EXAM - Econometrics 2

Exam Number: 241, 244, 235

Fall Semester 2020

Department of Economics

University of Copenhagen

Examinee 244 Examinee 241 Examinee 235

December 14, 2020

Character count

- Assignment 5 - 16471
- Assignment 1 - 11993
- Assignment 3 - 11862

Who wrote what

Assignment Exam
244: (1.1, 1.2, 1.3, 1.5, 2.5, 2.6)
241: (1.2, 1.3, 1.4, 2.1, 2.3)
235: (1.4, 2.2, 2.3, 2.4, 2.6)

Assignment 1
244: 3. Theory: "Our model" and "the estimator and it's properties", 4. Empirical analysis: "comparison to linear forecast" , 5. Conclusion, 6. Figure 2
241: 1. Introduction, 3. Theory: "AR model", "ARMA model", "Point Estimate and forecast error variance", "Model assumptions", "Diebold-Mariano-test", 6. Appendix: Figure 2, Figure 3
235: 2. Description of data, 3. Theory: "Misspecification tests", 4. Empirical analysis: "Empirical results", "Model selection", "Interpretation of the empirical results", 6. Figure 1 and Figure 3, Table 1

Assignment 3
244: 2. Description of the data, 4. Empirical Analysis, 5. Discussion and Conclusion, 6. Figure 2 and Figure 3
241: 3. Theory: entire section, excluding "Misspecification tests, 6. Appendix: all tables
235: 1. Introduction, 3. Theory: "Misspecification tests", 6. Figure 1 and Figure 3

1 GMM Estimation of Moving Average Models

1.1

(1) A time series $\{y_t\}_{t=1}^{T}$ is weakly stationary, given the following conditions (p.88, Nielsen, 2020[1]):

$$\begin{aligned} E(y_t) = & & = \mu_t \\ V(y_t) = E((y_t-\mu)^2) & & = \gamma_0 \\ cov(y_t, y_{t-i}) = E((y_t-\mu)(y_{t-i}-\mu)) & & = \gamma_i \quad \text{for } i \in \mathbb{N} \end{aligned} \tag{1}$$

for all values of t. All these expressions are derived in the following parts (**(2)-(4)**.

(2) The MA(2) model is stated as:

$$y_t = \epsilon_t + \alpha_1\epsilon_{t-1} + \alpha_2\epsilon_{t-2} \quad \epsilon_t|I_{t-1} \stackrel{d}{=} N(0,\sigma^2), \quad \sigma^2 > 0 \tag{2}$$

We derive:

$$\begin{aligned} E(y_t) &= E(\epsilon_t + \alpha_1\epsilon_{t-1} + \alpha_2\epsilon_{t-2}) \\ &= E(\epsilon_t) + \alpha_1 E(\epsilon_{t-1}) + \alpha_2 E(\epsilon_{t-2}) \quad \text{with } \epsilon_t \text{ i.i.d.} \\ &= 0 + \alpha_1 0 + \alpha_2 0 \quad \text{with LIE} \\ &= 0 \end{aligned} \tag{3}$$

Where the law of iterated expectations (LIE) $E[\epsilon_t] = E[E(\epsilon_t|\mathbf{I}_{t-1})]$ and $\epsilon_t|\mathbf{I}_{t-1}\ N(0,\sigma^2)$

(3)

$$\begin{aligned} \gamma_0 &= E[(y_t-\mu)^2] \qquad \text{with } \mu = E(y_t) = 0 \\ \gamma_0 &= (1+\alpha_1^2+\alpha_2^2)\sigma^2 \end{aligned} \tag{4}$$

With ϵ_t i.i.d, the LIE and the conditional distribution of ϵ_t

(4)

$$\begin{aligned} \gamma_i &= E[(y_t-\mu)(y_{t-i}-\mu)] \qquad \text{with } \epsilon_t \text{ iid} \\ &= E(\epsilon_t\epsilon_{t-i}) + \alpha_1 E(\epsilon_{t-1}\epsilon_{t-1-i}) + \alpha_2 E(\epsilon_t\epsilon_{t-2-i}) \\ &+ \alpha_1^2 E(\epsilon_{t-1}\epsilon_{t-1-i}) + \alpha_1 E(\epsilon_t\epsilon_{t-i}) + \alpha_1\alpha_2 E(\epsilon_{t-1}\epsilon_{t-2-i}) \\ &+ \alpha_2 E(\epsilon_{t-2}\epsilon_{t-i}) + \alpha_1\alpha_2 E(\epsilon_{t-2}\epsilon_{t-1-i}) + \alpha_2^2 E(\epsilon_{t-2}\epsilon_{t-2-i}) \end{aligned} \tag{5}$$

The expected values are equal to 0 when the subscripts of ϵ do not match ($E(\epsilon_t\epsilon_s) = 0 \forall s \neq t$). So we get

$$\gamma_i = \begin{cases} \sigma^2 + \alpha_1^2\sigma^2 + \alpha_2^2\sigma^2 & \text{if } i = 0 \\ \alpha_1\sigma^2 + \alpha_1\alpha_2\sigma^2 & \text{if } i = 1 \\ \alpha_2\sigma^2 & \text{if } i = 2 \\ 0 & \text{if } i > 2 \end{cases} \tag{6}$$

[1] Nielsen, H.B, (2020).*A Course in Time Series Econometrics.*University of Copenhagen.

1.2

See also chapter 11.3. by Nielsen (2020[2]). We use the moments γ_i derived in part 1.1 to formulate moment conditions $g(\theta) = E(f(\cdot)) = 0$ with $\hat{\gamma}$ and γ being $R \times 1$ vectors containing the empirical, respective theoretical, moments of y_t. As show in the equation above, γ_i is equal to 0 for all $i > 2$ and can therefore not be used to set a moment condition, since their theoretical value is zero. With the remaining $\gamma_0, \gamma_1, \gamma_2$ we can set up three moment conditions $g(\theta)$, depending on $\theta = (\alpha_1, \alpha_2, \sigma^2)'$:

Moment conditions:

$$g(\theta_0) = E[f(y,\theta)] = \begin{pmatrix} g_0 \\ g_1 \\ g_2 \end{pmatrix} = \begin{pmatrix} E(f_0(y,\theta)) \\ E(f_1(y,\theta)) \\ E(f_2(y,\theta)) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{7}$$

With f_i being defined as:

$$\begin{aligned} f_0(y,\theta) &= y_t^2 - \sigma^2(1+\alpha_1^2+\alpha_2^2) &&= y_t^2 - \gamma_0 \\ f_1(y,\theta) &= y_t y_{t-1} - (\alpha_1\sigma^2 + \alpha_1\alpha_2\sigma^2) &&= y_t y_{t-1} - \gamma_1 \\ f_2(y,\theta) &= y_t y_{t-2} - \alpha_2\sigma^2 &&= y_t y_{t-2} - \gamma_2 \end{aligned} \tag{8}$$

Thereof we derive the sample conditions:

$$g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} f(y,\theta) \tag{9}$$

The empirical solution for $g_T(\theta)$ is defined as:

$$\begin{aligned} g_{0,T}(\theta) &= \frac{1}{T}\sum_{t=1}^{T} f_0(y,\theta) &&= \frac{1}{T}\left(\sum_{t=1}^{T}(y_t^2) - \sigma^2(1+\alpha_1^2+\alpha_2^2)\right) &&= \hat{\gamma}_0 - \gamma_0 \\ g_{1,T}(\theta) &= \frac{1}{T}\sum_{t=1}^{T} f_1(y,\theta) &&= \frac{1}{T}\left(\sum_{t=1}^{T}(y_t y_{t-1}) - (\alpha_1\sigma^2 + \alpha_1\alpha_2\sigma^2)\right) &&= \hat{\gamma}_1 - \gamma_1 \\ g_{2,T}(\theta) &= \frac{1}{T}\sum_{t=1}^{T} f_2(y,\theta) &&= \frac{1}{T}\left(\sum_{t=1}^{T}(y_t y_{t-2}) - (\alpha_2\sigma^2)\right) &&= \hat{\gamma}_2 - \gamma_2 \end{aligned} \tag{10}$$

That's why we follow the intuition of minimising the sum of squares $g_T(\theta)'g_T(\theta)$. We account for the scaling of the moments by using a positive definite weight matrix $W_{T R\times R}$ that attaches a weight to each moment. (p.261, Nielsen,2020[3])).W_T should be positive definite, as negative weights would not turn the minimization problem into a maximization (which we do not need). Full rank in W_T ensures that no information from $g_T(\theta)$ is spilled or excluded. For simplicity, we use the identity or unity matrix of dimension R=3 (R = number of moment conditions), and weigh all conditions equally. This weighted sum of squares is called the quadratic form or also the criteria function Q_T:

$$\begin{aligned} Q_T(\theta) &= g_T(\theta)'W_T g_T(\theta) \\ &= \begin{pmatrix} g_0 & g_1 & g_2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} g_0 \\ g_1 \\ g_2 \end{pmatrix} \\ &= g_0^2 + g_1^2 + g_2^2 \end{aligned} \tag{11}$$

$$\hat{\theta}_{GMM}(W_T) = arg\ min(\theta)\quad \{g_T(\theta)'W_T g_T(\theta)\}$$

[2]Nielsen, H.B, (2020).*A Course in Time Series Econometrics.*University of Copenhagen.
[3]Nielsen, H.B, (2020).*A Course in Time Series Econometrics.*University of Copenhagen.

With use of the identity matrix, we can calculate a first estimate of $\hat{\theta}_{GMM}$ which is most likely far from optimal, as it ignores potential interaction between g_0, g_1, g_2. This first estimator can now be used however through $g_T(\hat{\theta}_{GMM})$ to obtain a better weight matrix, with wich we can estimate $\hat{\theta}_{GMM}$ once again thus performing a "2-step GMM", or multiple times as an "iterated GMM" until we find convergence in $\hat{\theta}_{GMM}$.

