


Figure 21: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line), the values are in 1000 SEK. *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of average wage.

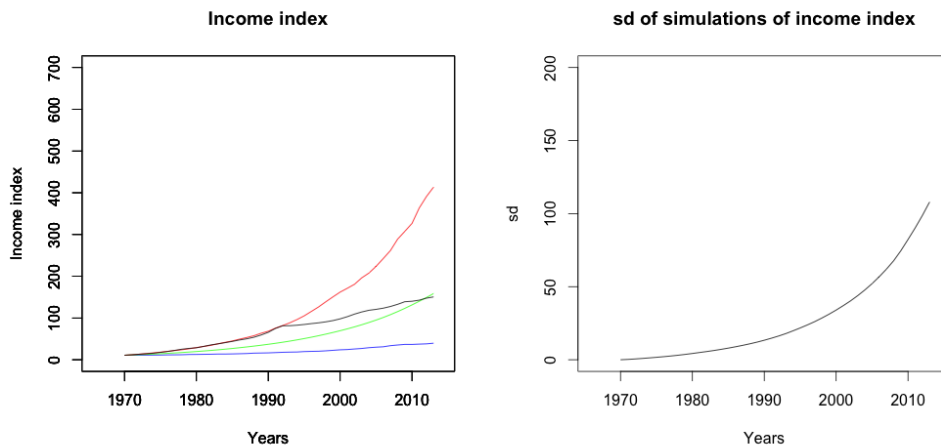


Figure 22: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line). *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of income index.

4.4.4 High inflation

In the scenario high inflation a drift $\mu = 0.07$ is added to log returns of inflation.

$$\mathbf{X}_t = \begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix} = \begin{bmatrix} 0.01 \\ 0.04 \\ 0.03 \end{bmatrix} + \begin{bmatrix} 0.74 & 0.05 & 0.00 \\ -0.25 & 0.02 & 0.42 \\ 0.50 & 0.03 & 0.50 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t^1 \\ \epsilon_t^2 \\ \epsilon_t^3 \end{bmatrix} \quad (28)$$

Coefficients	I_l	I_u
a_1	-0.02	0.03
a_2	-0.14	0.09
a_3	0.00	0.04
c_{11}	0.45	1.08
c_{21}	-1.16	1.66
c_{31}	0.18	0.70
c_{12}	-0.01	0.15
c_{22}	0.15	0.68
c_{32}	-0.02	0.10
c_{13}	-0.31	0.32
c_{23}	-0.78	1.65
c_{33}	0.33	0.79

Table 6: Confidence interval for the scenario low wage. A drift of $\mu = -0.07$ is added to the log returns of inflation from the years 1967-2013.

Figures 23-26 are obtained from simulations of equation (28).

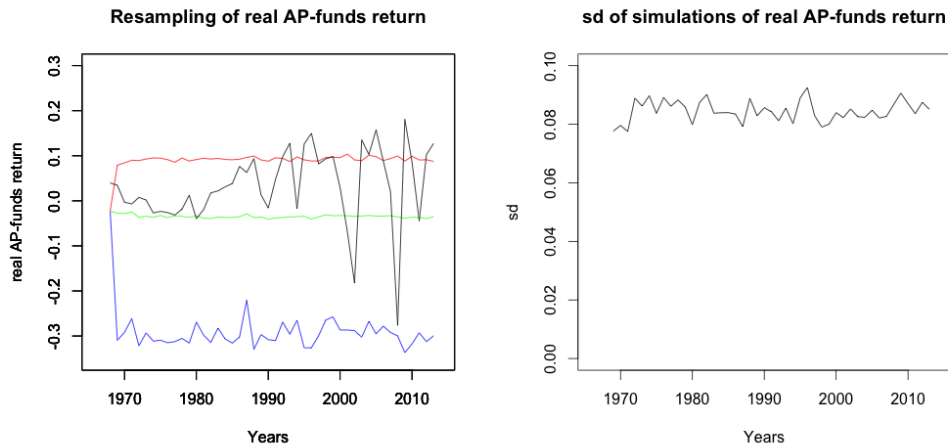


Figure 23: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line). *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average AP-funds return.

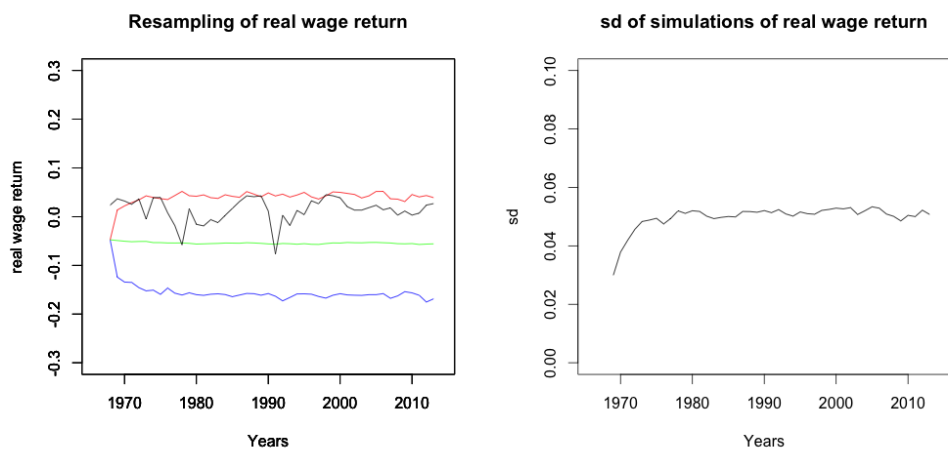


Figure 24: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line). *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average wage returns.

