



Stockholms
universitet

Vad hade vi tjänat utan covid-19?

Tidsserieanalys av löneutveckling per utbildningsnivå i Sverige

Turkinaz Örkash

Kandidatuppsats 2026:9
Matematisk statistik
Maj 2026

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Vad hade vi tjänat utan covid-19?

Tidsserieanalys av löneutveckling per utbildningsnivå i Sverige

Turkinaz Örkash*

Maj 2026

Sammanfattning

I detta arbete utförs en tidsserieanalys för att studera löneutvecklingen för olika utbildningsnivåer i Sverige under perioden 2003 till 2023. Reallönerna visar en generellt uppåtgående trend fram till 2020, därefter avtar lönerna tydligt, vilket sammanfaller med covid-19-pandemin. Detta väcker frågan hur löneutvecklingen hade sett ut utan pandemin.

Data transformeras genom differensiering, log-transformering och log-differensiering för att genomföra enhetsrotstester, i form av Augmented Dickey–Fuller(ADF)-test och Kwiatkowski–Phillips–Schmidt–Shin(KPSS)-test. Modelleringen baseras därefter på den stationära tidsserien, vilket blir den differensierade reallönen. Autokorrelationsfunktion (ACF) och partiell autokorrelationsfunktion(PACF) används för att identifiera modellens ordning, där en Autoregressive Integrated Moving Average ARIMA(0,1,0) visar sig vara en lämplig modell. Den utvärderas med hjälp av Ljung–Box test.

Prediktioner av reallöner för varje utbildningsnivå över en treårsperiod indikerar att utvecklingen, i frånvaro av covid-19, hade fortsatt i den uppåtgående trenden. Därmed dras slutsatsen att pandemin har haft en negativ påverkan på löneutvecklingen. Denna effekt verkar vara genomgående över utbildningsnivåerna och blir statistiskt tydlig från och med 2022 samt förstärks ytterligare under 2023.

*Postadress: Matematisk statistik, Stockholms universitet, 106 91, Sverige.
E-post: turkinaz.orkash@gmail.com. Handledare: Ola Hössjer, Leo Levenius.

Abstract

In this thesis, a time series analysis is conducted to examine wage development across different educational levels in Sweden between 2003 and 2023. The real wage data shows a generally increasing trend up until 2020, followed by a sharp decline, coinciding with the COVID-19 pandemic. This raises the question of how real wages would have evolved in the absence of the pandemic.

The data is transformed using differencing, log-transformation, and log-differencing to perform unit root tests, specifically the Augmented Dickey–Fuller (ADF) test and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test. The modeling is then based on a stationary dataset, which is identified as the differenced real wage series. The autocorrelation function (ACF) and partial autocorrelation function (PACF) are used to determine the model order, where an ARIMA(0,1,0) model is found to be the most appropriate. The model is evaluated through residual analysis, using the Ljung–Box test.

Forecasts of real wages for each educational level over a three-year horizon suggest that, in the absence of COVID-19, the upward trend would have continued. It is therefore concluded that the pandemic has a negative impact on wage development. This effect appears to be consistent across educational levels, becoming statistically evident from 2022 onwards and intensifying in 2023.

Förord

Stort tack till min handledare Ola Hössjer, som har visat stort intresse och uppmuntrat mig sedan dag ett. Tillsammans med handledaren Leo Levenius har jag fått möjlighet att diskutera idéer och vidareutveckla uppsatsen.

Jag vill också tacka min familj, mina vänner och kollegor som har motiverat mig och trott på mig genom hela arbetet.

Även ChatGPT har varit som ett stöd, särskilt vid programmering och språkgranskning.

Innehåll

1	Inledning	6
2	Teori	7
2.1	Tidsserie	7
2.1.1	Tidsserieanalys	7
2.2	Stationäritet	8
2.2.1	Differensiering och logtransformering	9
2.3	Enhetsrotttest	9
2.3.1	AR(1)-process	10
2.3.2	ADF-test	10
2.3.3	KPSS-test	11
2.4	Tidsseriemodellering	11
2.4.1	MA(1)- och ARMA(1,1)-process	12
2.4.2	ARIMA(1,1,1)-process	12
2.4.3	Identifiering av modellordning	13
2.4.4	Residualanalys av ARIMA-modell	13
2.5	Prediktion	14
3	Data	15
3.1	Initialdata	15
3.1.1	Grundlön	15
3.1.2	Reallön	16
3.2	Datatransformationer och enhetsrotttester	17
4	Resultat	19
4.1	Stationäritet av tidsserierna	19
4.2	ARIMA-modellering	20
4.2.1	Val av modellordning	20
4.2.2	Modellutvärdering	22
4.3	Prediktion av reallön	23
4.3.1	Prognosavvikelser	24
4.3.2	Prediktiva p -värden	24
4.3.3	Prediktionsintervall	25
5	Diskussion	27
5.1	Begränsningar av modellen	27
5.2	Vidare studier	27
6	Referenser	28
A	Appendix: ACF- och PACF-diagram	29

1 Inledning

Både lön och utbildning har länge varit centrala samhällsfrågor, och sambandet mellan dessa utgör en viktig del inom ekonomi och arbetsmarknadsanalys. Särskilt relevant är hur löner utvecklas över tid för olika utbildningsnivåer, eftersom utbildning ofta betraktas som en avgörande faktor bakom inkomstskillnader.

I detta arbete används lönedata från Statistiska centralbyrån(SCB) för åtta olika utbildningsnivåer i Sverige. Löneutvecklingen visar en långsiktig uppåtgående trend. För reallöner bryts dock denna utveckling omkring år 2020, där en tydlig nedgång istället observeras. Detta sammanfaller med perioden då covid-19-pandemin pågick mellan 2020 och 2023. Det väcker frågan hur löneutvecklingen för olika utbildningsnivåer hade sett ut utan pandemin.

För att kunna prediktera reallöner i frånvaro av pandemin används tidsserieanalys. Utöver prediktionen är syftet med arbetet även att modellera och förklara den historiska utvecklingen. Eftersom de flesta tidsseriemodeller kräver stationaritet, det vill säga att en tidsserie har konstanta statistiska egenskaper över tid, tillämpas transformationer på datan. Dessa utgörs av differensiering, log-transformationer och log-differensiering. Med hjälp av enhetsrotstester, såsom ADF- och KPSS-test, identifieras att den differensierade reallönen är stationär för samtliga utbildningsnivåer och därmed lämplig för vidare modellering.

Då tidsserierna uppnår stationaritet efter differensiering är en $ARIMA(p, d, q)$ -modell relevant. Eftersom en differensierad tidsserie har integrationsordningen $d = 1$. Parametrarna p och q anger antalet autoregressiva respektive glidande medelvärdeskomponenter. Dessa identifieras med hjälp av ACF- och PACF-diagram, vilka beskriver autokorrelation och partiell autokorrelation. Efter residualanalys med hjälp av Ljung–Box-testet fastställs att en $ARIMA(0, 1, 0)$ med drift är en lämplig modell.

Prediktioner av reallöner för perioden 2021–2023 indikerar att löneutvecklingen, i frånvaro av pandemin, hade fortsatt den uppåtgående trenden. Pandemin tycks ha påverkat samtliga utbildningsnivåer på ett liknande sätt, då inga signifikanta skillnader kunde påvisas mellan dem. Med hjälp av p -värden och prediktionsintervall visas effekten av covid-19 inte är statistiskt synbar under det år då pandemin inträffade, utan blir tydlig först under 2022 och förstärks ytterligare under 2023.

2 Teori

I detta avsnitt introduceras de centrala begreppen inom tidsserieanalys. Metoden syftar inte enbart till att beskriva historiska mönster, utan även till att modellera beroendestrukturen i en tidsserie för att möjliggöra prediktion av framtida värden.

Vi fokuserar på modeller av ordning 1, vilket innebär att modellen endast tar hänsyn till den närmast föregående observationen och/eller feltermen. Detta eftersom det är mer pedagogiskt att introducera en modell genom dess enklaste fall, men också eftersom det i detta arbete troligtvis inte kommer att behövas modeller av högre ordning.

2.1 Tidsserie

En *tidsserie* kännetecknas av en sekvens av data som är registrerade över en period med jämna mellanrum, i jämförelse med andra typer av data som kan vara insamlade diskontinuerligt och slumpmässigt. Tidsseriedata finns överallt eftersom tid är en del av allt som är observerbart. I denna alltmer digitaliserade värld finns exempelvis sensorer och system som ständigt lagrar data över tid, vilket gör den lättillgänglig för olika typer av analyser, såsom väderprognoser.

