MFE 230E Problem Set 4 - Solutions

Spring 2020

Note: This document only presents answers. Cf. the accompanying Jupyter Notebook for details and the corresponding code. Please email mfe230e@gmail.com if there are any corrections you would like to make to these solutions.

Problem 1

Remark: Note that we are in the case where $R_{t+j} = R$ for all j, i.e. we are under the assumption that returns are constant ($\approx$ returns are random walk, the "old-school" model for market efficiency).

One can check that the following equality holds:

$$\frac{1}{R}\sum_{i=0}^{K}\left[\frac{D_{t+i+1}}{(1+R)^{i+1}} - \frac{D_{t+i}}{(1+R)^{i}}\right] = \frac{1}{R}\left(\frac{D_{t+K+1}}{(1+R)^{K+1}} - D_t\right) \tag{1}$$

To verify this, notice as you write the sum explicitly that all terms cancel out except the two above. Apply a conditional expectation on both sides:

$$\mathbb{E}_t\left[\frac{1}{R}\sum_{i=0}^{K}\left[\frac{D_{t+i+1}}{(1+R)^{i+1}} - \frac{D_{t+i}}{(1+R)^{i}}\right]\right] = \frac{1}{R}\left(\mathbb{E}_{\approx}\left[\frac{D_{t+K+1}}{(1+R)^{K+1}}\right] - D_t\right) \tag{2}$$

where we used the fact that $-D_t$ and R are known at time t.

Take the limit as $K \to \infty$:

$$\mathbb{E}_t\left[\frac{1}{R}\sum_{i=0}^{\infty}\left[\frac{D_{t+i+1}}{(1+R)^{i+1}} - \frac{D_{t+i}}{(1+R)^{i}}\right]\right] = -\frac{D_t}{R} + \frac{1}{R}\lim_{K\to\infty}\mathbb{E}_t\left[\frac{D_{t+K+1}}{(1+R)^{K+1}}\right] \tag{3}$$

As long the last term converges to zero as $K \to \infty$ (i.e. transversality-type condition), which we most of the time assume, we get:

$$-\frac{D_t}{R} = \frac{1}{R}\mathbb{E}_t\sum_{i=0}^{\infty}\left[\frac{D_{t+i+1}}{(1+R)^{i+1}} - \frac{D_{t+i}}{(1+R)^{i}}\right]. \tag{4}$$

Our assumption about prices can be written in the following way:

$$P_t = \frac{1}{(1+R)}\mathbb{E}_t\sum_{i=0}^{\infty}\frac{D_{t+i+1}}{(1+R)^{i}}. \tag{5}$$

Adding Equations (4) and (5), we get the desired result:

$$P_t - \frac{D_t}{R} = \frac{1}{R(1+R)}\mathbb{E}_t\sum_{i=0}^{\infty}\left[\frac{RD_{t+i+1}}{(1+R)^{i}} + \frac{D_{t+i+1}}{(1+R)^{i}} - \frac{(1+R)D_{t+i}}{(1+R)^{i}}\right] = \frac{1}{R}\mathbb{E}_t\sum_{i=0}^{\infty}\frac{\Delta D_{t+i+1}}{(1+R)^{i}} \tag{6}$$

Problem 2

From the definition of log returns, we get:

$$\begin{aligned}
r_{t+1} &= \log\left(1 + R_{t+1}\right) \\
&= \log\left(\frac{P_{t+1} + D_{t+1}}{P_t}\right) \\
&= \log\left(\frac{P_{t+1}}{P_t}\left(1 + \frac{D_{t+1}}{P_{t+1}}\right)\right) \\
&= p_{t+1} - p_t + \log(1 + e^{d_{t+1} - p_{t+1}}) \qquad (1)
\end{aligned}$$

Now consider the linear approximation (i.e. first-order Taylor expansion ignoring higher-order terms) around $x = 0$:

$$\log(1 + e^x) \approx k + (1 - \rho)x \qquad (2)$$

Plugging (2) into (1), we get:

$$r_{t+1} \approx k + \rho p_{t+1} + (1 - \rho)d_{t+1} - p_t \Rightarrow p_t = k + \rho p_{t+1} + (1 - \rho)d_{t+1} - r_{t+1} \qquad (3)$$

where:

$$\begin{cases} \rho = \frac{1}{1 + e^{\overline{d - p}}} \\ k = -\log(\rho) - (1 - \rho)\log\left(\frac{1}{\rho} - 1\right) \end{cases}$$

Equation (5) tells us that p_t is (approximately) a linear function of p_{t+1}, d_{t+1} and r_{t+1}. Using the lag notation we get:

$$(1 + \rho L^{-1})p_t \approx k + (1 - \rho)d_{t+1} - r_{t+1} \Rightarrow p_t = \frac{k}{1 - \rho} + \sum_{j=0}^{\infty} \rho^j [(1 - \rho)d_{t+1+j} - r_{t+1+j}] \qquad (4)$$

Taking expectations from Equation (6), we prove the first part of the problem. To prove the second part, note that, following the same log as in **Problem 1**:

$$-d_t = \sum_{j=0}^{\infty} \rho^j [\rho d_{t+1+j} - d_{t+j}] \qquad (5)$$

Adding Equations (6) and (7), we get:

$$p_t - d_t \approx \frac{k}{1 - \rho} + \sum_{j=0}^{\infty} \rho^j [(1 - \rho)d_{t+1+j} - r_{t+1+j} + \rho d_{t+1+j} - d_{t+j}] = \sum_{j=0}^{\infty} \rho^j [\Delta d_{t+1+j} - r_{t+1+j}] \qquad (6)$$

Taking expectations in Equation (8), we get:

$$p_t - d_t \approx \frac{k}{1 - \rho} + \mathbb{E}_t \sum_{j=0}^{\infty} \rho^j [\Delta d_{t+1+j} - r_{t+1+j}]$$

Constants are unimportant so most of the time we ignore them and get:

$$p_t - d_t \approx \mathbb{E}_t \sum_{j=0}^{\infty} \rho^j [\Delta d_{t+1+j} - r_{t+1+j}]$$

Problem 3

Let $\boldsymbol{X}_t = (x_{1,t}, x_{2,t})'$ and $x_{1,t}, x_{2,t} \sim \mathcal{I}(1)$. $x_{1,t}$ and $x_{2,t}$ are cointegrated with cointegration vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)'$, so that $z_t = \boldsymbol{\alpha}'\boldsymbol{X}_t \sim \mathcal{I}(0)$.

