Chapter 2

1. In the proof of Theorem 2.10, orthogonality proved differently. Take $l \neq 0$ (otherwise nothing to prove) with $||l|| = 1$. Let $a = \langle f - \Pi f, l\rangle$. Then
$$||f - \Pi f||^2 \leq ||f - (\Pi f + al)||^2 = ||f - \Pi f||^2 - |\langle f - \Pi f, l\rangle|^2$$
shows that $\langle f - \Pi f, l\rangle = 0$.

2. In Example 2.25 it is said that the formula for the prediction is wrong for $\phi > 1$. The background of this statement is a different AR(∞) representation of X_t. Exercise: use this to compute the predictor. Alternatively, one may use Equation (2.4). Interesting to compute the $\gamma_X(h)$ here to see what the solution of the equation is. This approach should also be investigated for the non-stationary $X_t = Z_t + Z_{t-1}$, with independent Z_t. Possibility for new exercises.

3. Typo in Section 2.4, line 2: $Pi_{n-1}X_n$.

4. Page 34, definition of $\alpha_X(h)$: add $\alpha(0) = 1$ and $\Pi_0 Y = 0$ for Y with $\mathrm{E}Y = 0$.

5. Proof of Lemma 2.33. The last four lines give an ultra short argument, although the conclusion is correct. Careful here: for instance, the coefficients of $\Pi_{2,h}X_1$ are those of $\Pi_{2,h}X_{h+1}$, but in reversed order. One also needs $||X_{h+1} - \Pi_{2,h}X_{h+1}|| = ||X_h - \Pi_{h-1}X_h||$, but this is obvious by stationarity.

Chapter 3

1. Page 41: finding the path x_{n,l_n} has become homework. Details of the proof of Lemma 3.10 spelled out during class. Not mentioned that the "if and only if" in line 6 of the proof is actually a statement concerning the product topology.

2. Line 4 of the proof of Theorem 4.5: X_t^m is actually (also) $2m$-dependent.

Chapter 4

1. Page 53, line -4: this formula for $\alpha^{-1}(u)$ is very nice, but I miss the equivalence $\alpha^{-1}(u) = h \Leftrightarrow \alpha(h) \leq u \leq \alpha(h-1)$.

2. Page 53, line –1: Why not simply defining first $F^{-1}(y)$ and using $F^{-1}(1-u)$ later? I also missed $x \geq F^{-1}(y) \Leftrightarrow F(x) \geq y$, which is used in the proof of ?? (Can't find it back.)

3. On the proof of Theorem 4.7, page 56: I don't see how the second inequality in the estimate of $|\gamma_X(h)|$ is used and why one would need it. "Consequently" in the next line seems to refer to the first inequality only. Besides, there should be $F^{-1}_{|X_0|}$ in the middle of the displayed inequalities.

4. Section 4.7, first line: Theorem 4.7 instead of Theorem4.7.

Chapter 5

1. Page 73, line 2: in the last term a factor $\frac{n-h}{n}$ is missing.

2. Page 74, middle of the page: Should be "the variance of $\sqrt{n}\overline{Y}_n$ converge to etc."

3. Proof of Theorem 5.7: Y_t^m is $(2m+h)$-dependent.

Chapter 6

1. Page 84, line 3: integration should be w.r.t. $F_{n'}$ (n' instead of n).

2. Made a remark that uniqueness of F would be known from MTP, if we had been working with characteristic functions. But here the situation is different, since h is integer valued. Referred to the theory of Fourier series in some analysis course.

3. Page 94, line 23: time series'

4. Page 95, in or above Lemma 6.26: good place to introduce the notation $Z(f)$.

5. Page 96, line 10: $\int f d\mu$ should be $\int f dZ$.

6. Page 96, line 10: the convergence is convergence in $L_2(\Omega, \mathcal{U}, \mathrm{P})$.

7. Page 96, display (line 13): on the left the integrals should be w.r.t. Z.

8. Page 96, line 17: uniqueness can also be established by looking at mixed sequences, as on page 25.

Chapter 8

1. Page 114, footnote: This is a bit confusing. In the example $Y_t = X_{-t}$, the second interpretation would imply that B is the forward shift. I don't understand the last sentence, by the way. Perhaps better to emphasize (more) that B acting on any time series X gives a new times series X' with $X'_t = X_{t-1}$ (and then we write BX_t instead of X'_t, resulting in the example that then indeed $BY_t = Y_{t-1} = X_{-(t-1)}$.

