Macroeconometrics
Problem Set 1 [*]

Thor Donsby Noe [†] [a]

[a]Universitat de Barcelona, Spain

December 29, 2018

CONTENTS

1. Introduction

This problem set covers topics such as simulation of stochastic processes, ARIMA modelling and order of integration analysis of time series.

All of the estimates, predictions and outputs are created using Stata.[1]

2. Questions

2.1. Ergodic Stochastic Process

A stochastic progress x_t is given by the DGP

$$x_t = f(t) + \phi x_{t-1} + \varepsilon_t,$$
$$\text{with } f(t) = \mu \neq 0, \quad |\phi| < 1 \text{ and } \varepsilon_t \sim iid(0, \sigma_\varepsilon^2) \tag{2.1}$$

x_t is the realization of an ergodic stochastic process if it is stationary and for any point in time t satisfies

$$\lim_{T \to \infty} \left(T^{-1} \sum_{s=1}^{T} Cov(x_t, x_{t+s}) \right) = 0 \tag{2.2}$$

That is, on average the memory of the process is bounded. Thus, covariances goes towards 0 for two different point in time so that two parts of the same stochastic process can be independent from each other.

By iterative substitution we see that equation 2.1 can be rewritten as

$$
\begin{aligned}
x_t &= \mu + \phi x_{t-1} + \varepsilon_t \\
&= \mu + \phi(\mu + \phi x_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\
&= \mu(1 + \phi) + \phi^2 x_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t \\
&= \mu(1 + \phi) + \phi^2(\mu + \phi x_{t-3} + \varepsilon_{t-2}) + \phi \varepsilon_{t-1} + \varepsilon_t \\
&= \mu(1 + \phi + \phi^2) + \phi^3(x_{t-3}) + \phi^2 \varepsilon_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t \\
&\;\;\vdots \\
&= \phi^T x_{t-T} + \sum_{s=1}^{T} \left(\mu \phi^{s-1} + \phi^{s-1} \varepsilon_{t-s+1} \right)
\end{aligned}
\tag{2.3}
$$

rewriting the sum of the geometric series:

$$= \phi^T x_{t-T} + \mu \left(\frac{1 - \phi^T}{1 - \phi} \right) + \sum_{s=1}^{T} \phi^{s-1} \varepsilon_{t-s+1},$$

where T is the number of periods s prior to t.

---

[1] The Stata do-file is available as the GitHub repository
https://raw.githubusercontent.com/thornoe/ub/master/Macroeconometrics/PS1/stata_code/do_PS_1.do

If the process was a unit root the constant terms with μ would accumulate and produce a deterministic trend, but as $|\phi| < 1$ then $\phi^T \xrightarrow[T \to \infty]{} 0$ such that for a high number of periods T equation 2.3 converges to

$$x_t \xrightarrow[T \to \infty]{} \frac{\mu}{1-\phi} + \sum_{s=1}^{T} \phi^{s-1} \varepsilon_{t-s+1} \tag{2.4}$$

And as ε_t is a stochastic term the expected realization of $_t$ will be

$$E(x_t) = \frac{\mu}{1-\phi} \tag{2.5}$$

As the expected mean is a constant then $\{x_t\}_{t=1}^{T}$ is an integrable *iid* stochastic process and for $\mu < 0$ the Law of large numbers implies that the mean converge in probability

$$\bar{x}_T \xrightarrow{p} \frac{\mu}{1-\phi} \tag{2.6}$$

Furthermore, we assume $T \to \infty$ then x_t can be rewritten to a sequence of random variables $\{\tilde{x}_T\}_{T=1}^{\infty}$ that are drawn from a population with a probability distribution with finite mean $\frac{\mu}{1-\phi}$ and finite variance σ_ε^2. The Central Limit Theorem then implies convergence to the normal distribution

$$\sqrt{T}\left(\bar{x} - \frac{\mu}{1-\phi}\right) \xrightarrow{d} N(0, \sigma_\varepsilon^2) \tag{2.7}$$

The derivations above shows that the memory of x_t is bounded as the stationarity condition $|\phi| < 1$ is fullfilled, thus the memory of the initial value x_{t-T} and chock ε_{t-T+1} decreases to zero the further away in time t is from the initial period $t - T$.

Furthermore, the deterministic term given by the constant $\mu \neq 0$ does not give rise to a linear trend, as $|\phi| < 1$.

In conclusion, x_t in equation 2.1 indeed describes an ergodic stochastic process.

2.2. Simulation of DGPs

    a) Generate a realization of a Gaussian stochastic process of sample size T = 500, and plot the time series against time

As the data generating process (DGP) is stochastic without any autoregressive element, we simply perform 500 draws from a standard normal distribution. Panel A in figure 1 shows the sime series, whereas the histogram in panel B confirms that the realizations are approximately standard normal distributed.

Figure 1: Gaussian stochastic process
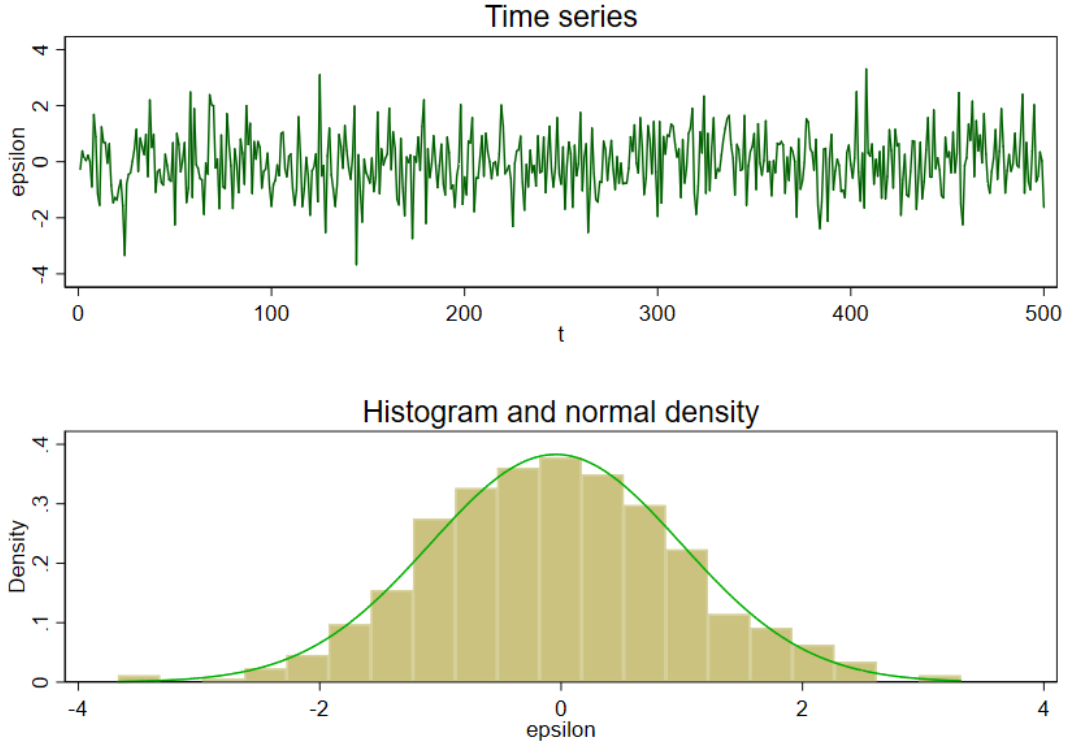
b) Generate AR(1) processes with three different autoregressive coefficients $\phi = \{0.5, 0.8, 0.95\}$

Keeping the sample size at 500 and including a burn-in period of 2.000 iterations, the time series for the autoregressive models are plotted in figure 2. The autocorrelation functions (ACFs) and partial autocorrelation functions (PACFs) are plottes in figure 3.

While all three processes are mean-reverting as the stationarity condition $|\phi| < 1$ is fulfilled, it is clear that persistence increases as $\phi \to 1$. This translates into longer lasting and more extreme cycles in figure 2 as the autoregressive element dominates the stochastic error term. Likewise, the ACFs in figure 3 show that temporal dependence is significant three periods back for $\phi = 0.5$ while the autocorrelation is significant for the past 12-13 periods for $\phi = 0.95$. While I have only shown 25 lags in the figures, extending it even shows that the latter process is near-borderline significantly autocorrelated with period 30-37 as well, but negatively.

Figure 3 shows for the ACFs as well as for the PACFs that the coefficient of correlation with the first lag is close to the actual ϕ-values in the DGP. Except for a few outliers, the PACFs are inconsistent with AR(p) models of order $p > 1$.

Furthermore, the number of lags that are significantly autocorrelated are much higher in the ACFs as opposed to the PACFs this indicate that the processes are AR(1) and not MA(q) or ARMA(1,q) processes.