Consider the VECM(1):

$$\begin{aligned}
\Delta x_{1,t} &= \gamma_1\, z_{t-1} + \phi_{11}\Delta x_{1,t-1} + \phi_{12}\Delta x_{2,t-1} + \epsilon_{1,t} \\
\Delta x_{2,t} &= \gamma_2\, z_{t-1} + \phi_{21}\Delta x_{1,t-1} + \phi_{22}\Delta x_{2,t-1} + \epsilon_{2,t} \\
\Leftrightarrow\ \Delta \boldsymbol{X}_t &= \boldsymbol{\gamma}\boldsymbol{\alpha}'\boldsymbol{X}_{t-1} + \boldsymbol{\Phi}_1\Delta\boldsymbol{X}_{t-1} + \boldsymbol{\epsilon}_t
\end{aligned}$$

$$\boldsymbol{\gamma} = \begin{pmatrix}\gamma_1\\ \gamma_2\end{pmatrix},\ \boldsymbol{\Phi}_1 = \begin{pmatrix}\phi_{11} & \phi_{12}\\ \phi_{21} & \phi_{22}\end{pmatrix}.$$

Remark: By writing the VECM this way, in particular the $\boldsymbol{\gamma}\boldsymbol{\alpha}'\boldsymbol{X}_{t-1}$ part, we are already implicitly imposing the restriction that the system is cointegrated with cointegration vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)'$. If we did not impose this restriction, by deriving a VECM from a standard VAR(p) representation in levels, the first term would be $\boldsymbol{\Pi}X_{t-1}$ with $\boldsymbol{\Pi} \equiv -\boldsymbol{\psi}(1) = -(\boldsymbol{I}-\boldsymbol{\Psi}_1-\boldsymbol{\Psi}_2-\ldots-\boldsymbol{\Psi}_p)$. Writing $\boldsymbol{\Pi}X_{t-1}$ as $\boldsymbol{\gamma}\boldsymbol{\alpha}'\boldsymbol{X}_{t-1}$ imposes the restriction that we have one cointegration vector $\boldsymbol{\alpha}$ (formally this can be written as a condition on the rank of the matrices of the VAR or equivalent MA representation, which implies that it can be written in this way). This is the equivalent of Equation (19.1.35) p. 579 in Hamilton: $\boldsymbol{\Phi}(1) = \boldsymbol{B}\boldsymbol{A}'$ in our specific case with one cointegration vector (note that $\boldsymbol{\Phi}(1)$ in Hamilton is the VAR matrix, i.e. our $\boldsymbol{\Psi}(1)$ here). All of this is the essence of the Granger representation theorem. More details can be found in Hamilton (1994, Chapter 19, p. 579-580 and p. 574-575), as well as in the textbook by Lütkepohl (2005, Ch. 6, p. 248-249), among many others. The exposition in Cochrane's time series textbook (2005) is also quite approachable. The consequence is that whatever restrictions we end up finding in the expression as we derive the VAR representation will directly stem from the cointegration relationship that we implicitly embedded.

(a) From above:

$$\begin{aligned}
x_{1,t} &= x_{1,t-1} + \gamma_1\, z_{t-1} + \phi_{11}\Delta x_{1,t-1} + \phi_{12}\Delta x_{2,t-1} + \epsilon_{1,t} \\
&= x_{1,t-1} + \gamma_1\left(\alpha_1 x_{1,t-1} + \alpha_2 x_{2,t-1}\right) + \phi_{11}\left(x_{1,t-1} - x_{1,t-2}\right) + \phi_{12}\left(x_{2,t-1} - x_{2,t-2}\right) + \epsilon_{1,t} \\
&= \left(1 + \gamma_1\alpha_1 + \phi_{11}\right) x_{1,t-1} + \left(\gamma_1\alpha_2 + \phi_{12}\right) x_{2,t-1} - \phi_{11}x_{1,t-2} - \phi_{12}x_{2,t-2} + \epsilon_{1,t}
\end{aligned}$$

Similary:

$$\begin{aligned}
x_{2,t} &= x_{2,t-1} + \gamma_2\, z_{t-1} + \phi_{21}\Delta x_{1,t-1} + \phi_{22}\Delta x_{2,t-1} + \epsilon_{2,t} \\
&= x_{2,t-1} + \gamma_2\left(\alpha_1 x_{1,t-1} + \alpha_2 x_{2,t-1}\right) + \phi_{21}\left(x_{1,t-1} - x_{1,t-2}\right) + \phi_{22}\left(x_{2,t-1} - x_{2,t-2}\right) + \epsilon_{2,t} \\
&= \left(\gamma_2\alpha_1 + \phi_{21}\right) x_{1,t-1} + \left(1 + \gamma_2\alpha_2 + \phi_{22}\right) x_{2,t-1} - \phi_{21}x_{1,t-2} - \phi_{22}x_{2,t-2} + \epsilon_{2,t}
\end{aligned}$$

Therefore:

$$\boldsymbol{\Psi}(L)\boldsymbol{X}_t = \left(\boldsymbol{I} - \boldsymbol{\Psi}_1 L - \boldsymbol{\Psi}_2 L^2\right)\boldsymbol{X}_t = \boldsymbol{\epsilon}_t$$

where:

$$\boldsymbol{\Psi}_1 = \begin{pmatrix} 1+\gamma_1\alpha_1+\phi_{11} & \gamma_1\alpha_2+\phi_{12} \\ \gamma_2\alpha_1+\phi_{21} & 1+\gamma_2\alpha_2+\phi_{22} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Psi}_2 = \begin{pmatrix} -\phi_{11} & -\phi_{12} \\ -\phi_{21} & -\phi_{22} \end{pmatrix}$$

This is a $\underline{\text{VAR}(p) \text{ with } p=2}$, as expected from the Granger representation theorem.

(b) Let us now assume that $\boldsymbol{\alpha} = (1,-1)'$, i.e. $\alpha_1 = 1$ and $\alpha_2 = -1$. The matrices for the VAR in levels become:

$$\boldsymbol{\Psi}_1 = \begin{pmatrix} 1+\gamma_1+\phi_{11} & -\gamma_1+\phi_{12} \\ \gamma_2+\phi_{21} & 1-\gamma_2+\phi_{22} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Psi}_2 = \begin{pmatrix} -\phi_{11} & -\phi_{12} \\ -\phi_{21} & -\phi_{22} \end{pmatrix}$$

Adding all elements from the first rows of $\boldsymbol{\Psi}_1$ and $\boldsymbol{\Psi}_2$ yields:

$$1+\gamma_1+\phi_{11}-\gamma_1+\phi_{12}-\phi_{11}-\phi_{12}=1$$

Adding all elements from the second rows of $\boldsymbol{\Psi}_1$ and $\boldsymbol{\Psi}_2$ yields:

$$\gamma_2+\phi_{21}+1-\gamma_2+\phi_{22}-\phi_{21}-\phi_{22}=1$$

Conclusion: the restrictions imposed on the VAR coefficients by the VECM(1) when $\boldsymbol{\alpha} = (1,-1)'$ are that the rows of $\boldsymbol{\Psi}_1 + \boldsymbol{\Psi}_2$ both sum to 1.

Two remarks:

1. This remains true as long as $\alpha_2 = -\alpha_1$.
2. Again, as explained in the remark above: there is no need to look for additional restrictions to impose of the form of Equation (19.1.35) in Hamilton (p. 579). Indeed, by writing the VECM with the $\boldsymbol{\gamma}\boldsymbol{\alpha}'\boldsymbol{X}_{t-1}$ as first term, we already implicitly imposed the cointegration restriction so that anything that we derived stemmed from that.

Problem 4

Figures 1 and 2 below show the simulated samples. Note that even when $\boldsymbol{\gamma} = (0, 0)$, the variables end up close to each other by pure luck in this case.