Enligt Brockwell & Davis (1991) definieras en tidsserie som en följd av observationer x_t som mäts vid specifika tidpunkter t . Mätningar som observeras vid tidpunkter som tillhör en diskret mängd T_0 , exempelvis vid bestämda tidsintervaller, utgör en diskret tidsserie. I detta arbete behandlas endast diskreta tidsserier, då datamaterialet består av årliga observationer. En kontinuerlig tidsserie innebär att observationerna registreras kontinuerligt, till exempel när $T_0 = [0, 1]$.

Vidare beskrivs det i ovannämnda bok att varje observation x_t är ett realiserat värde av en stokastisk variabel X_t . En tidsserie $\{x_t, t \in T_0\}$ är därmed en realisation av $\{X_t, t \in T_0\}$. Termen tidsserie avser både den observerade datan och den process varav dessa utgör en realisation.

2.1.1 Tidsserieanalys

I en tidsserieanalys studeras sambanden mellan variabler som förändras över tid. Den är ett verktyg för att identifiera mönster, prediktera framtida utvecklingar och förstå den underliggande dynamiken i datan. Tidsserieanalys används inom olika områden, såsom ekonomi, teknik och sjukvård.

I Brockwell & Davis (2016) beskrivs en allmän metod för tidsserieanalys med syftet att hitta en lämplig modell, vilken är:

1. Visualisera tidsserien i en graf och undersök om den innehåller något av följande

- Trend,
- Säsongsvariation,
- Skarp förändring,
- Extremvärden.

Enligt Rocha, Oliveira & Viana (2024) kännetecknas säsongsvariationen av återkommande mönster över tidsintervall, trenden representeras av en systematisk linjär eller icke-linjär komponent och extremvärden är observationer som avviker från övriga datapunkter.

2. Ta bort treden och säsongsvariationen för att erhålla stationära residualer. För att uppnå detta mål är det ibland vara nödvändigt att tillämpa en preliminär transformation av data, exempelvis genom logtransformering eller differensiering
3. Välj en modell som anpassas till residualerna.
4. Prognostisera residualerna och invertera därefter de transformationer som eventuellt har tillämpats, för att få fram en prognos för den ursprungliga tidsserien X_t .

2.2 Stationäritet

Huvudsyftet med tidsserieanalys är att undersöka den okända strukturen hos den stokastiska processen $\{X_t\}$ med de tillgängliga observationerna $\{x_1, x_2, \dots, x_n\}$. Analysen blir dock mycket svår, och i vissa fall omöjlig, om processens statistiska egenskaper varierar över tiden eller är beroende av den. Därför krävs en viss form av stabilitet, vilket inom tidsserieanalys kallas *stationäritet*.

Det finns många definitioner av stationäritet. Vi följer Brockwell & Davis (2016) som beskriver att en tidsserie $\{X_t, t = 0, \pm 1, \dots\}$ är stationär om dess statistiska egenskaper gäller även för "den tidsförskjutna serien" $\{X_{t+h}, t = 0, \pm 1, \dots\}$, för varje heltal h . Observera att dessa egenskaper för svag stationäritet endast beror på första och andra ordnings moment av $\{X_t\}$, vilka definieras som nedan.

Definition 2.1. Låt $\{X_t\}$ vara en tidsserie sådan att $E(X_t^2) < \infty$. Medelvärdesfunktionen för $\{X_t\}$ är

$$\mu_X(t) = E(X_t).$$

Kovariansfunktionen för $\{X_t\}$ är

$$\gamma_X(r, s) = \text{Cov}(X_r, X_s) = E[(X_r - \mu_X(r))(X_s - \mu_X(s))],$$

för alla heltal r och s .

Definition 2.2. En tidsserie $\{X_t\}$ sägs vara *svagt stationär* om

1. medelvärdesfunktionen $\mu_X(t)$ är konstant, och

2. kovariansfunktionen $\gamma_X(t+h, t)$ endast beror på laggen h och inte på t .

En svagt stationär tidsserie har alltså ett konstant medelvärde och en konstant varians. En serie sägs vara strikt stationär om den, utöver dessa villkor, även har en konstant fördelning över tid. I detta arbete används dock termen stationär enligt Definition 2.2.

2.2.1 Differensiering och logtransformering

För att kunna utföra en analys är det ofta nödvändigt att konvertera en icke-stationär tidsserie till en stationär. Detta kan göras på flera olika sätt. I Hyndman & Athanasopoulos (2021) nämns bland annat differensiering, vilket innebär att beräkna differensen mellan två på varandra följande observationer, det vill säga $\nabla X_t = X_t - X_{t-1}$. Observera att den differensierade serien endast har $T-1$ observationer, eftersom det inte är möjligt att beräkna en differens för den första observationen. Differensiering bidrar till att stabilisera medelvärdet i en tidsserie genom att ta bort nivåförändringar i serien, vilket i sin tur eliminerar eller reducerar trend och säsongsvariation.

Differensiering kan även definieras med hjälp av *bakåtoperatoren* B , där

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t. \quad (1)$$

Detta innebär att $BX_t = X_{t-1}$, det vill säga att bakåtoperatoren förskjuter serien ett steg bakåt i tiden.

En annan vanlig metod för att uppnå stationäritet är logtransformering av tidsserien, $\log(X_t)$, vilket hjälper till att stabilisera variansen. Den är särskilt effektiv för tidsserier med exponentiell tillväxt.

Genom att kombinera differensiering och logtransformering, alltså $\log(X_t) - \log(X_{t-1})$, leder det ofta till att både medelvärdet och variansen i tidsserien stabiliseras.

2.3 Enhetsrotttest

Många tidsseriemodeller förutsätter stationäritet, och för att undersöka detta är ett naturligt första steg att visualisera serien. Om grafen exempelvis visar en tydlig trend eller upprepande mönster över regelbundna tidsintervall tyder detta på att tidsserien inte är stationär. Visuellt analys är dock inte alltid tillräcklig för att identifiera alla egenskaper hos serien. Ett mer systematiskt tillvägagångssätt är därför att utföra ett *enhetsrotttest*, vilket är ett statistiskt hypotestest som används för att undersöka om en tidsserie är stationär och därmed avgöra om någon form av transformation krävs.

2.3.1 AR(1)-process

För att kunna fördjupa oss i vad en enhetsrot innebär är det nödvändigt att introducera en *autoregressiv process*. Den beskriver en genereringsmekanism där värdet på en serie vid en given tidpunkt är en funktion av dess närmast föregående värde samt en felterm.

I Sundberg (2023) beskrivs en autoregressiv process av första ordningen, AR(1), som det enda enkla AR-fallet då processen är Markovsk. Det innebär att all information om framtiden finns i det senaste värdet och att tidigare observationer inte tillför något ytterligare. Antag att $X_1, X_2, \dots, X_t$ följer en AR(1)-process. Då ges modellen av

$$X_t = \phi X_{t-1} + \varepsilon_t, \quad (2)$$

där ε_t är oberoende likafördelade slumpfel med fördelning $N(0, \sigma^2)$.

När parametern ϕ är lika med 1, då har vi en enhetsrot, vilket indikerar icke stationaritet. En AR(1) tidsserie är stationär när ϕ ligger mellan -1 och 1.

Ett speciellt fall av en AR(1)-process är en *slumpvandring* (random walk), $X_t = X_{t-1} + \varepsilon_t$, där slumptermen är vitt brus med varians σ^2 . En slumpvandring är dock inte stationär utan kännetecknas av ett väntevärde lika med startvärdet X_0 , linjärt växande varians och en kovariansfunktion enligt Definition 2.1 som ges av $\gamma_X(r, s) = \sigma^2 \min(r, s)$.

2.3.2 ADF-test

I Enders (2014) beskrivs Dickey–Fuller-testet som ett enhetsrotest som bygger på en AR(1)-process, där X_{t-1} subtraheras från båda sidor av modellen. Genom att sätta $\gamma = \phi - 1$ kan modellen skrivas om till $\Delta X_t = \gamma X_{t-1} + \varepsilon_t$. Att undersöka om $\phi = 1$ i Dickey–Fuller-testet är ekvivalent med att testa nollhypotesen $\gamma = 0$, vilket innebär att tidsserien innehåller en enhetsrot.

Augmented Dickey–Fuller(ADF)-testet är en utvidgning av Dickey–Fuller-testet där laggade differenser av högre ordning inkluderas för att kontrollera för seriell korrelation. Detta leder till följande tre regressionslikvationer:

$$\Delta X_t = \gamma X_{t-1} + \sum_{i=1}^p \beta_i \Delta X_{t-i} + \varepsilon_t, \quad (3)$$

$$\Delta X_t = a_0 + \gamma X_{t-1} + \sum_{i=1}^p \beta_i \Delta X_{t-i} + \varepsilon_t, \quad (4)$$

$$\Delta X_t = a_0 + \gamma X_{t-1} + a_2 t + \sum_{i=1}^p \beta_i \Delta X_{t-i} + \varepsilon_t. \quad (5)$$

Skillnaden mellan ekvationerna ligger i vilka deterministiska termer som inkluderas. Den första motsvarar en slumpvandring och saknar sådana termer, den andra innehåller dessutom drifttermen a_0 , medan den tredje involverar både drift och trend a_2t .