Figure 2: AR(1) processes with different autocorrelation coefficients
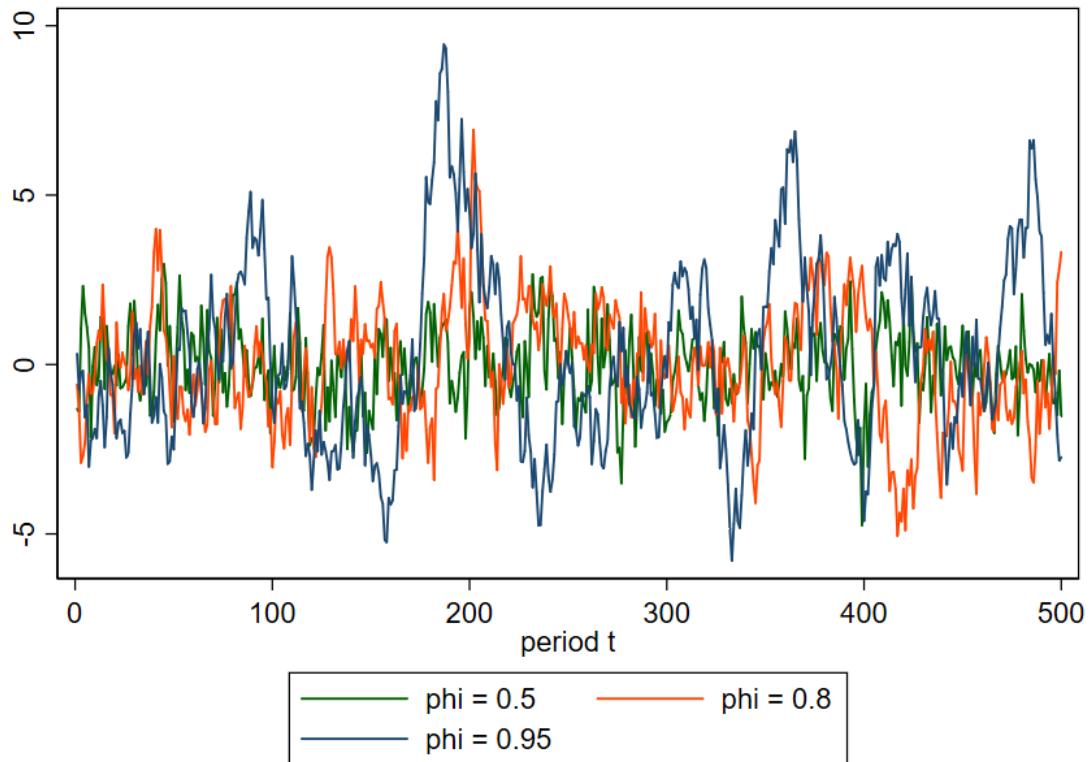
Figure 3: Autocorrelation and Partial Autocorrelation functions for AR(1) processes
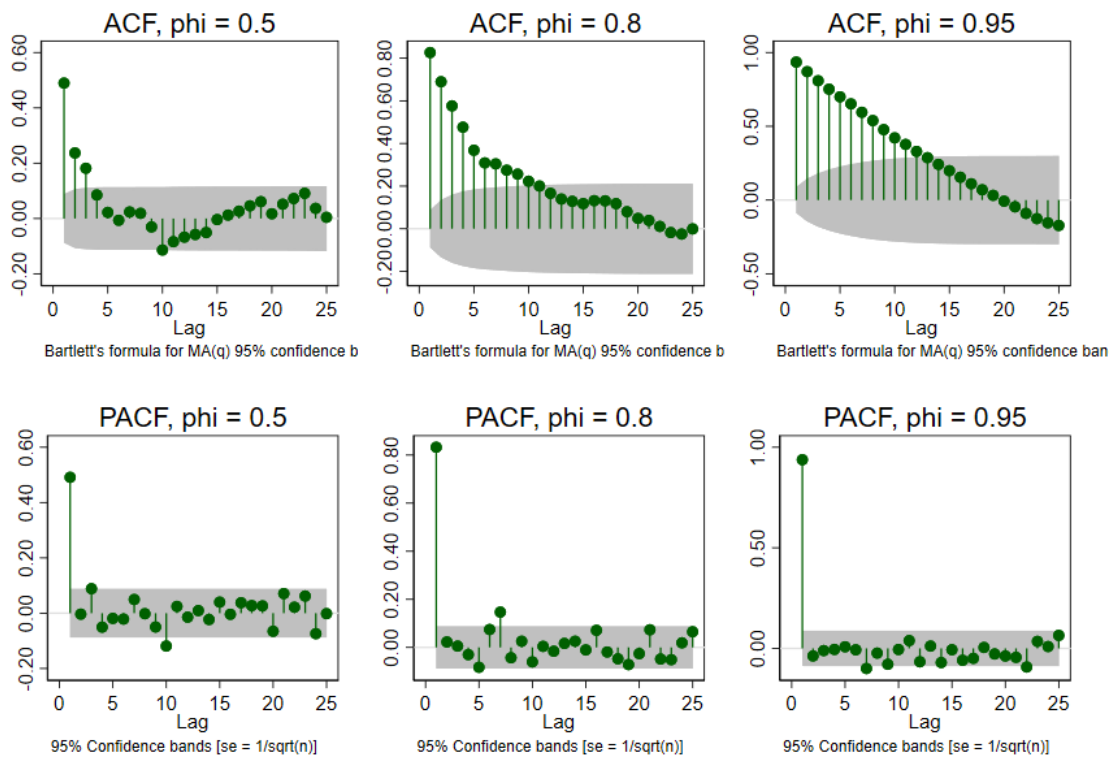
> **b) Generate MA(1) processes with three different moving average coefficients $\theta = \{0.5, 0.8, 0.95\}$**

The Moving Average processes are likewise generated with a sample size of 500 including a burn-in period of 2.000 iterations. The time series are plotted in figure 4 and the autocorrelation functions and partial autocorrelation functions are plottes in figure 5 on the next page.

As opposed to the AR(1) processes there are as good as no signs of persistence in the time series for the MA(1) processes. Though a bit difficult to tell from comparison of figure 1 and 4, the summary statistics show that the standard deviation is a little higher for the MA(1) processes than for the standard normal distribution, and increase slightly with $\theta \rightarrow 1$. This is due to an increased persistence of prior shocks.

The ACFs in figure 5 confirms that there are not any persistence after the 1st lag, while the PACFs show significant lags besides the first but with alternating sign. For $\theta = 0.5$ x_t is linearly correlated with x_{t-1}, x_{t-2}, and x_{t-3} and arguably with x_{t-4} (borderline significant). The MA(1) process with $\theta = 0.95$ on the other hand shows significant correlation with all lags 1-8.

The number of significant lags are much higher in the PACFs as opposed to the ACFs where only the first lag is significant which is consistent with an MA(q) model of order $q = 1$ while inconsistent with the processes being AR(p) or ARMA(p,q) processes.

Figure 4: MA(1) processes with different autocorrelation coefficients
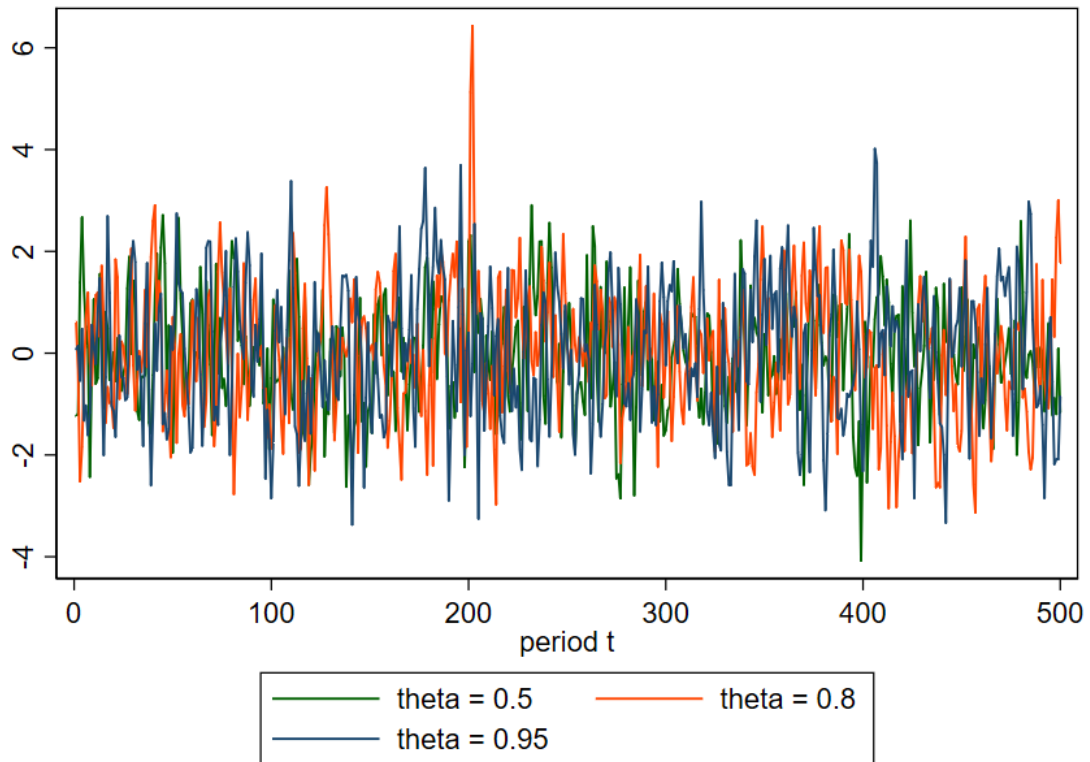
Figure 5: Autocorrelation and Partial Autocorrelation functions for MA(1) processes
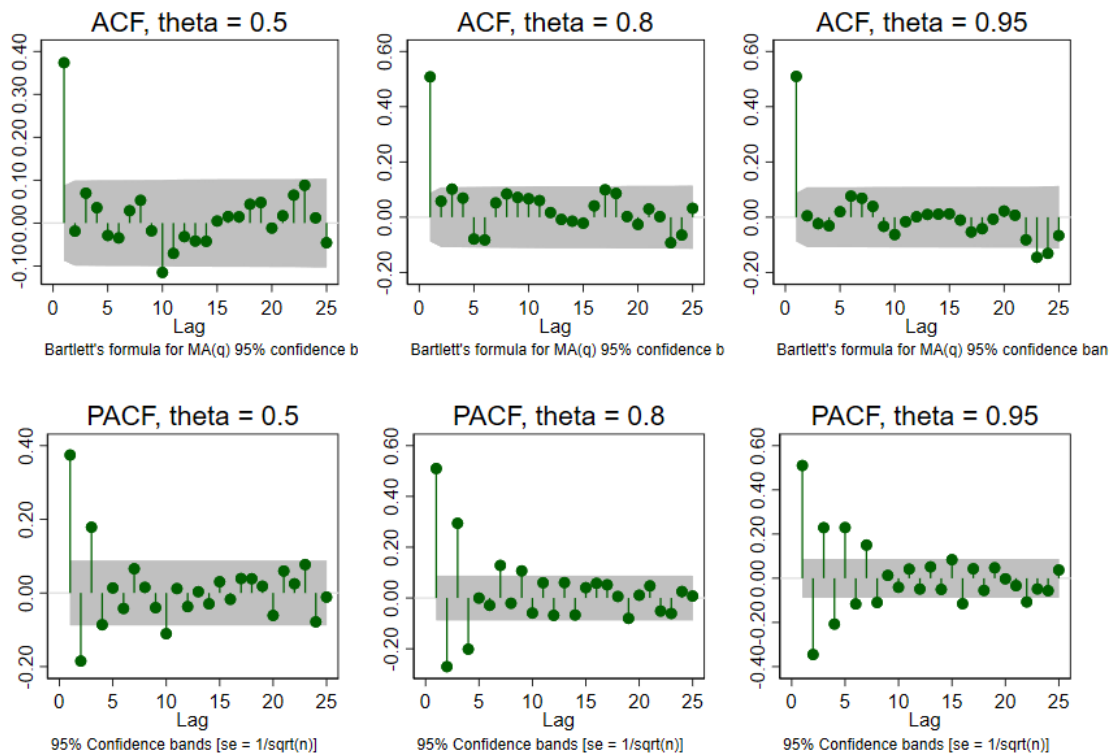
2.3. ARIMA modelling

