ECON 706: PROBLEM SET 6

DANIEL PFEFFER

Problem 1. Consider the class of AR(p) models with intercept:

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma^2) \tag{1}$$

with prior distribution

$$\phi \mid \sigma^2 \sim N(\phi_0, \sigma^2 V_0), \qquad p(\sigma^2) \propto \frac{1}{\sigma^2}. \tag{2}$$

(i) What is $\log p(y \mid \sigma^2)$?

Solution. Let $X = (1, y_{t-1}, \ldots, y_{t-p})'$ and $\phi = (c, \phi_1, \ldots, \phi_p)'$, and let $y = (y_1, y_2, \ldots, y_T)'$ be a $T \times 1$ vector of observations. Then combining likelihood with the prior gives the conditional posterior:

$$\begin{aligned}
p(\phi \mid y, \sigma^2) &\propto p(y \mid \phi, \sigma^2) p(\phi \mid \sigma^2) \\
&\propto \exp\left(-\frac{1}{2\sigma^2}(y - X\phi)'(y - X\phi)\right) \exp\left(-\frac{1}{2\sigma^2}(\phi - \phi_0)' V_0^{-1} (\phi - \phi_0)\right) \\
&\propto \exp\left(-\frac{1}{2\sigma^2}(\phi' V_0^{-1} \phi - 2\phi' X' y - \phi' X' X \phi - 2\phi' V_0 \phi_0)\right) \\
&\propto \exp\left(-\frac{1}{2\sigma^2}(\phi'(X'X + V_0^{-1})\phi - 2\phi'(X'y + V_0^{-1}\phi_0) + \phi_0' V_0^{-1} \phi_0)\right).
\end{aligned}$$

If we define

$$\begin{aligned}
\phi_1 &= (X'X + V_0^{-1})^{-1}(X'y + V_0^{-1}\phi_0) \\
V_1 &= (X'X + V_0^{-1})^{-1},
\end{aligned}$$

then,

$$\begin{aligned}
p(\phi \mid y, \sigma^2) &\propto \exp\left(-\frac{1}{2\sigma^2}\left((\phi - \phi_1)' V_1^{-1} (\phi - \phi_1) + \phi_0' V_0^{-1} \phi_0 - \phi_1 V_1^{-1} \phi_1\right)\right) \\
&\propto \exp\left(-\frac{1}{2\sigma^2}(\phi - \phi_1)' V_1^{-1} (\phi - \phi_1)\right).
\end{aligned}$$

Hence the posterior is Gaussian:

$$\phi \mid (y, \sigma^2) \sim N(\phi_1, V_1)$$

To derive the marginal likelihood, note that

$$\begin{aligned}
p(\phi \mid \sigma^2) &\overset{d}{=} N(\phi_0, \sigma^2 V_0) \\
p(\phi \mid y, \sigma^2) &\overset{d}{=} N(\phi_1, \sigma^2 V_1),
\end{aligned}$$

and "invert" Bayes' theorem to obtain

$$\begin{aligned} p(y \mid \sigma^2) &= \frac{p(y \mid \phi, \sigma^2) p(\phi \mid \sigma^2)}{p(\phi \mid y, \sigma^2)} \\ &= \frac{(2\pi\sigma^2)^{-T/2} \exp((2\sigma^2)^{-1} y'y)(2\pi)^{-k/2} |\sigma^2 V_0|^{-1/2} \exp((2\sigma^2)^{-1} \phi_0' V_0^{-1} \phi_0)}{(2\pi)^{-k/2} |\sigma^2 V_1|^{-1/2} \exp((2\sigma^2)^{-1} \phi_1' V_1^{-1} \phi_1)} \\ &= (2\pi\sigma^2)^{-T/2} \frac{|V_0|^{-1/2}}{|V_1|^{-1/2}} \exp\left(-\frac{1}{2\sigma^2} \left(y'y + \phi_0' V_0^{-1} \phi_0 - \phi_1' V_1^{-1} \phi_1 \right) \right), \end{aligned}$$

which follows from the fact that the ϕ terms must cancel. The log marginal likelihood is then

$$\log p(y \mid \sigma^2) = -\frac{T}{2} \log(2\pi\sigma^2) - \frac{1}{2} \log \frac{|V_0|}{|V_1|} - \frac{1}{2\sigma^2} \left(y'y + \phi_0' V_0^{-1} \phi_0 - \phi_1' V_1^{-1} \phi_1 \right).$$