ADF-testet utgår från att estimerar en (eller flera) av ovanstående ekvationer för att kunna skatta parametern γ och dess standardfel (standard error, SE). Därefter jämförs den erhållna t -statistikan

$$t = \frac{\hat{\gamma}}{\text{SE}(\hat{\gamma})} \quad (6)$$

med värden från Dickey–Fuller–tabellerna, vilket avgör om nollhypotesen ska förkastas eller ej.

2.3.3 KPSS-test

Kwiatkowski–Phillips–Schmidt–Shin (KPSS)-testet skiljer sig från de flesta enhetsrottester huvudsakligen på två sätt. För det första har KPSS-testet stationäritet som nollhypotes. För det andra innebär frånvaro av en enhetsrot inte nödvändigtvis stationäritet, utan *trendstationäritet*, vilket innebär att den underliggande trenden kan avlägsnas för att erhålla en stationär tidsserie.

I Kwiatkowski m. fl. (1992) formuleras en tidsserie som en summa av en deterministisk trend, en slumpvandring och en stationär felterm:

$$X_t = \xi t + r_t + \varepsilon_t. \quad (7)$$

Slumpvandringen r_t kan skrivas som

$$r_t = r_{t-1} + u_t, \quad (8)$$

där u_t är oberoende och likafördelade $N(0, \sigma_u^2)$. Under KPSS-testets nollhypotes gäller det att $\sigma_u^2 = 0$, vilket innebär trendstationäritet eftersom ε_t redan är stationär. Testet baseras på en Lagrangemultiplikator-statistika.

2.4 Tidsseriemodellering

En *tidsseriemodell* representerar den mekanism som genererar eller förklarar en observerad tidsserie, där varje observation uttrycks som en matematisk funktion av en kombination av tid, andra variabler, kända värden av intresse samt en eller flera slumpkomponenter. Efter att en lämplig modell har valts följer parameterskattning och utvärdering av modellens anpassning till data. Tidsseriemodellering möjliggör även prediktion av framtida värden.

2.4.1 MA(1)- och ARMA(1,1)-process

Vi har i avsnitt 2.3.1 introducerat AR(1)-process, vilken är en av de mest grundläggande modellerna inom tidsSerieanalys. En annan enkel och vanlig modell är *glidande-medelvärdes-processer*, förkortat MA från engelskans Moving Average. I en MA-process beror X_t inte på tidigare värden av processen, som i AR-fallet, utan på nuvarande och tidigare feltermen. Dessa feltermen antas vara sinsemellan oberoende, och processen är stationär.

En MA(1)-modell går ett tidssteg bakåt och definieras av

$$X_t = \varepsilon_t - \theta\varepsilon_{t-1}, \quad (9)$$

där $\{\varepsilon_t\}$ är vitt brus ($WN(0, \sigma^2)$), vilket innebär att $\varepsilon_t \sim N(0, \sigma^2)$ är oberoende. Parametern θ beskriver beroendet mellan slumpvarter

AR-modeller baseras på föregående värden i en tidsSerie, medan MA-modeller tar hänsyn till tidigare feltermen. En modell som kombinerar båda dessa komponenter är ARMA-modellen, det vill säga Autoregressive Moving Average. Vi begränsar oss återigen till ordning 1, och en ARMA(1,1)-modell kan då skrivas som

$$X_t = \phi X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}, \quad (10)$$

där $\varepsilon_t \sim WN(0, \sigma^2)$. ARMA-modeller förutsätter att tidsSerien är stationär, vilket kräver $|\phi| < 1$.

2.4.2 ARIMA(1,1,1)-process

I verkliga data är det sällan en tidsSerie är stationär från början, vilket gör det olämpligt att direkt tillämpa en ARMA-modell. En utvidgning av denna modell är ARIMA, vilket står för Autoregressive Integrated Moving Average. AR- och MA-komponenterna är vi nu bekanta med, medan I-komponenten representerar differensiering. I en ARIMA(p, d, q)-modell anger p och q ordningen, medan d betecknar antalet differensieringssteg som krävs för att göra tidsSerien stationär.

Brockwell & Davis (2016) definierar en tidsSerie som en ARIMA(p, d, q)-process om $(1 - B)^d X_t$ följer en ARMA(p, q)-process. Vi formulerar därför om AR(1), MA(1) och ARMA(1,1) med hjälp av bakåtoperatoren B , där $BX_t = X_{t-1}$. Detta ger:

$$X_t = \phi BX_t + \varepsilon_t, \quad (11)$$

$$X_t = (1 - \theta B)\varepsilon_t, \quad (12)$$

$$X_t = \phi BX_t + (1 - \theta B)\varepsilon_t. \quad (13)$$

ARIMA(1,1,1) härleds genom att applicera ARMA(1,1)-modellen på den differensierade serien. Med hjälp av den tidigare definierade $\nabla X_t = (1 - B)X_t$ erhålls

$$(1 - B)X_t = \phi B(1 - B)X_t + (1 - \theta B)\varepsilon_t. \quad (14)$$

2.4.3 Identifiering av modellordning

Valet av modellordning baseras på tidsseriens autokorrelationsstruktur. För att identifiera lämpliga värden på p och q analyseras därför autokorrelationsfunktionen (ACF) och den partiella autokorrelationsfunktionen (PACF).

ACF beskriver förhållandet mellan en tidsserie och en laggad version av sig själv över successiva tidsintervall. För en svagt stationär tidsserie $\{X_t\}$ definieras den som

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \frac{\text{Cov}(X_t, X_{t-k})}{\text{Var}(X_t)}, \quad k = 0, 1, 2, \dots \quad (15)$$

där $\gamma(k) = \text{Cov}(X_t, X_{t-k})$ är autokovariansfunktionen vid lag k .

PACF mäter det direkta sambandet mellan observationer vid ett visst lagg, efter att den linjära effekten av mellanliggande laggar har eliminerats. Detta kan beskrivas genom modellen

$$X_t = \phi_{k1}X_{t-1} + \phi_{k2}X_{t-2} + \dots + \phi_{kk}X_{t-k} + \varepsilon_t, \quad (16)$$

där ε_t är en vitbrusprocess och koefficienten ϕ_{kk} utgör den partiella autokorrelationen vid lagg k .

ACF används främst för att identifiera ordningen q i en MA-modell, medan PACF används för att bestämma ordningen p i en AR-modell. Båda funktionerna analyseras via diagram där de skattade korrelationerna plottas tillsammans med konfidensgränser. Staplar utanför dessa gränser betraktas som signifikanta och används som vägledning vid bestämning av modellens ordning.

2.4.4 Residualanalys av ARIMA-modell

Ett diagnostiskt anpassningstest (goodness-of-fit-test) för en ARIMA-modell utvärderar om modellen har fångat korrelationsstrukturen i datan genom analys av residualerna från den anpassade modellen. Dessa bör uppvisa egenskaper hos vitt brus, och som nämnts ovan innebär det att residualerna är okorrelerade, har konstant varians och väntevärde noll.

Ljung-Box-testet är en metod som används vid residualanalys av tidsserier. Enligt Sundberg (2023) definieras teststatistikan som

$$Q = T(T+2) \sum_{k=1}^K \frac{r_k^2}{T-k}, \quad (17)$$

vilken asymptotiskt är χ^2 -fördelad med $K - p - q$ frihetsgrader, där p och q är antalet autoregressiva respektive glidande medelvärdesparametrar och K är antalet laggar. Nollhypotesen är att residualerna r_k är okorrelerade.

ACF-diagram kan användas för en visuell analys av residualerna. Om inga systematiska avvikelser från konfidensbanden observeras tyder detta på frånvaro av autokorrelation hos residualerna.

2.5 Prediktion

En prediktion av ett framtida värde X_{t+h} baseras på information tillgänglig fram till tidpunkt t . Prediktionen kan tolkas som det betingade väntevärdet av X_{t+h} givet den historiska informationen, vilket motsvarar det värde som minimerar det förväntade kvadratiske prognosfelet, det vill säga

$$\hat{X}_{t+h} = E[X_{t+h} | X_1, \dots, X_t]. \quad (18)$$

Om $\{X_t\}$ är icke-stationär innebär detta att tidsserien saknar konstant väntevärde och varians, vilket försvårar prediktion. Därför baseras prediktionen av en ARI-MA(p, d, q)-process istället på den differensierade processen $(1 - B)^d X_t$, som senare återförs till nivåserien X_t .

Vid prediktion antas framtida feltermerna ha väntevärde noll eftersom de är okända. Detta innebär att prognoserna bestäms av modellens struktur och tidigare observationer.