2. Page 116, Definition 8.4: Perhaps to include the comments on page 119 (*Warning*) right after the definition. This has immediate relevance for the proof (and perhaps the formulation) of Theorem 8.6. By the way, I like these remarks about the subtleties involved. In the past, ARMA processes were for me always stationary.

3. Page 119, line -10: time series'

4. Page 120, just above the display: This remark could be made just after the proof of Theorem 8.8.

5. On the proof of Lemma 8.15(i): If $\sum_{j=0}^{\infty} |\pi_j| < \infty$, the function $z \mapsto \pi(z)$ has no poles on or inside the unit circle. But the poles of π coincide with the zeros of θ, since ϕ, θ are relatively prime. Do we really need the spectral argument?

6. Page 122, middle: It follows that the two $\overline{\mathrm{lin}}$ spaces are the same for causal invertible ARMA processes (just a remark).

7. Page 122, line -14: "independently" might arise confusion, although Exercise 8.18 corrects for that.

8. Page 123, line -7 (display): here infinite sum and projections are interchanged, which is justified by the continuity of the projection and that the infinite sum is L_2-convergent by Lemma 1.28.

9. Page 124, line 14: replace 'cannot be predicted' with 'is predicted by zero': $\Pi_t \theta(B) Z_{t+s} = 0$.

10. Page 124, line 17: 'vector-from'.

11. Page 124, displayed equation in the Proof: you also need $s \geq p$, which leads to minor modifications in the proof.

12. Page 124, line -9: I guess that $\|\Phi^s\|$ refers to the operator norm (also in the proof of Theorem 8.32).

13. Page 124, line -8: there should be $Y_s^2 \leq C c^{s-p} \sum_{i=1}^{p} Y_i^2$.

14. Page 129, line 10: 'ofr' should be 'for'.

15. Page 129, Corollary 8.33: Maybe good to emphasize that 'the' stationary distribution depends on the distribution of the given white noise sequence, otherwise readers may think that any stationary ARMA(p, q) process for given polynomials ϕ, θ has the same distribution.

Chapter 9

1. Page 134, line 2: ... is A white noise sequence ...

2. Page 135, line 1: ... all measurable functions $f(Y)$ of Y for which $E|Xf(Y)| < \infty$...

3. Page 135, line -8: ... has zero first *conditional* moment...

4. Page 136, line 6: Perhaps good to mention somewhere earlier the definition of conditional variance (σ_t^2) and then to show that $\sigma_t^2 = \mathrm{E}(X_t^2|\mathcal{F}_{t-1})$ for a MDS X_t. Also, $\mathrm{E}X_t^2 < \infty$ should be included in Definition 9.2.

5. Page 136, line 9: Perhaps useful to explicitly mention $\theta_0 = 0$ (implicitly in line 12).

6. Page 136, line -5: I still find the conventional terminology a bit confusing. Look at $\sigma_t^2 = \alpha + \theta(B)X_t^2$. I would either call σ_t^2 a moving average (although X_t^2 is not a white noise), or X_t^2 auto-regressive (although σ_t^2 isn't a white noise either).

7. Page 136, line -3: SinEce.

8. Page 136, line -1: Perhaps good to emphasize that $\sigma_t^2 \in \mathcal{F}_{t-1}$.

9. Page 137, line 11: There should be $\sigma_t^2 = E(X_t^2|\mathcal{F}_{t-1})$.

10. Page 137, line 15: Why not simply $\sigma_t \in \mathcal{F}_0 \vee \sigma(X_s, 1 \le s \le t-1)$?

11. Page 137, line -7: The GARCH equation is (9.1), better to explicitly state this.

12. Page 138, first two displays: The derivations are a bit unclear (but completely correct). Why not the following computations? First, mention that $\kappa(Z_t) = \mathrm{E}Z_t^4$ and repeat $\mathrm{E}(X_t^2|\mathcal{F}_{t-1}) = \sigma_t^2$. Similarly one has $\mathrm{E}X_t^4 = \mathrm{E}(\sigma_t^4\, \mathrm{E}(Z_t^4|\mathcal{F}_{t-1})) = \mathrm{E}\sigma_t^4\, \mathrm{E}Z_t^4$. Then something like

$$\begin{aligned}\frac{\kappa(X_t)}{\kappa(Z_t)} &= \frac{\mathrm{E}X_t^4}{(\mathrm{E}X_t^2)^2\, \mathrm{E}Z_t^4} \\ &= \frac{\mathrm{E}\,\sigma_t^4}{(\mathrm{E}X_t^2)^2} \\ &= \frac{\mathrm{var}(\sigma_t^2) + (\mathrm{E}X_t^2)^2}{(\mathrm{E}X_t^2)^2} \\ &= \frac{\mathrm{var}(\sigma_t^2)}{(\mathrm{E}X_t^2)^2} + 1,\end{aligned}$$

from which $\kappa(X_t) \ge \kappa(Z_t)$ follows.