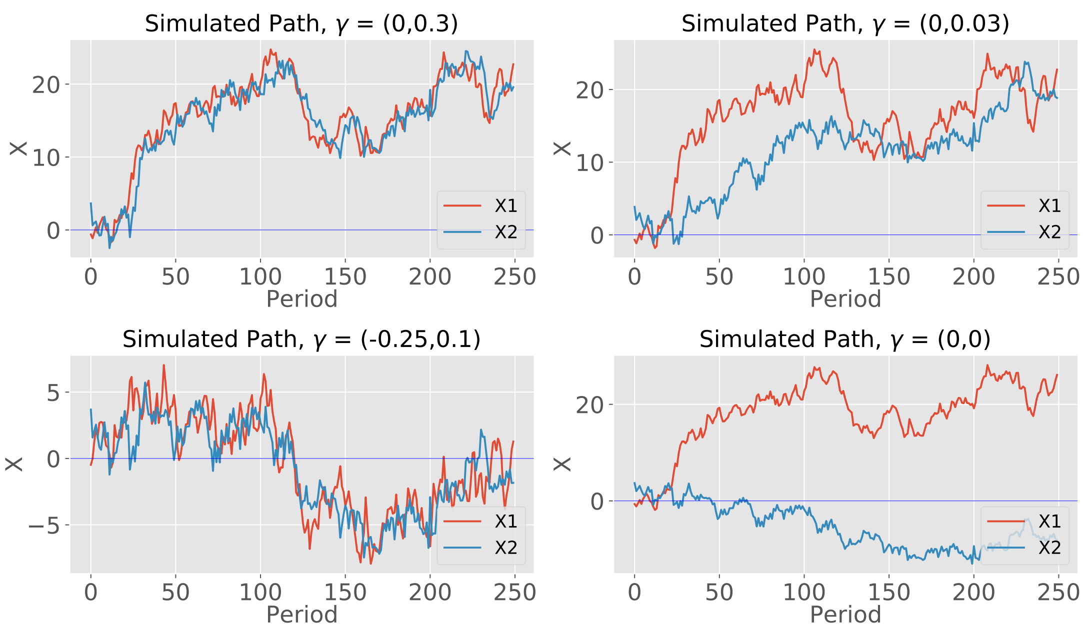

Figure 1: Sample of $x_{1,t}, x_{2,t}$ for $T = 250$

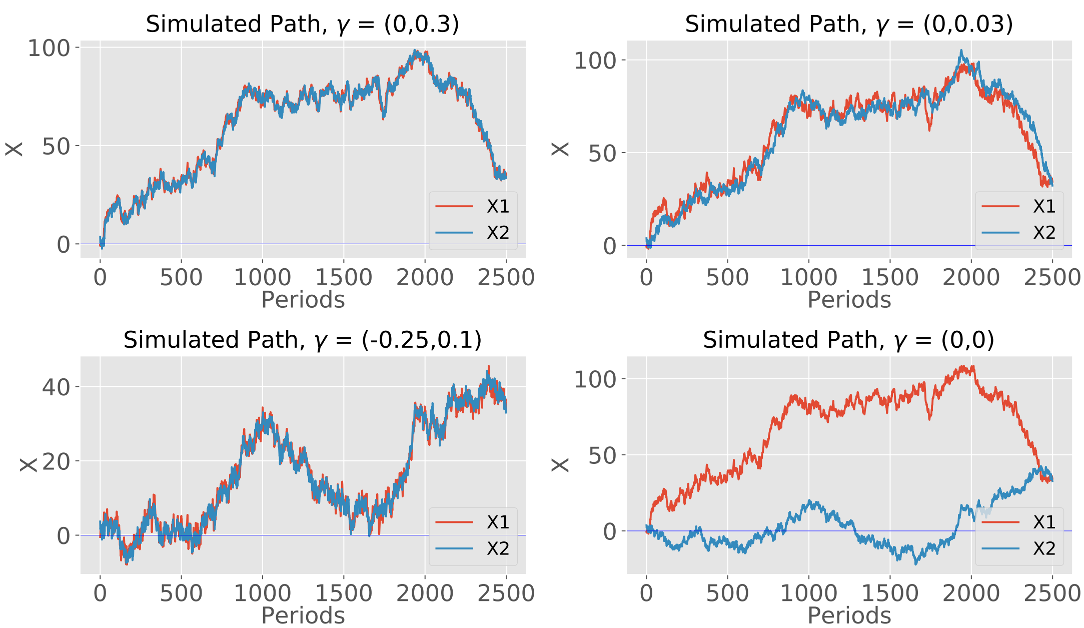

Figure 2: Sample of $x_{1,t}, x_{2,t}$ for $T = 2500$

(a) General discussion Regardless of the values of $\boldsymbol{\gamma}$, the estimation of a VAR in levels yields consistent estimates. This is true despite $x_{1,t}, x_{2,t}$ being $\mathcal{I}(1)$ because a VAR always includes the own lags of each variable, which are cointegrated with the variable themselves, and therefore deal with the unit root (with $\mathcal{I}(1)$ variables, the VAR in levels is in fact super-consistent). Whether it is also efficient (i.e. yielding the smallest standard errors) depends on the specific case. Note also that with $\mathcal{I}(1)$ variables, the OLS estimator of the VAR has a non-standard distribution (typically something like a Dickey-Fuller distribution). Despite those two caveats, the consistency is the key property, and the VAR remains well-defined in all cases. This makes it a robust estimator.

For $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3$, $x_{1,t}, x_{2,t}$ are cointegrated, so estimating a VECM is theoretically the most appropriate approach. Indeed, in that case, a VECM is consistent and efficient. The reason why the VAR in levels (estimated in the usual way, i.e. by OLS) is not as efficient in this case is that as we have seen, cointegration implies some restrictions on the VAR coefficients. If those restrictions are not imposed in the estimation (e.g. by estimating the VAR with restrictions by MLE), then the VAR in levels will be less efficient because we do not use all the information about the structure of the problem that we could. On the other hand, in that case, the VECM is efficient because it already integrates those restrictions, and is also easy to estimate (OLS). In practice, for $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_3$, a VECM is likely to be the best approach given that the cointegration relationships are sufficiently strong to not be due just to noise[1]. For $\boldsymbol{\gamma}_2$, this is less clear cut, because if we find estimates $\hat{\boldsymbol{\gamma}}_2$ in this ballpark, the difference from $\mathbf{0}$ could be mostly noise. In that case, the VAR in levels, even though theoretically less efficient, is safer because it is robust to both cointegration or no cointegration. For those three values of $\boldsymbol{\gamma}$, the VAR in differences is misspecified and inconsistent. Indeed, the existence of a cointegration vector means that there exists no finite-order VAR representation of the process in first difference. Because $\boldsymbol{\gamma}_2$ is close to $\mathbf{0}$, it might be that in that case the VAR($\Delta \boldsymbol{X}_t$) yields acceptable estimates, but this should theoretically be avoided.