Anta att vi dessutom vill testa om X_{t+h} är signifikant mindre än vad vi kan förvänta oss från observerade värden på tidsserien fram till tiden t . Om x_{t+h} är det observerade värdet av X_{t+h} kan vi då använda oss av det prediktiva p -värdet

$$p = P(X_{t+h} \leq x_{t+h} | X_1, \dots, X_t). \quad (19)$$

Detta p -värde är kvantilen för x_{t+h} i den prediktiva fördelningen för X_{t+h} givet observationer av $X_1, \dots, X_t$.

3 Data

I detta avsnitt presenteras två essentiella dataset för arbetet, grundlön och reallön. Här visas datans källa, filtrering och visualisering. Därefter tillämpas tre olika transformationer på tidsserierna för att senare kunna undersöka stationaritet.

3.1 Initialdata

3.1.1 Grundlön

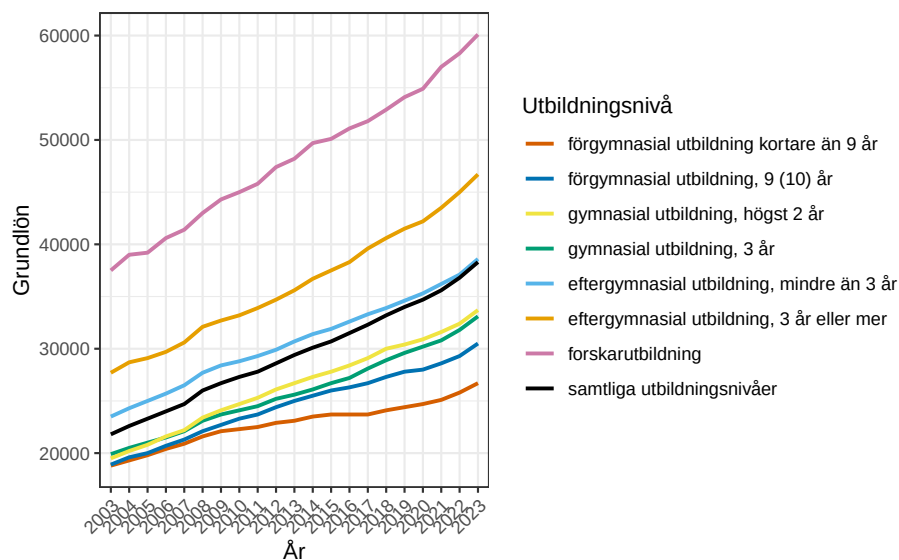
Detta arbete grundas på data om den årliga löneutvecklingen för olika utbildningsnivåer. I Statistiska centralbyråns(SCB), statistikdatabas finns kategorin "Lön efter yrkesgrupp (3-siffrig SSYK)". Förkortningen står för Standard för svensk yrkesklassificering och är ett system för att sortera individers yrken eller arbetsuppgifter. SCB tillhandahåller både 3- och 4-siffrig SSYK, där den senare innebär en mer detaljerad sortering. I detta arbete används den 3-siffriga klassificeringen, dels eftersom yrke inte analyseras i studien, men framförallt för att denna kategori innehåller data över fler år.

Under "Lön efter yrkesgrupp (3-siffrig SSYK)" finns bland annat dataset "Genomsnittlig grund- och månadslön samt kvinnors lön i procent av mäns lön efter sektor, yrkesgrupp (SSYK 2012), kön och utbildningsnivå (SUN)" för år 2003-2013, 2014-2022 och 2023. Vi sammanställer dessa och filtrerar på grundlön för alla utbildningsnivåer, vilka är:

- Förgymnasial utbildning kortare än 9 år
- Förgymnasial utbildning, 9 (10) år
- Gymnasial utbildning, högst 2 år
- Gymnasial utbildning, 3 år
- Eftergymnasial utbildning, mindre än 3 år
- Eftergymnasial utbildning, 3 år eller mer
- Forskarutbildning
- Samtliga utbildningsnivåer

Observera att "Samtliga utbildningsnivåer" ovan avser en specifik kategori i datamaterialet. I övriga delar av texten används uttrycket samtliga utbildningsnivåer i sin allmänna betydelse, det vill säga som samlingsbegrepp för alla åtta utbildningsnivåer.

Med grundlön för de åtta utbildningsnivåerna över 21 år får vi totalt 168 observationer. Löneutvecklingen visas i Figur 3.1.

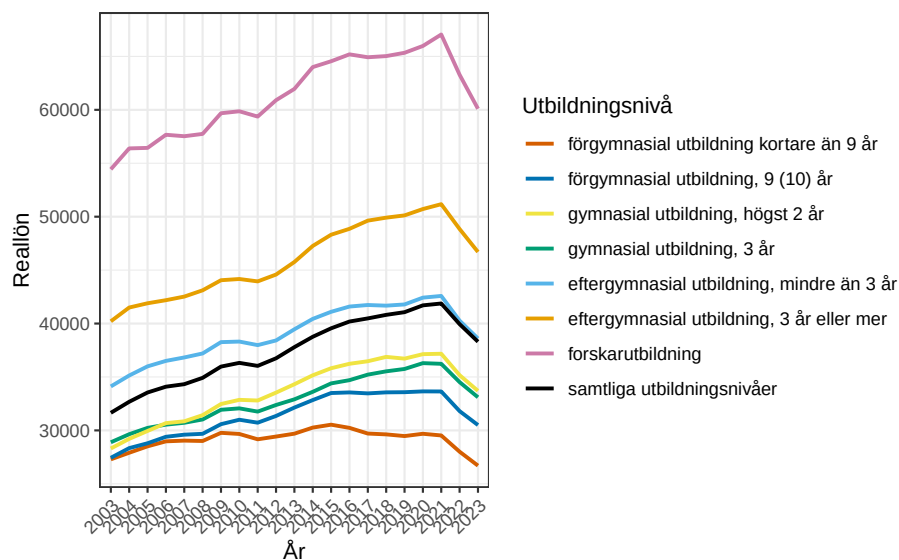


Figur 3.1: Utveckling av grundlön per utbildningsnivå under perioden 2003–2023

3.1.2 Reallön

Grundlön är ett fast belopp som betalas ut för de arbetade timmarna, utan något rörligt tillägg. Inflation innebär en ökning av den allmänna prisnivån i samhället och påverkar därmed köpkraften. En löneförhöjning tyder därför inte alltid på en real ökning, eftersom inflationen medför att pengarnas värde minskar. För att ta hänsyn till detta används även reallön i denna studie, vilken är inflationsjusterad. Detta görs med hjälp av konsumentprisindex, KPI, vilket är det vanligaste måttet för inflationsberäkning som mäter den genomsnittliga prisutvecklingen för hela den privata inhemska konsumtionen.

I denna studie används SCBs data ”KPI fastställda årsmedeltal, totalt, 1980=100” för att få inflationsvärden mellan år 2003 och 2023. Genom att sätta indexet från det senaste året i datasetet, 2023, som vår bas underlättar det för oss att förstå penningvärdet efter inflationsjusteringen. I praktiken innebär detta att indextalet för 2023 divideras med indextalet för respektive år och sedan multipliceras kvoten med grundlön från det aktuella året. Med reallön för samma åtta utbildningsnivåer mellan år 2003 och 2023 får vi ytterligare 168 observationer. Se Figur 3.2.



Figur 3.2: Utveckling av reallön per utbildningsnivå under perioden 2003–2023

3.2 Datatransformationer och enhetsrotttester

Figur 3.2 visar en tydlig och kraftfull nedgång i reallönen efter 2021. När data uppvisar en så pass stor förändring tyder det ofta på en större händelse. I detta fall är det mest sannolikt covid-19, som var en pandemi som i Sverige pågick från 2020 till 2023. Detta ger upphov till en naturlig frågeställning: hur hade lönen utvecklats om pandemin inte inträffat?

För att besvara frågeställningen inleds analysen med att representera grundlön respektive reallön, för perioden 2003–2020, som tidsserier. Varje utbildningsnivå utgör i sig en egen tidsserie, men då samma ARIMA-modell tillämpas på samtliga betraktas de som en sammanhängande mängd av tidsserier.

Stationaritet undersöks genom att tillämpa de tre transformationer som beskrivs i avsnitt 2.2.1, nämligen differensiering, logaritmering samt en kombination av dessa. Detta resulterar i åtta klasser av tidsserier (fyra baserade på grundlön och dess transformationer och fyra baserade på reallön och dess transformationer) av tidsserier som analyseras med enhetsrotttester. Både ADF- och KPSS-test används, vilket totalt ger 16 tester för respektive utbildningsnivå.