1.3

We decide to use an MM approach to find closed form solutions. From part 1.2, we have the following moments:

$$\begin{aligned}\gamma_0 &= \sigma^2(1+\alpha_1^2+\alpha_2^2)\\ \gamma_1 &= \alpha_1\sigma^2 + \alpha_1\alpha_2\sigma^2\\ \gamma_2 &= \alpha_2\sigma^2\end{aligned} \tag{12}$$

As $\hat{\sigma}^2$ is given, we only have to estimate α_1 and α_2. As we are formulating a MM, we only use 2 out of the 3 moments, namely γ_1 and γ_2. We derive $g_T(\theta)_{2\times 1} = (g_1, g_2)'$, starting with $g_{2,T}$ first, as it has only 1 unknown parameter

$$\begin{aligned}g_2 &= \frac{1}{T}\sum_{t=1}^{T}\left(y_t y_{t-2} - \hat{\alpha}_2\hat{\sigma}^2\right) = 0\\ \hat{\alpha}_2\hat{\sigma}^2 &= \frac{1}{T}\sum_{t=1}^{T}(y_t y_{t-2})\\ \Rightarrow \hat{\alpha}_2 &= \frac{1}{\hat{\sigma}^2 T}\sum_{t=1}^{T}(y_t y_{t-2})\end{aligned}$$

$$\begin{aligned}g_1 &= \frac{1}{T}\sum_{t=1}^{T}\left(y_t y_{t-1} - \hat{\alpha}_1\hat{\sigma}^2 - \hat{\alpha}_1\hat{\alpha}_2\hat{\sigma}^2\right) = 0\\ \frac{1}{T}\sum_{t=1}^{T}(y_t y_{t-1}) &= \hat{\alpha}_1\hat{\sigma}^2 + \hat{\alpha}_1\hat{\alpha}_2\hat{\sigma}^2 \qquad \hat{\alpha}_2 = (\cdot) \text{ from } g_2\\ \frac{1}{T}\sum_{t=1}^{T}(y_t y_{t-1}) &= \hat{\alpha}_1\hat{\sigma}^2 + \hat{\alpha}_1\frac{\hat{\sigma}^2}{\hat{\sigma}^2 T}\sum_{t=1}^{T}(y_t y_{t-2})\\ \frac{1}{T}\sum_{t=1}^{T}(y_t y_{t-1}) &= \hat{\alpha}_1\left(\hat{\sigma}^2 + \frac{1}{T}\sum_{t=1}^{T}(y_t y_{t-2})\right)\end{aligned} \tag{13}$$

$$\Rightarrow \hat{\alpha_1} = \frac{\frac{1}{T}\sum_{t=1}^{T}(y_t y_{t-1})}{\hat{\sigma}^2 + \frac{1}{T}\sum_{t=1}^{T}(y_t y_{t-2})}$$

To utilize all available information we can also make use of the third moment mentioned in part 1.2 and formulate a GMM. As shown previously in the last equation of part 1.2, we minimize the criteria function $Q_T(\theta)$ in θ. Thus a closed-form estimator can be derived by taking the first order condition. As $\hat{\sigma}^2$ is given for σ^2, we only have to solve for the parameters α_1 and α_2:

$$\begin{aligned}&\frac{\partial Q_T(\theta)}{\partial \theta'} = 0\\ \Longrightarrow\ &\frac{\partial Q_T(\theta)}{\partial \alpha_1} = 0; \quad \frac{\partial Q_T(\theta)}{\partial \alpha_2} = 0\end{aligned} \tag{14}$$

1.4

We use 3 moment conditions $g(\theta_0)$ and their sample conditions $g_{0,T}, g_{1,T}, g_{2,T}$, mentioned in part 1.2, to estimate 3 parameters $\alpha_1, \alpha_2, \sigma^2$, using the GMM module in OxMetrics. As shown in part 1.2, the first and second lag of y_t is needed in the conditions, and therefore also included in the setup of the model. We continue to formulate an iterated GMM, using heteroscedastic robust standard errors (HCSE). As it is a non-linear model, we choose the following formulation, shown in figure 1:

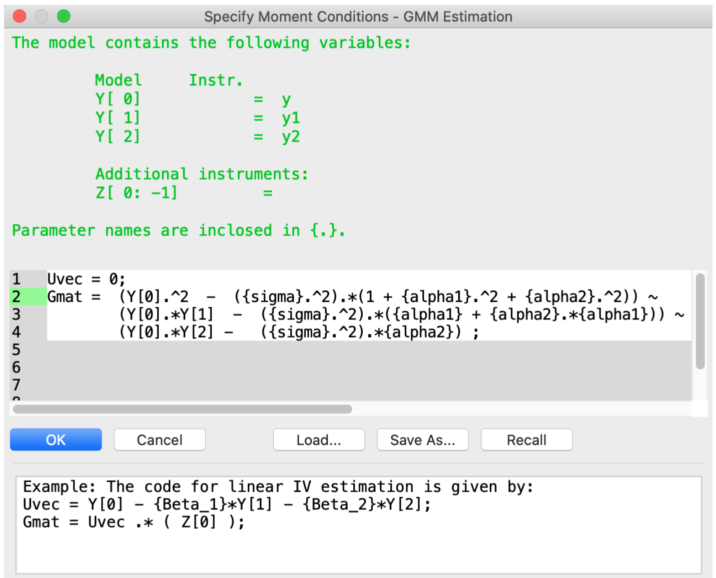

Figure 1: Formulation of the non-linear model in the GMM package

OxMetrics Output		
	ARFIMA(0,0,2)	**GMM**
MA-1 / α_1	0.485 (15.9)	0.481 (12.2)
MA-2 / α_2	0.264 (8.43)	0.229 (5.18)
$\hat{\sigma}$	1.041	1.046 (35.3)
Log-lik.	-1459.082	$\cdot$
Criteria function (Q_T)	$\cdot$	0.000
AIC	2.924	$\cdot$
HQ	2.930	$\cdot$
SC/BIC	2.939	$\cdot$
No autocorr. 1-31	[0.88]	$\cdot$
No hetero.	[0.32]	$\cdot$
Normality	[0.23]	$\cdot$
J-statistic ($T \cdot Q_T$), $\chi^2(0)$	$\cdot$	0.00
Parameters (K)	$\cdot$	3
Moment Conditions (R)	$\cdot$	3
Initial Weight Matrix (W_T)	$\cdot$	Unit-matrix (I_R)

Table 1: The table shows estimates of the MA(2) model using Gaussian maximum likelihood (ARFIMA(0,0,2)) and GMM respectively. t-ratios in $(\cdot)$ and p-values in $[\cdot]$ for misspecification tests.

Looking at the results in table 1, we see that GMM provides estimates for $\alpha_1, \alpha_2, \sigma^2$ which are very close to the estimates provided by the ARFIMA(0,0,2). This shows that an MA model can be stated as a special case of a GMM. We are further confident that the parameters of GMM are well estimated, as the value of the criteria function, which should be minimized, is actually 0 (after enough iterations when starting with the unity-matrix as an initial weight matrix).

To estimate the ARFIMA(0,0,2) we use an maximum likelihood estimator (MLE), which is explained in detail in section "The estimator and its properties" of assignment 1 (assignment1, 2020[4]).

1.5

The true data generating process y_t is an MA(1). The model we are trying to estimate is an AR(1). An MA(1) can be reformulated into an AR(∞) when the MA-Parameter is below 1 in absolute values.

$$\begin{aligned} y_t^{TRUE} &= \epsilon_t + \alpha_1 \epsilon_{t-1} && (15)\\ y_t^{TRUE} &= \epsilon_t + \alpha_1 (y_{t-1} - \epsilon_{t-2}) && (16)\\ &\vdots && (17)\\ y_t^{TRUE} &= \epsilon_t + \alpha_1 y_{t-1} - \alpha_1^2 y_{t-2} + \alpha_1^3 y_{t-3} - \alpha_1^4 y_{t-4} + \dots - \dots && (18) \end{aligned}$$

So we effectively try to estimate an AR(∞) with an AR(1) instead:

$$y_t^{MODEL} = \phi y_{t-1} + \eta_t \tag{19}$$

where ϕy_{t-1} corresponds to $\alpha_1 y_{t-1}$, so that $\phi = \alpha_1$. η_t the model residual is the whole rest of y_t^{TRUE}:

$$\eta_t = \epsilon_t - \alpha_1^2 y_{t-2} + \alpha_1^3 y_{t-3} - \dots + \dots \tag{20}$$

We find η_{t+1} the same way:

$$\eta_{t+1} = \epsilon_{t+1} - \alpha_1^2 y_{t-1} + \alpha_1^3 y_{t-2} - \dots + \dots \tag{21}$$

Some thoughts on the covariance: Because the structure of y_t is an MA(1) we have no autocovariance for any lag > 1 (analogous to the first part of task 5.1, but here with an MA(1) not MA(2)). For a 1 lag difference the autocovariance is $\alpha_1 \sigma^2$.

$$Cov(y_t, y_{t-1}) = \alpha_1 \sigma^2 \tag{22}$$

and the variance is:

$$Cov(y_t, y_t) = (1 + \alpha_1^2)\sigma^2 \tag{23}$$

Autocovariance of order 1 You can now think of each y_i within η_t having 3 partners in η_{t+1} that are y_{i-1}, y_i, y_{i+1} for the other terms the covariance is zero. We can split the covariance of η_t and η_{t+1} into the sum of covariances of each summand in η_t and η_{t+1}.