For $\boldsymbol{\gamma}_4$, $x_{1,t}, x_{2,t}$ are not cointegrated, i.e. there is no linear combination of the two variables that is $\mathcal{I}(0)$. In that case, a VECM is misspecified and inconsistent and should be avoided. The VAR($\Delta \boldsymbol{X}_t$) is consistent and efficient, and is therefore most appropriate. The VAR($\boldsymbol{X}_t$) is consistent, but not efficient. This is because in this case, we know that both variables are pure non-cointegrated random walks so that differencing them is the appropriate thing to do. The VAR($\Delta \boldsymbol{X}_t$) does it, which explains why it is most efficient, while the VAR($\boldsymbol{X}_t$) does not, and estimate the coefficient equal to 1 in front of each lag, instead of imposing it, and is therefore more robust but less efficient.[2]

A final point before turning to the results: in practice, we cannot necessarily distinguish for sure cases like $\boldsymbol{\gamma}_2$ and $\boldsymbol{\gamma}_4$, so that using a VECM or VAR($\Delta \boldsymbol{X}_t$) is more risky. If it turns out that we are in the case of $\boldsymbol{\gamma}_2$, then using a VECM is theoretically the most appropriate, while using a VAR($\Delta \boldsymbol{X}_t$) is misspecified and yields inconsistent estimates. If it turns out that we are in the case of $\boldsymbol{\gamma}_4$, a VECM would be misspecified and inconsistent, while VAR($\Delta \boldsymbol{X}_t$) would be most appropriate. Because in such cases there is often little way to be sure, turning to a VAR($\boldsymbol{X}_t$) is a safer option. This is the reason why several time-series econometricians, including Nobel laureate Chris Sims, often advocates for VAR in levels in practice.

[1] What we mean by that is that if we find $\hat{\boldsymbol{\gamma}}_1$ and $\hat{\boldsymbol{\gamma}}_3$ in this ballpark, they are likely to not be due just to noise.
[2] This is quite an heuristic argument. For formal details, cf. Hamilton (1994) or others.

(b) Cf. the enclosed Jupyter notebook for the details of the results, which are broadly in line with the remarks above. Here are the highlights:

- For all cases (i.e. all $\boldsymbol{\gamma}_i$), the estimates from the VAR in levels are broadly close to true values. This is even more true as T increases, consistent with the OLS estimator being super-consistent.

- For $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_3$, the estimates from the VECM are also broadly close to the truth, even though increasing T is sometimes needed to get more precise estimates. The estimated $\hat{\boldsymbol{\alpha}}$ are very close to $\boldsymbol{\alpha}$ even for small T, consistent with OLS being super-consistent for estimating the cointegration relationship. For those $\boldsymbol{\gamma}$, the estimates from the VAR in first differences are very off for some coefficients, and increasing T does nothing to change this, given that the OLS estimator of the VAR in first differences is not consistent.

- For $\boldsymbol{\gamma}_2$, the VECM estimation encounters more difficulties for $T = 250$. For instance $\hat{\alpha}_2 = -0.84$, somewhat far from -1, and the estimated $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\Phi}}$ are also not particularly on point. Increasing the sample size to $T = 2500$ solves those issues, consistent with the fact that the system is theoretically cointegrated, but with a somewhat weak cointegration relationship, which requires more data to be recovered precisely by the VECM. For this value of $\boldsymbol{\gamma}$, the VAR in first differences is theoretically misspecified and inconsistent, but actually turns out to do somewhat ok in practice in particular as T gets larger. Again, this is consistent with the system being cointegrated but with a weak cointegration relationship. Note that despite this case, it is often very dangerous to estimate VAR in first differences when there is a risk for the system to be cointegrated.

- $\boldsymbol{\gamma}_4$ is the flip-side of $\boldsymbol{\gamma}_2$. In that case, the VECM is theoretically misspecified and inconsistent, and it does do poorly when $T = 250$. In particular, it looks like that $\hat{\gamma}_1$ is statistically different from zero (p-value $= 0.01$), which is in fact not true. As the sample size increases to $T = 2500$, the VECM does slightly "better" in the sense that the $\hat{\boldsymbol{\Phi}}$ is estimated somewhat precisely, but even though $\hat{\gamma}_1 \approx 0$, it is still wrongly statistically different from zero. $\hat{\boldsymbol{\alpha}}$ is also very imprecisely estimated, not surprisingly given that there is no cointegration relationship in truth. In summary, the VECM should not be used in that case. Conversely, $\boldsymbol{\gamma}_4 = (0, 0)$ is exactly the case in which the VAR in first differences does well, as it is consistent and efficient. This is confirmed in the results, even though increasing T still helps quite a bit in getting estimates precisely close to the truth.

- All those statements were quite heuristic and could be made more precise in multiple ways, for instance by testing the hypothesis that each coefficient is equal to the truth (that we know here because we simulated the data ourselves).

- For the VARs, the appropriate number of lags was selected using the BIC. One could do that for the VECM as well. The resulting optimal number of lags is $p = 2$ for all VARs in levels, which is consistent with the true system. It varies quite a bit for the VAR in first differences between $p = 1$ and $p = 6$, consistent with this estimation method being inappropriate and instable in most cases, even though the BIC does pick $p = 1$ for $\boldsymbol{\gamma} = 4$, which is consistent with the truth in the only case in which $\text{VAR}(\Delta \boldsymbol{X}_t)$ is appropriate.

- The restrictions on the coefficients of the VAR in levels are verified for all cases. Note that this is no longer true when $\alpha_2 \neq -\alpha_1$.
- In terms of what each method implies for the cointegration mechanism:
  - The VAR in levels allows for any cointegration relationship, even though it does not use all restrictions when there exists one and is therefore less efficient (except if estimated by MLE with restrictions).
  - The VECM specifically embeds the cointegration mechanism, i.e. it includes relationships between variables both in the short-term and long-run. It is therefore efficient when such a cointegration mechanism exists, but misspecified and inconsistent when it does not.
  - The VAR in first differences restrict the system to not have any cointegration relationship. When this is the case, it is efficient, but it is clearly misspecified and inconsistent when a cointegration relationship exists in the data.

In conclusion:

- The VAR in levels does well in all cases.
- The VECM does well even for small T for cointegrated systems for which the cointegration relationship is strong enough, but requires larger T to do well for cointegrated systems with weak cointegration relationships.
- The VAR in first differences does well only for non-cointegrated non-stationary systems (but can still require large T), and does very poorly in all other cases.

A note on BIC: we can write it, for a given lag number m as

$$BIC(m) = \ln |\tilde{\Sigma}_\epsilon(m)| + \frac{\ln T}{T} m K^2$$

where $\tilde{\Sigma}_\epsilon(m)$ is the MLE estimator of the variance-covariance matrix Σ_ϵ obtained for a VAR(m) model. Cf. Lütkepohl (2005, p. 150) or Hamilton (1994), among others. $\tilde{\Sigma}_\epsilon(m)$ is also specified in those references. You can then compute the BIC manually. Note that this is the BIC for a VAR model for all equations taken together, and is therefore the relevant one. This is what is returned to you by *statsmodel*, and is different from the BIC of each equation taken separately.