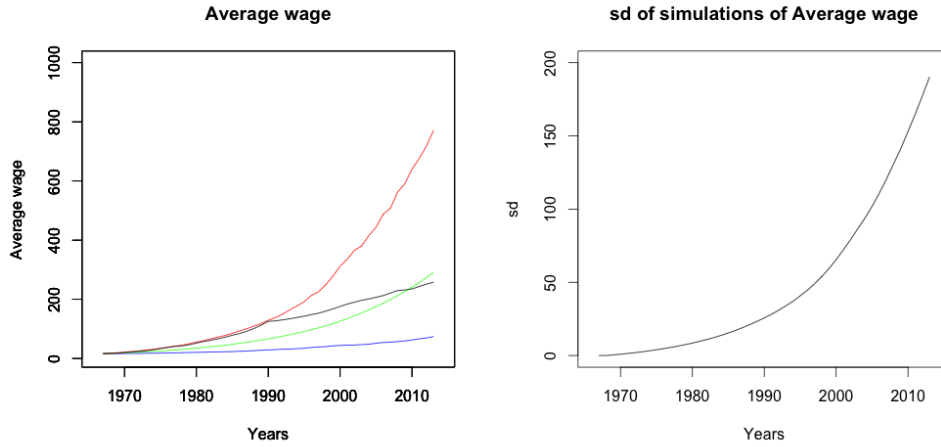


Figure 25: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line), the values are in 1000 SEK. *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of average wages.

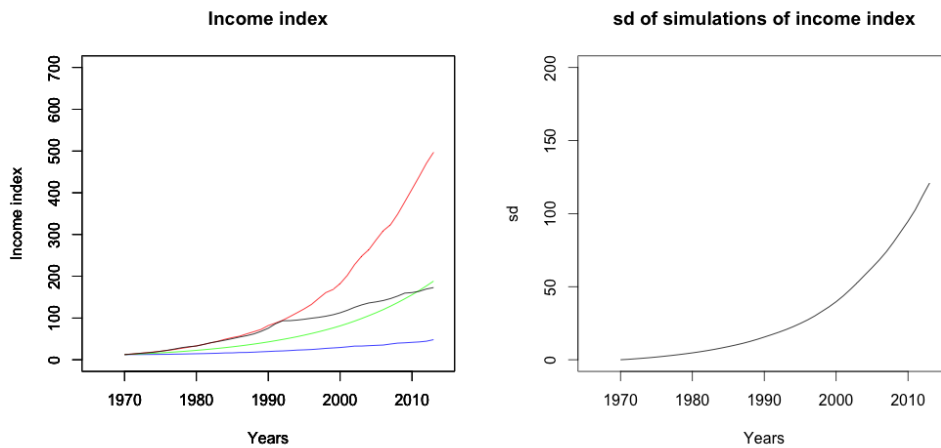


Figure 26: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line). *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average income index.

4.5 Scenarios with probability

In this section the results of the second scenario method is presented, where the residuals have been split up into a negative and positive part.

$$\mathbf{X}_t = \begin{cases} A_0 + A\mathbf{X}_{t-1} + \epsilon^+, & p \leq 0.3 \\ A_0 + A\mathbf{X}_{t-1} + \epsilon^-, & p > 0.3 \end{cases} \quad (29)$$

In (29) one can see that it is more likely to choose a negative value than a positive value because $p \in U(0, 1)$ where U denotes the uniform distribution. Hence this gives the low scenarios and if the inequalities are changed, i.e.,

$$\mathbf{X}_t = \begin{cases} A_0 + A\mathbf{X}_{t-1} + \epsilon^+, & p > 0.3 \\ A_0 + A\mathbf{X}_{t-1} + \epsilon^-, & p \leq 0.3 \end{cases} \quad (30)$$

the high scenarios are received.

4.5.1 Low average wage returns

In this scenario one has sorted the residuals from the log returns of average wage into a positive and a negative part, and it is more likely to draw a negative residual than a positive residual as described above.

Figures 27-30 are obtained from simulations of equation (29) when the residuals from average wage are sorted.

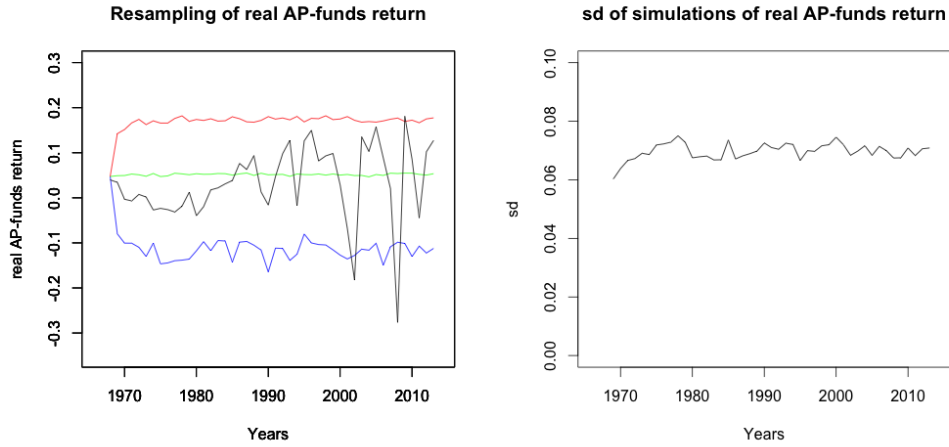


Figure 27: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line). *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average AP-funds return.

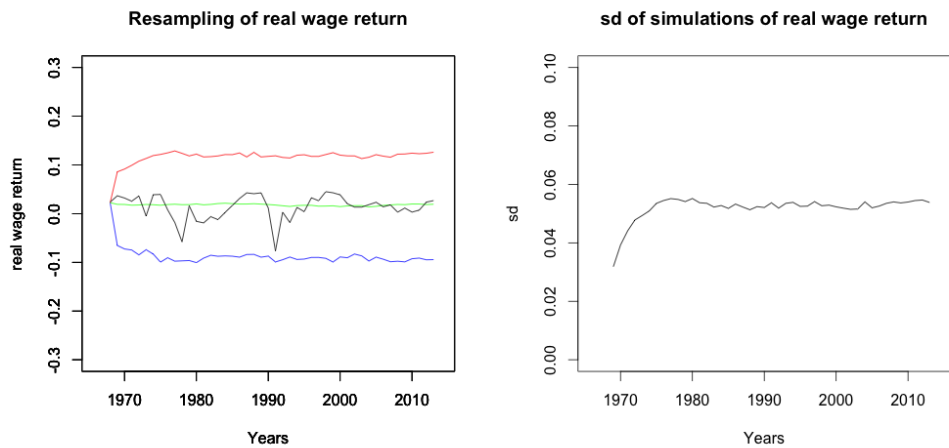


Figure 28: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line). *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average wage return.

