Exam Solution Guide

Econometrics II

December 2016

Part 1
What is Your Forecast of GDP Growth?

The Case The goal of this part of the exam is to estimate a univariate time series model for the growth in nominal GDP for Denmark over the period from 1980 to 2015 and use the estimated model to forecast the growth rate until 2025.

The Data Graphs of the data and relevant transformations must be shown in the exam. It must be noted that the level of nominal GDP is clearly non-stationary, but the log growth rate appears stationary. Moreover, it can be noted that there appears to be a level shift in the (log) growth rate around 1986-1987 and 2009-2010, while the negative growth rates in 2008-2009 appears to be extreme events.

Econometric Theory The econometric theory must include the following:

(1) A precise definition and interpretation of the model considered and its properties. Specifically, a univariate autoregressive (AR) or autoregressive moving average (ARMA) model must be presented. Furthermore, a precise definition of the out-of-sample forecasts and the forecast variance must be given.

(2) A precise description of the estimator used, in particular a precise account of the assumptions used to derive the estimator. Specifically, the method of moments (MM) or the maximum likelihood (ML) estimators can be used dependent on the model considered.

(3) A precise account of the necessary assumptions for consistent estimation and valid inference. This includes a precise definition of the null hypotheses, test statistics, and asymptotic distributions used to test relevant hypotheses.

(4) The theory must be presented in a logical order and with a consistent and correct notation.

Empirical Results The empirical results must include the following:

(1) A presentation and discussion of the relevant empirical results, so that the reader is able to understand the steps carried out in the process as well as the conclusions made.

(2) A description of the model selection process based on a general-to-specific approach, information criteria, or both.

(3) A discussion of whether the assumptions for consistent estimation and valid inference are satisfied for the estimated models. Specifically, this includes misspecification testing, which must be presented and discussed before statistical testing is carried out.

(4) A clear conclusion to the main question and a discussion of the limitations of the approach used to reach the conclusion. Specifically, the conclusion regarding the out-of-sample forecasts and the forecast variance must be presented and the limitations of the estimated models must be discussed in relation the forecasts.

Part 2
Money Demand and Interest Rates

The Case The goal of this part of the exam is to use cointegration techniques to test the empirical validity of a theoretical equilibrium relation between money velocity and interest rates.

The Data Graphs of the data and relevant transformations must be shown in the exam. It must be noted that the time series for Danish money velocity and interest rates appear to be unit root processes with some long-run co-movements indicating cointegration between them.

Econometric Theory The econometric theory must include the following:

(1) A precise definition and interpretation of the models considered and their properties. Specifically, an interpretation of cointegration must be presented along with a presentation of univariate autoregressive (AR) models used to test for unit roots and a single equation cointegration approach based on the Engle-Granger two-step procedure or the autoregressive distributed lag (ADL) and error-correction models (ECM).

(2) A precise description of the estimator used, in particular a precise account of the assumptions used to derive an estimator.

(3) A precise account of the necessary assumptions for consistent estimation and valid inference. This includes a precise definition of the null hypotheses, test statistics, and asymptotic distributions used to test relevant hypotheses.

(4) The theory must be presented in a logical order and with a consistent and correct notation.

Empirical Results The empirical results must include the following:

(1) A presentation and discussion of the relevant empirical results, so that the reader is able to understand the steps carried out in the process as well as the conclusions made.

(2) A description of the model selection process based on a general-to-specific approach, information criteria, or both.

(3) A discussion of whether the assumptions for consistent estimation and valid inference are satisfied for the estimated models. Specifically, this includes misspecification testing, which must be presented and discussed before statistical testing is carried out.

(4) A clear conclusion to the main question and a discussion of the limitations of the approach used to reach the conclusion. Specifically, the conclusion regarding cointegration between money velocity and interest rates must be presented and the limitations of the single-equation cointegration approach must be discussed in relation to the conclusion.

Part 3
Is there Empirical Evidence of a Taylor Rule with Interest Smoothing under Different Fed Chairs?

The Case The goal of this part of the exam is to use generalized method of moments to test if there is empirical evidence of the Federal Reserve following a Taylor rule with interest rate smoothing under three different chairs.

The Data Graphs of the data and relevant transformations must be shown in the exam. It must be noted that the time series appear quite persistent and potentially non-stationary. Moreover, there seems to be some long-run co-movements between the Federal Funds rate, the inflation rates, and proxies for the output gap.