$$Cov(\eta_t, \eta_{t+1}) = Cov(\epsilon_t, \eta_{t+1}) + Cov(-\alpha_1^2 y_{t-2}, \eta_{t+1}) + Cov(\alpha_1^3 y_{t-3}, \eta_{t+1}) + \dots \tag{24}$$

with:

$$Cov(\epsilon_t, \eta_{t+1}) = 0 \tag{25}$$

and

$$\begin{aligned} Cov(-\alpha_1^2 y_{t-2}, \epsilon_{t+1} - \alpha_1^2 y_{t-1} + \alpha_1^3 y_{t-2} - \alpha_1^4 y_{t-3} + \dots) &= && (26)\\ \alpha_1^4(\alpha_1\sigma^2) - \alpha_1^5((1+\alpha_1^2)\sigma^2) + \alpha_1^6 \alpha_1 \sigma^2 &= 0 && (27) \end{aligned}$$

[4](2020).*Econometrics II - Assignment 1.* Fall course 2020.

and

$$Cov(\alpha_1^3 y_{t-3}, \dots + \alpha_1^3 y_{t-2} - \alpha_1^4 y_{t-3} + \alpha_1^5 y_{t-4} - \dots) = \tag{28}$$

$$\alpha_1^6(\alpha_1\sigma^2) - \alpha_1^7((1+\alpha_1^2)\sigma^2) + \alpha_1^8\alpha_1\sigma^2 = 0 \tag{29}$$

$$\tag{30}$$

and so on. The single terms are zero, so that the sum of those single terms is zero as well.

$$Cov(\eta_t, \eta_{t+1}) = 0 \tag{31}$$

Autocovariance of order 2 For the autocovarinace of order 2 we use the same approach. As an example: the first term ($Cov(\epsilon_t, \eta_{t+1})$) is 0 and the second term of $Cov(\eta, \eta_t + 2)$ illustrates all other terms:

$$Cov(\eta, \eta_{t+2}) = Cov(-\alpha_1^2 y_{t-2}, \alpha_1^3 y_{t-1} - \alpha_1^4 y_{t-2} + \alpha_1^5 y_{t-3} + \dots) + \dots \tag{32}$$

$$= -\alpha_1^5\alpha_1\sigma^2 + \alpha_1^6(1+\alpha_1^2)\sigma^2 - \alpha_1^7\alpha_1\sigma^2 + \dots = 0 + \dots \tag{33}$$

$$\implies Cov(\eta, \eta_{t+2}) = 0 \tag{34}$$

For the autocovariances of higher order we get the same result. The dependencies cancel each other out so that autocovariances are zero.

OLS estimation

$$\hat{\phi} = \frac{\sum_{t=1}^{T}(\phi y_t + \eta_t) y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2} \tag{35}$$

$$= \frac{\phi y_{t-1}^2 \sum_{t=1}^{T} y_{t-1}\eta_t}{\sum_{t=1}^{T} y_{t-1}^2} \tag{36}$$

$$= \dots = \phi + \frac{\sum_{t=1}^{T}\eta_t\eta_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2} + \frac{\sum_{t=1}^{T}\eta_t\eta_{t-2}}{\sum_{t=1}^{T} y_{t-1}^2} + \dots + \frac{\sum_{t=1}^{T}\eta_t\eta_0}{\sum_{t=1}^{T} y_{t-1}^2} \tag{37}$$

We can multiply all fractions with T^{-1}/T^{-1} for the law of large numbers to apply. The law of large numbers is described in detail in p.87 of Nielsen (2020[5]). The autocoviariances are finite, so the estimator converges as all autocovariances are equal to zero:

$$plim(\hat{\phi}) = \phi + \frac{\sum_{j=1}^{t} Cov(\eta_{t-j}, \eta_t)}{(1+\alpha_1^2)\sigma^2} = \phi \tag{38}$$

The OLS estimator converges anyway because the dependencies within the autocovariances cancel each other out.

[5]Nielsen, H.B, (2020).*A Course in Time Series Econometrics.*University of Copenhagen.

2 ARCH Models

2.1

The squared residual of x_t can be decomposed into the conditional expectation and an error term $v - t$ (p.228, Nielsen, 2020[6]):

$$\begin{aligned}
\epsilon_t^2 &= E(\epsilon_t^2|\mathbf{I}_{t-1}) + E(v_t||\mathbf{I}_{t-1}) \\
&= E(\epsilon_t^2|\mathbf{I}_{t-1}) + E(\underbrace{E(v_t||\mathbf{I}_{t-1})}_{=0}) && \text{with LIE} \\
&= E(h_t z_t|\mathbf{I}_{t-1}) \\
&= E((\alpha_0 + \alpha_1\epsilon_{t-1}^2)z_t|\mathbf{I}_{t-1}) \\
&= (\alpha_0 + \alpha_1\epsilon_{t-1}^2)\underbrace{E(z_t|\mathbf{I}_{t-1})}_{=1} && \epsilon_t \text{ in } \mathbf{I}_{t-1};\ z_t \text{ Gaussian} \\
E(\epsilon_t^2) &= \alpha_0 + \alpha_1 E(\epsilon_{t-1}^2) && E(\epsilon_t^2) = E(\epsilon_{t-1}^2) \text{ if } |\alpha_1| < 1, \text{ (stationary)} \\
(1-\alpha_1)E(\epsilon_t^2) &= \alpha_0 \\
E(\epsilon_t^2) &= \frac{\alpha_0}{(1-\alpha_1)} \\
1 &= \frac{\alpha_0}{(1-\alpha_1)} && \text{with } E(\epsilon_t^2) = 1 \text{ from task} \\
1 &= \alpha_0 + \alpha_1
\end{aligned} \tag{39}$$

LIE denotes the law of iterated expectations which was previously presented in this paper.

As the variance in the ARCH model is structured as an AR process, we refer to the condition of stationarity for univariate autoregressive models in section "Model assumptions" of assignment 1 (assignment 1, 2020[7]).

With the condition of stationarity ($|\alpha_1| < 1$, we can assume $E(\epsilon_t^2)$ to be stable and thus equal to $E(\epsilon_{t-1}^2)$. The expected value for the squared residuals $E(\epsilon_t^2) = 1$ is therefore given for the conditions that:

- $\alpha_0 \geq 0$, $\alpha_1 > 0$, such that the variance of the ARCH model is strict positive.
- $\alpha_1 < 1$ for stationarity in the autoregressive variance and that $E(\epsilon_t^2) < \infty$.
- $\alpha_0 + \alpha_1 = 1$ for $E(\epsilon_t^2) = 1$.

2.2

$$\begin{aligned}
E(x_t|\mathbf{I}_{t-1}) &= E(\mu_t + \epsilon_t|\mathbf{I}_{t-1}) \\
&= E(\beta + \delta h_t + h_t^{1/2} z_t|\mathbf{I}_{t-1}) \\
&= E(\beta|\mathbf{I}_{t-1}) + E(\delta h_t|\mathbf{I}_{t-1}) + E(h_t^{1/2} z_t|\mathbf{I}_{t-1}) \\
&= \beta + \delta E(h_t|\mathbf{I}_{t-1}) + \underbrace{E(h_t^{1/2}|\mathbf{I}_{t-1})\underbrace{E(z_t|\mathbf{I}_{t-1})}_{=0}}_{=0} && \text{as } z_t \overset{d}{=} N(0,1) \\
&= \beta + \delta(\alpha_0 + \alpha_1\epsilon_{t-1}^2)
\end{aligned} \tag{40}$$

An asset holder of a risky asset fears higher uncertainty about future returns. In order to incentivise

[6] Nielsen, H.B, (2020).*A Course in Time Series Econometrics.*University of Copenhagen.

[7] (2020).*Econometrics II - Assignment 1.* Fall course 2020.

the holder to nevertheless hold such an asset, he or she must be paid a risk premium in return for this high volatility. By that, it is a compensation for the risk that a financial asset bears. In this particular ARCH-model, a higher volatility in the past period multiplied by the variable risk premium δ, has only a positive effect on the return, if it exceeds a certain "base risk" (e.g. could be a market risk in a Capital Asset Pricing Model) which is denoted as β, such that $\beta - \delta(\alpha_0 + \alpha_1\epsilon_{t-1}^2)$ is larger than 0.

2.3

$$\begin{aligned}
E(x_t) &= E(\mu_t + \epsilon_t) \\
&= E(\beta + \delta h_t + h_t^{1/2} z_t) \\
&= E(\beta) + E(\delta h_t) + E(h_t^{1/2} z_t) \\
&= \beta + \delta E(h_t) + \underbrace{E(h_t^{1/2})\underbrace{E(z_t)}_{=0}}_{=0} \qquad \text{as } z_t \overset{d}{=} N(0,1) \\
&= \beta + \delta E(\frac{\epsilon_t^2}{z_t^2}) \qquad \text{as } \epsilon_t = h_t^{1/2} z_t \implies h_t = \frac{\epsilon_t^2}{z_t^2} \\
&= \beta + \delta \frac{E(\epsilon_t^2)}{E(z_t^2)} \qquad \text{as } E(\epsilon_t^2) = 1 \text{ cf. 5.2.1 and } z_t \overset{d}{=} N(0,1) \implies E(z_t^2) = 1 \\
&= \beta + \delta E\left(\frac{1}{1}\right) \\
&= \beta + \delta = \begin{cases} -2+2=0 & \text{Case A, for } \beta=-2, \delta=2 \\ -1+1=0 & \text{Case B, for } \beta=-1, \delta=1 \end{cases}
\end{aligned} \tag{41}$$

As the two time series, formerly described as Case A and Case B are identical in their process and shocks. Therefore the unconditional expected value for h_t will be identical (=1 as in equation above) and thus also the unconditional expectation of x_t will also be identical (=0).

However throughout the observed time frame, h_t will take various different realisations. For each conditional expectation of h_t different from 1, the conditional expectation of x_t will deviate from 0 (its unconditional mean).