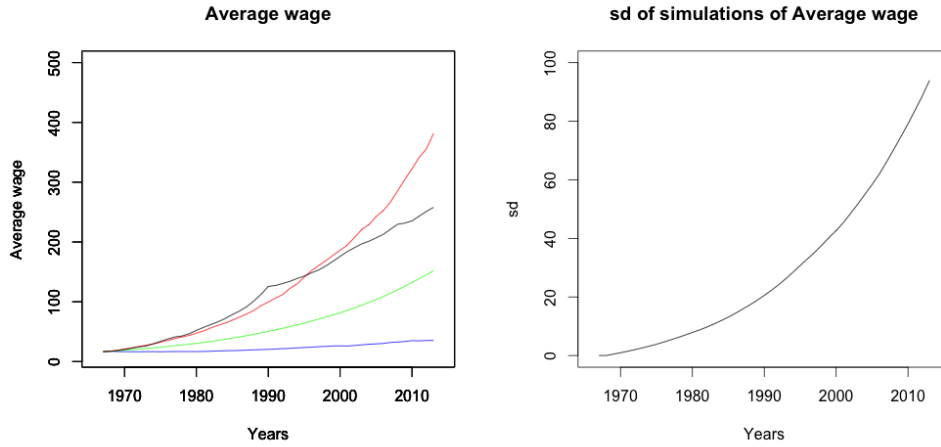


Figure 29: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line), the values are in 1000 SEK. *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average wage return.

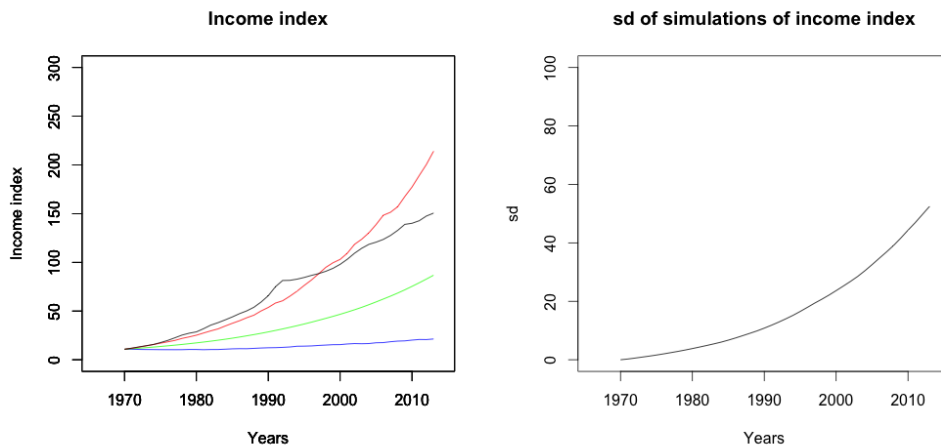


Figure 30: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line). *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average wage return.

4.5.2 High average wage returns

In the scenario one has sorted the residuals from the log returns of average wage into a positive and a negative part, and it is more likely to draw a positive residual than a negative residual the opposite of the low scenario case.

Figures 31-34 are obtained from simulations of equation (30) when the residuals from average wage are sorted.

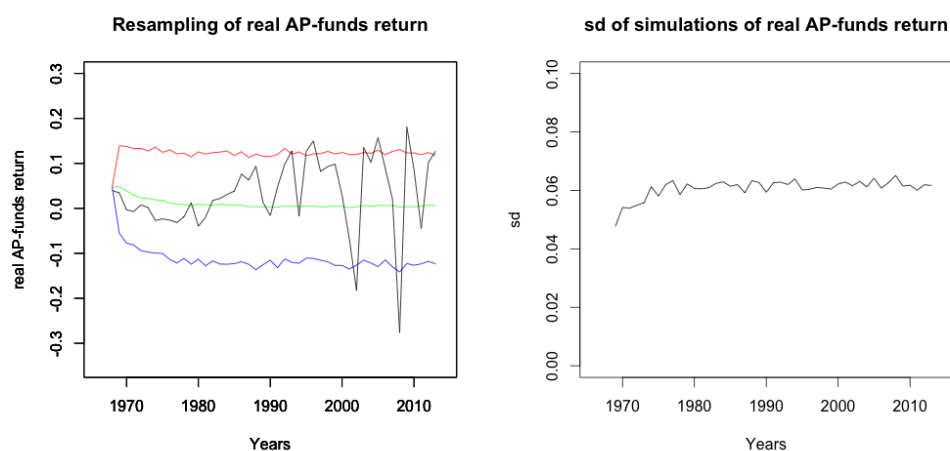


Figure 31: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line). *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average AP-funds return.

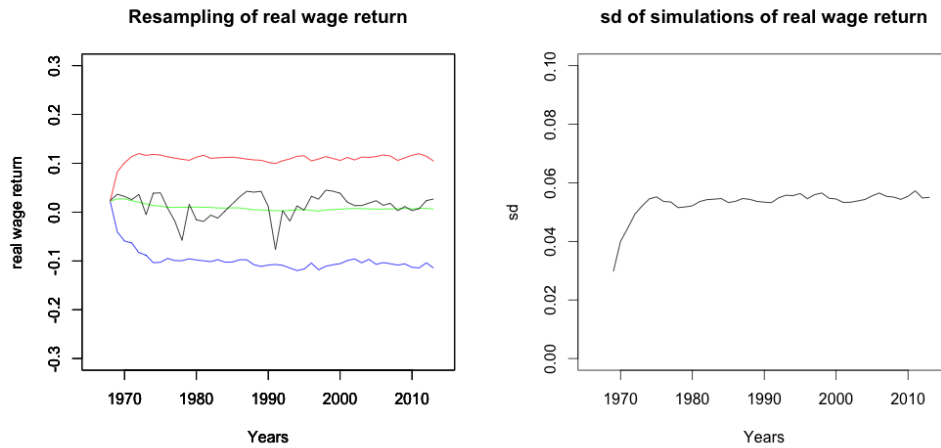


Figure 32: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line). *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average wage return.