(c)(d)(e) We run long-horizon regressions of the form:

$$\Delta x_{i,t+1} + \Delta x_{i,t+2} + \ldots + \Delta x_{i,t+h} = \alpha_h + \beta_h z_t + u_{i,t+h,h}$$

for $i = 1, 2$ and h ranging from 1 to 20. Several specifications of the variance-covariance matrix are used: standard OLS errors, White-corrected standard errors, Newey-West standard errors, and

Hansen-Hodrick standard errors. OLS standard errors are obviously not adapted as they would impose homoskedasticity, and White-corrected standard errors are not sufficient either because even though they allow for heteroskedasticity, they do not allow for serial correlation in the errors. Even though those are not showed here in the interest of space, your answer for Question (d) should show that errors are serially correlated, which is expected due to the fact that we use overlapping observations. As a result, only Newey-West or Hansen-Hodrick (or their variations) are adequate here. In the attached code, you can see in the Tables that not using the latter two would lead to significant mistakes in the inference. For instance, for $T = 250$, it would often wrongly seem that x_1 is statistically predictable, while it is not when we use corrected standard errors.

Figures 3, 4, 5 and 6 show the resulting long-horizon regression coefficients, together with 95% confidence interval built from Hansen-Hodrick standard errors, while Figures 7, 8, 9 and 10 show the corresponding R^2.

Results are as expected:

- Variable x_1 is mostly not predictable for $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_4$ with coefficients being close to zero (zero is well within the confidence interval), and R^2 being low. Note that for $\boldsymbol{\gamma}_2$ and $\boldsymbol{\gamma}_4$, for which the system is either close to being or actually is composed of two independent random walks, the estimation has difficulties for the smaller sample ($T = 250$) and find evidence of predictability even though the true γ are close or equal to 0, which should mean no predictability whatsoever (e.g. the R^2 are quite high for $\boldsymbol{\gamma}_4$). This is both a result of the smaller sample, as this disappears when $T = 2500$, and of the fact that even though the true γ are zero, there can still appear some apparent (even though illusory) patterns of predictability in a given sample, in particular of small size.

- For γ_3, variable x_1 is strongly predictable, as expected, and stays so for many periods. This is apparent both from the coefficients, which are significantly different from zero (both statistically and economically), and the R^2. This is for both $T = 250$ and $T = 2500$.

- For variable x_2, we find clear patterns of predictability both in terms of the coefficients and the R^2 for $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3$. This is so for the two sample sizes $T = 250$ and $T = 2500$, and the coefficient magnitudes are consistent with the value of the true $\boldsymbol{\gamma}$. In particular, coefficients are largest for $\boldsymbol{\gamma}_1$ for which the γ for x_2 is 0.3, followed by $\boldsymbol{\gamma}_3$ and $\boldsymbol{\gamma}_2$. The R^2s broadly follow. Interestingly, the value for the coefficient for $\boldsymbol{\gamma}_2$ increases somewhat gradually over time, in particular for the long sample.

- For $\boldsymbol{\gamma}_4$, there is broadly no sign of predictability for x_2 for either of the sample sizes.

- Note that all those results were obtained by using the true cointegration vector $\alpha = (1, -1)$. They could be repeated using the estimated $\hat{\alpha}$ from before. Because the estimation of $\hat{\alpha}$ was broadly accurate in practice, this should not make a big difference, but it can be an interesting further exercise to consider.

Figure 3: Regression coefficients (β_h) for x_1 ($T = 250$)

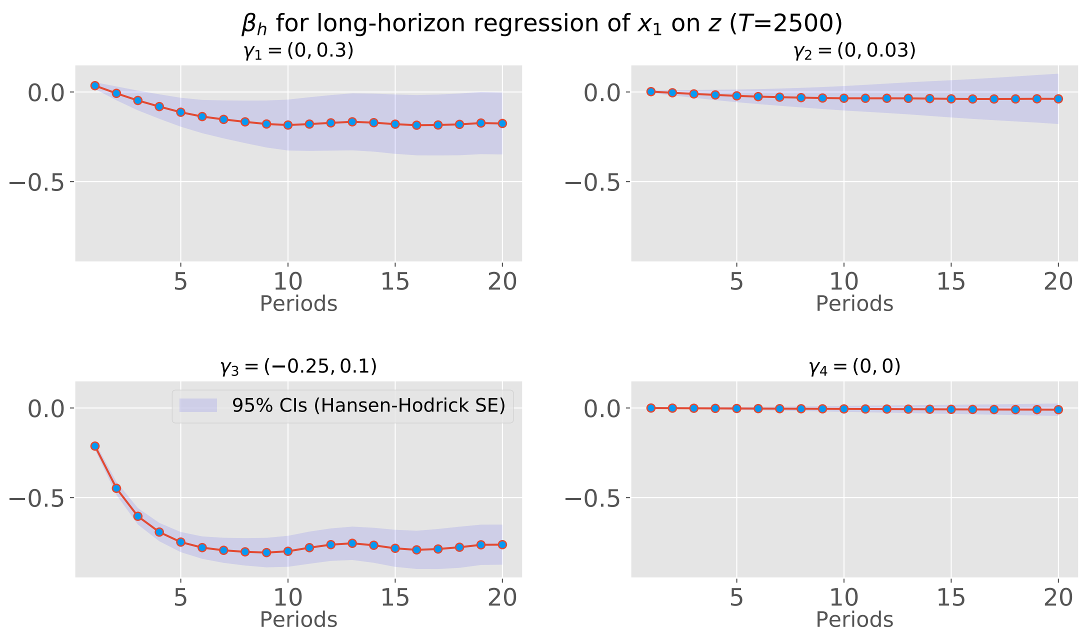

Figure 4: Regression coefficients (β_h) for x_1 ($T = 2500$)

Figure 5: Regression coefficients (β_h) for x_2 ($T = 250$)

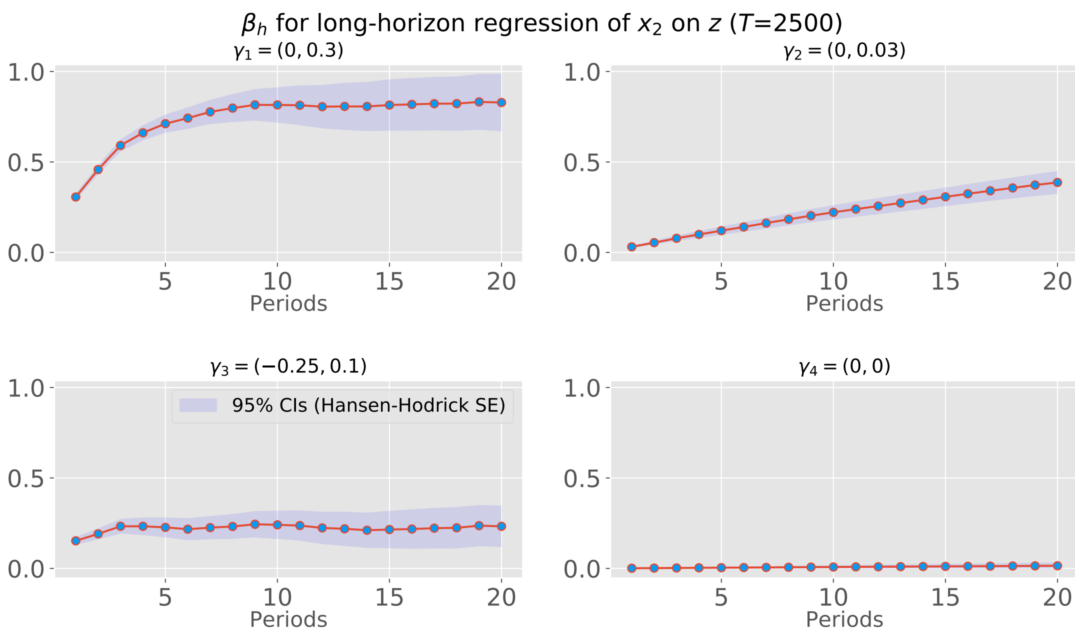

Figure 6: Regression coefficients (β_h) for x_2 ($T = 2500$)