Using the "Penn World Tables v. 9.0"[2] I analyze the development in the logarithm of real GDP per capita (at 2011 national prices) from 1950-2014 for Spain, (Western) Germany, and Denmark[3].

a) Plot the time series.

From figure 6 below it is obvious that the series for actual values of log GDP are not stationary as should be expected while the transformed series might be. Therefore, we will work with these series of first-differences instead leading to not looking to model ARMA(p,q) models but ARIMA(p,1,q) models indicating the 1st order transformation.

Figure 6: Time series and first differences of log real GDP per capita
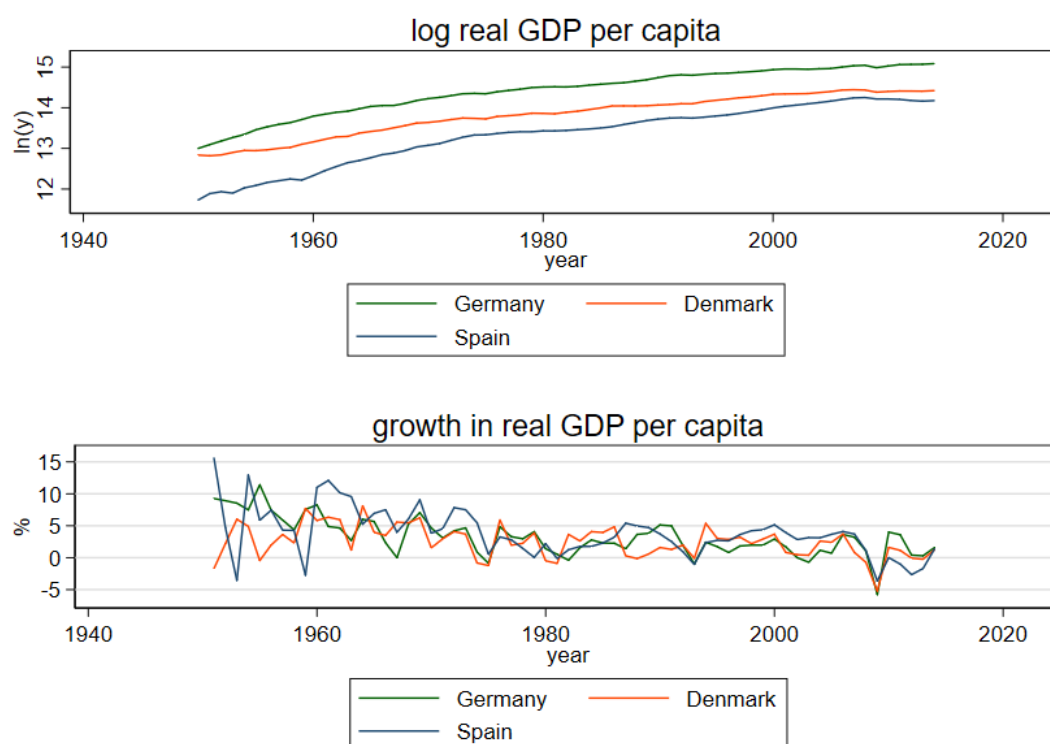
However, the much higher growth rates in the 50s and 60s does raise the question of functional form even for the transformed processes of 1st order. Is there a negative trend or a level shift? Also business cycles and especially negative shocks are evident. All three countries experience negative growth due to the global recessions in 1993 and 2009. Furthermore Denmark was hit significantly by the oil crisis in 1974-75 and

---

[2]Downloadable from `https://www.rug.nl/ggdc/productivity/pwt/`

[3]Before taking the log, the Danish Kroner (DKK) is transformed to Euros by multiplication with 7.46038 which is the central exchange rate at which the DKK is pegged to the Euro within a tiny 2.5% band, according to the European Exchange Rate Mechanism (ERM).

Germany in 1975. Denmark also started the 80s with negative growth rates due to the global recession, while Germany again only experienced detraction for one year in 1982 and likewise briefly in 2003 and 2009. The financial crisis showed persistent negative growth in Denmark in 2008-2009 and 2012-2013 as in Spain with detraction through 2011-13. Germany had a single year with zero growth in 1967 and Denmark and Spain also experienced short recessions during the 50s like the 1959 reaction to the Stabilization Plan in Spain.

b) Identification: using autocorrelation and partial autocorrelation functions, iden-
tify the orders of the ARIMA model. Compute both the numerical and graphical
autocorrelation functions (including confidence bands)

The Autocorrelation Functions (ACFs) and Partial Autocorrelation Functions (PACFs) in figure 7 below show first of all that we do not need to worry about unit roots as none of the lags have an autocorrelation coefficient above 0.6.

Figure 7: ACFs and PACFs of growth in real GDP per capita in Germany, Denmark, and Spain
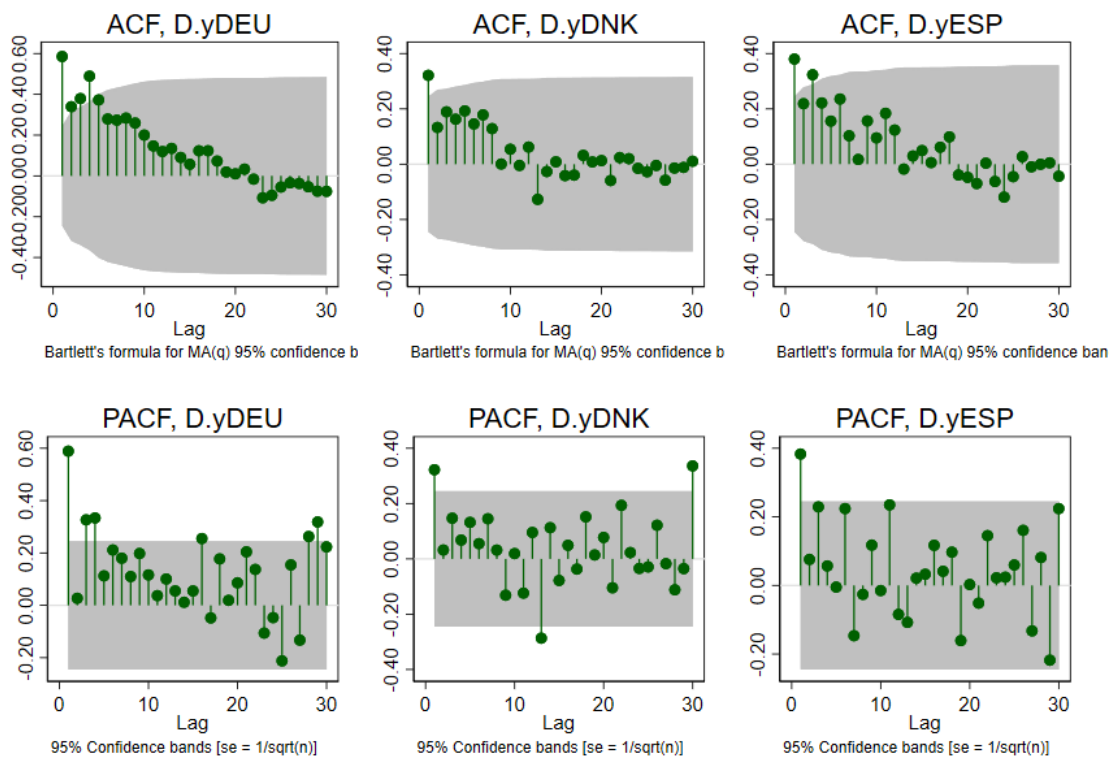
The real GDP growth in Germany shows clear signs of autocorrelation for lags 1-4 and close to significant for autocorrelation for lag 5. In the partial autocorrelation function lag 1, 3-4, 16, and 28-30 are significant while 6-7 and 9 are close to significant. Computing the numerical correlogram Portmanteau's Q test for white noise clearly rejects that real GDP growth in Germany is a white noise process, on the contrary up to 30+ lags