The this deviation in $E(h_t)$ will lead to a larger deviation from 0 for Case A, since $E(h_t)$ gets multiplied by 2 (compared to 1 in Case B). This leads to a higher variance in x_t for the Case A, despite both cases having the same unconditional expectation of x_t.

2.4

We start by calculating x_{t+i} for i = 0,1,2, ... letting $\epsilon_s = 0$ for all s $\neq$ t. We impose the latter restriction below at each (*).

$$
\begin{aligned}
x_t &= \mu_t + \epsilon_t \\
&= \beta + \delta h_t + \epsilon_t \\
&= \beta + \delta(\alpha_0 + \alpha_1 \underbrace{\epsilon_{t-1}^2}_{=0}) + \epsilon_t \qquad (*) \\
&= \beta + \delta\alpha_0 + \epsilon_t \\
x_{t+1} &= \beta + \delta(\alpha_0 + \alpha_1 \epsilon_t^2) + \underbrace{\epsilon_{t+1}}_{=0} \qquad (*) \\
x_{t+2} &= \beta + \delta(\alpha_0 + \alpha_1 \underbrace{\epsilon_{t+1}^2}_{=0}) + \underbrace{\epsilon_{t+2}}_{=0} \qquad (*) \\
&\vdots \\
x_{t+i} &= \beta + \delta(\alpha_0 + \alpha_1 \underbrace{\epsilon_{t+i-1}^2}_{=0}) + \underbrace{\epsilon_{t+i}}_{=0} \qquad (*)
\end{aligned}
\tag{42}
$$

We then trace out the impulse-reponse function for a single shock ϵ_t. We see that the effect only affect the process in x_t and x_{t+1}.

$$
\frac{\partial x_t}{\partial \epsilon_t} = 1 \qquad \frac{\partial x_{t+1}}{\partial \epsilon_t} = 2\delta\alpha_1\epsilon_t \qquad \frac{\partial x_{t+2}}{\partial \epsilon_t} = 0 \qquad \dots \qquad \frac{\partial x_{t+i}}{\partial \epsilon_t} = 0 \;\text{ for } i > 1 \tag{43}
$$

2.5

Unconditional Variance of η_t

$$
\begin{aligned}
E(\eta_t^2) &= E(k_t h_t z_t^2) = E(k_t)E(h_t)E(z_t^2) &(44)\\
&= (\gamma_0 + \gamma_1 E(Greed_{t-1}))(\alpha_0 + \alpha_1 E(\epsilon_{t-1}^2)) &(45)
\end{aligned}
$$

Because Greed is stationary it has a constant mean. The mean of ϵ_{t-1}^2 follows directly from task 5.2.1. With $E(Greed_{t-1}) = E(Greed_t) = \mu_{GR}$ and $E(\epsilon_{t-1}^2) = 1$ we get

$$
E(\eta_t^2) = (\gamma_0 + \gamma_1\mu_{GR})(\alpha_0 + \alpha_1) \tag{46}
$$

Conditional Variance of η_t We get

$$
\begin{aligned}
E(\eta_t^2|I_{t-1}) &= E(k_t h_t z_t^2|I_{t-1}) = E(k_t|I_{t-1})E(h_t|I_{t-1})E(z_t^2|I_{t-1}) &(47)\\
&= (\gamma_0 + \gamma_1 Greed_{t-1})(\alpha_0 + \alpha_1\epsilon_{t-1}^2), &(48)
\end{aligned}
$$

as both $Greed_{t-1}$ and ϵ_{t-1} are part of the information set.

The extended model assumes that the sentiment of greed on the financial market in a specific period affects the volatility of financial asset x. The assumption of the variable *Greed* being stationary makes sense, as the sentiment of greed is unlikely to explode but rather to oscillate around a certain constant. A low level of *Greed* can also mitigate volatility (for $\gamma_0 + \gamma_1 Greed_t < 1$). *Greed* is assumed to not affect the risk premium μ_t. Hence, the risk premium would not reward greed induced volatility.

Measuring *Greed* among the actors on the financial market might be the wrong approach, as many of those actors are profit maximising entities per se (investment companies, fund managers, insurance companies, etc.). Another way to account for recklessness on financial markets could be measuring a level of regulation.

2.6

a) We decide to use the transformation "DlogNasdaq" for our model, as this series is stationary from a graphical inspection (cf. Figure 2). We observe that there is a volatility clustering, which supports the use of a GARCH-model. We decide to use a GARCH(1,1) with a constant for the mean and a non-normal error distribution. We report this estimation in Table 2, specification (1). Unfortuntelly, we have still autocorrelation in this specification (cf. c)). We also tried several other versions (EGARCH, Threshold-GARCH) and/or different number of lags (also for the mean), but did not come up with a version that is not misspecified. Hence we decided to keep it simple, especially because a GARCH(1,1) is not misspecified when testing on a 5%-level.

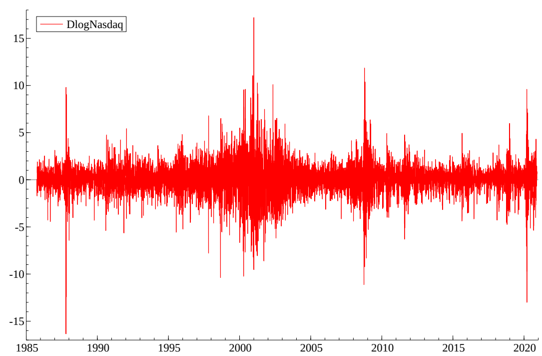

Figure 2: DlogNasdaq

b) We use are model from a) but include "log(h_t) in mean" to analyze if there are signs of a risk premium in the daily return. We report this estimation in Table 2, specification (2).

We find no significant risk premium in the NASDAQ returns. That means that the returns of the NASDAQ are not significantly higher when the NASDAQ is more volatile. This result is be be looked at with caution as our GARCH model is misspecified (cf. c)).

Another way to study the presence of risk premium could be a cross-sectional one. One could investigate if there is a significant risk premium when comparing Tech-Indices to governmental treasury bills.

c) For a GARCH to be well-specified we need no-autocorrelation as well as no-ARCH effects in the residuals. This is because autocorrelation in ϵ_t always implies autocorrelation in ϵ_t^2 but not vice-versa (assignment 4, 2020[8]). We see that both of our model indeed have no significant ARCH-effects, while autocorrelation (Portmanteau, 1-94) is a problem, at least when testing at the 1%-level. On the 5%-level, autocorrelation is not a problem. By testing for normality (which is rejected clearly), we further see that our use of a non-normal error distribution is justified.

[8](2020).*Econometrics II - Assignment 4.* Fall course 2020.

	(1)	(2)
Constant (X)	0.1140 (0.0111)	0.1142 (0.0112)
log(h_t) (X)	.	0.03347 (0.0188)
alpha_0 (H)	0.01859 (0.00375)	0.01891 (0.00377)
alpha_1 (H)	0.09566 (0.00791)	0.09625 (0.00793)
beta_1 (H)	0.9009 (0.00781)	0.9001 (0.00783)
student-t df	7.106 (0.522)	7.114 (0.524)
alpha(1)+beta(1)	0.997	0.996
Log-lik.	-14835.721	-14833.998
AIC	3.349	3.348
HQ	3.350	3.350
SC/BIC	3.353	3.353
Portmanteau, 1-94	[0.03]	[0.03]
No ARCH(1)	[0.13]	[0.14]
Normality	[0.00]	[0.00]
T	8864	8864
Sample start	1985-10-02	1985-10-02
Sample end	2020-11-30	2020-11-30

Table 2: The table shows estimates of a GARCH(1,1) with (2) and without (1) including "log(h_t) in mean" to analyze risk premium effects. Standard errors in $(\cdot)$ and p-values in $[\cdot]$ for misspecification tests.

1 Introduction

For our assignment we have been given US GDP data and are asked to provide an estimate, using a univariate time series model, and to produce a forecast for the recovery and later development of 2009(3)-2019(4). Auto-regressive (moving average) models allow to work with non-iid data without requiring additional observations or explanatory variables. Using an AR(2) model we can estimate a very close fit to the actual GDP observations and also a decent forecast to predict the near future development of US GDP.
We compare our forecast to a linear forecast applying a Diebold-Mariano test. Our forecast turns out to be more accurate for a short forecast horizon but less accurate in the long run. For mid- or far future developments a more sophisticated version such as a vector auto regressive model might be worth considering to improve the forecast.

2 Description of data

The data set provided for this assignment contains quarterly data for the real US GDP from 1975(1) to 2019(4) as well as a linear conservative forecast (LinFor). The data (GDPC1) has been downloaded from the FRED database maintained by the Federal Reserve Bank of St. Louis[1]. We use the given transformation described in the assignment task[2], i.e. D4log(GDP) being the yearly growth rate. This is clearly beneficial for our analysis, cf. Figure 1 where one can see that the time series now is stationary to an acceptable extent. Hence, we continue our analysis with D4logGDP, i.e. if not stated differently, we mean the latter transformation by y_t. Despite having an overall stationary appearance, we notice shifts in the GDP which might imply level changes. Further it is clearly visible that certain periods include considerable outliers, most prominent amongst them are the subprime crisis (2008), the second oil crisis (1979) and the US-Recession of 1981 and 1982. These have to be kept in mind when estimating our model as they might be a source of potential bias or misspecification.