Enligt definitionen i avsnitt 2.3.2 finns tre olika specifikationer av ADF-testet, beroende på vilken deterministisk komponent tidsserien innehåller. I vårt fall, för både grundlön och reallön i originalform, tillämpas ekvation (5), då serierna kan innehålla både trend och drift. Efter differensiering, samt log-differensiering, eliminerades trenden i teorin, vilket innebär att (4), med enbart drift, används.

Logaritmering stabiliserar variansen, men eliminerar varken trend eller drift, varför (5) återigen tillämpas.

För KPSS-testet, avsnitt 2.3.3, finns det två varianter, där ekvation (7) definierar test med deterministiskt trendterm och (8) utan. Med samma resonemang som för ADF-testet, för de tidsserier med differensiering tillämpas (8) och de utan (7).

4 Resultat

4.1 Stationäritet av tidsserierna

Målet med enhetsrottesterna är att identifiera en mängd av stationära tidsserier, så att samma ARIMA-modell kunde tillämpas på alla åtta utbildningsnivåerna. Tabell 4.1 visar antalet utbildningsnivåer som klassificeras som stationära för respektive transformation.

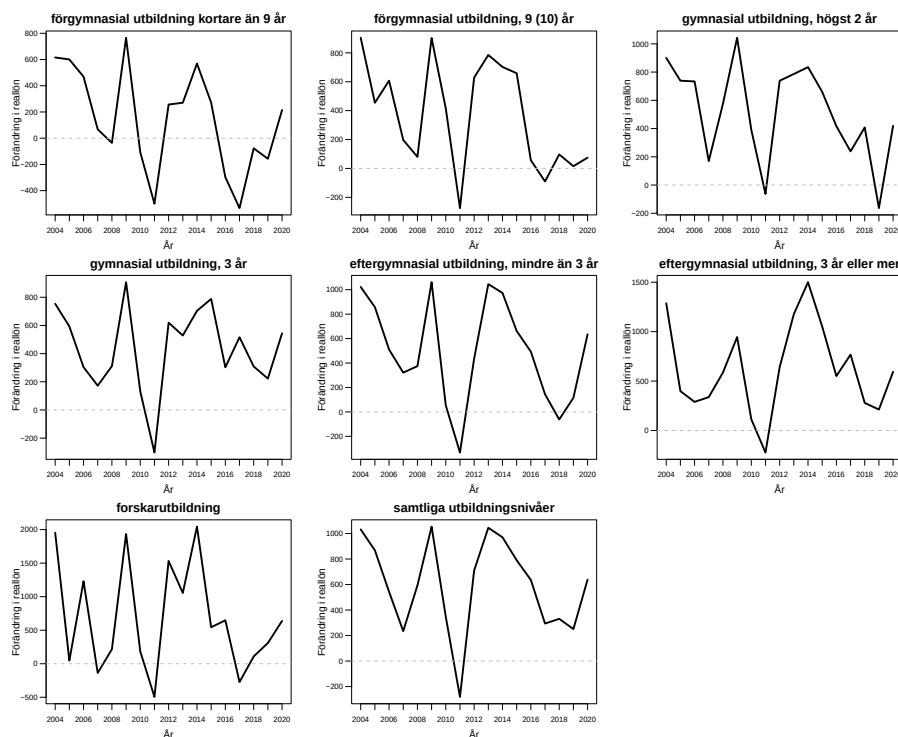
Tabell 4.1: Antalet utbildningsnivåer som ger stationära tidsserier enligt ADF-respektive KPSS-testet för olika transformationer av grundlön och reallön.

	ADF	KPSS
Grundlön	0	5
Differensierad grundlön	4	7
Logaritmerad grundlön	0	2
Log-differensierad grundlön	1	5
Reallön	2	7
Differensierad reallön	6	8
Logaritmerad reallön	1	6
Log-differensierad reallön	5	8

För ADF-testerna identifieras inget scenario där samtliga utbildningsnivåer är stationära. Däremot indikerar KPSS-testerna att både den differensierade och den log-differensierade reallönen är stationära.

I tabellen kan även effektiviteten hos de olika metoderna analyseras utifrån antalet tidsserier som har blivit stationära efter respektive transformation. Log-transformeringen ger ingen tydlig förbättring, medan både grundlön och reallön efter differensiering uppvisar flest antal stationära tidsserier.

Utav de två dataseten som uppfyller stationaritetsskravet väljs den differensierade reallönen, se Figur 4.1, för vidare modellering. Detta val motiveras av att differensiering innebär en mindre omfattande transformation än log-differensiering, vilket gör att tidsserien i högre grad bevarar sin ursprungliga skala. Dessutom uppvisar den differensierade reallönen flest antal stationära tidsserier även enligt ADF-testet.



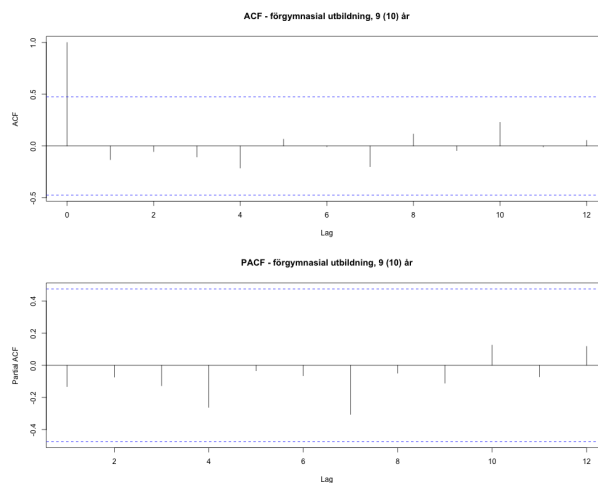
Figur 4.1: Differensierad reallön för respektive utbildningsnivå, under perioden 2003–2020.

4.2 ARIMA-modellering

4.2.1 Val av modellordning

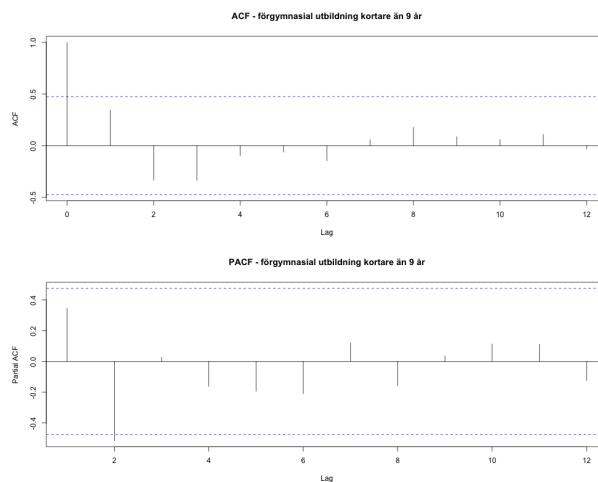
Varje tidsserie i den differensierade reallönen plottas i ett ACF- respektive PACF-diagram. Genom att observera vilka laggar som uppvisar signifikanta autokorrelationer respektive partiella autokorrelationer kan lämpliga värden på p respektive q identifieras. Signifikans bedömdes utifrån om stolparna ligger utanför de approximativa 95%-konfidsbands. Tolkningen baseras inte enbart på enskilda stolpar, utan även på det övergripande mönstret i diagrammen, såsom förekomsten av ett tydligt cutoff, det vill säga att korrelationen avtar till noll efter en viss lagg.

Inga av ACF-graferna uppvisar några signifikanta autokorrelationer. Sex av åtta utbildningsnivåer visar inte heller några tydliga tecken på en autoregressiv struktur i PACF. Figur 4.2 illustrerar detta med tidsserien förgymnasial utbildning, 9 (10) år som ett representativt exempel.



Figur 4.2: Exempel på ACF- och PACF-diagram utan signifikant autokorrelation.

Avvikande från övriga PACF-grafer är tidsserierna förgymnasial utbildning kortare än 9 år och forskarutbildning. För dessa observeras en signifikant partiell autokorrelation vid lag 2, där stolpen ligger utanför konfidensbandet. Figur 4.3 visar ett av dessa två fall.



Figur 4.3: Exempel på ACF- och PACF-diagram med möjlig autoregressiv struktur vid lag 2.