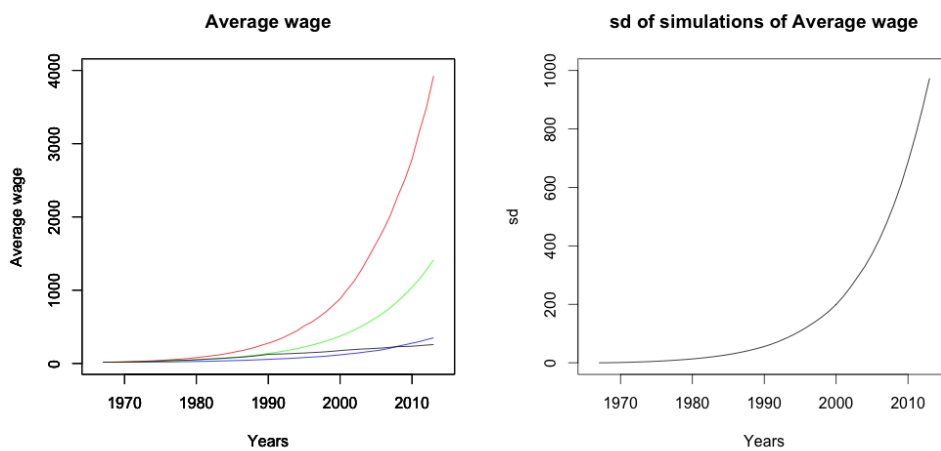


Figure 33: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line), the values are in 1000 SEK. *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average wage return.

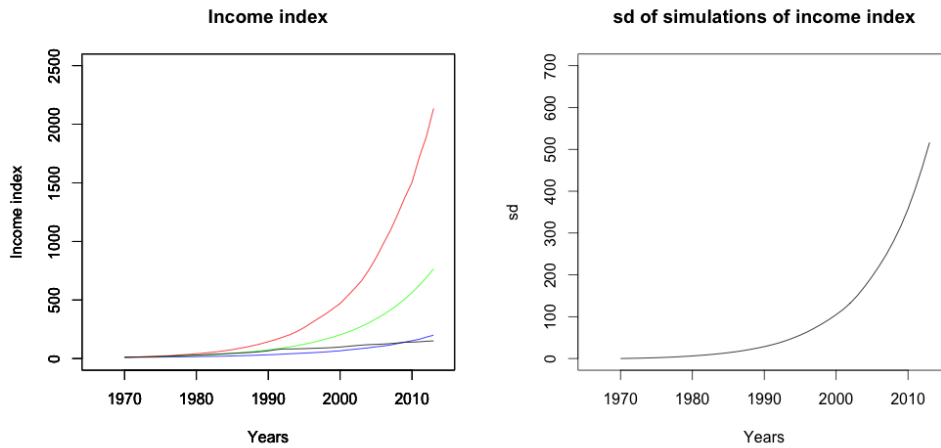


Figure 34: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line). *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average wage return.

4.5.3 Low inflation

In the scenario one has sorted the residuals from the log returns of inflation into a positive and a negative part, and it is more likely to draw a negative residual than a positive residual.

Figures 35-38 are obtained from simulations of equation (29) when the residuals from inflation are sorted.

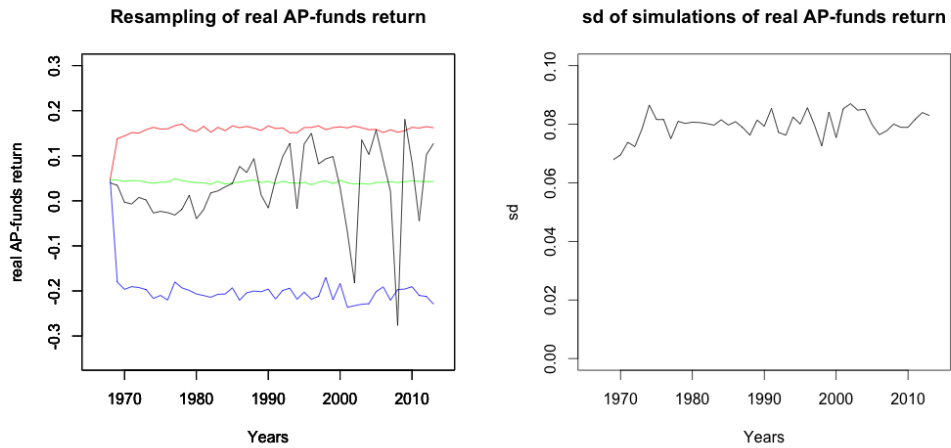


Figure 35: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line). *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average AP-funds return.

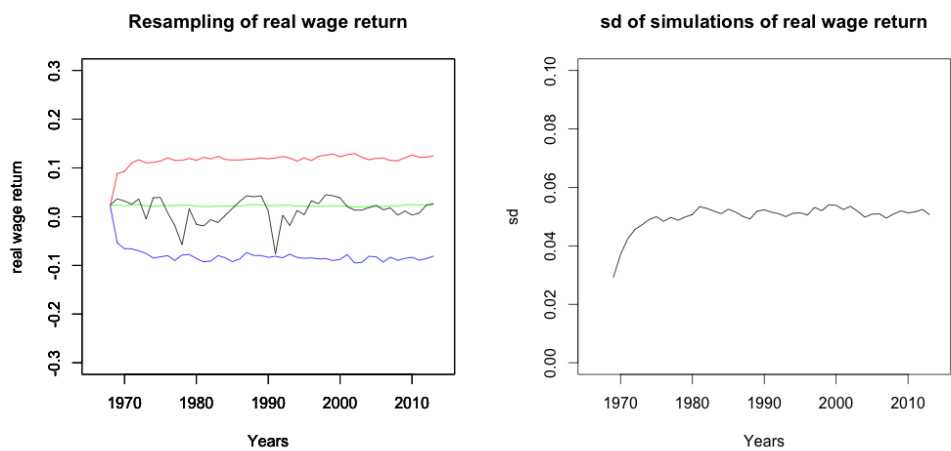


Figure 36: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line). *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average wage return.