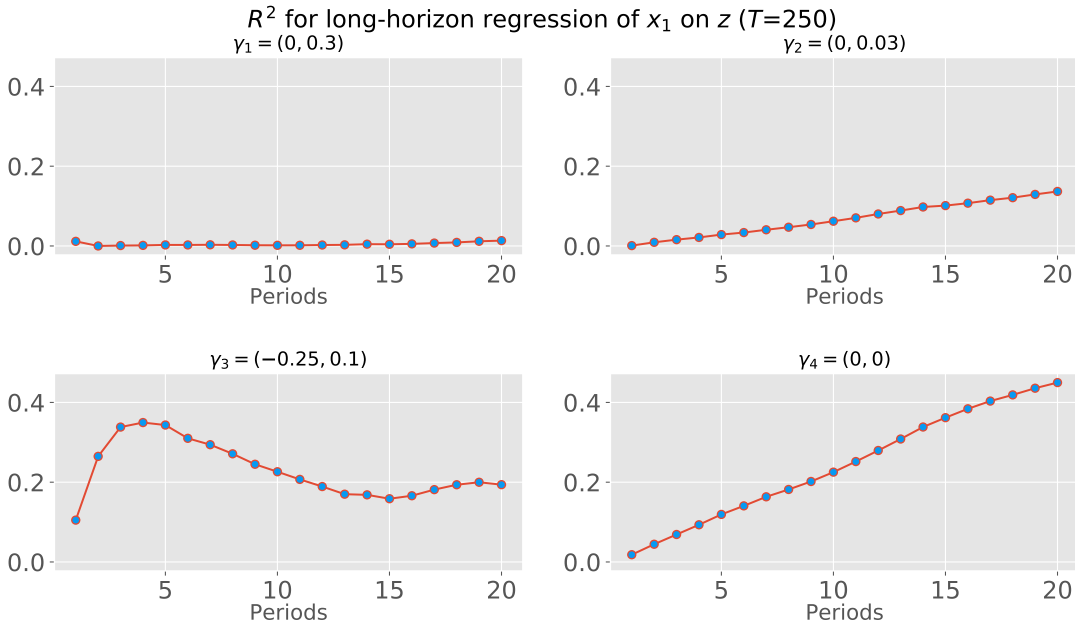

Figure 7: R^2 for x_1 ($T = 250$)

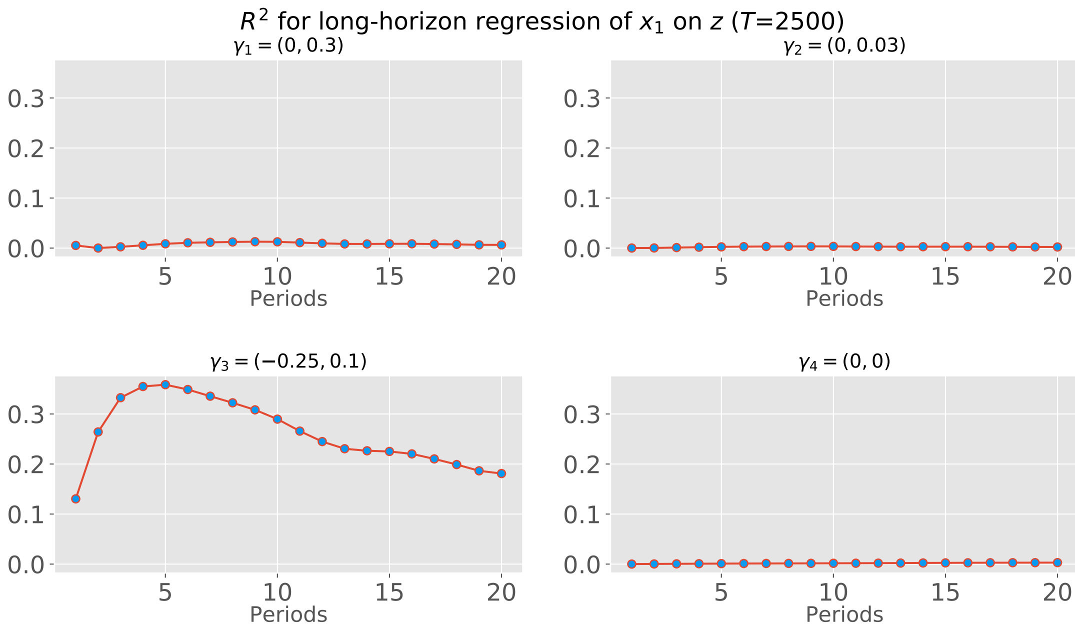

Figure 8: R^2 for x_1 ($T = 2500$)

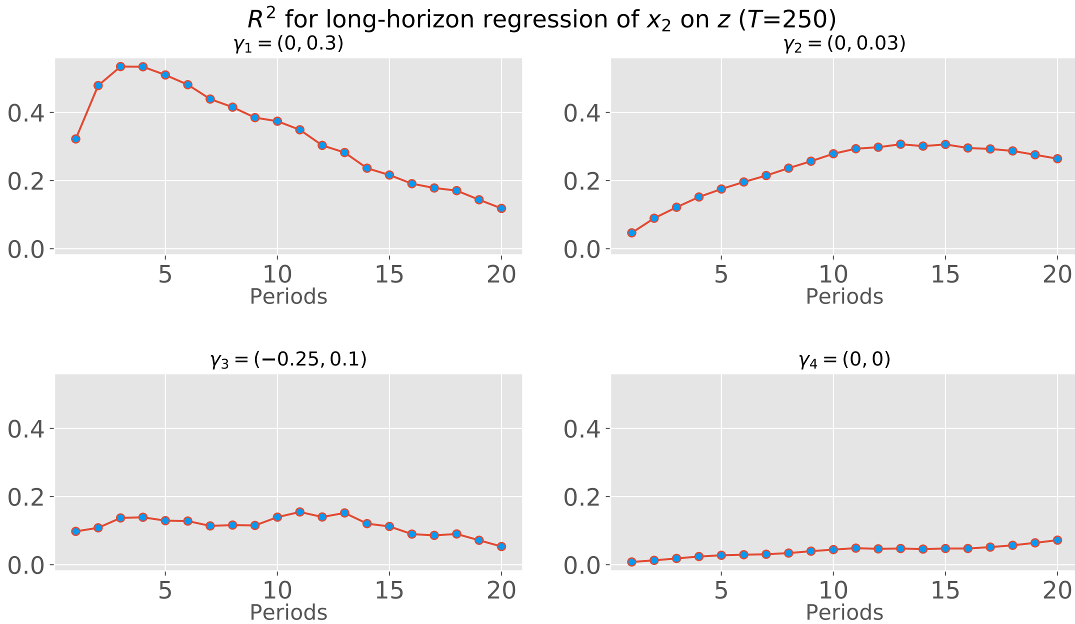

Figure 9: R^2 for x_2 ($T = 250$)