Econometric Theory The econometric theory must include the following:

(1) A precise definition and interpretation of the model considered and its properties. Specifically, the general method of moments (GMM) used to estimate the parameters of the economic theory.

(2) A precise description of the estimator used, in particular a precise account of the assumptions used to derive an estimator.

(3) A precise account of the necessary assumptions for consistent estimation and valid inference. This includes a precise definition of the null hypotheses, test statistics, and asymptotic distributions used to test relevant hypotheses.

(4) The theory must be presented in a logical order and with a consistent and correct notation.

Empirical Results The empirical results must include the following:

(1) A presentation and discussion of the relevant empirical results, so that the reader is able to understand the steps carried out in the process as well as the conclusions made.

(2) A discussion of whether the assumptions for consistent estimation and valid inference are satisfied for the estimated models.

(3) A robustness analysis of the estimated model.

(4) A conclusion to the main question. Specifically, if there is empirical evidence of a Taylor rule with interest rate smoothing.

(5) A clear conclusion to the main question and a discussion of the limitations of the approach used to reach the conclusion. Specifically, the conclusion regarding empirical evidence of a Taylor rule with interest rate smoothing and the limitations of the single-equation cointegration approach used must be discussed in relation to the conclusion drawn.

Part 4
Theoretical Problems

#4.1 Cointegration and Error-Correction

Consider the system for $x_t = (x_{1t}, x_{2t}, x_{3t})'$, given by,

$$x_{1t} = \rho x_{1t-1} + \delta + \epsilon_{1t} \tag{4.1}$$
$$x_{2t} = b_1 x_{1t-1} + b_3 x_{3t-1} + \epsilon_{2t} \tag{4.2}$$
$$x_{3t} = x_{3t-1} + \epsilon_{3t}, \tag{4.3}$$

for $t = 1, 2, ..., T$, where $|\rho| < 1$, the error terms are uncorrelated, independent over time and normally distributed, i.e. $\epsilon_{it} \sim N(0, \sigma_i^2)$ for $i = 1, 2, 3$, and the initial values $x_0 = (x_{10}, x_{20}, x_{30})'$ are given.

Question 1
By recursive substitution the moving average representation for x_{1t} is found as,

$$x_{1t} = \rho^t x_{10} + (1 + \rho + \rho^2 + ... + \rho^{t-1})\delta + \epsilon_{1t} + \rho\epsilon_{1t-1} + \rho^2\epsilon_{1t-2} + ... + \rho^{t-1}\epsilon_{11}.$$

As $|\rho| < 1$ the process for x_{1t} is stationary, so $x_{1t} \sim I(0)$. First, the effect of the initial value is given by, $\rho^t x_{10}$, which converges to zero as t increases as $\rho^t x_{10} \to 0$ for $t \to \infty$. Second, the effect of the constant term is given by, $(1 + \rho + \rho^2 + ... + \rho^{t-1})\delta$, which converges to $\mu = \frac{\delta}{1-\rho}$ for $t \to \infty$. Finally, the effect of shocks is temporary as $\rho^i \to 0$ for $i \to \infty$. The expectation of x_{1t} conditional on the initial value, x_{10}, is given by, $E[x_{1t}|x_{10}] = \rho^t x_{10} + (1 + \rho + \rho^2 + ... + \rho^{t-1})\delta$, which converges to the unconditional mean, $\mu = E[x_{1t}] = \frac{\delta}{1-\rho}$, as t increases.

Next, the moving average representation for x_{3t} is found as,

$$x_{3t} = x_{30} + \epsilon_{3t} + \epsilon_{3t-1} + ... + \epsilon_{31} = x_{30} + \sum_{i=1}^{t} \epsilon_{3i},$$

which is a unit root process without a drift (or a pure random walk), so $x_{3t} \sim I(1)$ as $\Delta x_{3t} = \epsilon_{3t} \sim I(0)$. First, the initial value stays in the process for x_{3t}. Second, the shocks ϵ_{3t} accumulate into a stochastic trend, $\sum_{i=1}^{t} \epsilon_{3i}$, and they have permanent effects on x_{3t} as $\partial x_{3t}/\partial\epsilon_{3t-k} = \partial x_{3t+k}/\partial\epsilon_{3t} = 1$ for all $k \geq 0$. The expectation of x_{3t} conditional on the initial value, x_{30}, is given by, $E[x_{3t}|x_{30}] = x_{30}$, which does not converge as t increases.