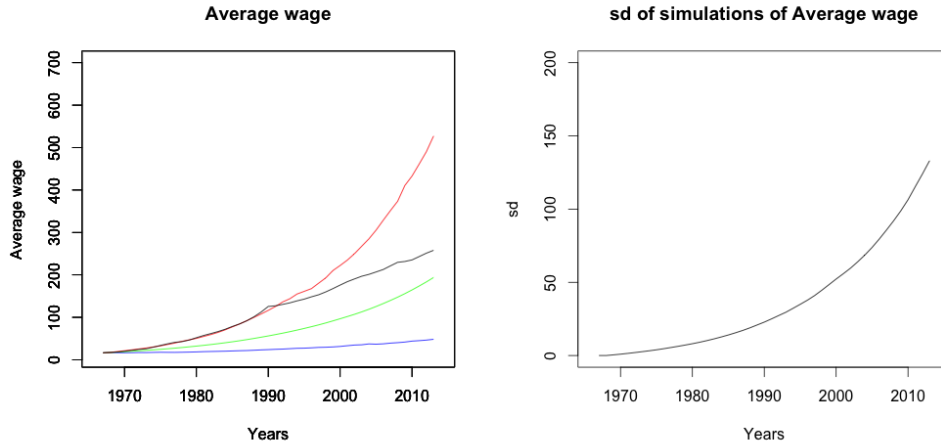


Figure 37: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line), the values are in 1000 SEK. *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average wage return.

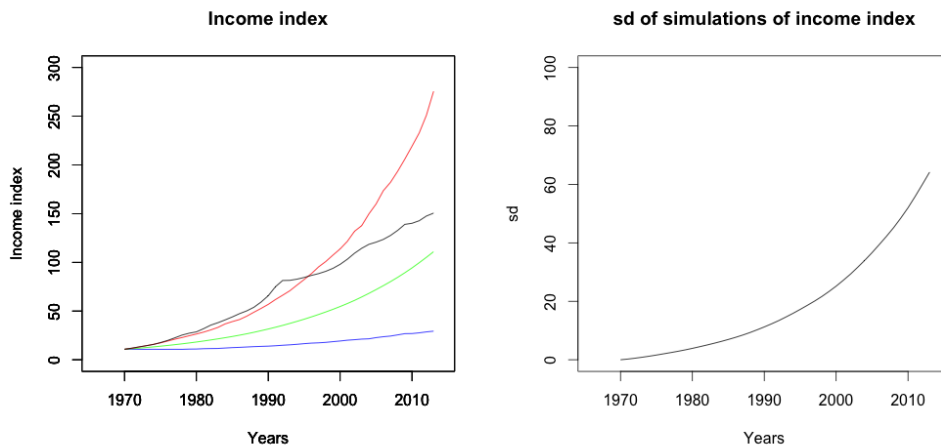


Figure 38: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line), the values are in 1000 SEK. *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average wage return.

4.5.4 High inflation

In the scenario one has sorted the residuals from the log returns of inflation into a positive and a negative part, and it is more likely to draw a positive residual than a negative residual as described above.

Figures 39-42 are obtained from simulations of equation (30) when the residuals from inflation are sorted.

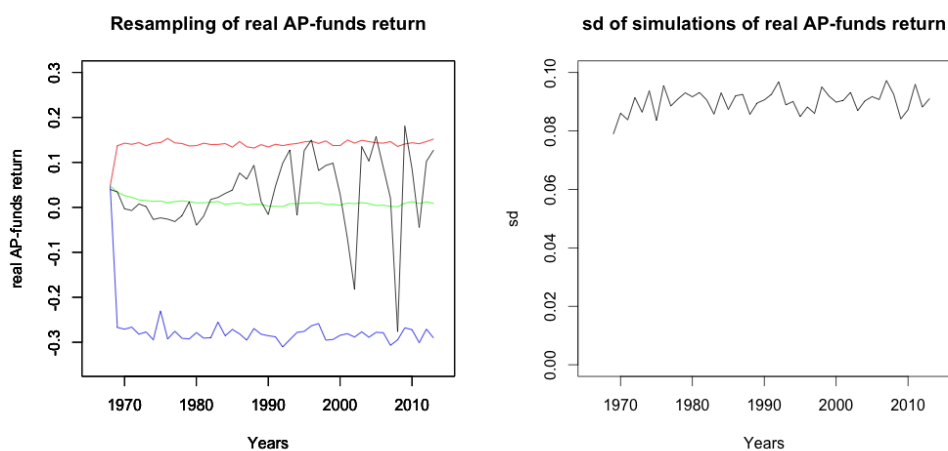


Figure 39: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line). *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average AP-funds return.