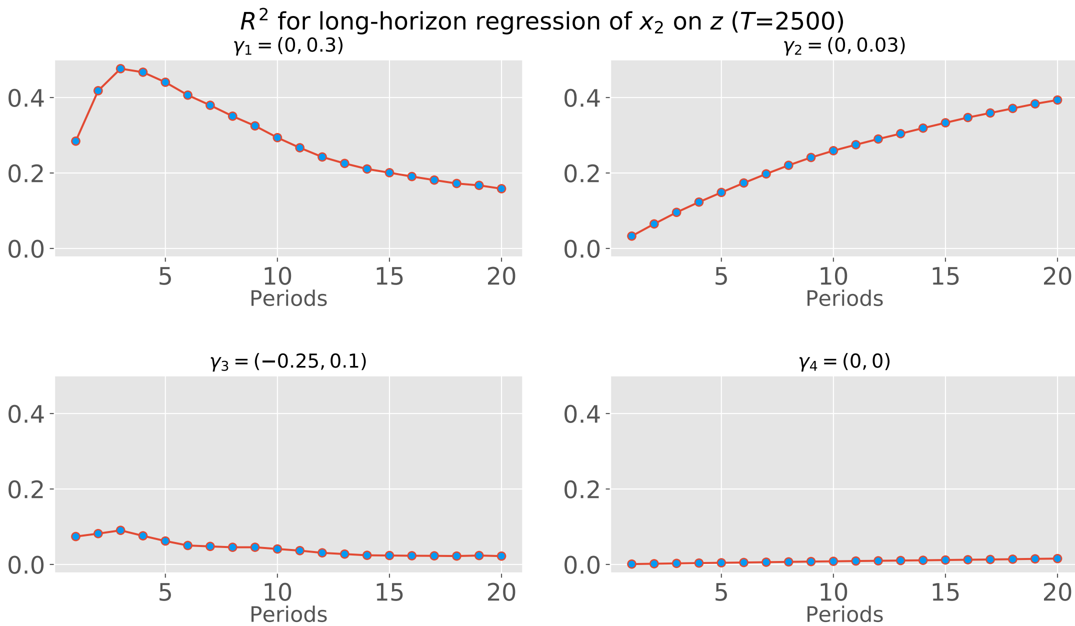

Figure 10: R^2 for x_2 ($T = 2500$)

Problem 5

(a) Using an ADF-test, we cannot reject the null hypothesis that $p_t \equiv \log(P_t)$ and $e_t \equiv \log(E_t)$ have a unit root. The statistics are -0.227 and -0.279, respectively, well above the critical value of -2.595 at the 10% level. The corresponding p-values are 0.935 and 0.928. Clearly, p_t and e_t are not stationary (we could use different versions of the ADF-test for similar results).

For reference, Figure 11 shows the evolution of both variables.

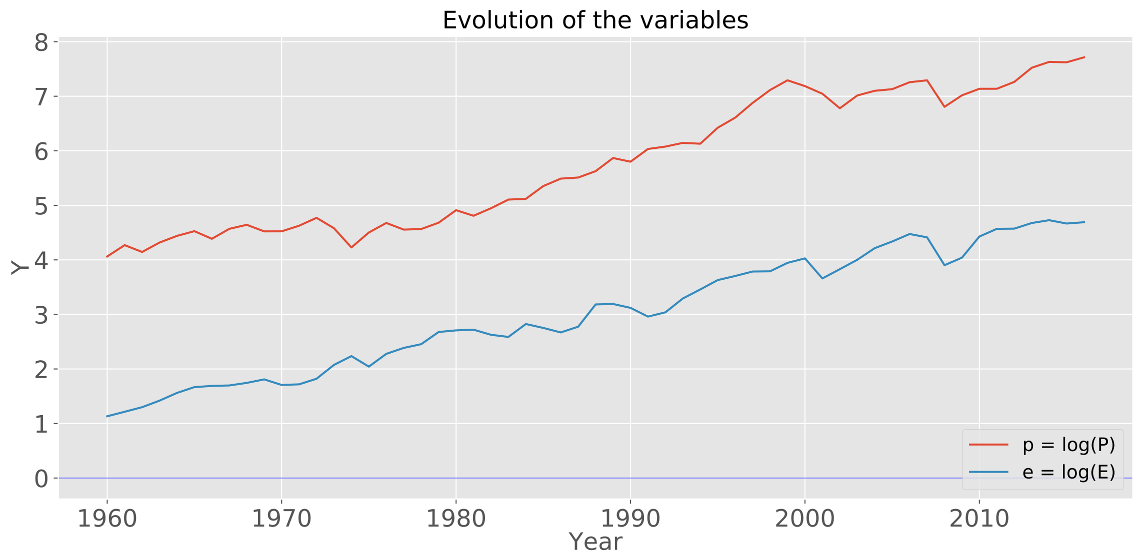

Figure 11: p_t, e_t

When using the ADF-test on $\Delta p_t, \Delta e_t$, we can strongly reject the unit root (p-value = 0.000 for both). This suggests that $p_t, e_t \sim \mathcal{I}(1)$.

(b) Using the ADF-test on $p_t - e_t$, we formally fail to reject the null hypothesis of a unit root, even though for some versions of the test, we are not *too* far from rejection. The p-values for the several versions are: $0.153, 0.270, 0.500, 0.614$ (respectively: constant only, which is the most standard, constant & trend, constant & linear and quadratic trend, no constant & no trend). In words, it appears that the log price-earnings ratio is not necessarily stationary, at least from a statistical perspective. This is an issue we encountered in Lecture 4 with the log price-dividend ratio, and is not surprising given the persistent dynamics of $p_t - e_t$ shown in Figure 12. Still, given the overall evolution of the variable, there is no clear evidence of non-stationarity either and the failure to reject might also be due to the shortness of the sample (data is yearly). We therefore proceed with the analysis assuming that the log price-earnings ratio is broadly $\mathcal{I}(0)$, but keeping this caveat in mind. Note that, like for $p_t - d_t$, some slow-moving evolutions might be at play in driving the low frequency moves, which have nothing to do with predictability, e.g. demographics and increased participation in asset markets.