3 Economic theory

AR model As mentioned, univariate models are particularly useful for characterizing dependence and for computing simple forecasts. We first consider an AR(p) model:

$$y_t = \delta + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \theta_3 y_{t-3+...} + \theta_p y_{t-p} + \epsilon_t \tag{1}$$

where our residuals $\epsilon \sim iid(0, \sigma^2)$, and we can observe the $y_{-(p-1)}, y_{-(p-2)}, ..., y_0$ initial values. Using the lag-operator L which has the following property $Ly_t = y_{t-1}$, we can reformulate the AR(p) model, so that all the y terms can be combined with a single coefficient:

$$\underbrace{(1 - \theta_1 L - \theta_2 L^2 - ... - \theta_p L^p)}_{\theta(L)} y_t = \theta(L) y_t = \delta + \epsilon_t \tag{2}$$

For the polynomial $\theta(L)$ to be invertible, it must hold that $|\theta| < 1$. Because the polynomial can also have complex solutions, we use the characteristic equation to then factorize it and obtain the inverse roots: $\phi_1 = z_1^{-1}, \phi_2 = z_2^{-2}, ..., \phi_p = z_p^{-p}$. In Equation 3, $\theta(z)$ is now invertible if each of its factors is invertible, meaning if $|\phi_j| < 1$ for all $j = 1, ..., p$. The mentioned inverse polynomial is written as:

$$\theta(z) = 1 - \theta_1 z - \theta_2 z^2 - ... - \theta_p z^p = 0 \quad \xrightarrow{\text{if } |\phi_j| < 1} \quad \theta(z)^{-1} = 1 + c_1 z + c_2 z^2 + ... + c_p z^p \tag{3}$$

[1]Federal Reserve of St. Louis, (2020). *https://fred.stlouisfed.org*
[2]Econonmetrics II, (2020. *Assignment 1*

As $\theta(L)$ is invertible under the mentioned conditions, we take the results of Equation 2 and reformulate it to:

$$\begin{aligned} y_t = \theta(L)^{-1}(\delta + \epsilon_t) \quad &= (1 + c_1 L + c_2 L^2 + ...c_p L^p)(\delta + \epsilon_t) \\ &= (1 + c_1 + c_2 + ... + c_p)\delta + \epsilon_t + c_1\epsilon_{t-1} + c_2\epsilon_{t-2} + ... + c_p\epsilon_{t-p} \\ &\implies \frac{\partial y_t}{\partial \epsilon_t} = 1, \quad \frac{\partial y_t}{\partial \epsilon_{t-1}} = c_1, \quad \frac{\partial y_t}{\partial \epsilon_{t-2}} = c_2, ... \end{aligned} \tag{4}$$

In the last line of Equation 4 we see that c_j measures the dynamic impact of a shock on y_t and is therefore also called impulse response.

ARMA model An AR process can also be extended as an autoregressive model including a moving average. In Equation 5, the MA-term with α depicts how strong an influence a past shock has for y_t do deviate from the mean:

$$y_t = \underbrace{\theta_1 y_{t-1} + ... + \theta_p y_{t-p}}_{AR:\ \theta(L)} + \underbrace{\delta}_{mean} + \underbrace{\epsilon_t + \alpha_1\epsilon_{t-1} + ... + \alpha_p\epsilon_{t-p}}_{MA:\ \alpha(L)} \tag{5}$$

Point estimate and forecast error variance When using univariate dynamic models, estimating turning points is very difficult, but it is still possible to predict the direction of future movement. The optimal point estimate given the information set I_T is shown with (I) in Equation 6, which converges to the unconditional mean, with an increasing prediction window. The forecast error is (II). The forecast error variance (FEV) (III) increases over time, also eventually converging to the unconditional variance of y_t. For the FEV of $q+1$ periods, we reformulate the ARMA model to the MA(∞) representation and take the variance:

$$\begin{aligned} &(I) \quad & y_{T+q+1|I_T} &= E(y_{T+q+1|I_T}) \to \delta \qquad \text{for } q > \text{the number of lags} \\ &(II) & \eta_{q+1} &= y_{T+q+1} - y_{T+q+1|I_T} \\ &(III) & FEV(q+1) &= E[\eta_{q+1}] = E[(\epsilon_{T+q+1} + \alpha_1\epsilon_{T+q} + ... + \alpha_q\epsilon_{T+1})^2 | I_T] \\ & & &= (1 + \alpha_1^2 + ... + \alpha_q^2)\sigma^2 \end{aligned} \tag{6}$$

Model assumptions

- **Stationarity**: AR(p) is stationary if $\theta(z)$ is invertible, or put differently if $|\phi_j| < 1$ for $j = 1, 2, ..., p$.
- **Weak dependence**: The stationarity condition implies that the impulse response will decrease over time and eventually die out. Looking at Figure 1, we clearly see that the ACF is decreasing and eventually converging to 0. This observation strengthens the assumption of weak dependence, with decreasing autocorrelation with increasing lags.

Misspecification tests

- Residual autocorrelation (x periods): Breusch-Godfrey LM test, H_0: No AC, $\xi_{AR} = T \cdot R^2 \xrightarrow{d} \chi^2(x)$
- Residual normality: Jarque-Bera test, H_0: Error terms are normally distributed, $\xi_{JB} = \xi_S + \xi_K \xrightarrow{d} \chi^2(2)$
- Heteroskedasticity (k regressors): Standart LM-test, H_0: Errors are homoscedastic, $\xi_{HET} = T \cdot R^2 \xrightarrow{d} \chi^2(2k)$

Diebold-Mariano-test: The test statistic uses $\bar{d}$, the average and the long-run variance of the difference of the squared forecast error (η^2) between two different forecasts, across all the forecast periods: $H_0: \ E(\bar{d}) = 0$ for $\xi_{DM} = \frac{\bar{d}}{\sqrt{V(\bar{d})}} \xrightarrow{d} N(0,1)$

Our model Our final model (cf. Part 4) is an AR(2) with two lags and a control vector $\overrightarrow{\gamma}$ which includes dummy variables to capture outliers with extremely large residuals as well as a dummy variable ("*Regime*") indicating the suspected level shift in the period 2001(1)-2009(3). This control vector helps our model to receive approximately normally distributed residuals, a vital condition for the consistency of the OLS-estimator.

$$\begin{aligned} y_t &= \delta + \theta_1 y_{t-1} + \theta_4 y_{t-4} \\ &\quad + \gamma_1 \mathbf{1}_{1979(1)} + \gamma_2 \mathbf{1}_{1981(4)} + \gamma_3 \mathbf{1}_{1982(1)} + \gamma_4 \mathbf{1}_{2008(4)} + \gamma_5 \mathbf{1}_{Regime} + \epsilon_t \end{aligned} \tag{7}$$

The estimator and its properties We use the OLS estimator $\hat{\beta}$, which conveniently is also the closed-form maximum likelihood estimator (MLE) for autoregressive models, given that their residuals are normally distributed. The vector $\hat{\beta}$ estimates the model parameters by minimising the sum of squared observed residuals. The closed form solution of the minimisation problem is

$$\hat{\beta} = (\sum_{t=1}^{T} x_t x_t')^{-1} x_t y_t. \tag{8}$$

with x_t traditionally being a vector of the explanatory variables. In our univariate time series x_t consists of lagged values of y_t and the indicator variables.
Under certain conditions the OLS estimator is unbiased:

- The conditional mean of the residuals has to be zero: $E(\epsilon_t|x) = 0$

and consistent:

- Random variable Y, from which the observations y_t are drawn, must have a well defined fourth moment: $E(Y^4) < \infty$.
- Y must be i.i.d., so that the residuals have identical variance and are not auto correlated.
  - If (conditionally) not identically distributed, the estimator $\hat{\beta}$ is consistent but we must use a robust standard error estimator.

Our estimation will indicate that the residuals in our model are not correlated. We can therefore rely on consistent point estimates for δ, θ_1, θ_4 and the dummy coefficients.
Under conditional homoscedasticity:

$$Var(\hat{\beta}|X) = E((\hat{\beta} - \beta)(\hat{\beta} - \beta)'|X) = (\sum_{t=1}^{T} x_t x_t')^{-1} \sigma_\epsilon^2 \tag{9}$$

Homoscedasticity is not strictly necessary as there are standard error estimators that account for heteroscedasticity. In our model the residuals are in fact homoscedastic. As all necessary criteria are fulfilled $\hat{\beta}$ is a consistent estimator with a variance estimated in a correct way. It's asymptotic distribution is normal.

4 Empirical analysis

Empirical results Our final forecasting model is displayed in Table 1, in the second column. The other columns show an intermediate step in our model selection (1) and the comparison to another model group (3), both explained more detailed in the next paragraph.

Model selection Looking at the ACF and PACF of D4logGDP in Figure 1, we see a clear convergence towards zero, which indicates weak dependence. The graph shows (nearly) significant PACF terms still at lags of over ten periods. Following the general-to-specific (GETS) approach we start our model estimation with an AR(13) model.
As the AR(13) is over-specified and contains various insignificant coefficients, we use automatic model selection for a first shrinking (cf. first column of Table 1). The misspecification test for no autocorrelation is however (borderline) rejected at the 5%-level. Further the inverse roots are scattered very close to the unit circle. Adjusting for these flaws, we decrease the number of lags (to lower the general size of the inverse roots) and introduce four dummies, each canceling out a quarter with a particularly large residual. They correspond to the time periods mentioned in Part 2. This correction should help to improve the distribution of the residuals to be approximately normal distributed. Similarly, we correct for the suspected level shift in the later half of the observed period, i.e. after the first quarter of 2001 the GDP seems to oscillate around a lower level than before, which violates our assumption of stationarity.
Even though we followed the GETS approach, the few significant ACF terms of the time series leave us in doubt about potential MA terms which could further improve the model. Given the first ACF term is the largest, we extend our AR(2) model with an additional MA term to an ARMA(2,1). As the number of AR terms is very low, we are not concerned about potential cancelling roots. At a first glance the MA inclusion seems to be successful as all information criteria show a greater value. However, the misspecification test for no-autocorrelation in the residuals, is now rejected (alongside with heteroscedasticity), which makes the estimator inconsistent and therefore worse than the simpler AR model.

Misspecification tests and link to economic theory Looking at the misspecification tests of our final AR(2) model, we can consider our estimator to be consistent. The AR model shows no autocorrelation and its residuals are normal distributed. Heteroscedasticity is also no problem. The inverse roots plotted in Figure 3 are reasonably far away from the unit circle, with all roots having a modulus smaller than 0.88.