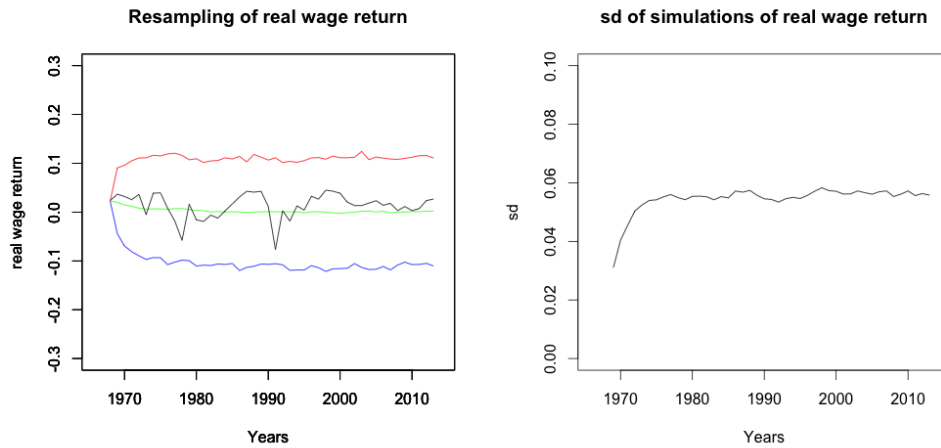


Figure 40: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line). *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average wage return.



Figure 41: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line), the values are in 1000 SEK. *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average wage return.

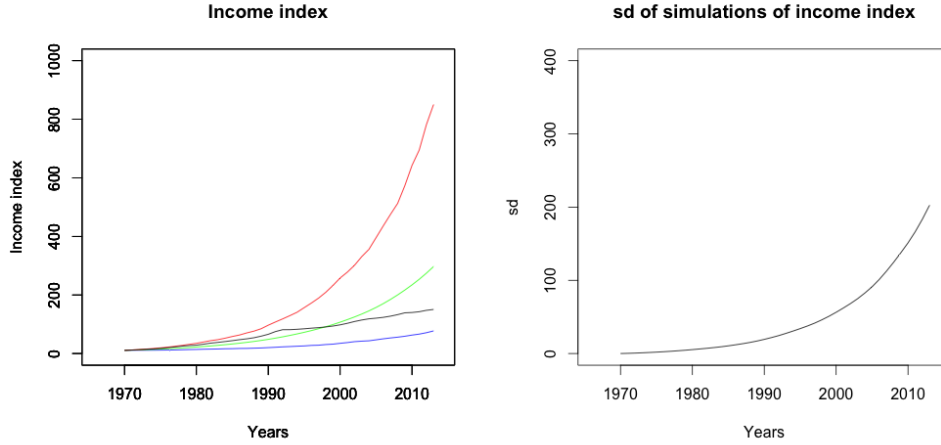


Figure 42: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values (blue line) and the data (black line). *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average wage return.

5 Results pension balance

To get a better understanding of what the risks are in the pension system the pension balance, see equation (1) is calculated for the years 1970-2013 (the years that it is possible calculate income index). The pension balance is approximated using equation (31) from [13]. The balance ratio is not considered (it is considered in [13]) see equation (31). It is assumed that 16% of the average wages are the premiums. The maximum premium is 7.5 income bases (inkomstbasbelopp) for respective year and the product $ACF_t \times IGF_t \approx 1$. Since the second scenario model (scenario with probability) gave less extreme scenarios it is chosen to be presented here.

$$PB_t^I = (PB_{t-1}^I + P_t) \frac{I_t}{I_{t-1}} \times ACF_t \times IGF_t \quad (31)$$

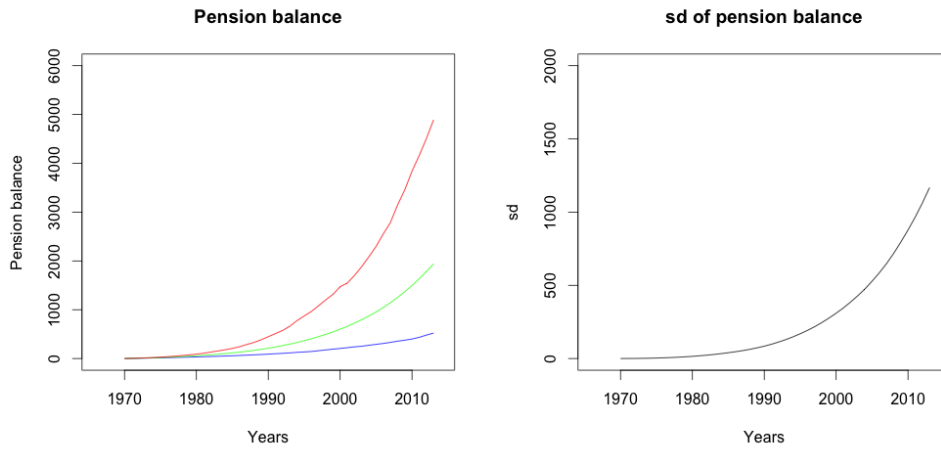


Figure 43: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values and (blue line), the values are in 1000 SEK. *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average wage return.

Figure 43 shows the pension balance from simulated data. One can see that there is a big standard deviation which seems to increase with respect to time.

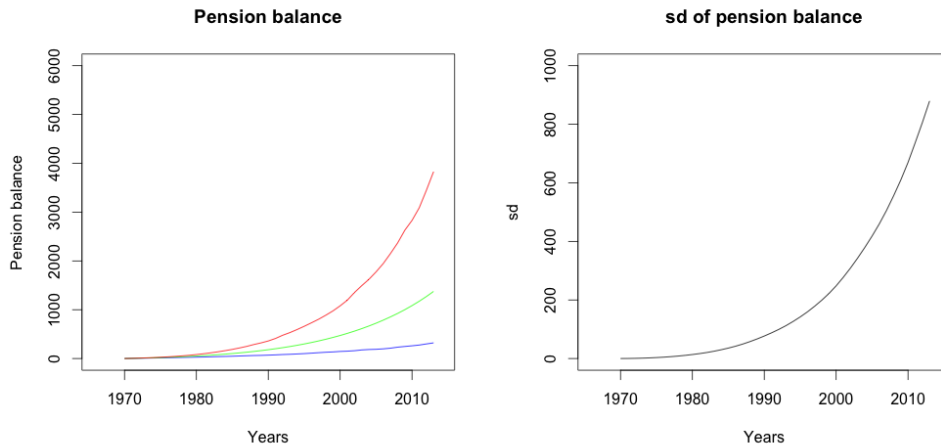


Figure 44: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values and (blue line), the values are in 1000 SEK. *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average wage return.