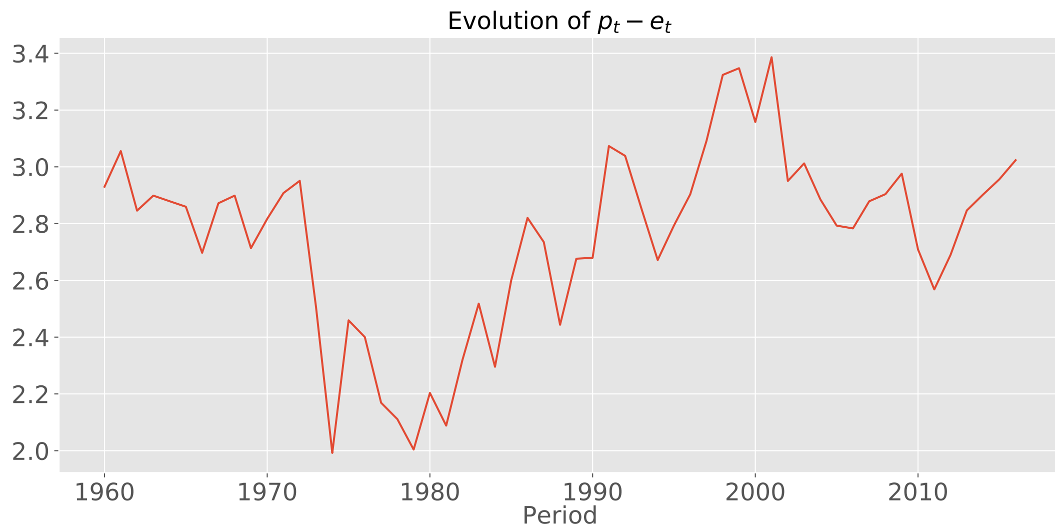

Figure 12: $p_t - e_t$

(c) The results from the VECM estimation are shown in Table 1 below, and the equation that we estimate is as follows:

$$\begin{pmatrix} \Delta p_t \\ \Delta e_t \end{pmatrix} = \gamma \alpha' \begin{pmatrix} p_{t-1} \\ e_{t-1} \end{pmatrix} + \boldsymbol{\mu} + \boldsymbol{\Phi}_1 \begin{pmatrix} \Delta p_{t-1} \\ \Delta e_{t-1} \end{pmatrix} + \boldsymbol{\epsilon}_t$$

Δp_t	**coef**	**std err**	**z**	**P$>\|z\|$**	**[0.025**	**0.975]**
const.	0.3519	0.142	2.471	0.013	0.073	0.631
Δp_{t-1}	0.0019	0.134	0.014	0.989	-0.260	0.264
Δe_{t-1}	-0.0181	0.144	-0.126	0.900	-0.299	0.263
Δe_t	**coef**	**std err**	**z**	**P$>\|z\|$**	**[0.025**	**0.975]**
const.	-0.1734	0.136	-1.272	0.203	-0.441	0.094
Δp_{t-1}	0.1160	0.128	0.907	0.364	-0.135	0.367
Δe_{t-1}	0.0973	0.137	0.709	0.479	-0.172	0.367
	coef	**std err**	**z**	**P$>\|z\|$**	**[0.025**	**0.975]**
γ_p	-0.1351	0.065	-2.065	0.039	-0.263	-0.007
γ_e	0.1044	0.063	1.668	0.095	-0.018	0.227
	coef	**std err**	**z**	**P$>\|z\|$**	**[0.025**	**0.975]**
α_p	1.0000	0	0	0.000	1.000	1.000
α_e	-1.2013	0.099	-12.186	0.000	-1.395	-1.008

Table 1: VECM estimation for p_t, e_t

From the Table, the coefficients on the lagged vector of differences $(\Delta p_{t-1}, \Delta e_{t-1})'$ are mostly statistically insignificant and small in magnitude, beyond the constant. However, we find noticeable evidence of error correction in prices, with a coefficient $\hat{\gamma}_p = -0.1351$ somewhat large and statistically significant at the 5% level (95% confidence interval = $[-0.263, -0.007]$). This is consistent with evidence that we discussed from the price-dividend ratio. The coefficient on earnings is also of broadly similar magnitude and significant at the 10% level (95% confidence interval = $[-0.018, 0.227]$), while that on dividend growth was not in the VECM discussed in Lecture 4. The sign of both coefficients is as expected: a higher $p-e$ value today, (with respect to its long-run average), can come from higher expected fundamental growth ($\gamma_e > 0$), or lower expected future returns, i.e. lower expected future price growth ($\gamma_p < 0$).

The estimated cointegration vector is $\hat{\alpha} = (1, -1.2013)$, somewhat close to the "true" vector implied by theory: $\alpha = (1, -1)$. Even though not particularly precisely estimated, the true value is close to the 95% confidence interval of $[-1.395, -1.008]$. Estimating the VECM without estimating the cointegration vector, i.e. assuming that it is know and equal to $\alpha = (1, -1)$ would potentially yield stronger results for error correction.

(d)(e)(f)(g) We now estimate the following long-horizon regressions:

$$\Delta p_{t+1} + \ldots + \Delta p_{t+k} = \alpha_k + \beta_k(p_t - e_t) + u_{p,t+k,k}$$
$$\Delta e_{t+1} + \ldots + \Delta e_{t+k} = \alpha_k + \beta_k(p_t - e_t) + u_{e,t+k,k}$$

The heteroskedasticity and serial correlation patterns of the errors are omitted in the interest of space, but you should find that errors are indeed heteroskedastic, which would make White standard errors a bare minimum, but also serially correlated due to the overlapping data, which makes Newey-West or Hansen-Hodrick standard errors absolutely necessary in practice.

Tables 2 and 3 below show the results for Δp_{t+h} and Δe_{t+h}, with the different specifications for the standard errors (Hansen-Hodrick 2 was computed manually while Hansen-Hodrick 1 is from *statsmodel*). They confirm the results above, and those in the Lecture obtained with the log price-dividend ratio.

- There is strong evidence of error correction, and therefore predictability, in prices. This is equivalent to predictability in returns, as we discussed. The coefficient is significant and large at most lags, and grows substantially as the horizon increases, as does the R^2 of the regression. You can verify that both increase further as we continue to widen the horizon. The sign on the coefficient is negative as expected from theory: a higher $p - e$ value today, (with respect to its long-run average), predicts lower expected returns going forward, i.e. lower expected prices and price growth.

- The evidence of predictability for earnings is mostly inexistent with coefficients staying relatively small in magnitude, statistically insignificant, and with R^2 staying very low (again this is even more the case as we increase k). This squares back better with the evidence we had from the price-dividend ratio: there is little evidence of predictability in fundamentals. This suggests that the results of mild predictability of earnings that we found in the VECM estimation is not borne out as we increase the horizon of predictability, and may have been due to the fact that the cointegration vector was somewhat imprecisely estimated.

⇒ **Executive summary:** strong evidence of predictability in returns/prices, mostly no evidence of predictability in fundamentals (earnings).

Horizon k	β_k	SE(white)	SE(nw)	SE(hh1)	SE(hh2)	R^2
1	-0.12	0.06	0.06	0.06	0.06	0.05
2	-0.22	0.07	0.08	0.09	0.08	0.10
3	-0.28	0.08	0.10	0.11	0.09	0.12
4	-0.31	0.09	0.11	0.12	0.11	0.12
5	-0.38	0.09	0.11	0.12	0.11	0.14

Table 2: Long-horizon regression for Δp_{t+h}

Horizon k	β_k	SE(white)	SE(nw)	SE(hh1)	SE(hh2)	R^2
1	0.07	0.06	0.06	0.06	0.05	0.02
2	0.12	0.09	0.10	0.11	0.10	0.03
3	0.14	0.10	0.12	0.13	0.11	0.04
4	0.17	0.10	0.12	0.12	0.09	0.05
5	0.17	0.09	0.10	0.08	0.05	0.06

Table 3: Long-horizon regression for Δe_{t+h}