might be autocorrelated. The high number of lags in both the ACF and PACF suggests that German log real GDP per capita can be approximated by an ARIMA(p,1,q) model, that is, estimating growth in real GDP by an ARMA(p,q) model. While it is possible to estimate a model where only lag $q = 1, 3, 4, 16, 28, 29, 30$ are included, it is economically unrealistic and the later lags are obviously 'false signals' that can be due to model shifting caused by to structural breaks, e.g. the recessions in the German economy highlighted above. That aside, the graphical representation with the confidence bands is more consistent with up to 4 ACF lags and 4 PACF lags, thus initially suggesting an ARMA(4,4) model though different combinations of included lags $p \leq 4$ and $q \leq 4$ could be feasible as well.

While the graphical plots of both the ACF and PACF for real GDP growth in Denmark only show that the firsts lag are within the 95% confidence several lags up to 7 gets close to the confidence bands and lag 13 is significantly autocorrelated with time t. Even more so the correlogram still show signs of autocorrelation and partial autocorrelation. In fact it is not until all the 12 first lags are included, that the Q test rejects the null-hypothesis that there at a 5% significance level is no autocorrelation in the sample. This gives confidence that we can try an ARMA(1,1) model. Especially if controlling for some of the no less than 12 out of the 64 periods with negative growth in Denmark that might cause 'false signals'.

The stochastic process for Spain show less evidence of autocorrelation and partial autocorrelation than for Germany but a little more than for Denmark. We will likewise use an ARMA(1,1) model but also try out modeling growth in real GDP by an AR(1), AR(3), ARMA(3,3), ARMA(3,1), or ARMA(1,3) and test all four against each other, as lag 3 is about borderline significantly autocorrelated and partially autocorrelated as well.

For all three countries, however, the PACFs in figure 7 seem very close to zero for lag 2 which could cause issues trying to estimate ARIMA(p,q) models of order $q > 1$.

c) Estimation: fit the identified ARIMA specification, using MLE estimation procedures

The ACFs and PACFs in figure 7 showed that for each country several different model specifications could be reasonable suggestions for modelling real GDP growth. Therefore, all of the more likely model specifications are estimated and compared both in terms of the within-sample fit according to the Information Criteria (IC) and on the realism of the model based on the significance of the estimated parameters according to the z test statistic.

For Germany the baseline ARMA(4,4) is tested against an ARMA(4,1) and ARMA(1,4) model. The latter clearly surpasses the first two in terms of both IC and due to the parameter estimates actually being significant besides from the parameters for lag

$q = 3, 4$ that are insignificant for all. The ARMA(1,2) and the ARMA(2,2) model do not even converge (not surprisingly, keeping the PACF in figure 7 in mind). Instead the ARMA(1,4) is instead tested against the ARMA(1,1), AR(1) and MA(1) model. As the Akaike Information Criterion (AIC) points towards the ARMA(1,4) while the Bayesian IC (BIC) points towards the ARMA(1,1) specification we still have a subjective choice to make.

In the spirit of Milton Friedman's methodology of positive economics we could go with the ARIMA(1,1,4) model in order to maximize predictive power at the cost of using a less realistic model specification including insignificant parameters, especially $\hat{\theta}_3 \sim 0$ related to $q = 3$. Oppositely we could follow the Parsimonious principle of estimating the more simple model in the ARMA(1,1). Though we should be well-knowing about the lost information from not including lag 3 and 4 that showed signs of significant in the PACF in figure 7.

All of the estimated ARMA(p,q) models have an estimate of the 1$^\text{st}$ order autoregressive parameter $\hat{\phi}_1$ that is not significantly different from unity. Thus, only the persistence of shocks from the moving average elements prevents the model from acting as a random walk. An alternative further step would be to consider the AR(1) model that performs almost as good as the ARMA models and much better than the MA(1) model but with $\hat{\phi}_1 = .62$ being far from unit root. In this autoregressive process there would only be indirect persistence of shocks, though. We will continue investigating the performance and implications of each of the three models throughout and the Maximum Likelihood (ML) estimates are shown in column (1)-(3) of table 2.

Table 1 contain the estimates of the three models I choose to proceed with for Denmark and Spain. For Denmark the AR(1) model is clearly preferred as it has a lower AIC and BIC than both the ARMA(1,1) and the MA(1) that would be the two other reasonable specifications following the ACF and PACF in figure 7.

Table 1: Estimates of ARMA models for growth in real GDP per capita in Denmark and Spain

	(1) Denmark, AR(1) b/se	(2) Spain, ARMA(1,1) b/se	(3) Spain, AR(1) b/se
cons	0.024***	0.038	0.039***
	(0.004)	(0.024)	(0.008)
ARMA			
L.ar	0.332***	0.923***	0.443***
	(0.123)	(0.117)	(0.087)
L.ma		-0.688***	
		(0.164)	
sigma	0.023***	0.034***	0.034***
	(0.002)	(0.002)	(0.002)
aic	-292.6	-244.5	-243.4
bic	-286.1	-235.8	-236.9

Standard errors are in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

For Spain none of the models including $p = 3$ or $q = 3$ perform well, so the two preferred models are the ARMA(1,1) model with the lowest AIC and the AR(1) model with the lowest BIC. As for Germany it is not significant that $\hat{\phi}_1 \neq 1$ for the ARMA(1,1) model but it is for the AR(1) model.

d) Validation: check the estimated residuals for misspecification errors

The residuals are computed and investigated for the estimated models. For Germany graphs for both the time series of the residuals, and their Autocorrelation Functions (ACFs) and Partial Autocorrelation Functions (PACFs) are shown in figure A.2 in the appendix. First looking at the ARMA(1,4) none of the lags in the ACF for the residuals are significant. In the PACF only lag 15, 22, 25 are significant but it would not make sense to include these as lags of q in the ARMA(1,q). Thus, they must be so-called 'false signals' which could be explained by model shifting due to one of the structural breaks that we found earlier. The more sizable negative residuals are for the 2009-crisis followed by the 1974-75 oil-crisis and the 1993-recession. For the residuals of the ARMA(1,1) model the AC and PAC of lag 2 is significant, however, we know that the model does not converge when $q = 2$ and that ϕ_2 is not significant in an ARMA(2,1) model. The Q test is rejected for lags 2-6 meaning that they might be autocorrelated or partially autocorrelated That is, there is something left to be explained in the residuals and the model might be misspecified. For the AR(1) lag 4 of the residuals is significant and lag 2 close to significant for both the ACF and PACF the Q test is likewise rejected for lag 4.

Some of the stronger burst and busts might have disproportionately big impact on the size of the estimates and explain the mixed signals in the PACFs. Thus to improve the estimated models time dummies are included for the stronger structural breaks. To

create a common criteria time dummies are included if the residual is 1.5 times bigger than the estimated standard error of the model $\hat{\sigma}$.[4] Table 2 show the estimates of the ARMA(1,4), ARMA(1,1), and AR(1) models both with and without the inclusion of time dummies. Column 1-3 show the original models and column 4-6 show the same models but including time dummies for major structural breaks. The different models are compared to each other in terms of Information Criteria (IC) and simply by size and significance of the estimates. As expected both IC improve when controlling for 'outliers'. However the MA coefficients become highly insignificant for the ARIMA(1,1,4) where the standard errors explode to unacceptable dimensions for $q = 1, 4$ and $\hat{\theta} = 1$ (column 4). Though still significant at a 5% level the sign of $\hat{\theta}_1$ changes from negative to positive for the ARMA(1,1) model (column 5), an improvement however is that $\hat{\phi}_1$ is now significantly different from unity.

---

[4]For the AR(1) model the same year dummies are included as for the ARMA(1,1) model since the standard error $\hat{\sigma}$ is relative high such that 2009 otherwise would be the only included negative outlier.