Figure 44 shows the pension balance from simulated data when the scenario low inflation with probability is used. One can see that the standard deviation is smaller than than from simulations with the original data.

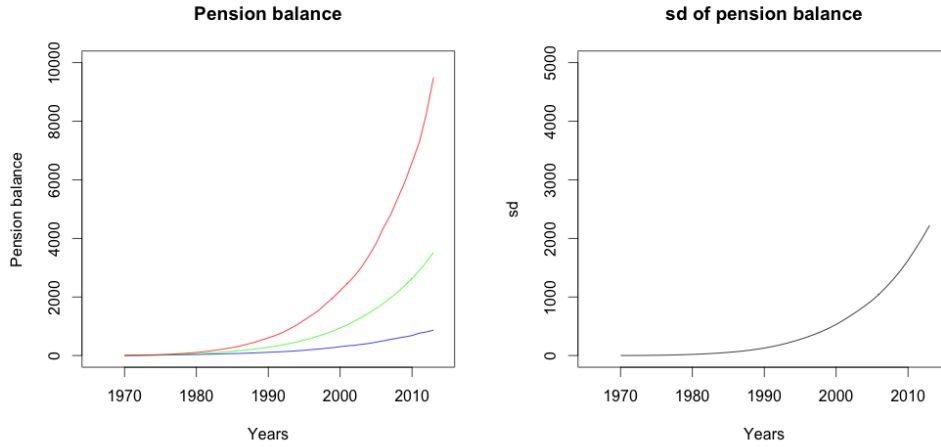


Figure 45: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values and (blue line), the values are in 1000 SEK. *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average wage return.

Figure 45 shows the pension balance from simulated data when the scenario high inflation with probability is used. One can see that the standard deviation is much larger than from simulations with the original data.

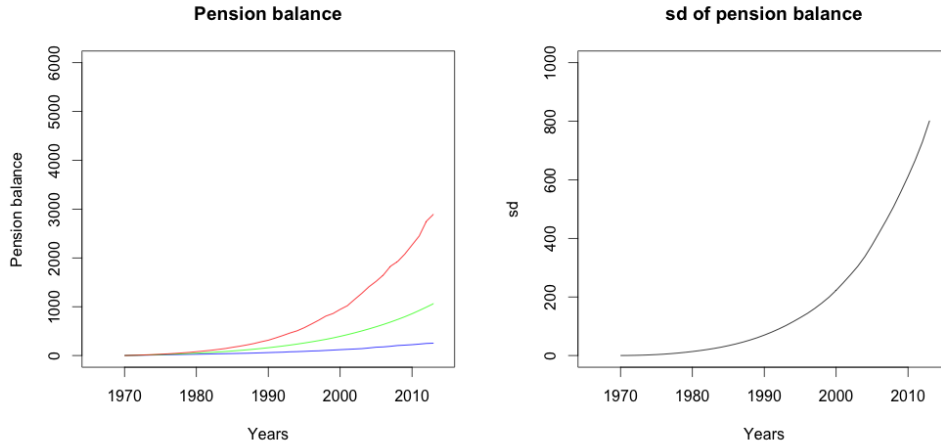


Figure 46: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values and (blue line), the values are in 1000 SEK. *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average wage return.

Figure 46 illustrates the pension balance from simulated data when the scenario low average wage with probability is used. One can see that the standard deviation is much smaller than from simulations with the original data.

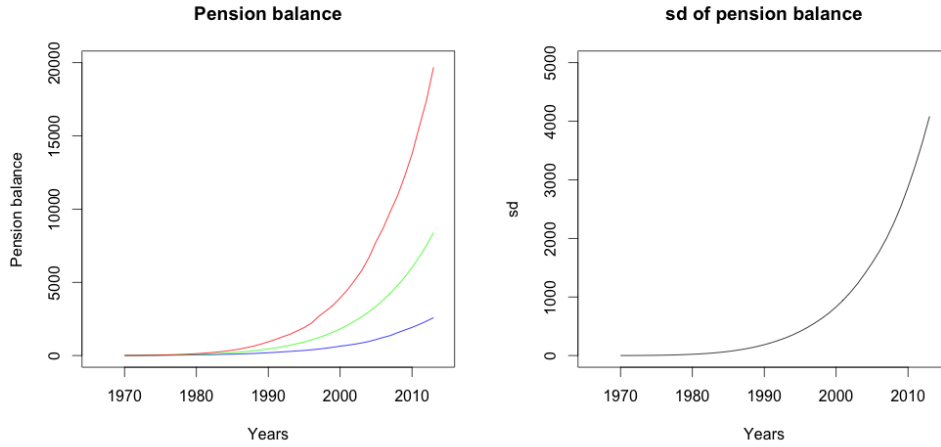


Figure 47: *Left:* Plot shows the 5% largest values (red line), the mean values (green line), the 5% smallest values and (blue line), the values are in 1000 SEK. *Right:* Plot shows standard deviations from each time step in the bootstrapped resampling procedure of real average wage return.

Figure 47 shows the pension balance from simulated data when the scenario high wage with probability is used. One can see that the standard deviation is much larger than from simulations with the original data.

6 Conclusion

The investigation of lag, variance and correlation structures between the input variables are presented in the results. It can be seen in equation (22) that there are correlations between average wage return and inflation, but small correlation between average wage return and the AP-funds return. Equation (22) also suggest small correlation between AP-funds return and inflation.

From the VAR(1) model one can conclude that inflation depends on average wage and itself at one time lag. The model also show that average wage depends on itself at one time lag. It could also be seen that the returns from the AP-funds does not depend on either inflation, average wage return or itself.

When the simulated parameters are put back in the equations (16), (12), (13) and (14), the return variables seem to have a stable constant standard deviation, see figures 7 and 8. However, the results suggest that average wage and income index have an increasing standard deviation, see figures 9 and 10.