□

(ii) Collect data on real GDP growth from FRED for the period from 1984:Q1 to 2015:Q4. Figure 1 plots the data.

```python
import numpy as np
import matplotlib.pyplot as plt
import pandas_datareader.data as web
import datetime

# Query data
start = datetime.datetime(1984, 10, 1)
end = datetime.datetime(2015, 12, 31)
ts = web.DataReader('GDPC1', 'fred', start, end)
ts = 4*np.log(ts).diff().dropna()

# Plot data
plt.figure(figsize=(9,6))
plt.plot(ts)
plt.ylabel('GDP Growth')
plt.title('Annualized Quarter-on-Quarter GDP Growth Rates')
plt.show()
```

(iii) The goal is to determine the "correct" number of lags for the AR model. We will regard each lag length p as a separate model $\mathcal{M}_p$. The computation of posterior probabilities requires $\log p(y \mid \mathcal{M}_p)$. To simplify the calculations a bit, we use $\log p(y \mid \hat{\sigma}^2, \mathcal{M}_p)$ instead, where

$$\hat{\sigma}^2 = (y - X\phi_1)'(y - X\phi_1)/T.$$

Here ϕ_1 is the posterior mean. Choose numerical values for ϕ_0, and V_0 for models $\mathcal{M}_1$ to $\mathcal{M}_4$ and compute $\log p(y \mid \hat{\sigma}^2, \mathcal{M}_p)$ for y ranging form 1985:Q1 to 2015:Q4. (Note that you need the 1984 observations to initialize lags).

Solution. Fit the model.

```python
def ar_bayes(df, lags, tau):
    y = df['GDPC1'][lags:]
    X = df
```