Finally, the moving average representation for x_{2t} is found as,

$$x_{2t} = b_1 \left(\rho^{t-1} x_{10} + \sum_{i=0}^{t-2} \rho^i \delta + \sum_{i=0}^{t-2} \rho^i \epsilon_{1t-1-i} \right) + b_3 \left(x_{30} + \sum_{i=1}^{t-1} \epsilon_{3i} \right) + \epsilon_{2t}$$
$$= \left(b_1 \rho^{t-1} x_{10} + b_3 x_{30} \right) + \left(b_1 \sum_{i=0}^{t-2} \rho^i \delta + b_1 \sum_{i=0}^{t-2} \rho^i \epsilon_{1t-1-i} + \epsilon_{2t} \right) + \left(b_3 \sum_{i=1}^{t-1} \epsilon_{3i} \right),$$

which is a unit root process as it depends on x_{3t-1}, so $x_{2t} \sim I(1)$. The first parenthesis is the effect of the initial values. The effect of the initial value x_{10} goes to zero as t increases, the effect of the initial value x_{30} stays in the process for x_{2t}, and the initial value x_{20} has no impact on x_{2t}. The second parenthesis is a stationary component, which consists of the effect from the constant term in x_{1t} given by δ, the shocks to x_{1t} given by $\epsilon_{11}, ..., \epsilon_{1t-1}$, and the shock ϵ_{2t}. It can be noted that the past shocks $\epsilon_{11}, ..., \epsilon_{2t-1}$ do not have an impact on x_{2t}, so the shocks to x_{2t} only have a temporary effect at time t. The final parenthesis is a stochastic trend given by the accumulated shocks to x_{3t}. The expectation of x_{2t} conditional on the initial values x_{10} and x_{30} is $E[x_{2t}|x_0] = b_1 \rho^{t-1} x_{10} + b_3 x_{30} + b_1 \sum_{i=0}^{t-2} \rho^i \delta$.

Question 2

Two or more unit root processes are cointegrated if there exists a linear combination of the variables which is stationary. Formally, let x_t be a unit root process of dimension k, $x_t \sim I(1)$. The variables in x_t are cointegrated if a k-dimensional vector β exists, $\beta \neq 0$, such that the linear combination $\beta' x_t$ is stationary, $\beta' x_t \sim I(0)$. Thus, the common stochastic trends in the unit root processes cancel out in the linear combination $\beta' x_t$. We refer to β as a cointegration vector.

The cointegration relation β can be interpreted as defining a long-run equilibrium. The variables themselves wander arbitrarily up and down due to the presence of stochastic trends, but they never deviate too much from equilibrium.

Above we showed that both x_{2t} and x_{3t} are unit root processes, i.e. $x_{2t} \sim I(1)$ and $x_{3t} \sim I(1)$, while x_{1t} is a stationary process. Hence, x_{2t} and x_{3t} are cointegrated if there exists a linear combination between them which is stationary. Consider the linear combination,

$$\beta' x_t = \begin{pmatrix} 0 & 1 & -b_3 \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} = x_{2t} - b_3 x_{3t}$$
$$= b_1 x_{1t-1} + b_3 x_{3t-1} + \epsilon_{2t} - b_3 x_{3t}$$
$$= b_1 x_{1t-1} + \epsilon_{2t} - b_3 \Delta x_{3t}$$
$$= b_1 x_{1t-1} + \epsilon_{2t} - b_3 \epsilon_{3t},$$

where we have plugged in from (4.2) and (4.3). As x_{1t} is a stationary process (and ϵ_{2t} and ϵ_{3t} are stationary by definition) the linear combination $\beta' x_t$ is stationary, $\beta' x_t \sim I(0)$, and the variables are cointegrated. The cointegration relation $\beta' x_t$ defines the deviation from the long-run equilibrium between x_{2t} and x_{3t}.

It can be noted that the cointegration vector β is only unique up to a constant factor. If $\beta' x_t \sim I(0)$ then it also holds that $B' x_t \sim I(0)$, where $B = a \cdot \beta$ for any non-zero constant a.

Question 3

The representation theorem by Engle and Granger (1987) states that the variables x_{2t} and x_{3t} cointegrate if and only if there exists an error correction model for either x_{2t}, x_{3t}, or both. Formally a variable is error correcting if its first-difference reacts to the lagged cointegration relation given by $\beta' x_{t-1}$, so that whenever the variables are away from their long-run equilibrium values there are forces pulling them back towards the equilibrium.