13. Page 138, line -16: I would start by defining $W_t = X_t^2 - \mathrm{E}(X_t^2|\mathcal{F}_{t-1}) = X_t^2 - \sigma_t^2$.

14. Page 141, below (9.5): ...tends to zero *in probability*...
15. Page 141, add immediately after (9.6): where the infinite sum is convergent in probability.
16. Page 141, line -12: product of expectations $\mathrm{E}A_t \cdots \mathrm{E}A_{t-n}$.
17. Page 141, line -11: there should be Z_{t-j}^2.
18. Page 141, lines 2–4: spent a few minutes in class on the Jordan decomposition.
19. Page 142, line 1: you also need $\phi_j + \theta_j > 0$ (which follows from the model assumptions) to apply the argument.
20. Theorem 9.15: I treated only the first item, using a slightly different argument at the end of the proof. We have $\mathrm{E}|X_t - \tilde{X}_t|^2 = \mathrm{E}|\sigma_t - \tilde{\sigma}_t|^2 \leq \mathrm{E}|\sigma_t^2 - \tilde{\sigma}_t^2|$, since $\sigma_t, \tilde{\sigma}_t \geq 0$, and this was shown to converge to zero.

Chapter 10

1. Page 152, line 4: Perhaps "the process *randomly* chooses".
2. Page 153, beginning of Section 10.1: I also like another definition, in which a HMM is defined as (X_t, Y_t) Markov with $\mathcal{L}(X_t, Y_t | X_s, Y_s,\, s \leq t-1) = \mathcal{L}(X_t, Y_t | X_{t-1})$, or alternatively with (X_{t+1}, Y_t) Markov.
3. Page 155, Example 10.4: There should be ARMA$(q+1, q)$ or in the next line the choice $p = r + 1$ (preferred in view of the text that follows).
4. Page 155, line –9 and further down: I gave the students a homework exercise to find a state space realization with X_t having dimension $\max\{p, q\}$.
5. Page 158: The presentation of the Kalman filter is a bit messy. In view of the huge importance of this filter in practice, it deserves a Theorem stating that $\Pi_t X_i$ satisfies the recursion as in (10.5), together with the other relevant quantities. The computations can then be given in the proof.
6. Page 163, line 4 etc.: I couldn't find a definition of μ_{t-1}, although it is clear what is meant.
7. Page 167, line -2: chisquare

Chapter 11

1. Page 171: I used a bit different notation, ϕ for the vector of the ϕ_j, and $\hat{\phi}_n$ for an estimator of it (no arrows).

2. Page 172, line –7: I guess that after the $\approx$ there should just be $\hat{\gamma}_n(t)$.

3. Page 173, line 3 (display): on the RHS the terms with $\overline{X}_n$ are missing.

4. Page 173, line 11: I don't understand the reference to Chapter 8. Perhaps Theorem 5.6 instead? In lines –4,–3 there seems to be referred to the proof of that theorem.

5. Page 173, line –8: I even get $O_P(1/n)$ by rearranging a bit.

6. Page 174, line 1: funny black box at the end of the line.

7. Page 174, line –1: I asked the students to compute the difference.

8. Page 175, end of Example 11.7: I think that (loss in) efficiency is not formally defined.

9. Page 176, line –6: The 2 in the confidence intervals come a bit out of the blue; why not 1.96?

10. Page 186, Lemma 11.16: The map ϕ takes values in $\mathbb{R}^{q+1}$.

11. Page 192, line –8: There should be $V_t = \sum_{j=0}^{\infty}(-\theta)^j Z_{t-j}$.

Chapter 12

1. Page 195, Exercise 12.3: I was wondering whether 1 as a vector containing ones only has been introduced before.

2. Page 195, line –7: fucntion $\rightarrow$ function.

3. Page 196, line 2: There should be $I_{n,X-1\overline{X}}(\lambda_j)$.

4. Page 196, Eq (12.3): should have $e^{-ih\lambda_j}$ on the RHS.

5. Page 197, line 5: It seems to me that verification of the Lindeberg condition is a bit awkward (I made some sketch computations only).