Comparison to linear forecast We formulate a Diebold-Mariano test to test if our forecast is significantly better than the linear forecast. We use the package "forecast" by R. Hyndman and the included "dm.test"-function[3]. The test reveals, that our forecast seems to be worse for the time span of 42 quarters. We cannot reject that our forecast is less accurate (for both squared errors and absolute value errors) but we can reject that our forecast is more accurate (for absolute value errors, with a significance level of 10%). We think that although our forecast is very accurate for the rest of 2009 and the year 2010, the discrepancy in 2011 is too high for our forecast to be better than the linear forecast. This is emphasized by the sum of squared errors in the first five quarters of forecasting. The sum of squared errors of our forecast is 0.0007071481 and the one of the linear forecast is 0.001280855 and almost twice as large. Until 2011(3) our forecast has a lower sum of squared errors than the linear forecast.

Interpretation of the empirical results A one percent increase in the yearly-growth rate in the quarter before increases the yearly-growth rate of the dependent quarter by 0.96%. For the last years yearly-growth rate (D4log(GDP)_4) it is a decrease by 0.23%. As expected, our time dummies control for the extraordinary negative shocks and the Regime variable takes the level-shift into account. All parameters are highly significant.

[3]RDocumentation, (2020).*https://www.rdocumentation.org/packages/forecast/versions/8.13/topics/dm.test*

5 Conclusion

We derive a well thought out linear model to describe the quarterly GDP growth of the USA. It fulfills all requirements for a consistent and precise estimation with the OLS estimator. We compared it to other possible model choices and can conclude that ours is less prone to errors as there is no misspecification and no possible threat of a unit root.
With this model, we are able to compute a forecast that predicts the GDP growth very precise until the end of 2010. Considering forecasts of longer intervalls however, we would advise for a more sophisticated model like a vector auto regression, which is also capable of controlling for other factors.

6 Appendix

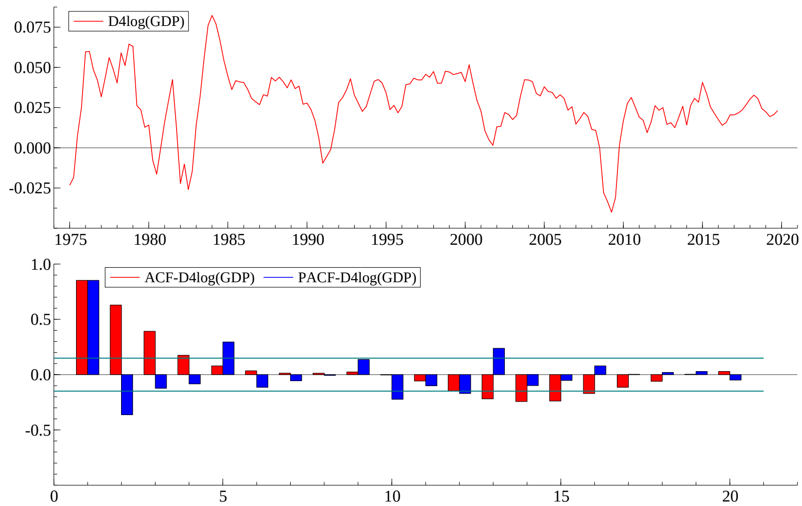

Figure 1: ACF and PACF for the D4logGDP variable

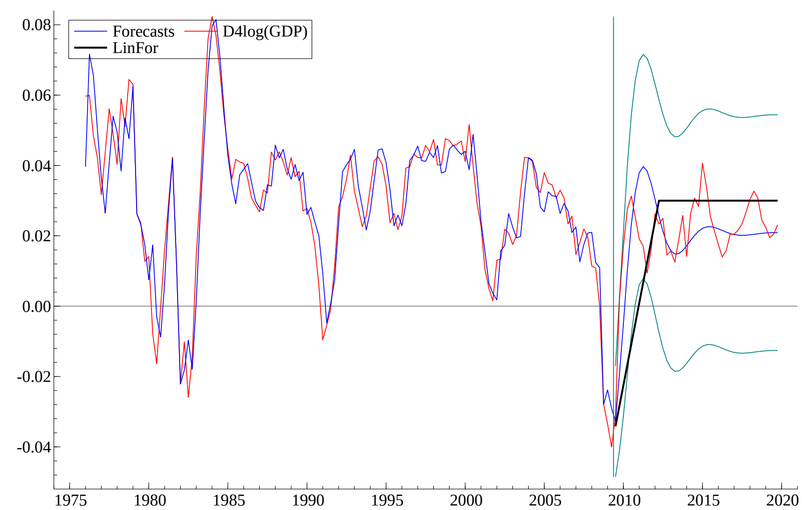

Figure 2: Fitted values of AR(2) model to actual time series, including forecast for period after 2009(3) with 95% confident-bands and the linear forecast

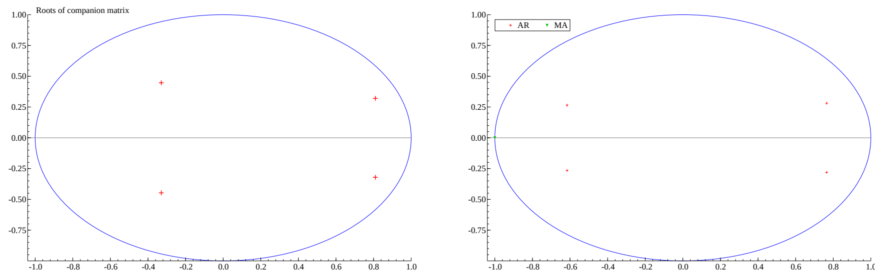

Figure 3: Plot of inverse roots for AR(2) (left) and ARMA(4,1) (right).

	AMS (First attempt)	**AR** (Our model)	ARMA (Comparison)
Constant	.	0.01 (5)	0.003 (6.49)
D4log(GDP)_1	1.167 (28.8)	0.960 (21.7)	1.159 (55.6)
D4log(GDP)_4	−0.714 (−4.65)	−0.234 (−5.17)	−0.262 (−12)
D4log(GDP)_5	0.569 (3.33)	.	.
D4log(GDP)_8	−0.483 (−3.34)	.	.
D4log(GDP)_9	0.617 (4.22)	.	.
D4log(GDP)_12	−0.650 (−5.62)	.	.
D4log(GDP)_13	0.467 (5.87)	.	.
MA-1	.	.	−1 (−26.3)
I:1979(2)	.	−0.031 (−21.8)	−0.014 (−2.74)
I:1981(4)	.	−0.038 (−19.5)	−0.019 (−1.49)
I:1982(1)	.	−0.041 (−29.1)	0.009 (0.77)
I:2008(4)	.	−0.029 (−19.2)	−0.044 (−5.35)
Regime	.	−0.004 (−2.83)	−0.001 (−3.18)
$\hat{\sigma}$	0.00779	0.007876	0.0118
Log-lik.	433.105	463.075	402.368
AIC	-6.818	-6.792	-5.856
HQ	-6.753	-6.722	-5.768
SC/BIC	-6.659	-6.619	-5.640
No autocorr.	[0.05]	[0.24]	[0.00]
No hetero.	[0.01]	[0.16]	[0.00]
Normality	[0.01]	[0.09]	[0.02]
T	125	134	134
Sample start	1978(2)	1976(1)	1976(1)
Sample end	2009(2)	2009(2)	2009(2)

Table 1: The table shows estimates of the model in equation (X) with various restrictions imposed. T-ratios in $(\cdot)$ and p-values in $[\cdot]$ for misspecification tests. AMS for Automatic Model Selection in OxMetrics

1 Introduction

For this assignment we are asked to examine to which extent the purchasing power parity (PPP) relation between Mozambique (MZM) and South Africa (ZAR) holds. According to the PPP theory, the exchange rate between two currencies should be such that an identical good has the same price in different locations. It is unlikely that PPP holds in a strict sense, as transportation costs and trade barriers are likely to be present. The consensus view about the PPP is that it does not hold in the short-run and may hold in the long-run PPP[1]. We use co-integrated vector autoregression and vector error-correction with a model that includes the first three lags of each variable and has a co-integration rank one. We can confirm that the PPP is constant but we can reject the strict version. We find that the variable that error corrects in equilibrium is the exchange rate.

2 Description of Data

We have been given monthly observations of the respective consumer price indices of Mozambique (CPI_MZM) and South Africa (CPI_ZAR) and the exchange rate (MZM_ZAR, Metical per Rand) from 2004(1) to 2020(9). We note that Mozambique is one of the poorest countries in the world, ranks 180/189 concerning human development[2] and 137/141 regarding global competitiveness[3]. South Africa in contrast is by far more developed (rank 113) and globally competitive (rank 60). Its GDP of 351.4 billion USD[4] is a multiple of Mozambique's 14.9 billion USD GDP. According to Corbae and Ouliaris[5], it is usually the smaller economy that adapts and hence these insights suggest that if one economy adjusts it should be Mozambique. We also note that the CPI_ZAR seems to follow a stable trend while the CPI_MZM is much more unstable, cf. Figure 1. The exchange rate is also quite unstable, ranging from roughly 0.9 to 1.7. All data is not seasonally adjusted.

3 Economic Theory

3.1 Co-integration of Univariates

We can try to co-integrate two unit-root non-stationanry time series x_1, x_2 (also called I(1)), to find a "common" random walk. In Equation 1 we use a co-integration vector β' to get a linear combination for I(1) x_1 and x_2. The variables co-integrate if the deviation u_t is stationary.

$$\beta' X_t = \begin{pmatrix} 1 & -\beta_2 \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} \implies x_{1t} = \mu + \beta_2 x_{2t} + u_t \tag{1}$$

We choose a vector error correction model (VECM) because there is no assumption on the causality between the variables (as in the simple vector autoregressive (VAR) model). This leaves the most leeway in the interpretation of the results, Furthermore, using the vector setup of the model as well as the trace statistic allows to test for the co-integration of various different variables at once.