As x_{3t} is a pure random walk it is not error correcting. By rewriting (4.2) we get,

$$\begin{aligned} x_{2t} &= b_1 x_{2t-1} + b_3 x_{2t-1} + \epsilon_{2t} \\ \Delta x_{2t} &= -1(x_{2t-1} - b_3 x_{3t-1}) + b_1 x_{2t-1} + \epsilon_{2t}, \end{aligned}$$

which shows that x_{2t} is error-correcting as Δx_{2t} reacts to the lagged deviation from the long-run equilibrium given by $\beta' x_{t-1} = x_{2t-1} - b_3 x_{3t-1}$, as found above. The coefficient in front of the parenthesis is the error-correction coefficient (often referred to as α), which must satisfy $-1 \leq \alpha < 0$ for x_{2t} to error-correct. We note that x_{2t} is instantly error-correcting as $\alpha = -1$.

Question 4

We rewrite the autoregressive distributed lag (ADL) model in (4.4) into an error correction model as,

$$\begin{aligned} x_{2t} &= \delta_2 + \theta_1 x_{2t-1} + \theta_2 x_{2t-2} + \phi_0 x_{1t} + \phi_1 x_{1t-1} + \phi_2 x_{1t-2} \\ &\qquad + \psi_0 x_{3t} + \psi_1 x_{3t-1} + \psi_2 x_{3t-2} + \varepsilon_t && (4.4) \\ \Delta x_{2t} &= \delta_2 + (\theta_1 + \theta_2 - 1) x_{2t-1} - \theta_2 \Delta x_{2t-1} \\ &\qquad + \phi_0 \Delta x_{1t} + (\phi_0 + \phi_1 + \phi_2) x_{1t-1} - \phi_2 \Delta x_{1t-1} \\ &\qquad\quad + \psi_0 \Delta x_{3t} + (\psi_0 + \psi_1 + \psi_2) x_{3t-1} - \psi_2 \Delta x_{3t-1} + \varepsilon_t && (4.5) \\ \Delta x_{2t} &= \delta_2 + \gamma_1 x_{1t-1} + \gamma_2 x_{2t-1} + \gamma_3 x_{3t-1} \\ &\qquad - \theta_2 \Delta x_{2t-1} + \phi_0 \Delta x_{1t} - \phi_2 \Delta x_{1t-1} + \psi_0 \Delta x_{3t} - \psi_2 \Delta x_{3t-1} + \varepsilon_t && (4.6) \\ \Delta x_{2t} &= \delta_2 + \gamma_1 x_{1t-1} + \alpha \left(x_{2t-1} - \beta_3 x_{3t-1} \right) \\ &\qquad - \theta_2 \Delta x_{2t-1} + \phi_0 \Delta x_{1t} - \phi_2 \Delta x_{1t-1} + \psi_0 \Delta x_{3t} - \psi_2 \Delta x_{3t-1} + \varepsilon_t, && (4.7) \end{aligned}$$

where

$$\gamma_1 = \phi_0 + \phi_1 + \phi_2, \quad \gamma_2 = \theta_1 + \theta_2 - 1, \quad \gamma_3 = \psi_0 + \psi_1 + \psi_2, \quad \alpha = \gamma_1, \quad \beta_3 = \frac{\gamma_3}{\gamma_2}.$$

The representation in (4.6) is the linear error correction model and the model in (4.7) is the non-linear error correction model. In (4.7), the parenthesis is the cointegration relation between x_{2t} and x_{3t}, while α is the error-correction coefficient for x_{2t}. It should be noted that (4.4)-(4.7) are different representations of the same model.

The ADL model in (4.2) is a restricted version of the ADL model in (4.4). By imposing the restrictions,

$$\theta_1 = \theta_2 = \phi_0 = \phi_2 = \psi_0 = \psi_2 = 0, \qquad \phi_1 = b_1, \qquad \psi_1 = b_3,$$

the ADL model in (4.4) reduces to (4.2), and likewise the error correction model in (4.7) reduces to the error correction model for x_{2t} found in Question 3.

Question 5

The *PcGive test for no error-correction/no cointegration* is a test of the null $\mathcal{H}_0 : \gamma_1 = 0$ in the linear ECM model in (4.6) against the alternative $\mathcal{H}_A : \gamma_1 < 0$. Under the null x_{2t} is not error-correcting, so the null corresponds to no cointegration. Under the alternative x_{2t} is error-correcting, so that the variables are cointegrated. The test statistics is given by the usual t-ratio, $t_{\gamma_1=0} = \frac{\widehat{\gamma}_1}{s.e.(\widehat{\gamma}_1)}$, but the test statistics asymptotically follows a Dickey-Fuller type distribution which depends on the number of $I(1)$ variables and the deterministic terms in the model.