6. Page 199, displayed formula: There should be $\frac{\sigma^2}{n}$, and the result is also zero if $(k+l)/n$ or $(k-l)/n$ are integers. To see the latter, we consider the case $(k-l)/n$ an integer. Write $l = np + k$ with p integer. Then

$$\begin{aligned}\cos(\frac{2\pi}{n}kt)\sin(\frac{2\pi}{n}lt) &= \cos(\frac{2\pi}{n}kt)\sin(\frac{2\pi}{n}kt)\\ &= \frac{1}{2}\sin(\frac{4\pi}{n}kt).\end{aligned}$$

Write $2k = mn + q$ with m, q integers and $0 \le q < n$. Then $\sin(\frac{4\pi}{n}kt) = \sin(\frac{2\pi}{n}qt)$. Note that $\lambda_q = \frac{2\pi}{n}q$ is like a natural frequency, although $\lambda_q > \pi$ also happens. Hence, summing over t one obtains, up to a factor $\sqrt{n}$, the imaginary part of $d_1(\lambda_q)$, which is zero. An explicit computation is as follows. If $q=0$, then $\sin(\frac{2\pi}{n}qt)=0$ and there is nothing to prove. For $1\le q<n$ and $\beta=\exp(\mathrm{i}\frac{2\pi}{n}q)\neq 1$ one has

$$\sum_{t=1}^n \sin(\frac{2\pi}{n}qt)=\mathrm{Im}\sum_{t=1}^n \beta^t=\mathrm{Im}\frac{\beta}{1-\beta}(1-\beta^n)=0.$$

7. Page 200, Eq (12.6): I was just wondering whether this is the original Parzen-Rosenblatt estimator.

8. Page 201, display: First equality should be $\approx$.

9. Page 204, line 3: Isn't the difference between the estimators approximately equal to $\overline{X}^2$?

10. Page 204, line 4: "calculations of Chapter 5": (similar to) those in the proof of Theorem 5.6, I guess.

11. Page 206, line 3 (display): I was not very fond of the notation $I_n(a)$, since we also have $I_n(\lambda)$.

Chapter 13

1. Page 210, line –13: Mean zero applies to the deviation of the estimator. Perhaps better to write something like $\sqrt{n}(\hat\theta_n-\theta)\rightsquigarrow N(0,I_\theta^{-1})$.

2. Page 211, line 10 (and further down): I prefer the likelihood with capital arguments, $p_\theta(X_1,\ldots,X_n)$. This happens in most of the following text.

3. Page 213, line –3: Also here I would write $X_t,X_{t-1},\ldots$.

4. Page 214, (13.5): Perhaps useful to add on the right a term with the conditional expectation, so $\mathrm{E}_{\theta_0}\mathrm{E}_{\theta_0}[\log p_\theta(X_1|\overrightarrow{X_0})|\overrightarrow{X_0}]$ in view of the next displayed relation, where it should also explained what μ is.

5. Page 214, line 8: I would write $p_\theta(\cdot|\overrightarrow{X_0})$ and $p_{\theta_0}(\cdot|\overrightarrow{X_0})$.

6. Page 214, Lemma 13.4: strange formulation. There should be something with two conditional distributions that are identical. E.g. that the conditional laws $\mathcal{L}_\theta(X_1|\overrightarrow{X_0})$ and $\mathcal{L}_{\theta_0}(X_1|\overrightarrow{X_0})$ are identical a.s. under the distribution of $\overrightarrow{X_0}$ with parameter θ_0. (Brrrr, what a sentence!)

7. Page 214, Proof of Lemma 13.4: Isn't it nicer to write $\int\log(\frac{p}{q})p\,d\mu\ge\int(\sqrt{q}-\sqrt{p})^2\,d\mu$?

8. Page 215, line –9: Maybe good to mention that we assume that we can differentiate under the integral to get the derivatives.

9. Page 215, line –1: There should be $\mathrm{Cov}_\theta(\cdot) = \mathrm{E}_\theta(\cdot)$ (Cov and E are swapped).

10. Page 216, line 6: Maybe $\xrightarrow{P_\theta}$.

11. page 216, line –7: see ?? (reference missing).

12. Page 219, line 6: Maybe add something on the uniqueness of θ_0 as a maximizer under identifiability conditions, as on page 214 (line –7).

13. Page 220, line –4: There should be $t > p$. Then in the sequel, there is some asymmetry, the first t terms (those where the formula for the prediction is not valid) are left out from the summation in the bottom line. One could therefore replace in the first two terms replace n with $n - p$ in order to reflect that only the conditional distributions of the X_t given their past for $t > p$ are used.