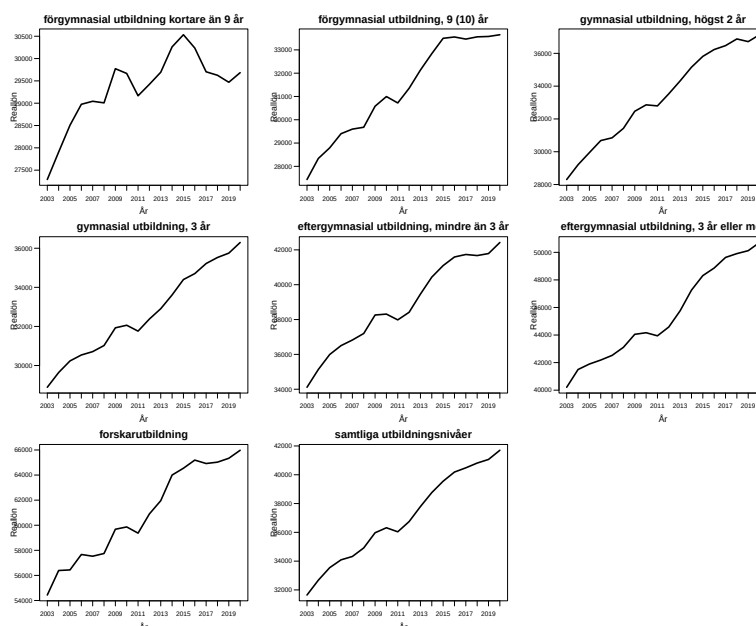
För ACF- och PACF-diagram för alla åtta tidsserier, se Appendix A.

Baserat på resultaten från ACF- och PACF-graferna bedöms ARIMA(0,1,0) som den mest lämpliga modellen. Även om en signifikant partiell autokorrelation observeras vid lagg 2 för två av tidsserierna (se en av dem i Figur 4.3), saknas motsvarande struktur vid lagg 1. Eftersom signifikansen för lagg 2 är så pass liten är det inte motiverat att anpassa en ARIMA(2,1,0)-modell.

4.2.2 Modellutvärdering

En ARIMA(0,1,0)-modell är ekvivalent med en slumpvandring. En prediktion baserad på modellen är dock inte särskilt informativ och realistisk, eftersom frånvaron av autoregressiva och glidande medelvärdeskomponenter innebär att prognoserna blir konstanta och lika med det senast observerade värdet. Vid prediktion av reallönen för år 2021 med hjälp av ARIMA(0,1,0) erhålls därför samma värde som observeras år 2020.

En driftkomponent bör inkluderas när den underliggande tidsserien uppvisar en trend. Trots att de differensierade reallönerna identifieras som stationära innebär detta inte att det är dessa som modelleras direkt. ARIMA(0,1,0) appliceras istället på de ursprungliga tidsserierna, eftersom modellen redan tar hänsyn till differensiering. Graferna i Figur 4.4 visar att reallönerna uppvisar en uppåtgående trend.



Figur 4.4: Differensierad reallön för respektive utbildningsnivå under perioden 2003-2020.

Vår slutgiltiga modell är således ARIMA(0,1,0) med drift. Modellen utvärderas med hjälp av residualanalys. Ljung–Box-testet ger höga p -värden för samtliga tidsserier, vilket innebär att vi inte kan förkasta nollhypotesen om att residualerna är okorrelerade. Som komplement används även ACF-diagram för varje tidsserie, vilka inte visar några systematiska autokorrelationer i residualerna. Modellen bedöms därmed ha fångat de beroenden som finns mellan observationer över tid.

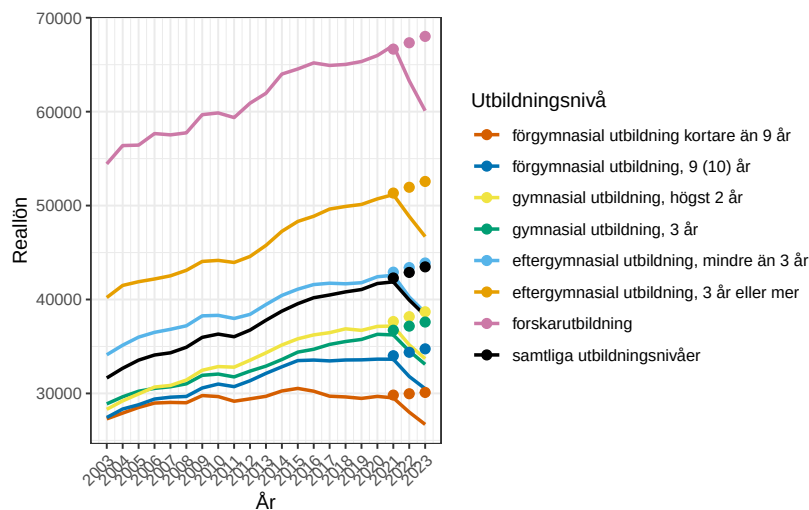
4.3 Prediktion av reallön

För att besvara den centrala frågan om hur lönerna hade sett ut utan covid-19 predikteras reallönen för åren 2021, 2022 och 2023, det vill säga de år då pandemin skedde. Prediktionerna baseras på skattade parametrar för ARIMA-(0,1,0) med drift, där framtida värden beräknas rekursivt utifrån modellens struktur och tidigare observationer. De prickade linjerna representerar den predikterade utvecklingen av reallönen för varje utbildningsnivå, som visas i Figur 4.5. Prediktionerna kan skrivas som

$$\hat{X}_{t+h} = X_t + h\hat{\mu}, \quad (20)$$

där $\hat{\mu}$ är en skattning av tidsseriens driftparameter.

Prognoserna visar en uppåtgående trend i reallönerna för samtliga utbildningsnivåer. Denna utveckling följer den historiska trend som observerats i datan fram till 2020.



Figur 4.5: Observerade reallöner per utbildningsnivå under perioden 2003–2020 tillsammans med predikterade värden för 2021–2023.

4.3.1 Prognosavvikelser

Figur 4.5 illustrerar att skillnaden mellan de predikterade och observerade reallönerna ökar för varje år mellan 2021 och 2023. Detta framgår även i Tabell 4.2, där både observerade och predikterade reallöner presenteras, och skillnaden (predikterad subtraheras med observerad) markeras i fet stil.

Tabell 4.2: Observerade och predikterade reallöner samt absoluta och procentuella skillnader per utbildningsnivå för 2021–2023.

Utbildningsnivå	2021	2022	2023
Förgymnasial utbildning kortare än 9 år	29526 / 29825 (-299, -0.98%)	28005 / 29965 (-1960, -6.85%)	26700 / 30106 (-3406, -11.91%)
Förgymnasial utbildning, 9 (10) år	33643 / 34015 (-372, -1.00%)	31804 / 34381 (-2577, -6.54%)	30500 / 34746 (-4246, -11.31%)
Gymnasial utbildning, högst 2 år	37172 / 37654 (-482, -1.09%)	35169 / 38173 (-3004, -7.50%)	33700 / 38693 (-4993, -12.22%)
Gymnasial utbildning, 3 år	36231 / 36729 (-498, -1.28%)	34518 / 37165 (-2647, -7.87%)	33100 / 37601 (-4501, -12.90%)
Eftergymnasial utbildning, mindre än 3 år	42583 / 42911 (-328, -1.36%)	40271 / 43400 (-3129, -7.12%)	38600 / 43889 (-5289, -11.97%)
Eftergymnasial utbildning, 3 år eller mer	51170 / 51333 (-163, -0.76%)	48846 / 51951 (-3105, -7.21%)	46700 / 52569 (-5869, -12.05%)
Forskarutbildning	67050 / 66656 (394, 0.59%)	63283 / 67335 (-4052, -5.98%)	60100 / 68014 (-7914, -11.64%)
Samtliga utbildningsnivåer	41877 / 42293 (-416, -0.98%)	39946 / 42885 (-2939, -6.85%)	38300 / 43476 (-5176, -11.91%)

Samtliga utbildningsnivåerna visar en negativ skillnad, med ett undantag för forskarutbildning år 2021. Vidare ökar skillnaderna tydligt över tid för alla utbildningsnivåer. År 2021 är avvikelserna relativt små och ligger inom några hundratal kronor, medan de blir avsevärt större under 2022 och framför allt 2023. De procentuella löneskillnaderna bekräftar den ökande trenden.

Det finns ingen signifikant skillnad i löneökningarna mellan de olika utbildningsnivåerna. Med försiktighet kan detta tolkas som att covid-19 påverkade löneutvecklingen på ett liknande sätt mellan utbildningsnivåerna.

4.3.2 Prediktiva p -värden

Observera att i och med att vi predikterar tre år i rad byggs prediktionen för år 2021 på de observerade värdena från 2020, medan de uppskattade värdena för 2022 och 2023 baseras på tidigare predikterade värden. Detta innebär att säkerheten i prediktionerna minskar med prognoshorisonten, vilket kan vara en förklaring till varför skillnaderna mellan de predikterade och observerade värdena ökade över tid.