Question 6

Consider first the test of the null hypothesis $\mathcal{H}_0 : \phi_1 = 0$ against the alternative $\mathcal{H}_A : \phi_1 \neq 0$ in the ADL model (4.4). As ϕ_1 is the coefficient to the stationary variable x_{1t-1}, the null hypothesis can be tested with a t-test and standard inference applies in the sense that the test statistics asymptotically follows a standard normal distribution under the null.

Next, the null hypothesis $\mathcal{H}_0 : \psi_2 = 0$ is a test on a coefficient to the unit-root process x_{3t-2}. But in (4.6), ψ_2 is a coefficient to a mean-zero stationary variable, Δx_{3t-1}, we can test the null hypothesis with a t-test in (4.4) and standard inference applies in the sense that the test statistics asymptotically follows a standard normal distribution under the null. This is an implication of the general result by Sims, Stock, and Watson (1990) that the estimated parameter follows a normal distribution asymptotically if the parameter is a coefficient to a mean zero stationary variable, possibly after a linear transformation of the model.

Question 7

There are three major limitations of using the single-equation cointegration approach based on the ADL model in (4.4).

First, there could in principle exist an error-correction model for all the variables consider, but we only considered the equation for x_{2t}. Thus, the single equation ADL model is, in general, inefficient as the cointegration parameters also enter in the other equations which are not considered. Only in the special case where only x_{2t} is error-correcting is it sufficient to consider only the ADL/ECM model for x_{2t}. But the assumption that only x_{2t} error-corrects can only be empirically tested in a vector error-correction model for all variables. In can be noted that in the model given by (4.1)-(4.3), only x_{2t} is error-correcting, so in that case it is sufficient to estimate a single-equation ADL/ECM model for x_{2t}.

Second, the single-equation ADL/ECM approach implicitly assumes that only one cointegration relation exists between the variables in the model. In general, p unit root variables can have up to $p-1$ cointegration relations between them. If there are more cointegration relations the single equation ADL/ECM model for x_{2t} is not able to separately identify the cointegration relations.

Third, in the single-equation ADL/ECM model we condition on x_{1t} and x_{3t}, so we assume that they are predetermined, $E[x_{1t}\epsilon_{2t}] = 0$ and $E[x_{3t}\epsilon_{2t}] = 0$, which rules out contemporaneous feedback from x_{2t} to x_{1t} and x_{3t}.

#4.2 Forecasting Volatility

Consider the GARCH-X model for y_t given the stationary exogenous variables x_t,

$$y_t = \delta + \epsilon_t \tag{4.8}$$
$$\epsilon_t = \sigma_t z_t \tag{4.9}$$
$$\sigma_t^2 = \varpi + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \phi x_{t-1}^2, \tag{4.10}$$

for $t = 1, 2, ..., T$, where $\varpi > 0$, $\alpha \geq 0$, and $\beta \geq 0$, the innovation z_t is assumed independent over time and standard normally distributed, i.e. $z_t \sim N(0, 1)$, and the initial values are given. The exogenous variable x_t influences the conditional variance of y_t in (4.10). Assume that x_t is given by a stationary first-order autoregressive process,

$$x_t = \rho x_{t-1} + \eta_t, \tag{4.11}$$

where $|\rho| < 1$ and the error term η_t is independent of z_t, independent over time and normally distributed, i.e. $\eta_t \sim N(0, \sigma_x^2)$.

Question 1

The process for y_t is weakly stationary when $0 \leq \alpha + \beta < 1$ given that $|\rho| < 1$ as stated in the exam (so that x_t is a mean-zero stationary process).

As $E[y_t] = \delta$ is constant for all t and $cov(y_t, y_{t-k}) = 0$ for all $k \neq 0$, the process y_t is weakly stationarity if the unconditional variance of ϵ_t, σ^2, is finite, which is fulfilled if ϵ_t^2 has a stationary solution.