Table 2: Estimation of ARMA models for growth in real GDP per capita in Germany

	(1) ARMA(1,4) b/se	(2) ARMA(1,1) b/se	(3) AR(1) b/se	(4) ARMA(1,4) b/se	(5) ARMA(1,1) b/se	(6) AR(1) b/se
D.a51				0.001	0.005	0.002
				(0.015)	(0.007)	(0.010)
D.a55				0.023***	0.019***	0.017**
				(0.008)	(0.006)	(0.007)
D.a63				-0.007*	-0.013**	-0.015
				(0.004)	(0.007)	(0.016)
D.a74				-0.025***	-0.014	-0.013
				(0.004)	(0.009)	(0.008)
D.a75				-0.044***	-0.030**	-0.033**
				(0.005)	(0.013)	(0.015)
D.a93				-0.014***	-0.014	-0.017
				(0.004)	(0.017)	(0.013)
D.a09				-0.045**	-0.064***	-0.066***
				(0.020)	(0.020)	(0.021)
D.a10				-0.020	-0.030*	-0.031
				(0.018)	(0.018)	(0.022)
D.a67					-0.012***	-0.018***
					(0.004)	(0.006)
cons	0.043*	0.042*	0.034***	0.042	0.034***	0.035***
	(0.023)	(0.024)	(0.007)	(0.030)	(0.009)	(0.012)
ARMA						
L.ar	0.970***	0.980***	0.620***	0.986***	0.691***	0.840***
	(0.044)	(0.031)	(0.073)	(0.030)	(0.131)	(0.085)
L.ma	-0.535***	-0.731***		-0.665	0.410**	
	(0.148)	(0.096)		(.)	(0.162)	
L2.ma	-0.384**			-1.272***		
	(0.175)			(0.149)		
L3.ma	0.021			-0.665***		
	(0.175)			(0.232)		
L4.ma	0.294*			1.000		
	(0.155)			(.)		
sigma	0.020***	0.022***	0.023***	0.008***	0.015***	0.015***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
aic	-301.9	-297.0	-293.3	-342.5	-331.3	-330.7
bic	-286.8	-288.4	-286.8	-314.4	-303.2	-304.8

Standard errors are in parentheses. * p<0.10, ** p<0.05, *** p<0.01

As both estimated ARMA models seem to be quite unrobust to the ex- or inclusion of time dummies, this at first could seem to be where the AR(1) model comes to its right (column 6). The value of $\hat{phi}_1$ increases with the inclusion of time controls which is to be expected as it would need to undershoot before to not be too vulnerable to burst and busts. When including dummies both IC are lowest for the ARMA(1,4) model. When comparing the ARMA(1,1) and the AR(1) they are about level where the AIC is lower for the ARMA(1,1) and the BIC is lower for the AR(1) model.

While the IC naturally has improved a lot by removing some of the unexplained extremes, figure A.2 in the appendix show that the ACFs, PACFs and the Q test for the residuals have only improved minimally. Nonetheless, f-or the ARMA(1,1) model it is just enough so that the Q test of no-autocorrelation of the residuals cannot be rejected for any lags. For the AR(1) without time dummies the Q test of the residuals was rejected for the 4$^{\text{th}}$ lag that was also significant in both the ACF and the PACF. Including the time dummies have actually made the lag 2 of the residuals significant instead and now the Q test is rejected for lags 5-7. Clearly there is something left unexplained for the model, even with inclusion of time dummies. Therefore we reject the AR(1) model for Germany as misspecified and continue with the two ARMA models only.

For Denmark and Spain none of the lags of the residuals are significantly autocorrelated except for a few pointless partial autocorrelations of high-order lags as shown in figure 8. This is confirmed as none of the Q tests of no-autocorrelation in the residuals can be rejected. Thus, there is nothing left to be explained and our models seem well-specified. Yet it should be kept in mind that the autocorrelation and partial autocorrelation of the residuals in the AR(1) model for Spain are about borderline significantly for both lag 3 and 6 however which might possibly change if controlling for structural breaks as done for Germany.

Figure 8: Residuals from modelling growth in real GDP per capita in Denmark and Spain
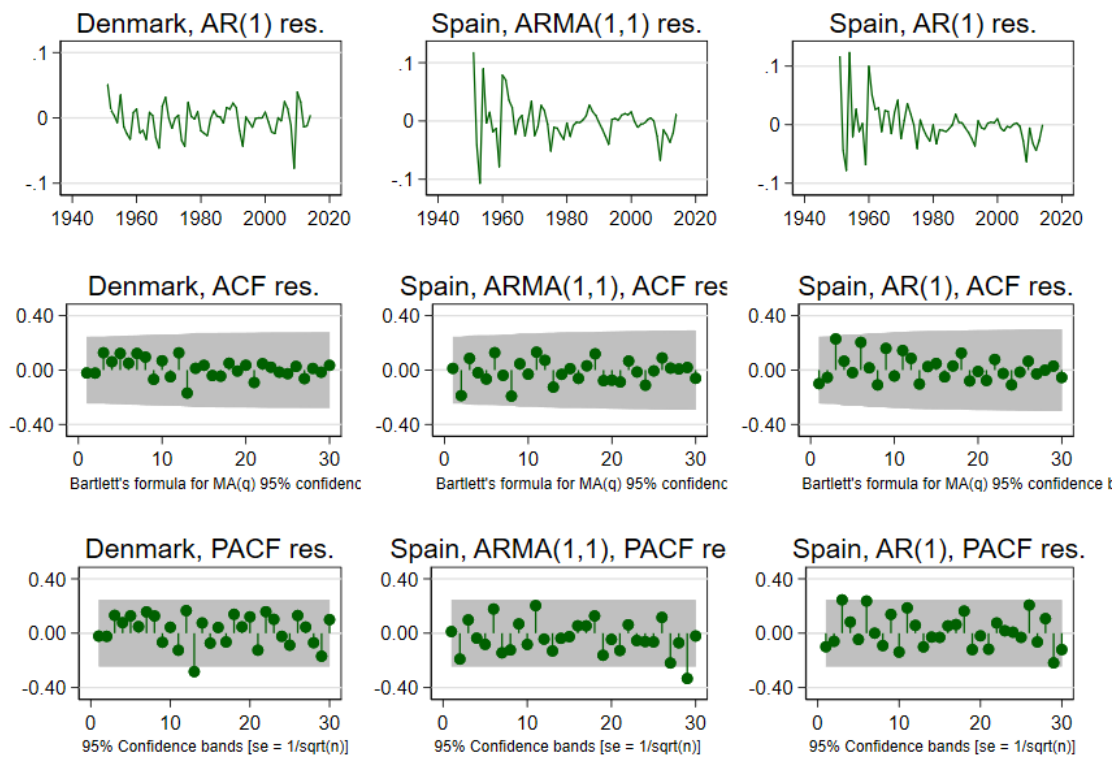
e) *Forecasting: compute the one-step ahead forecast of the log of real GDP per capita time series*

For Stata to be able to forecast even one-step ahead out of sample requires that the time dummies are set to zero, thus handling all of the sample as 'unknown' by disregarding beforehand-knowledge of which years will be extremes. The predictions of the growth rate in real GDP per capita in Germany of for each of the different models are shown in table 3. It becomes very clear that while including dummies a model will certainly improve in terms of the within-sample fit, but the out-of-sample performance is definitely not guaranteed to improve. For the two ARMA models estimated without use of time dummies the prediction is a growth rate of 1.4% for 2015 which is 0.2 percentage points less than the growth rate for 2014 in the sample. Reality has proven that the growth rate actually fell with exactly 0.2 percentage points.

The ARMA(1,1) model estimated using time dummies on the other hand predicts the growth rate to be a full percentage point higher. This reveals the nature of using time dummies for the estimation. As seen by the size and even the directions of the parameters in table 2 the cautions in the predictions are removed when ignoring extreme observations by regarding them as exogenous shocks that will point the autoregressive process in another direction when needed, instead of regarding burst and busts endogenous phenomena that should be regarded as part of the process an estimated model should aim at capturing to some degree.

The forecast of negative growth of no less than 5.5% by the ARMA(1,4) model estimated using times dummies is the direct results of the extreme values of the Moving Average terms where $\hat{\theta}_2 = -1.3$ and $\hat{\theta}_2 = 1.0$. In fact these parameter values show that I for model selection probably should have been putting more emphasize on the features of the actual estimated parameters and be putting less weight on the IC and whether there were signs of autocorrelation in the residuals.

Table 3: Predictions for 2015 growth in real GDP per capita, Germany

	DyDEU_pred1	DyDEU_pred2	DyDEU_pred3	DyDEU_pred4
Model	ARMA(1,4)	ARMA(1,1)	ARMA(1,4)	ARMA(1,1)
Time dummies	no	no	yes	yes
mean	1.42	1.40	-5.52	2.36

The AR(1) models predicts the growth in real GDP per capita for 2015 to be 2.0% for Denmark and for 2.8% for Spain. However the forecast of the ARMA(1,1) for Spain is more conservative at 0.7%. For both countries the 2014 growth rate was around 1.3% with negative growth rates in the years before that. Like the ARMA(1,1) model estimated without time dummies for Germany, the MA term $\hat{\theta}_1 < 0$ which is a mean reversion mechanism to correct for shocks. That is, going from negative growth rates in the previous years to a positive growth rate in 2014 implies very sizable positive residual which leads to the ARMA(1,1) prediction that the growth rate for Spain will drop significantly in 2015 as such a big residual should not be expected in the subsequent year as well. For 2015 this turned out to be untrue however, as the growth rate in Spain sky-rocketed to 3,4% according to the World Bank.