14. Page 223, first display: X as subscript in f_X is missing, and a λ in the exponential.

15. Page 223, line 15: in the Whittle approximation I get $-n\log(2\pi)$ as the first term.

16. Page 223, line –7: Lemma ??

17. Page 224, line 6: Kolmogorov-Szegö formula (See ??).

18. Page 224, first display: the first integral should be $\int I_n(\lambda)\frac{|\phi(e^{-i\lambda})|^2}{\sigma^2}\,d\lambda$, and σ^2 has disappeared in what follows. Since σ^2 is not relevant for the maximization w.r.t. the ϕ_k, better to have it dropped from the integral right away.

19. Page 224, line –4: There should be $f_X = f_\theta$.

20. Page 225, Example 13.16: Clash of notations, θ as a parameter and as a polynomial. I guess we have to live with that.

21. Page 226, line –2: Maybe add something like "in view of the zero blocks in J_θ".

Additional exercises

1. Consider an ARMA(p, q) process (Y_t), $\phi(B)Y = \theta(Z)$, and let $r = p \vee q$.

   (a) Show that (Y_t) can be put in state space form as follows.

$$X_{t+1} = AX_t + BZ_{t+1} \tag{1}$$
$$Y_t = CX_t + Z_t, \tag{2}$$

where

$$A = \begin{pmatrix} \phi_1 & 1 & 0 & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & & \\ \vdots & & & \ddots & \ddots & \\ \phi_{r-1} & 0 & & & 0 & 1 \\ \phi_r & 0 & & & \cdots & 0 \end{pmatrix},$$

$$B = \begin{pmatrix} \phi_1 + \theta_1 \\ \vdots \\ \phi_r + \theta_r \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix}.$$

Hint: Choose the elements of X_t as $X_{k,t} = \sum_{j\geq k} \phi_j Y_{t+k-j-1} + \sum_{j\geq k} \theta_j Z_{t+k-j-1}$.

(b) Give a condition on the ϕ-polynomial such that the matrix A is stable (all its eigenvalues are inside the unit disk).

(c) Show that the $r \times r$ matrix

$$\mathcal{O} = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{r-1} \end{pmatrix} \tag{3}$$

is invertible.

(d) Suppose that the equations (1) and (2) are given for certain A, B and C and that the matrix $\mathcal{O}$ in (3) is invertible. Show that we can represent (Y_t) in ARMA form.

2. Consider a partitioned matrix

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

and assume that A, D are square and M and D invertible.

(a) Show the decomposition

$$M = \begin{pmatrix} I & BD^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} I & 0 \\ D^{-1}C & I \end{pmatrix}.$$

(b) Compute M^{-1} in partitioned form, in terms of the inverse of the matrix $\Delta = A - BD^{-1}C$.

(c) Apply the preceding to quantify the loss of efficiency when estimating the AR-parameters with a too large order p as in Section 11.1.1. (Make the expression $(\Gamma_p)^{-1}_{s,t=1,\ldots,p_0} - \Gamma_{p_0} \geq 0$ precise.)

3. Prove the Kolmogorov-Szegö formula $\int_{-\pi}^{\pi} \log f_\theta(\lambda)\,\mathrm{d}\lambda = 2\pi \log(\sigma^2/2\pi)$, page 224, for f_θ as given. *Hint:* take first the case $p = 1$ and use the Taylor expansion $\log(1-z) = \sum_{k=1}(-z)^k/k$, valid for $|z| < 1$.
(*Remark:* In fact, the Kolmogorov-Szegö formula is valid in more generality for $\sigma^2 := \mathrm{E}(X_t - \Pi_{t-1}X_t)^2$ and f_θ replaced with f, the spectral density of a stationary time series. For the causal AR case, we know that $X_t - \Pi_{t-1}X_t = Z_t$.)

4. Consider a stationary, causal and invertible ARMA(p, q) process (X_t) given by $\phi(B)X_t = \theta(B)Z_t$. Write $\theta = (\phi, \theta) = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)$ and $I(\theta)$ for the Gaussian Fisher information matrix. Let J_p be the p-dimensional shift matrix with ij-element equal to $\delta_{i,j+1}$ and J_q its q-dimensional sister. Write e_p for the first standard column basis vector of $\mathbb{R}^p$ and likewise define e_q. Finally we have the $(p+q) \times (p+q)$ matrix

$$A(\theta) = \begin{pmatrix} J_p + e_p\phi & 0 \\ 0 & J_q - e_q\theta \end{pmatrix},$$

and $B^\top = \left(e_p^\top, e_q^\top\right)$. Show that $I(\theta)$ satisfies the linear equation

$$I(\theta) = A(\theta)I(\theta)A(\theta)^\top + BB^\top.$$