First, decompose ϵ_t^2 into the conditional expectation and a surprise in the squared innovation:

$$\epsilon_t^2 = E[\epsilon_t^2|\mathcal{I}_{t-1}] + v_t = E[\sigma_t^2 z_t^2|\mathcal{I}_{t-1}] + v_t = \sigma_t^2 E[z_t^2|\mathcal{I}_{t-1}] + v_t = \sigma_t^2 + v_t,$$

where $E[v_t|\mathcal{I}_{t-1}] = 0$, so v_t is uncorrelated over time. The last steps follow as σ_t^2 is in the information set $\mathcal{I}_{t-1}$ and $z_t \sim N(0, 1)$. Pluggin into (4.7) we get,

$$\begin{aligned}
\sigma_t^2 &= \varpi + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \phi x_{t-1}^2 \\
\epsilon_t^2 - v_t &= \varpi + \alpha\epsilon_{t-1}^2 + \beta(\epsilon_{t-1}^2 - v_{t-1}) + \phi x_{t-1}^2 \\
\epsilon_t^2 &= \varpi + (\alpha + \beta)\epsilon_{t-1}^2 + \phi x_{t-1}^2 + v_t - \beta v_{t-1}.
\end{aligned}$$

This is an ARMA(1,1) model extended with a squared stationary variable x_t, so the process for ϵ_t^2 is stationary when $|\alpha + \beta| < 1$. As $\alpha \geq 0$, $\beta \geq 0$, and $\phi \geq 0$ are required for a non-zero conditional variance, the stationary condition becomes $0 \leq \alpha + \beta < 1$.

Given that the stationarity condition holds, we use $E[\epsilon_t^2] = E[\epsilon_{t-1}^2]$ to derive the unconditional variance,

$$\begin{aligned}
\sigma^2 = E[\epsilon_t^2] &= E[\varpi + (\alpha + \beta)\epsilon_{t-1}^2 + \phi x_{t-1}^2 + v_t - \beta v_{t-1}] \\
&= \varpi + (\alpha + \beta)E[\epsilon_{t-1}^2] + \phi E[x_{t-1}^2] + E[v_t] - \beta E[v_{t-1}] \\
&= \varpi + (\alpha + \beta)E[\epsilon_t^2] + \phi E[x_{t-1}^2] \\
&= \frac{\varpi}{1 - \alpha - \beta} + \frac{\phi}{1 - \alpha - \beta} E[x_{t-1}^2].
\end{aligned}$$

Next, we use (4.8) to find an expression for $E[x_{t-1}^2]$, which we note is equal to the unconditional variance of x_t as it is a mean-zero stationary process. By recursive substitution we find,

$$E[x_{t-1}^2] = E[(\rho x_{t-2} + \eta_{t-1})^2] = ... = E\left[\left(\sum_{i=0}^{\infty} \rho^i \eta_{t-i}\right)^2\right] = (1 + \rho^2 + \rho^4 + ...)\sigma_x^2 = \frac{\sigma_x^2}{1-\rho^2},$$

as η_t is assumed to be independent over time and normally distributed with variance σ_x^2 (so that $E[\eta^2] = \sigma_x^2$ and $E[\eta_t \eta_{t-k}] = 0$ for all $k \neq 0$).

That gives the unconditional variance of the innovation (and hence y_t),

$$\sigma^2 = \frac{\varpi}{1-\alpha-\beta} + \frac{\phi}{1-\alpha-\beta} E[x_{t-1}^2] = \frac{\varpi}{1-\alpha-\beta} + \frac{\phi \sigma_x^2}{(1-\alpha-\beta)(1-\rho^2)}.$$

It should be noted that the unconditional variance of y_t depends on the unconditional variance of x_t.

Question 2

To derive the forecasts it is useful, but not necessary, to formulate the constant term ϖ in terms of the unconditional variance,

$$\begin{aligned}
\sigma^2 &= \frac{\varpi}{1-\alpha-\beta} + \frac{\phi \sigma_x^2}{(1-\alpha-\beta)(1-\rho^2)} \\
\varpi &= (1-\alpha-\beta)\sigma^2 - \frac{\phi \sigma_x^2}{(1-\rho^2)}.
\end{aligned}$$

The conditional variance can then be written as,

$$\begin{aligned}
\sigma_t^2 &= (1-\alpha-\beta)\sigma^2 - \frac{\phi \sigma_x^2}{(1-\rho^2)} + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 + \phi x_{t-1}^2 \\
&= \sigma^2 + \alpha(\epsilon_{t-1}^2 - \sigma^2) + \beta(\sigma_{t-1}^2 - \sigma^2) + \phi\left(x_{t-1}^2 - \frac{\sigma_x^2}{(1-\rho^2)}\right).
\end{aligned}$$