Table 4: Predictions for 2015 growth in real GDP per capita, Denmark and Spain

|	DyDNK_pred	DyESP_pred1	DyESP_pred
Country	Denmark	Spain	Spain
Model	AR(1)	ARMA(1,1)	AR(1)
Time dummies	no	no	no
mean	2.04	0.71	2.79

f) Plot both the actual and predicted time series for each country

Figure 9, 10, and 11 performs visual evaluations of the predictive power 'out-of-sample' of each models by plotting the values that each model would predict year-by-year

against the actual values. Furthermore, growth in real GDP per capita is forecasted the coming 10 years to see the inner workings of the models in the absence of new shocks.

Figure 9 evaluate the different models for Germany. While including year dummies for negative chocks clearly increases within-sample predictability, the downfall becomes very evident when predicting out of sample as well. The first panel show the predictions of the ARMA(1,4) model estimated with and without time dummies. The predictions of the dummy-model initially do a better job at capturing the full size of the fluctuations but then it seems that the oil-crisis of 1974-75, that is no longer controlled for, sparks a wild fluctuation where the model alternates between very excessive extremes. In panel two of 9 the ARMA(1,1) model estimated using dummies does a much better job. As seen in table 2 the estimate of $\hat{\theta}_1$ becomes positive as time dummies prevents the model from over-shooting too much within-sample. However, out-of-sample the model basically acts as a linear extrapolation of the tendency in the period before, thus consistently 'overshoots' and fails to predict when business cycles turn. On the contrary the considerable size of the negative parameter $\hat{\theta}_1$ in the ARMA(1,1) model estimated without dummies overpredicts the mean-reversion to a degree where it almost does not capture any fluctuations but mostly corrects for, namely, a moving average of the time series.

The ARMA(1,1) model estimated using time dummies predicts that the growth rate in real GDP per capita converges towards no less than 3.4& in 2014. While the trends of convergence are much more modest for the without-dummy estimations. The ARMA(1.4) and ARMA(1,1) models predict the growth rate in 2024 to be 2.0 and 1.9 respectively. This unrealistic level of optimism in the forecast is the result of both removing negative shocks from the training-sample of the model but also thereby reducing persistence of slumps.

Figure 9: Actual vs predicted growth in real GDP per capita, Germany
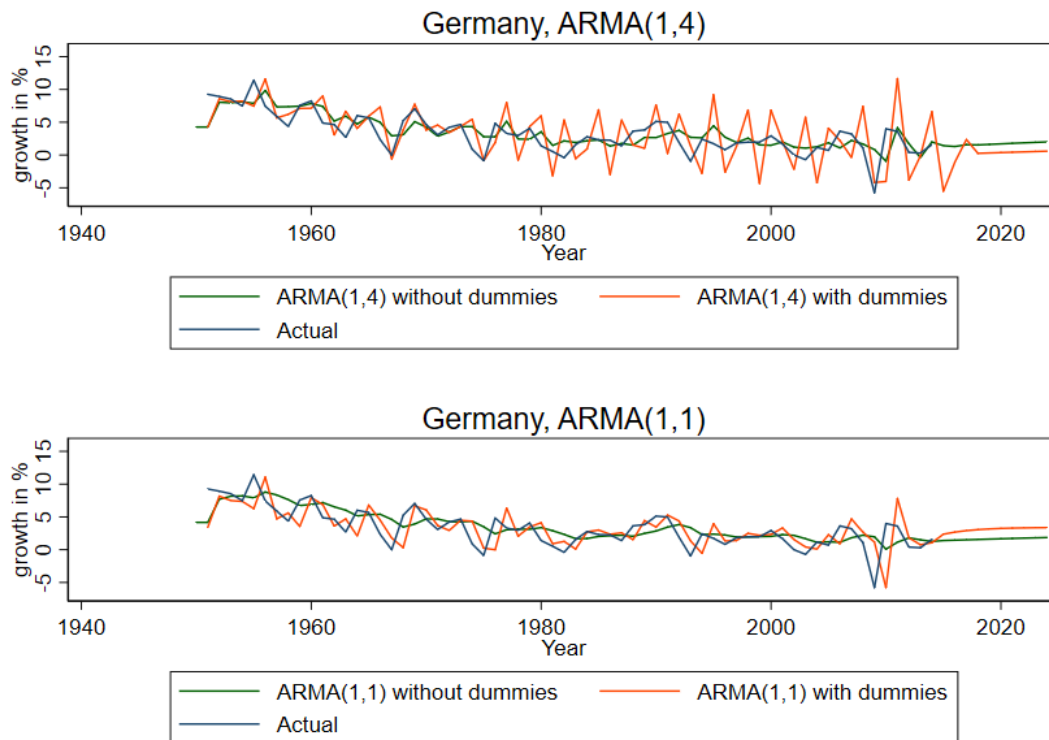
It is clear that it is preferable not to use time dummies, at least not as extensively but only for an event that actually with high confidence is not expected to happen in a similar magnitude ever again. When deciding between the ARMA(1,4) and the ARMA(1,1) models both estimated without dummies, the choice in the end becomes subjective and depends on which qualities are appreciated as seen in figure 10 below. Recalling table 2 the ARMA(1,4) has the lower value of the AIC as she model better fits the magnitude of business cycles when guessing right. On the downside, the size of wrong predictions and overshooting are more extreme as well. Literally and figuratively the predictive power of the ARMA(1,1) has both higher highs and lower lows. As an example, for 1995 the optimistic forecast was that the growth rate would go up by 1.8 percentage points. Instead it fell by 0.7 percentage points. Such a sizable prediction error could put an institution in a bad light. Thus, an economist that does not want to run such a risk could instead take the more cautious approach of leaning more towards the BIC that points towards the ARMA(1,1) model. This is reflected in the variance of the predicted time series being much smaller than for the actual time series. However, the model looses relevance if it at a too high degree just becomes an indication of the average of the growth rate in the recent years.

Figure 10: Actual vs predicted growth in real GDP, Germany
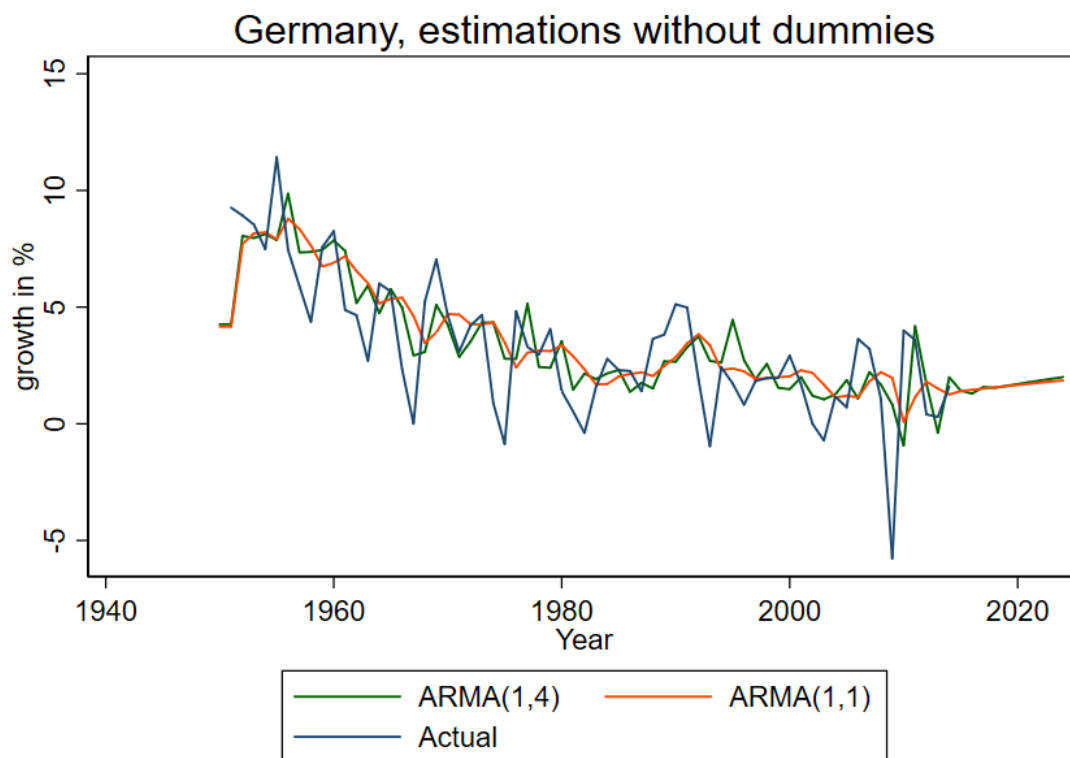
In figure 11 for Denmark and Spain we see that the ARMA(1,1) model for Spain naturally only does a moderately good job at predicting when a boom or bust will end. But even more so, the AR(1) model for both Denmark and Spain has the mechanic of simply extrapolating the growth rate in the former period. Both model types will cause prediction error whenever a business cycle turns which leads to smaller parameter sizes to reduce the size of these prediction errors weighted against the cost of deliberately undershooting when, after all, predicting the right direction for most of the periods.