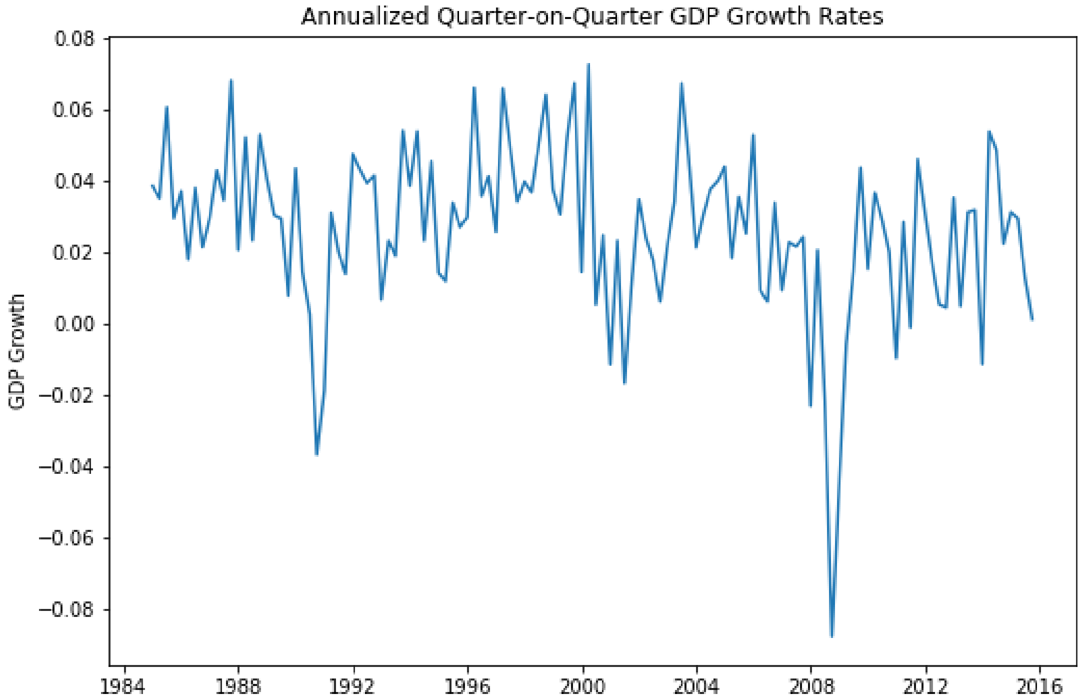

Figure 1

```
# Model i data
for col in X.columns:
    for l in range(1,lags+1):
        X.loc[:,col+"_L"+str(l)] = X[col].shift(l)
X = df.drop(['GDPC1'], axis=1).dropna()
from statsmodels.api import add_constant
X = add_constant(X)
T = len(ts) - lags

from numpy.linalg import inv, det
# Prior means and covariance matrices
phi_0 = np.zeros(lags+1)
V_0 = tau*np.eye(lags+1)

# Posterior means and covariance matrices
V_1 = inv(X.T @ X + inv(V_0))
phi_1 = V_1 @ (X.T @ y + inv(V_0) @ phi_0.T)
```

```
    sigma2 = ((y - X @ phi_1).T @ (y - X @ phi_1))/T

    mdd = -T/2*np.log(2*np.pi*sigma2) \
                    - 0.5*np.log(det(V_0)/det(V_1)) \
          - 0.5/sigma2*(y.T @ y + phi_0.T @ inv(V_0) \
             @ phi_0 - phi_1.T @ inv(V_1) @ phi_1)

    return {'phi_1':phi_1, 'V_1': V_1, 'mdd':mdd}

# Marginal likelihoods
mdds = []
for i in range(1, 5):
    mdds.append(ar_bayes(ts[:], i, tau=10)['mdd'])
mdds
```

The marginal data density values are 285.03, 284.28, 282.36, and 279.41 for models $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_3$, and $\mathcal{M}_4$, respectively. The optimal lag order is therefore one. □

(iv) Convert the $\log p(y \mid \hat{\sigma}^2, \mathcal{M}_p)$ into posterior model probabilities (you need to assume some prior model probabilities). If you had to select a lag order, which one would you select? If you would average predictions across models, which of the model specifications would receive non-trivial weight?

Solution. The following computes the posterior model probabilities, where the prior is such that the prior probability or each lag order is equal.

```
# Log posterior probabilities
mod_prior = [0.25, 0.25, 0.25, 0.25]
denom = 0.25*np.sum(np.exp(mdds))
mod_post_prob = []
for i in range(4):
    mod_post_prob.append(0.25*np.exp(mdds[i])/denom)
mod_post_prob
```

The posterior model probabilities are 0.65, 0.31, 0.05, and 0.002, for models $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_3$, and $\mathcal{M}_4$, respectively. □

(v) Now suppose that you condition on the model with the preferred lag order $\hat{p}$. Suppose that you introduce a hyperparameter τ and change your prior to

$$\phi \mid \sigma^2 \sim N(\phi_0, \sigma^2 \tau V_0). \tag{3}$$

Now compute $p(y \mid \tau, \hat{\sigma}^2, \mathcal{M}_{\hat{p}})$ for values of τ on the grid

$$\frac{1}{100}, \frac{1}{10}, 1, 10, 100.$$

Which scaling of the prior covariance matrix is preferred?

Solution. The following code computes the marginal data density (conditional on τ).

```
# Generate grid
hyper_mdds = []
taus = [1/1000] * 6
```

```
for i in range(1,5):
    taus[i+1] = taus[i]*10
    hyper_mdds.append(ar_bayes(ts[:], lags=1, tau=taus[i])['mdd'])
hyper_mdds
```

The marginal likelihoods conditional on τ are 240.35, 256.88, 278.64, and 282.93. So the preferred scaling is where $\tau = 0.01$. □

Problem 2. Consider the state-space model:

$$\begin{aligned} y_t &= \lambda s_t + \upsilon_t \\ s_t &= \phi s_{t-1} + \varepsilon_t. \end{aligned} \tag{4}$$

For now assume that $\upsilon_t \sim$ iid $N(0,1)$, $\varepsilon_t \sim$ iid $N(0,1)$, and $\upsilon_t \perp \varepsilon_t$.