För att ta hänsyn till att osäkerheten i prediktionerna ökar med prediktionshorisontens längd kan vi använda den prediktiva fördelningen för X_{t+h} givet observerade värden $X_1, \dots, X_t$. För en ARIMA(0,1,0)-process med drift får vi approximativt

$$X_{t+h} \mid X_1, \dots, X_t \sim N(\hat{X}_{t+h}, h\hat{\sigma}^2), \quad (21)$$

där $\hat{X}_{t+h}$ är det predikterade värdet från (20) och där $\hat{\sigma}^2$ är en skattning av feltermernas varians. Vi ser från (21) att prediktionsfelsvariansen ökar linjärt med tidshorisontens längd h . Med hjälp av den prediktiva fördelningen i (21)

kan vi räkna ut det prediktiva p -värdet i (19) för det observerade värdet x_{t+h} av X_{t+h} enligt

$$p = \Phi \left(\frac{x_{t+h} - \hat{X}_{t+h}}{\sqrt{h}\hat{\sigma}} \right), \quad (22)$$

där $\Phi(x) = P(\varepsilon \leq x)$ är fördelningsfunktionen för en standard normalfördelning $\varepsilon \sim N(0, 1)$. I Tabell 4.3 återges de prediktiva p -värdena för samtliga åtta utbildningsnivåer för 2021 ($h = 1$), 2022 ($h = 2$) och 2023 ($h = 3$).

Tabell 4.3: Prediktiva p -värden för observerade reallöner jämfört med predikterade värden under perioden 2021—2023.

Utbildningsnivå	2021	2022	2023
Förgymnasial utbildning kortare än 9 år	0.22	$1.35 \cdot 10^{-4}$	$1.18 \cdot 10^{-7}$
Förgymnasial utbildning, 9 (10) år	0.15	$1.99 \cdot 10^{-7}$	$4.45 \cdot 10^{-12}$
Gymnasial utbildning, högst 2 år	0.07	$3.83 \cdot 10^{-11}$	$5.20 \cdot 10^{-19}$
Gymnasial utbildning, 3 år	0.04	$4.85 \cdot 10^{-11}$	$1.29 \cdot 10^{-19}$
Eftergymnasial utbildning, mindre än 3 år	0.21	$2.70 \cdot 10^{-8}$	$3.08 \cdot 10^{-14}$
Eftergymnasial utbildning, 3 år eller mer	0.36	$3.66 \cdot 10^{-7}$	$1.04 \cdot 10^{-14}$
Forskarutbildning	0.69	$1.36 \cdot 10^{-4}$	$3.21 \cdot 10^{-9}$
Samtliga utbildningsnivåer	0.12	$1.89 \cdot 10^{-9}$	$1.18 \cdot 10^{-17}$

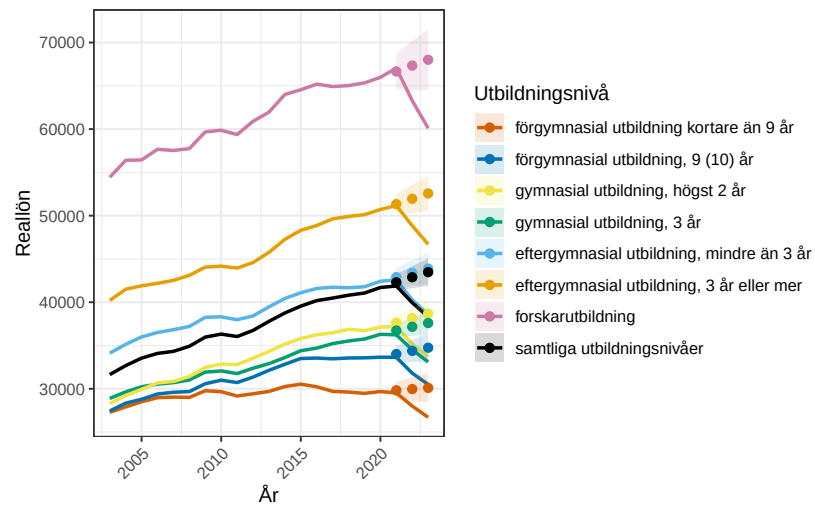
För år 2021 är p -värdena relativt stora, vilket innebär att inga signifikanta avvikelser från modellens prediktioner kan påvisas. För 2022 och 2023 är p -värdena däremot genomgående mycket små, vilket indikerar stark evidens mot nollhypotesen om att de observerade värdena inte avviker från modellens prediktioner utöver slumpmässig variation. Att p -värdena dessutom blir mer extrema mellan 2022 och 2023 tyder på att avvikelserna förstärks över tid. Mönstret är konsistent över samtliga utbildningsnivåer.

4.3.3 Prediktionsintervall

För att visuellt undersöka om de observerade värdena för perioden 2021–2023 avviker från modellens prediktioner konstrueras ett 99% prediktionsintervall. Givet prediktionen i (20) formuleras prediktionsintervallet som

$$\hat{X}_{t+h} \pm \lambda_{0.005} \sqrt{h}\hat{\sigma}. \quad (23)$$

Prediktionsintervallet för varje utbildningsnivå markeras som skugga i samma färg i Figur 4.6. Det tyder på samma resultat som p -värden, alltså inga signifikanta avvikelser mellan de observerade värden och modellens prediktioner i 2021, men i 2022 och 2023. Detta kan indikera att effekten av covid-19 på löneutvecklingen försköts i tiden och först blir statistiskt tydlig under 2022, för att därefter förstärkas under 2023.



Figur 4.6: 99% Prediktionsintervall för reallöner. Skuggade områden representerar prediktionsintervallen, medan punkterna visar modellens prediktioner.

5 Diskussion

5.1 Begränsningar av modellen

Modelleringen i detta arbete har byggts på antagandet att lönesutveckling kan beskrivas med hjälp av tidsserieanalys, det vill säga att den enbart kan uttryckas som en funktion av tid. I praktiken påverkas lönen av ett flertal faktorer, såsom kön och yrke. Dess utveckling påverkas exempelvis av arbetsmarknad och kollektivavtal. Modellen utgör därmed en förenklad beskrivning av verkligheten och fångade inte dessa underliggande samband.

Ett sätt att värdera en prediktion är att jämföra de uppskattade värdena med de observerade. Exempelvis sparas den sista delen av datan i vissa studier för validering, vilket inte är relevant i vårt fall. Vi kan inte validera prediktionen mot faktiska värden, eftersom analysen baseras på ett kontrafaktiskt scenario, det vill säga att covid-19 inte existerade.

Trots begränsningarna och bristerna i modelleringen och analysen finns det fortfarande resultat som stämmer överens med teorin och är rimliga. Dels har KPSS-testet gett fler stationära tidsserier än ADF-testet, vilket kan förklaras av deras motsatta nollhypoteser, där det är lättare att uppfylla nollhypotesen om att en tidsserie är stationär. Dels följer de predikterade värdena den trend som observerats före covid-19 och att resultaten tyder på att pandemin haft ett negativt effekt på löneutvecklingen. Dessutom är effekten inte statistiskt påvisbar under det år då pandemin inträffade, utan fördröjs med ett år och förstärks över tid.

5.2 Vidare studier

Eftersom löneutveckling är en komplicerad process skulle en multivariat tidsseriemodell, där fler förklarande faktorer än utbildningsnivå fångas, kunna förklara löneutvecklingen mer utförligt. Data för exempelvis yrkesgrupper och ålder finns tillgängliga från samma SCB-datakälla som användes i detta arbete.

Även vid fokus på enbart utbildningsnivåer kan vidare studier utföras för att analysera skillnaderna mellan utbildningarna mer ingående. Detta kan exempelvis göras genom att ta differenser mellan utbildningsnivåer och betrakta dessa som separata tidsserier.

Datan i sig visar generellt en växande trend i Figur 3.1 och 3.2. Utöver tidsserieanalys skulle även exempelvis regressionsmodeller kunna tillämpas, där flera variabler inkluderas för att ge en mer omfattande förklaring av löneutvecklingen.

6 Referenser

Brockwell, Peter J. & Davis, Richard A. (1991). *Time Series: Theory and Methods*. 2 uppl. Springer.

Brockwell, Peter J. & Davis, Richard A. (2016). *Introduction to Time Series and Forecasting*. 3 uppl. Springer.

Enders, Walter (2014). *Applied Econometric Time Series*. 4 uppl. Wiley.

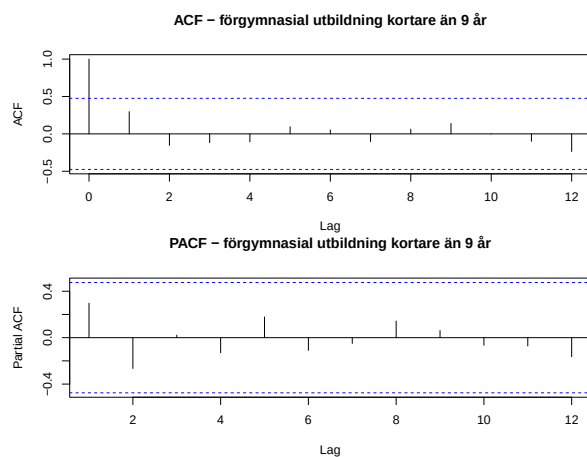
Hyndman, Rob J. & Athanasopoulos, George (2021). *Forecasting: Principles and Practice*. 3 uppl. OTexts.