Figure 11: Actual vs predicted growth in real GDP per capita, Denmark and Spain
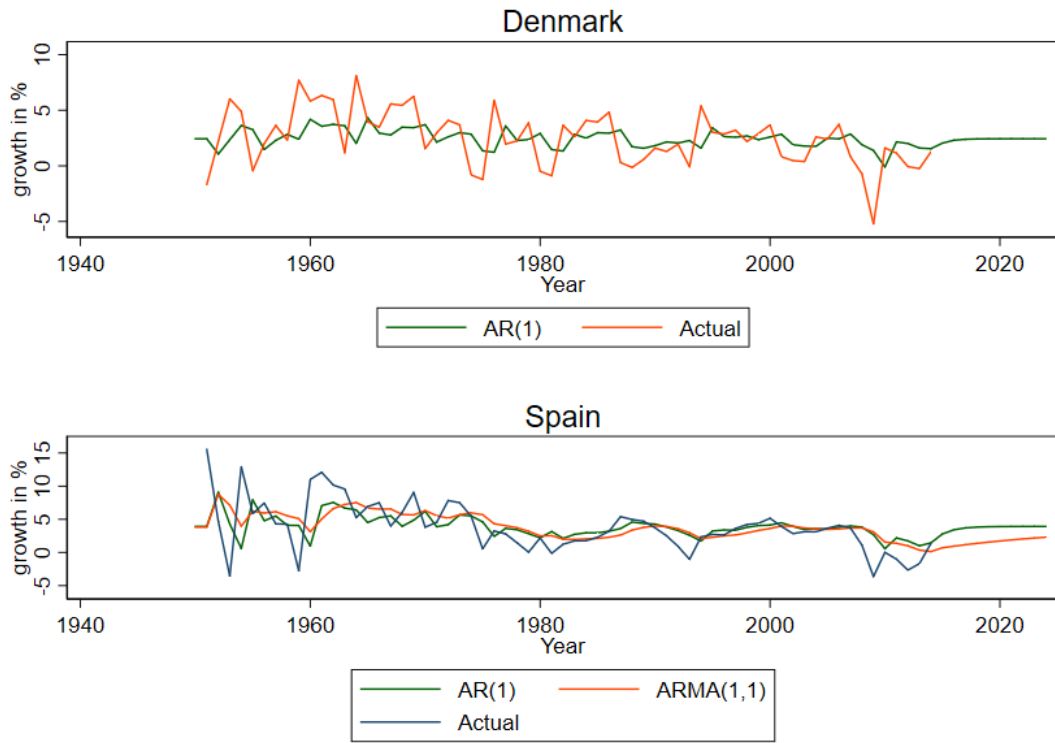
2.4. Order of integration analysis

Following the Dickey-Pantula strategy, determine the order of integration of the real per capita GDP time series that you have selected.

In both cases, describe how the order of integration analysis have been performed - null and alternative hypothesis of the different test statistics and the specification of the deterministic component, among other features.

Following the Dickey-Pantula strategy we should first test the more general situation that the series of log real GDP per capita is integrated of order 2 by testing for more than one unit root.

$$
\begin{cases}
H_0 : \log GDPpc_t \sim I(2) \\
H_1 : \log GDPpc_t \sim I(1) \text{ or } \sim I(0)
\end{cases}
\tag{2.8}
$$

A feasible way of testing this is to apply the 1st order transformation as equation 2.8 is equivalent to

$$
\begin{cases}
H_0 : \Delta \log GDPpc_t \sim I(1) \\
H_1 : \Delta \log GDPpc_t \sim I(0)
\end{cases}
\tag{2.9}
$$

Thus, if we cannot reject the H_0 hypothesis 2.9 that $\Delta \log GDPpc_t$ contains a unit root then equivalently we cannot reject the null-hypothesis 2.8 that $\log GDPpc_t$ would have more than one unit root. On the contrary, if we are able to reject that $\Delta \log GDPpc_t$ contains a unit root it should be tested whether $\log GDPpc_t$ has a single unit root or not.

a) *Analyze the order of integration of* $\log(GDPpc_t)$ *with a deterministic component defined by a constant term using, at least, the ADF, PP, ADF-GLS and KPSS test statistics.*

The Dickey-Fuller (DF) test assumes a first-order autoregressive process

$$x_t = f(t) + \phi x_{t-1} + \varepsilon_t \tag{2.10}$$

Here we first define the deterministic component as a constant term $f(t) = \mu$. Taking the first-difference of this would suppress the deterministic term such that $f(t) = 0$. The null-hypothesis is that the time series contain a unit-root against the alternative of no unit-root such that

$$\begin{cases} H_0 : |\phi| = 1 \equiv x_t \sim I(1) \\ H_1 : |\phi| < 1 \equiv x_t \sim I(0) \end{cases} \tag{2.11}$$

Which is decided on by the computation of the normalized bias test (2.12) and the pseudo t-ratio test statistics (2.13)

$$T(\hat{\phi} - 1) \tag{2.12}$$

$$t_{\hat{\phi}} = \frac{\hat{\phi} - 1}{std.err.(\hat{\phi})} \tag{2.13}$$

The DF test relies on the strong assumption of no autocorrelation in the disturbance term. This assumption can be loosened by either a parametric correction, the Augmented Dickey-Fuller (ADF) test, or a non-parametric correction, the Philips-Perron (PP) test. The ADF test tests for unit root in the regression equation

$$\Delta x_t = f(t) + \alpha x_{t-1} + \sum_{j=1}^{p-1} \gamma_j \Delta x_{t-j} + \varepsilon_t \tag{2.14}$$

The Philips-Perron (PP) test builds on the Dickey-Fuller test in equation 2.10 as well but performs a non-parametric correction of the DF test statistics. However, under the presence of Moving Average components in the disturbance term, the ADF test is less distorted than the PP test in finite samples.

To provide better results in finite samples an ADF-GLS test was proposed by Ng and Perron (1998) by modifying the PP test. Finally Kwiatkowski, Philips, Schmidt, and Shin

(1992) proposed a Lagrange multiplier test given by

$$KPSS = \frac{T^{-1}\sum_{t=1}^{T} S_t^2}{s^2} \tag{2.15}$$

As opposed to the prior tests, the KPSS test is a test for level stationarity such that the null-hypothesis and the alternative are

$$\begin{cases} H_0 : x_t \sim I(d-1) \\ H_1 : x_t \sim I(d) \end{cases} \tag{2.16}$$

Here, the order of integration in the alternative hypothesis is $d = 2$ for log real GDP per capita, equivalent to $d = 1$ for the first-differences. The size of the KPSS test statistics, like for the PP test, is distorted in the presence of MA components.

All four tests are applied to test the order of integration of the 1st order transformed of log real GDP per capita with a deterministic component given by a constant term. Taking the first-differences makes the constant terms cancel out, however, as the ADF-GLS nor the KPSS test implementation in Stata allows for suppressing the deterministic term, these are estimated including a constant term. Table 5 below show for each country the number of included lagged difference terms p for which the H_0 hypothesis of each test is rejected due to the test statistic being above the 5% critical value. For the ADF-GLS and the KPSS tests the Schwert criterion (1989) is applied by default such that the sample size implies that a maximum of $p = 10$ lags are included in the tests. Recalling the ACFs in figure 7 the autocorrelation of lags $p > 10$ approaches zero. To be on the safe side, for the ADF and PP the tests are performed using each number in $p \in 1, 2, \dots, 10$ as maximum lags as well.

It is evident that the bandwidth of the spectral window heavily affects the performance of the test, thus in the tables it is denoted which is the optimal number of included lags in the ADF-GLS test according to the Ng-Perron sequential t (seq t), the minimum Schwarz information criterion (SIC), and the Ng-Perron modified Akaike information criterion (MAIC) respectively. Likewise, for the KPSS test the optimal number of included lags is calculated based on the automatic bandwidth selection proposed by Newey and West (1994).

Table 5: Order of integration test results for growth in real GDP per capita

Country	Deterministic constant	Test	H_0 rejected	seq t	SIC	MAIC	Newey & West
Germany	no	ADF	w. 1-10 lags				
Germany	no	PP	w. 1-10 lags				
Germany	yes	ADF-GLS	-	3	3	3	
Germany	yes	KPSS	w. 1-10 lags				5
Denmark	no	ADF	w. 1 lag				
Denmark	no	PP	w. 1-10 lags				
Denmark	yes	ADF-GLS	w. 1-2 lags	2	2	6	
Denmark	yes	KPSS	w. 1-8 lags				5
Spain	no	ADF	w. 1, 10 lags				
Spain	no	PP	w. 1-10 lags				
Spain	yes	ADF-GLS	-	10	2	2	
Spain	yes	KPSS	w. 1-10 lags				5

Maximum number of included lags $p = 10$ in accordance with the Schwert criterion.