3.2 Vector Error Correction Model

We start by considering a vector autoregressive model conditional on (X_0, X_1), with $k = 2$ lags, p variables, $t = 1, 2, ..., T$ observations, and with $\epsilon_t | X_{t-1}, X_{t-2} \overset{d}{=} N(0, \Omega)$. The VAR($k$) model can be

[1] Taylor, A.M., Taylor, M.P., (2004). *The Purchasing Power Parity Debate*. Journal of Economics Perspectives 18(4)

[2] United Nations Development Programme, (2020). *http://hdr.undp.org/en/content/2019-human-development-index-ranking*. Human Development Reports

[3] World Economic Forum (2019). *The Global Competitiveness Report*

[4] World Bank, (2020). *https://data.worldbank.org/indicator/NY.GDP.MKTP.CD*

[5] Corbae, D., Ouliaris, S.,(1988).*Cointegration and tests of purchasing power parity*. The Review of Economics and Statistics 70(3)

written as shown in the top row of Equation 2. Using the lag-operator (L) we can reformulate it to:

$$\begin{aligned} X_t &= \Pi_1 X_{t-1} + \Pi_2 X_{t-2} + \mu + \epsilon_t \\ (I_p - \Pi_1 L - \Pi_2 L^2) X_t &= \mu + \epsilon_t \\ \Theta(L) X_t &= \mu + \epsilon_t \end{aligned} \tag{2}$$

$\Theta(z)_{p \times p}$ is a quadratic matrix with p dimensions of polynomials, which also gives the characteristic equation $|\Theta(z)| = 0$. For $\Theta(1) = I_p - \Pi_1 - \Pi_2$, the VAR model has a unit root if $|\Theta(1)| = 0$. Similarly to the univariate co-integration, the VAR can be rewritten to a VECM:

$$\begin{aligned} X_t &= \Pi_1 X_{t-1} + \Pi_2 X_{t-2} + \mu + \epsilon_t \\ X_t - X_{t-1} &= \underbrace{(\Pi_1 + \Pi_2 - I_p)}_{\Pi} X_{t-1} + \underbrace{\Pi_2}_{-\Gamma_1} (X_{t-2} - X_{t-1}) + \mu + \epsilon_t \\ \Delta X_t &= \Pi X_{t-1} + \Gamma_1 \Delta X_{t-1} + \mu + \epsilon_t \qquad \Rightarrow \Pi = \Pi_1 + \Pi_2 - I_p = -\Theta(1) \end{aligned} \tag{3}$$

Depending on the I(1) combination of the variables in X_t, Π has a reduced rank (r), so $rank(\Pi) = 1 = r < p$ with p number of variables. Π can therefore be decomposed into the vectors α and β (where β is normalized in β_1). Given the rank of Π we can derive three different cases, shown and explained in Equation 5.

3.3 Estimator

To formulate a VECM, we use the CATS module in OxMetrics. This module uses a reduced rank regression to estimate the parameters of Equation 5. As we did not further cover reduced rank regression in our course, we also cannot show its assumptions and composition.

For estimating a VAR we us the maximum likelihood estimator (MLE). Assuming a multivariate Gaussian density, the maximum likelihood estimation for the VAR model is: $logL(\theta) = \sum_{t=1}^{T} logf(Z_t|Z_{t-1};\theta)$ with $\theta = \{\mu, \Pi, \Omega\}$, $\hat{\Omega} = \frac{1}{T}\sum_{t=1}^{T} \hat{\epsilon}_t \hat{\epsilon}_t'$
If the eigenvalues $|\lambda_p| < 1$ then $\epsilon_t \sim iid(0, \Omega)$ then it holds that the $\hat{\theta}_{MLE}$ has approximately a Gaussian distribution, otherwise it is the consistent but not most efficient quasi maximum likelihood estimator ($\hat{\theta}_{QMLE}$).

3.4 Normalization

As shown in Equation 5, if Π has not full rank, α and β can be estimated to maximizes the likelihood. Beside these estimate vectors, there are also countless other combinations of α and β which yield the exact same result for Π. This process is called normalization and identification (setting a different $\beta_i = 1$). This is particularly of interest as we try to find an normalization of the parameters, which matches the economic theory.

Inference co-integration and error-correction vector $\hat{\beta}$ is considered super-consistent and $T(\hat{\beta} - \beta)$ converges in mean to zero with a mixed Gaussian distribution in which the variance is random. For the case of $\beta = (1 \quad -\beta_2)'$ this would result in Equation 4 where V_{β_2} is random. However we can estimate V_{β_2} from the data and then form the t-test statistic which will have a standard N(0,1) distribution.

$$T(\hat{\beta}_2 - \beta) \xrightarrow{d} N(0, V_{\beta_2}) \qquad t_{\beta_2 = b} = \frac{\hat{\beta}_2 - b}{se(\hat{\beta}_2)} = \frac{\hat{\beta}_2 - b}{\sqrt{T^{-1}\hat{V}_{\beta_2}}} \xrightarrow{d} N(0,1) \tag{4}$$

3.5 Tests

Misspecification tests

- Residual autocorrelation (x periods): Breusch-Godfrey LM test, H_0: No AC, $\xi_{AR} = T \cdot R^2 \xrightarrow{d} \chi^2(x)$
- Residual normality: Jarque-Bera test, H_0: Error terms are normally distributed, $\xi_{JB} = \xi_S + \xi_K \xrightarrow{d} \chi^2(2)$
- Heteroskedasticity (k regressors): Standart LM-test, H_0: Errors are homoscedastic, $\xi_{HET} = T \cdot R^2 \xrightarrow{d} \chi^2(2k)$

Trace test Given the VECM in Equation 3 and depending on the rank of Π we can estimate three different, nested models (H):

$$\begin{aligned}
H_0: \quad & \Delta X_t = & \Gamma_1 \Delta X_{t-1} + \mu + \epsilon \quad & \text{, for } rank(\Pi) = 0, \\
H_1: \quad & \Delta X_t = \alpha\beta' X_{t-1} + & \Gamma_1 \Delta X_{t-1} + \mu + \epsilon \quad & \text{, for } rank(\Pi) < p; \ \alpha_{p\times r}\beta'_{r\times p} \\
H_2: \quad & \Delta X_t = \Pi X_{t-1} + & \Gamma_1 \Delta X_{t-1} + \mu + \epsilon \quad & \text{, for full rank in } \Pi
\end{aligned} \tag{5}$$

We apply the LR test statistic to use the maximized log-likelihood values of the nested models to test for the co-integration rank of each. We start by comparing the model with the largest rank to the one with zero rank and descend from there, each model having one rank less.

$$H_0 \subset H_1 \subset H_2 \subset ... \subset ...H_p \quad \Rightarrow \quad LR(H_0|H_p) \to ... \to LR(H_0|H_2) \to LR(H_0|H_1)$$
$$\text{with} \quad LR(H_0|H_2) = -2(logL(H_0) - logL(H_2)) \xrightarrow{d} DF$$

These LR tests are the trace test statistic with a Dickey-Fuller type distribution (with $T \to \infty$), depending on the number of estimated parameters and p and r. Given the setup of the VECM, we do not have to previously check unit root in X_t. The test stops at the smallest model not rejected. For this model we characterize the vector for co-integration β', as well as for speed of adjustment α.

4 Empirical analysis

To investigate the relationship between exchange rate and price levels we will formulate a VAR model and a VECM. The advantage of a co-integrated VAR is that we allow all variables to error correct. Although we have argued that Mozambican parameters are more likely to change in an equilibrium relationship, we are on the safer side with a less restrictive model.

$$X_t = \Pi_1 X_{t-1} + \Pi_2 X_{t-2} + \Pi_3 X_{t-3} + \Lambda \times CS + \epsilon_t \tag{6}$$

As a well specified unrestricted VAR we find that a VAR(3) satisfies all conditions on consistent estimation, shown in Table 2. We include centred seasonal dummies. CS is the matrix including those dummies and Λ contains their coefficients. p denotes the log-price-level in Mozambique, $p*$ in South Africa and e the logarithmised exchange rate.

$$X_t = (p_t, p*_t, e_t)' \tag{7}$$

Considering Table 2 the hypothesis "no autocorrelation" can be rejected. We can account for heteroscedasticity with a robust standard error estimator. Non-normal residuals are also non-problematic because of the usage of the QMLE. We reformulate the consistently estimated model into a VECM for co-integration. We perform a trace test on the Π-matrix of our VECM, shown in Table 1. It reveals that Π has a rank of one. This means that all three variables are unit root processes that co-integrate to one non-stationary process. Figure 2 shows this graphically. We found two linear combinations of the three non-stationary processes that are stationary ($\beta_1' X_{t1}$ and $\beta_2' X_{t1}$). The reduced rank allows us to decompose the matrix into two vectors, the error correction vector α and the co-integration vector β, with

$$\alpha = (0.0161, -0.0099, -0.0553)' \text{ and } \beta = (-2.87, 3.64, 1)'. \tag{8}$$

We decided to normalise β_3 to unity (table 5). A possible argument against this is that α_3 is borderline insignificant (table 6). We still chose the exchange rate and base this decision on the fact that the magnitude of adjustment is by far the highest for the exchange rate. Naturally, the exchange rate can adjust quicker than price levels because less entities are involved with exchange rate adjustments compared to adjustments of price levels throughout an entire nation.
This suggests that the exchange rate is the variable that adjusts to maintain the equilibrium. We come up with a stationary equilibrium relation:

$$\begin{aligned} e_t &= 2.87 p_t - 3.64 p_t^* + u_t \\ \Leftrightarrow log(\frac{MZM}{ZAR}) &= log(\frac{2.87 \times CPIMZM}{3.64 \times CPIZAR}) + u_t \end{aligned} \tag{9}$$