The volatility forecast for $T+1$ conditional on the information set $\mathcal{I}_T$ is given by,

$$\begin{aligned}
\sigma_{T+1|T}^2 &= E[\epsilon_{T+1}^2 | \mathcal{I}_T] \\
&= E[\sigma_{T+1}^2 | \mathcal{I}_T] \\
&= E\left[\sigma^2 + \alpha(\epsilon_T^2 - \sigma^2) + \beta(\sigma_T^2 - \sigma^2) + \phi\left(x_T^2 - \frac{\sigma_x^2}{(1-\rho^2)}\right) \middle| \mathcal{I}_T\right] \\
&= \sigma^2 + \alpha(E[\epsilon_T^2|\mathcal{I}_T] - \sigma^2) + \beta(E[\sigma_T^2|\mathcal{I}_T] - \sigma^2) + \phi\left(E[x_T^2|\mathcal{I}_T] - \frac{\sigma_x^2}{(1-\rho^2)}\right) \\
&= \sigma^2 + \alpha(\epsilon_T^2 - \sigma^2) + \beta(\sigma_T^2 - \sigma^2) + \phi\left(x_T^2 - \frac{\sigma_x^2}{(1-\rho^2)}\right),
\end{aligned}$$

where the last equality holds as ϵ_T^2, σ_T^2, and x_T^2 are all in the information set $\mathcal{I}_T$.

The volatility forecast for $T+2$ conditional on the information set $\mathcal{I}_T$ is given by,

$$
\begin{aligned}
\sigma^2_{T+2|T} &= E[\epsilon^2_{T+2}|\mathcal{I}_T] \\
&= E[\sigma^2_{T+2}|\mathcal{I}_T] \\
&= E\left[\sigma^2 + \alpha(\epsilon^2_{T+1} - \sigma^2) + \beta(\sigma^2_{T+1} - \sigma^2) + \phi\left(x^2_{T+1} - \frac{\sigma^2_x}{(1-\rho^2)}\right)\middle|\mathcal{I}_T\right] \\
&= \sigma^2 + \alpha(E[\epsilon^2_{T+1}|\mathcal{I}_T] - \sigma^2) + \beta(E[\sigma^2_{T+1}|\mathcal{I}_T] - \sigma^2) + \phi\left(E[x^2_{T+1}|\mathcal{I}_T] - \frac{\sigma^2_x}{(1-\rho^2)}\right) \\
&= \sigma^2 + \alpha(\sigma^2_{T+1|T} - \sigma^2) + \beta(\sigma^2_{T+1|T} - \sigma^2) + \phi\left(E[x^2_{T+1}|\mathcal{I}_T] - \frac{\sigma^2_x}{(1-\rho^2)}\right)) \\
&= \sigma^2 + (\alpha+\beta)(\sigma^2_{T+1|T} - \sigma^2) + \phi\left(E[x^2_{T+1}|\mathcal{I}_T] - \frac{\sigma^2_x}{(1-\rho^2)}\right) \\
&= \sigma^2 + (\alpha+\beta)(\sigma^2_{T+1|T} - \sigma^2) + \phi\left(E[(\rho x_T + \eta_{T+1})^2|\mathcal{I}_T] - \frac{\sigma^2_x}{(1-\rho^2)}\right) \\
&= \sigma^2 + (\alpha+\beta)(\sigma^2_{T+1|T} - \sigma^2) + \phi\left(E[\rho^2 x^2_T + \eta^2_{T+1} + \rho x_T \eta_{T+1}|\mathcal{I}_T] - \frac{\sigma^2_x}{(1-\rho^2)}\right) \\
&= \sigma^2 + (\alpha+\beta)(\sigma^2_{T+1|T} - \sigma^2) + \phi\left(\rho^2 E[x^2_T|\mathcal{I}_T] + E[\eta^2_{T+1}|\mathcal{I}_T] + \rho E[x_T\eta_{T+1}|\mathcal{I}_T] - \frac{\sigma^2_x}{(1-\rho^2)}\right) \\
&= \sigma^2 + (\alpha+\beta)(\sigma^2_{T+1|T} - \sigma^2) + \phi\left(\rho^2 x^2_T + \sigma^2_x + \rho x_T E[\eta_{T+1}|\mathcal{I}_T] - \frac{\sigma^2_x}{(1-\rho^2)}\right) \\
&= \sigma^2 + (\alpha+\beta)(\sigma^2_{T+1|T} - \sigma^2) + \phi\left(\rho^2 x^2_T + \sigma^2_x - \frac{\sigma^2_x}{(1-\rho^2)}\right)
\end{aligned}
$$

where it has been used that $E[\eta^2_{T+1}|\mathcal{I}_T] = \sigma^2_x$ and $E[\eta_{T+1}|\mathcal{I}_T] = 0$. Compared to the standard GARCH(1,1) model, the forecasts also depend on the level of x_T.