When the model is tested in different scenarios for the parameter average wage return tab 3 and 4 show that a_3 (intercept for inflation) becomes significant from zero. For the case low average wage return a_3 becomes positive and for the other case high average wage return a_3 becomes negative. The other coefficients are unchanged.

When the model is tested in different scenarios for the variable inflation a_3 becomes negative for the scenario low inflation and zero for high inflation, see Table 6 and 5. The largest variation in income index seems to depend mostly on changes in average wage returns, see Figure 14 and Figure 18 for high and low scenario on average wage return. The scenarios for inflation do not seem to affect the income index as much as the average wage return scenarios, see figures 22 and 26.

In the variables real average wage return and real return from the AP-funds one can see changes in figures 11, 15, 19, 23, 13, 33, 21 and 24. Low inflation generates a higher real average wage return and high inflation generate a low real average wage return and similar for the real AP-funds return.

The second scenario method (scenario with probability) reveal similar results as the first scenario method (adding a drift), but not as extreme as when adding a drift compare figures 33 and 17.

One can see that the pension balance is affected in a similar way as income index, pension balance is mostly affected by the change in average wage see figure 47 and 46

7 Discussion

The results of the estimated VAR(1) model seems satisfying at first, but when looking at the confidence intervals for the estimated coefficients it can be seen that 8 of the 12 coefficients are not significantly separated from zero. If this is the case then the model is not very useful. From the scenario analysis one can see how real average wage returns and inflation depend on each other. The analysis shows drastic changes in income index and average income in scenarios high and low average wage return.

A reason why a change in the average wages affect the income index more than a change in inflation is that the VAR model suggests that inflation deepens on average wage returns but that average wage returns does not depend on inflation.

It is a challenge to model data over a 46 year time period because so much has happened under this time e.g. financial crises, negative repo rate. The large variations in the model is probably caused of large variations in the data

To make the model more realistic one could consider a VAR model that describes average wages, inflation and AP funds return better than the VAR model used in this master thesis. A suggested further research is to include demographic factors in the model, i.e. age structures.

References

- [1] Pensionsmyndigheten, “Temaartikel: Balanstalet inkomstpensionens stabila styråra?,” *Pension agency annual report*, 2002.
- [2] G. Englund, “Datorintensiva metoder i matematisk statistik, kth.” Unpublished lecture notes, 1995.
- [3] S. P. system, “Orange report,” 2013.
- [4] R. S. Tsay, *Multivariate Time Series Analysis With R and Financial Applications*. Wiley, 2014.
- [5] H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*. Springer, 2005.
- [6] H. Hult, F. Lindskog, O. Hammalid, and C. J. Rehn, *Risk and Portfolio Analysis Principles and methods*. Springer, 2012.
- [7] SCB, “<http://www.scb.se/pr0101/>,” 2015.
- [8] R. Cooper and A. A. John, *Theory and Applications of macroeconomics*. Unpublished, 2012.
- [9] D. Susanto, H. O. Zapata, and G. L. Cramer, “Bootstrapping in vector autoregressions: An application to the pork sector,” *American Agricultural Economics Association*, pp. 8,9, 2004.
- [10] SCB, “<http://www.statistikdatabasen.scb.se/pxweb/sv/ssd/>,” 2015.
- [11] A. funds, “<http://www.pensionsmyndigheten.se/4453.html>,” 2015.
- [12] P. Agency, “<https://secure.pensionsmyndigheten.se/basbeloppochvarderegler.html>,” 2015.
- [13] K. Birkholz, “Annuity divisor - comparison between different computational methods.” MSc Thesis KTH TRITA - MAT - E, 06 2013.
- [14] H. Lang, “Elements of regression analysis.” Unpublished lecture notes, 2014.

Appendices

A Algebra

In this section some Algebra concepts are presented. These concepts are the eigenvalues of a matrix, the vectorization of a matrix and the Kronecker product of two matrices.

Let λ be a scalar number real or complex and \mathbf{b} be a $m \times 1$ vector. If equation (32) holds, λ is called the eigenvalue of \mathbf{A} and \mathbf{b} is the corresponding eigenvector.

$$\mathbf{A}\mathbf{b} = \lambda\mathbf{b} \tag{32}$$

There are m eigenvalues and m eigenvectors for a $m \times m$ matrix \mathbf{A} [4]. Now some useful vector and matrix notations are defined. The vectorization of a matrix $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_n]$ where \mathbf{a}_i for $i \in \{1, \dots, n\}$ is $m \times 1$ is called $\text{vec}(A)$ and has the dimension $mn \times 1$

$$\text{vec}(A) = \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_n \end{bmatrix}.$$

The Kronecker product for two matrices $\mathbf{A}_{m \times n}$ and $\mathbf{C}_{p \times q}$ is defined as

$$\mathbf{A} \otimes \mathbf{C} = \begin{bmatrix} a_{1,1}\mathbf{C} & a_{1,2}\mathbf{C} & \dots & a_{1,n}\mathbf{C} \\ a_{2,1}\mathbf{C} & a_{2,2}\mathbf{C} & \dots & a_{2,n}\mathbf{C} \\ \vdots & \vdots & \dots & \vdots \\ a_{m,1}\mathbf{C} & a_{m,2}\mathbf{C} & \dots & a_{m,n}\mathbf{C} \end{bmatrix}.$$

The resulting matrix has dimension $mp \times nq$.

B Ordinary least squares

Let us consider the model

$$Y = X\beta + e$$

Here Y is an $n \times 1$ matrix and X is $n \times (k + 1)$ matrix and e is a $n \times 1$ matrix. From [14] the estimate $\hat{\beta}$ of β that minimises the sum of squares $\hat{e}^T \hat{e} = |\hat{e}|$ is received from the normal equations.

$$X^t \hat{e} = 0.$$

Hence

$$\hat{\beta} = (X^T X)^{-1} X^T Y.$$

$\hat{\beta}$ is the OLS estimate of β .