Kwiatkowski, Denis, Phillips, Peter C. B., Schmidt, Peter & Shin, Yongcheol (1992). Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root.” *Journal of Econometrics*, 54, 159–178.

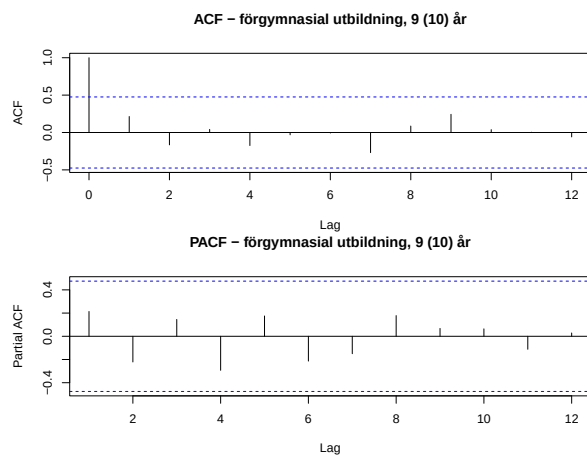
Rocha, Jorge, Oliveira, Sandra & Viana, Cláudia M. (2024). *Time Series Analysis: Recent Advances, New Perspectives and Applications*. IntechOpen.

Sundberg, Rolf (2023). *Lineära Statistiska Modeller*. Matematisk statistik, Stockholms universitet. Kompendium.

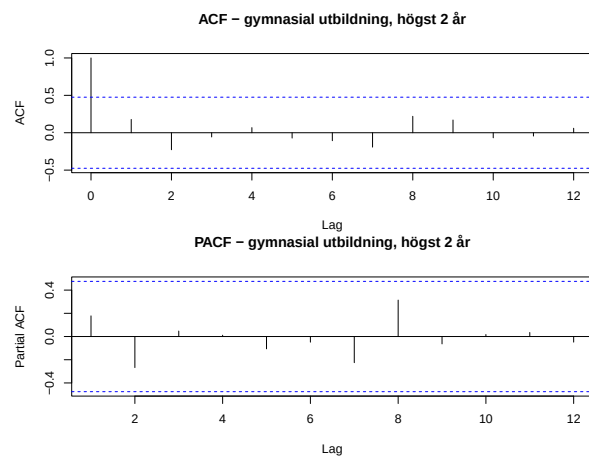
A Appendix: ACF- och PACF-diagram



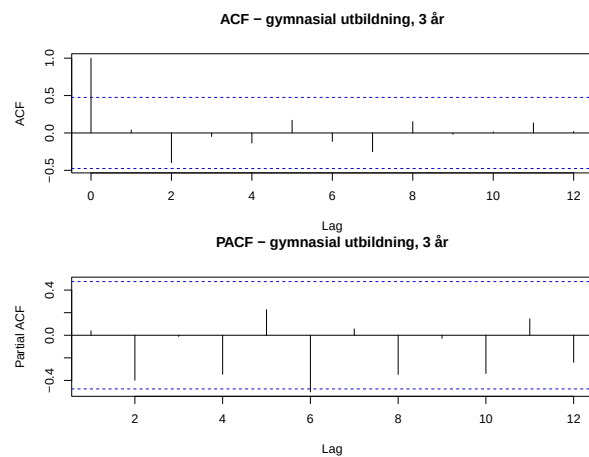
Figur A.1: Skattade korrelationsfunktioner för årsvisa differenser av reallönen för utbildningsnivån Förgymnasial utbildning kortare än 9 år, baserat på data under 2003–2020.



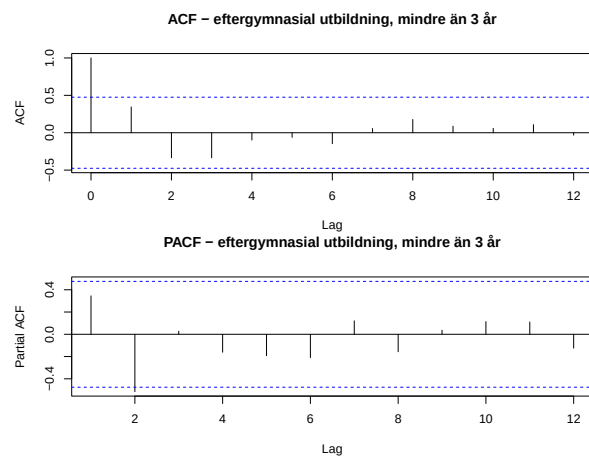
Figur A.2: Skattade korrelationsfunktioner för årsvisa differenser av reallönen för utbildningsnivån Förgymnasial utbildning, 9 (10) år, baserat på data under 2003–2020.



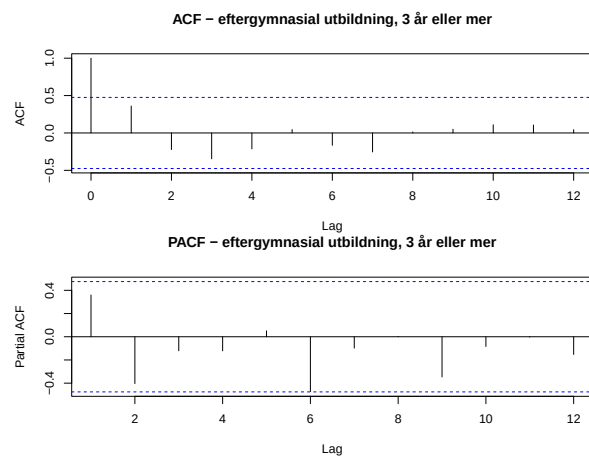
Figur A.3: Skattade korrelationsfunktioner för årsvisa differenser av reallönen för utbildningsnivån Gymnasial utbildning, högst 2 år, baserat på data under 2003–2020.



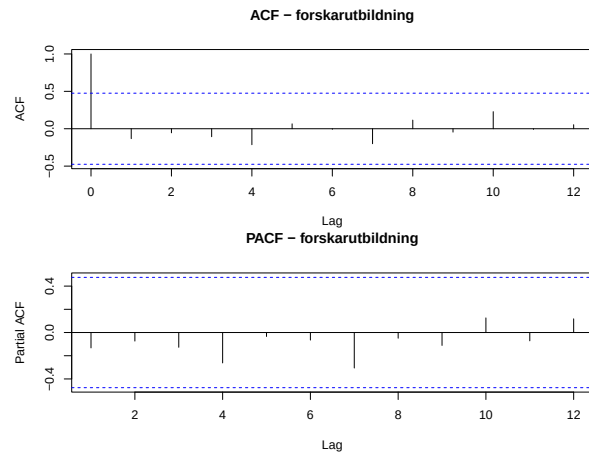
Figur A.4: Skattade korrelationsfunktioner för årsvisa differenser av reallönen för utbildningsnivån Gymnasial utbildning, 3 år, baserat på data under 2003–2020.



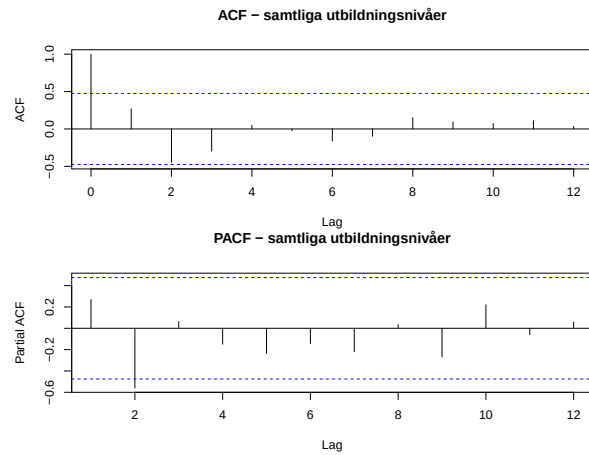
Figur A.5: Skattade korrelationsfunktioner för årsvisa differenser av reallönen för utbildningsnivån Eftergymnasial utbildning, mindre än 3 år, baserat på data under 2003–2020.



Figur A.6: Skattade korrelationsfunktioner för årsvisa differenser av reallönen för utbildningsnivån Eftergymnasial utbildning, 3 år eller mer, baserat på data under 2003–2020.



Figur A.7: Skattade korrelationsfunktioner för årsvisa differenser av reallönen för utbildningsnivån Forskarutbildning, baserat på data under 2003–2020.



Figur A.8: Skattade korrelationsfunktioner för årsvisa differenser av reallönen för utbildningsnivån Samtliga utbildningsnivåer, baserat på data under 2003–2020.