Question 3

The volatility forecast for $T+k$ conditional on the information set $\mathcal{I}_T$ is given by,

$$
\sigma^2_{T+k|T} = \sigma^2 + (\alpha+\beta)(\sigma^2_{T+k-1|T} - \sigma^2) + \phi\left(E[x^2_{T+k-1}|\mathcal{I}_T] - \frac{\sigma^2_x}{(1-\rho^2)}\right),
$$

which converges towards the unconditional variance σ^2 for $k \to \infty$.

It can be noted that,

$$
\begin{aligned}
E[x^2_{T+k-1}|\mathcal{I}_T] &= E[(\rho x_{T+k-2} + \eta_{T+k-1})^2|\mathcal{I}_T] \\
&= E\left[\left(\rho^{k-1}x_T + \sum_{i=1}^{k-1}\rho^{k-1-i}\eta_{T+i}\right)^2\middle|\mathcal{I}_T\right] \\
&= E\left[\rho^{2(k-1)}x^2_T + \sum_{i=1}^{k-1}\rho^{2(k-1-i)}\eta^2_{T+i}\middle|\mathcal{I}_T\right] \\
&= \rho^{2(k-1)}E[x^2_T|\mathcal{I}_T] + \sum_{i=1}^{k-1}\rho^{2(k-1-i)}E[\eta^2_{T+i}|\mathcal{I}_T] \\
&= \rho^{2(k-1)}x^2_T + \left(\sum_{i=1}^{k-1}\rho^{2(k-1-i)}\right)\sigma^2_x \\
&\to \frac{\sigma^2_x}{(1-\rho)^2} \text{ for } k \to \infty \text{ as } |\rho| < 1,
\end{aligned}
$$

where all cross-products with conditional expectation of zero have been left out in the third expression. Hence, $E[x^2_{T+k-1}|\mathcal{I}_T] - \frac{\sigma_x^2}{(1-\rho^2)} \to 0$, and as $0 \leq \alpha + \beta < 1$ the volatility forecast for $T+k$ conditional on the information set $\mathcal{I}_T$ converges towards the unconditional variance, σ^2. The intuition is that the information included in the information set $\mathcal{I}_T$ becomes less and less relevant as the forecasting horizon increases and therefore the volatility forecast converges towards the unconditional variance, which can be interpreted as the volatility forecast based on the empty information set.

Question 4

To estimate the parameters in the model (4.5)-(4.7), $\theta = (\delta, \varpi, \alpha, \beta, \phi)'$, by maximum likelihood conditional on $x_0, x_1, ..., x_T$, we use the assumption of conditional normality of ϵ_t,

$$\epsilon_t = \sigma_t z_t, \quad z_t \sim N(0,1),$$

or alternatively that $\epsilon_t|\mathcal{I}_{t-1} \sim N(0, \sigma_t^2)$.

The likelihood contribution can be written in terms of the observed data as,

$$\begin{aligned}
&L_t(\delta, \varpi, \alpha, \beta, \phi | y_t, x_t, y_{t-1}, x_{t-1}, ..., y_1, x_1, y_0, x_0) \\
&= \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{1}{2}\frac{\epsilon_t^2}{\sigma_t^2}\right) \\
&= \frac{1}{\sqrt{2\pi(\varpi + \alpha(y_{t-1}-\delta)^2 + \beta\sigma_{t-1}^2 + \phi x_{t-1}^2)}} \exp\left(-\frac{1}{2}\frac{(y_t-\delta)^2}{\varpi + \alpha(y_{t-1}-\delta)^2 + \beta\sigma_{t-1}^2 + \phi x_{t-1}^2}\right),
\end{aligned}$$

for $t = 1, 2, ..., T$, where σ_t^2 can be calculated recursively for a given set of parameters θ and given assumed initial values for σ_0^2 and ϵ_0^2.

The likelihood function is given by the product of the likelihood contributions over $t = 1, ..., T$. The maximum likelihood estimator is found by maximizing the (log-) likelihood function with respect to the parameters θ. As we cannot solve the likelihood equations analytically, the maximum likelihood estimator is found by numerical optimization.

It can be noted that the maximum likelihood estimator can be based on other conditional distributions of ϵ_t. For example, a fat-tailed distribution can be used, such as a student $t(v)$ distribution where the degrees of freedom, v, can be treated as a parameter that can be estimated from the data.