Table 5 indeed emphasizes that one should not rely on a single test or specification of the number of included lagged difference terms. For Germany the H_0 that $\Delta \log GDPpc_t$ contains a unit root is not rejected by the ADF-GLS test but is rejected by the ADF and the PP test. The contradicting test results could be due to disturbance from the Moving Average components in the disturbance term that we indeed found to be of significant relevance in section 2.2 as the ARMA(1,4) model is the better fitting model. After all the ADF-GLS test is generally regarded the best performing of the three in finite samples, thus, we conclude that we cannot reject that $\Delta \log GDPpc_t$ contains a unit root for Germany. Even more so, this is consistent with the KPSS test rejecting that the series is level-stationary.

Similarly, for Spain we found the ARMA(1,1) model to be the best performing which can be the reason that the PP test clearly rejects the H_0. As could be expected the MA components affect the ADF test less than the PP test, nonetheless, H_0 is still rejected when the ADF test specification includes a total of either 1 or 10 lagged difference terms. Like for Germany, the ADF-GLS and the KPSS test statistics are consistent with the series for growth in real GDP per capita in Spain containing a unit-root.

This is equivalent with the series of log real GDP per capita containing more than one unit root for Germany and Spain. Recalling the parameter estimates in table 1 and 2 the AR parameter $\hat{\phi}_1$ is indeed not found to be significantly different from unity in the ARMA(1,1) model for Spain nor any estimation of the ARMA(1,4) model for Germany or in the ARMA(1,1) model for Germany estimated without dummies.

In contrast, For Denmark estimates and tests carried out in section 2.2 found the AR(1) model to give the better fit of growth in real GDP per capita and with an parameter estimate $\hat{\phi} = 0.3$ clearly different from zero. Indeed the ADF-GLS test is able to reject the H_0, at least when including only 1 or 2 lagged difference terms. I decide to reject the

H_0 of a unit root, as it is optimal to include exactly 2 lags in the ADF-GLS test according to both the Ng-Perron sequential t and the minimum Schwarz information criterion. The lesser relevance of the MA components for Denmark increases the credibility of the PP and ADF tests where the former clearly rejects the H_0 of a unit root while the latter only rejects it when including a single lagged difference term. The conclusion is less consistent with the KPSS test that rejects stationarity, however, not when including 9 or 10 lagged difference terms in the test specification.

This is equivalent with rejecting that the non-transformed series for log real GDP per capita in Denmark contains two or more unit roots. Following the Dickey-Pantula strategy we should proceed to test whether the series contain 1 or 0 unit roots, that is

$$\begin{cases} H_0 : \log GDPpc_t \sim I(1) \\ H_1 : \log GDPpc_t \sim I(0) \end{cases} \tag{2.17}$$

Letting the deterministic component be given by a constant term we apply the same tests in table 6. The ADF-GLS and the KPSS tests are consistent with $\log GDPpc_t$ not being a stationary but a stochastic process containing a unit root when the deterministic component is restricted to be a constant term. Even the PP test does not reject the H_0 of a unit root according to the Philips-Perron ρ test statistic, however, it is rejected according to Philips-Perron τ test statistic and by the ADF test when including 3, 4, 9, or 10 lagged difference terms. The size of the PP and ADF tests statistics can again be disturbed by MA components in the disturbance term.

Table 6: Order of integration test results for log real GDP per capita, with a constant term

Country	1st order transformed	Test	H_0 rejected	seq t	SIC	MAIC	Newey & West
Denmark	no	ADF	w. 3,4,9,10 lags				
Denmark	no	PP	(w. 1-10 lags)*				
Denmark	no	ADF-GLS	-	3	2	3	
Denmark	no	KPSS	w. 1-10 lags				6

Maximum number of included lags $p = 10$ in accordance with the Schwert criterion.
*H_0 rejected by the Philips-Perron τ test statistic, but not by the Philips-Perron ρ test statistic.

b) **Analyze the order of integration of** $\log(GDPpc_t)$ **with a deterministic component defined by a linear time trend using, at least, the ADF, PP, ADF-GLS and PSS test statistics.**

In the following I test the order of integration of the series for log real GDP per capita with a deterministic component given by a linear time trend

$$f(t) = \mu + \beta t \tag{2.18}$$

Testing whether the non-transformed time series contains two or more unit roots is equivalent to testing whether the 1st order transformed contains a single unit root. Taking the first-differences implies transforming the deterministic component with a trend

term as given by equation 2.18 to a constant term, thus, this is exactly the same test specifications for the ADF-GLS and the KPSS tests for which the results are shown in table 5. Moreover, the time series for log real GDP per capita in Germany and Spain contain a unit root both with and without a constant term. Though the ADF and PP tests sizes are likely to be distorted, it is worth considering the test results in table 7 where a deterministic constant term is included in their specification. The PP tests are still rejected for all numbers of included lags across the countries. The H_0 of the ADF test is now only rejected for Germany when the test specification includes a total of exactly 1, 4, or 9 lagged difference terms and for Spain only for exactly one lagged difference term. For Denmark it is now also rejected for 2 lagged difference terms.

Table 7: Order of integration test results for growth in real GDP per capita, with a constant

Country	Deterministic constant	Test	H_0 rejected
Germany	yes	ADF	w. 1, 4, 9 lags
Germany	yes	PP	w. 1-10 lags
Denmark	yes	ADF	w. 1-2 lags
Denmark	yes	PP	w. 1-10 lags
Spain	yes	ADF	w. 1 lag
Spain	yes	PP	w. 1-10 lags

Maximum number of included lags $p = 10$ in accordance with the Schwert criterion.

On a close to identical basis I reject that the non-transformed series for Danish log real GDP per capita with a time trend contains two or more unit roots. I follow the Dickey-Pantula strategy and apply the ADF, PP, and ADF-GLS tests of the null-hypothesis that the series contains a unit root, that is, can be described by a random walk. The table 8 show that the H_0 cannot be rejected by either unit-root test while stationarity around a linear trend is clearly rejected by the KPSS test. The unambiguity of the different tests is in contrast to the results in table 6 where the test specifications were constrained to not include a deterministic component with a linear trend term. In conclusion, of the tested specifications the time series of log real GDP per capita in Denmark can best be described by a random walk with a positive drift and is integrated of order 1, that is, the first-order differentiated is a stationary series with a constant term.

Table 8: Order of integration test results for log real GDP per capita, with a time trend

Country	1st order transformed	Test	H_0 rejected	seq t	SIC	MAIC	Newey & West
Denmark	no	ADF	-				
Denmark	no	PP	-				
Denmark	no	ADF-GLS	-	1	1	1	
Denmark	no	KPSS	w. 1-10 lags				5

Maximum number of included lags $p = 10$ in accordance with the Schwert criterion.

A. APPENDIX

Figure A.1: Residuals from modelling GDP growth in Germany, without time dummies
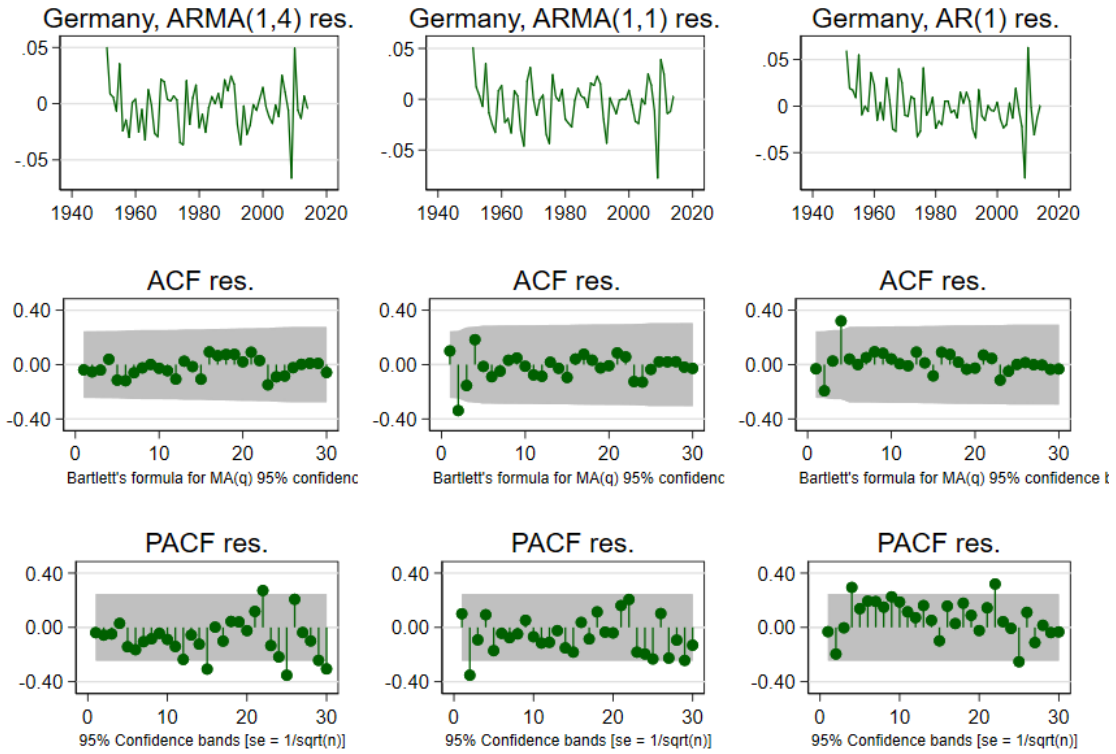
Figure A.2: Residuals from modelling GDP growth in Germany, using time dummies