This equilibrium relationship allows us to answer the original research questions. Our findings indicate that the PPP is indeed constant in the long run, as we find the exchange rate to be a stationary process that error corrects deviations of prices.
For the strict PPP to hold we need both CPI's to have the same influence on the exchange rate. This implies that both coefficients, β_1 and β_2 are equal with opposite signs. Although those coefficients are non-normal we can still test this using a t-test.

$$\begin{aligned} & H_0 : \beta_1 = -\beta_2 \Leftrightarrow \beta_1 + \beta_2 = 0 \qquad \text{vs.} \qquad H_1 : \beta_1 \neq -\beta_2 \\ \text{with} \qquad & t = \frac{\hat{\beta}_1 + \hat{\beta}_2}{se(\hat{\beta}_1 + \hat{\beta}_2)} \overset{d}{=} N(0,1) \qquad \Longrightarrow \qquad t = \frac{-2.87 + 3.64}{0.00665974} = 115.62 \end{aligned} \tag{10}$$

We can reject the null on any level of significance. The standard errors are computed from the residual standard errors and correlation in Table 4.
The strict PPP does not hold. The impulse response functions (Figure 3) show the reactions of the variables to shocks of other variables. Unlike our original hypothesis, that South African measures are rather independent from Mozambican, South African prices react more sensitive to shocks in Mozambican prizes than the other way around. This is quite surprising and raises the question: Why would the huge economy be more sensitive?

The impulse response functions (Figure 3) also show how the exchange rate keeps the equilibrium when price levels change. The exchange rate is more sensitive to price changes in South Africa in a sense that the shock is equaled out faster. One can see that shocks in p have an effect for 6 months and shocks in p^* for about 4 months. This was already implied by the error correction vector α in Table 6. This is evidence for the PPP to hold in the long run, but not in the short run. The exchange rate needs months to account for changes in price levels. Despite paying attention to a correct computation of the impulse response functions (order in OxMetrcis) we miss contemporaneous effects of price levels on the exchange rate.

5 Discussion and Conclusion

In our empirical analysis we found statistical evidence that the PPP between Mozambique and South Africa is constant in the long run. The co-integration relationship we found suggests that the exchange rate adjusts to maintain a stable equilibrium. It needs between four and six months to adjust when price levels change. Therefore we can say that the PPP is not constant in the short run.

In section two we argued that Mozambican measures adjust rather than South African. Combining this thesis with our findings, it is the value of Mozambican metical that adjusts. This should be further investigated by looking into the exchange rate of metical and other currencies. An argument against this is that we found that South African prices react more sensitive to changes In Mozambican prices than the other way around.

The gap between the two economies is also shown in the coefficient estimates of the equilibrium. We reject the hypothesis of a strict PPP because the effect on the exchange rate of South African prices is significantly higher than the effect of Mozambican prices. Other authors[6][7][8] use perfect commodity arbitrage as an argument for the theoretical existence of the strict PPP. In the case of South Africa and Mozambique the underdeveloped infrastructure of Mozambique[9] might hinder the use of perfect commodity arbitrage.

[6] Corbae, D., Ouliaris, S.,(1988).*Cointegration and tests of purchasing power parity.* The Review of Economics and Statistics 70(3)

[7] Kim, Y.,(1990). *Purchasing Power Parity in the Long Run: A Cointegration Approach.*Journal of Money, Credit and Banking, 22(4)

[8] Taylor, A.M., Taylor, M.P., (2004). *The Purchasing Power Parity Debate.* Journal of Economics Perspectives 18(4)

[9] BorgenProject, (2020). *https://borgenproject.org/infrastructure-in-mozambique/*

6 Appendix

Trace test VECM(3)						
p-r	r	Trace	Trace	crit. 5% value	p-value	pvalue
3	0	54.32	50.70	29.80	[0.0]**	[0.0]**
2	1	9.16	8.24	15.41	[0.357]	[0.448]
1	2	0.70	0.67	3.85	[0.403]	[0.413]

Table 1: Trace test for VECM(3) model with the variables p, p^* and e.

Misspecification tests VAR(3)		
	t-statistics	p-value
AR 1-7 test:	1.2850	[0.0796]
Normality test:	42.097	[0.0000]**
Hetero test:	1.3711	[0.0023]**
Sample start / end	2004(1)	2020(9)

Table 2: The table shows t-statistics and p-values of standard misspecification tests for our unrestricted VAR(3) model.

Π-parameters			
	p	p^*	e
p	-0.046 (-3.6)	0.0584 (3.6)	0.0161 (3.8)
p^*	0.0284 (4.7)	-0.036 (-4.6)	-0.0099 (-5.2)
e	0.159 (1.8)	-0.201 (-1.8)	-0.0553 (-1.8)

Table 3: Π-parameters of the CATS module for the VECM(3)

residuals correlation and standard errors			
	p	p^*	e
p	1	.	.
p^*	0.13	1	.
e	-0.0002	0.00963	1
SE	0.00593	0.00267	0.0435

Table 4: Correlation and standard errors of the variables of the CATS module for the VECM(3)

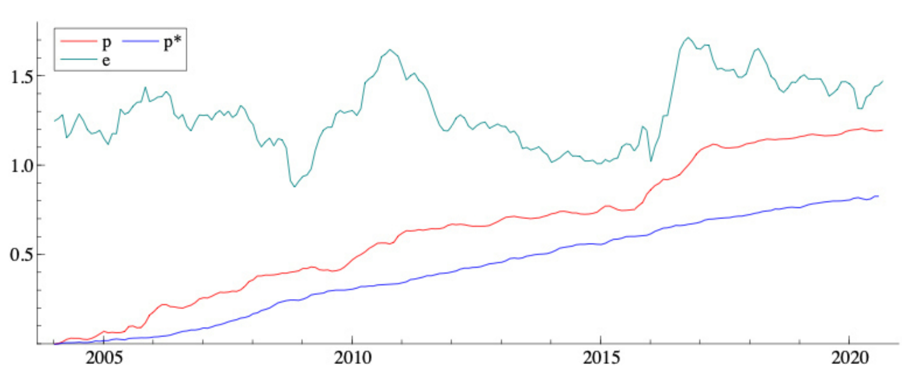

Figure 1: Time series of the log-price level in Mozambique (p), South Africa (p^*) and their exchange rate metical per rand (e); monthly observations 2004(1) - 2020(9)

β-parameters			
	β_p	β_{p^*}	β_e
β:	-2.87	3.64	1
	(-10.7)	(9.9)	

Table 5: β-parameters of the CATS module for the VECM(3)

α-parameters			
	α_p	α_{p^*}	α_e
α:	0.0161	-0.0099	-0.0553
	(3.8)	(-5.2)	(-1.8)

Table 6: α-parameters of the CATS module for the VECM(3)

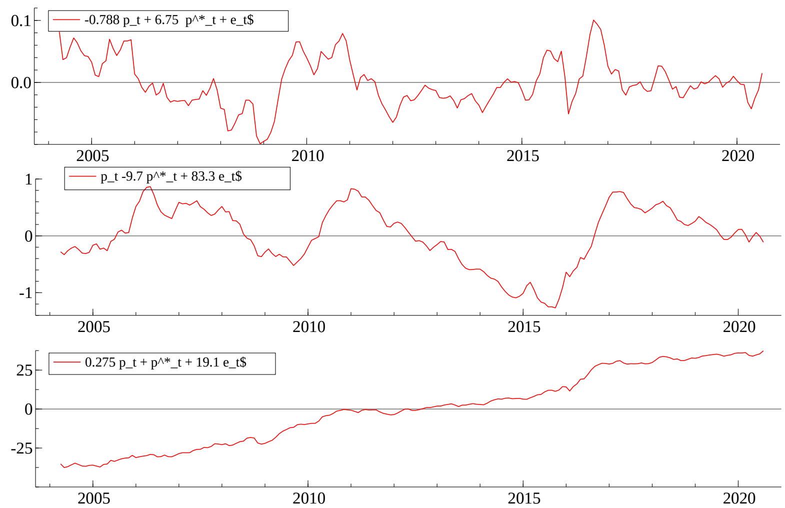

Figure 2: Linear combination of the p_t, p_t^* and e_t

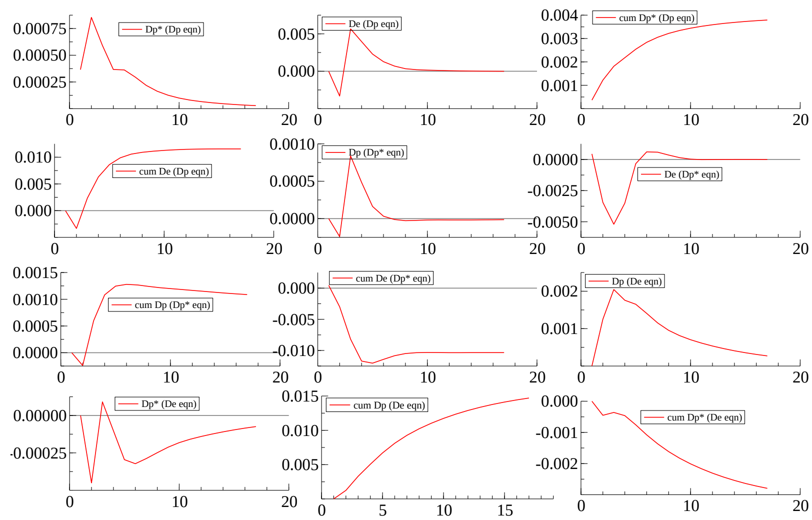

Figure 3: Impulse response functions