(i) Derive the autocovariance function for y_t.

Solution. Since state equation is an AR(1) process, its autocovariance function is

$$\gamma_s(h) = \begin{cases} 1/(1-\phi^2) & \text{if } h = 0 \\ \phi^{|h|}/(1-\phi^2) & \text{if } h \neq 0. \end{cases}$$

Since υ_t and ε_t are orthogonal at all leads and lags,

$$\gamma_y(0) = \text{var}(y_t) = \text{var}(\lambda s_t + \upsilon_t) = \gamma_s(0)\lambda^2 + 1 = \frac{\lambda^2}{1-\phi^2} + 1$$

$$\gamma_y(h) = \text{cov}(\lambda s_t + \upsilon_t, \lambda s_{t-h} + \upsilon_{t-h}) = \gamma_s(0)\lambda^2\phi^{|h|} = \frac{\lambda^2}{1-\phi^2}\phi^{|h|}, \qquad h \neq 0.$$

□

(ii) Are the coefficients of the state-space model identified?

Solution. No. The factor loading λ is unique only up an orthogonal transformation and therefore is not sign-identified. To see this point, let $\tilde{s}_t = -s_t$ and let $\tilde{\varepsilon}_t = -\varepsilon_t$, and

$$y_t = -\lambda \tilde{s}_t + \upsilon_t, \qquad \tilde{s}_t = \phi \tilde{s}_{t-1} + \tilde{\varepsilon}_t.$$

Then

$$\begin{aligned} y_t &= -\lambda(\phi \tilde{s}_{t-1} + \tilde{\varepsilon}_t) + \upsilon_t \\ &= -\lambda(\phi(-s_{t-1}) + (-\varepsilon_t)) + \upsilon_t \\ &= \lambda(\phi s_{t-1} + \varepsilon_t) + \upsilon_t, \end{aligned}$$

which is observationally equivalent to the state-space representation (4) since it has the same autocorrelation function. □

(iii) Find an observationally equivalent ARMA representation for the state-space model. Express the ARMA parameters as functions of (λ, ϕ).

Solution. Subtract ϕy_{t-1} from both sides of the observation equation to obtain

$$y_t - \phi y_{t-1} = \lambda s_t + \upsilon_t - \lambda\phi s_{t-1} - \phi\upsilon_{t-1}$$

or

$$y_t = \phi y_{t-1} + \lambda\varepsilon_t + \upsilon_t + \phi\upsilon_{t-1}.$$

This resembles an ARMA(1,1) process:

$$y_t = \rho y_{t-1} + \eta_t + \theta\eta_{t-1}, \qquad \eta_t \sim \text{iid } N(0, \sigma_\eta^2),$$

which has covariance structure

$$\gamma_y(0) = \sigma_\eta^2 \frac{1+\theta^2+2\theta\rho}{1-\rho^2}$$
$$\gamma_y(1) = \sigma_\eta^2 \frac{(1+\theta\rho)(\rho+\theta)}{1-\rho^2}$$
$$\gamma_y(h) = \gamma_y(1)\rho^{|h|-1}, \qquad |h| \geq 1.$$

To obtain an observationally equivalent ARMA representation, we will match autocovariance functions. Evidently, for $h > 1$, we require $\rho = \phi$, which gives the following 2×2 system:

$$\gamma_y(0)(1-\phi^2) = (1+\theta^2+2\theta\phi)\sigma_\eta^2 = 1-\phi^2+\lambda^2$$
$$\gamma_y(1)(1-\phi^2) = (\theta-\theta\phi+1+\theta^2+2\theta\phi)\sigma_\eta^2 = \phi.$$

Subtracting the corresponding equations gives

$$\gamma_y(0)(1-\phi^2) = (1+\theta^2+2\theta\phi)\sigma_\eta^2 = 1-\phi^2+\lambda$$
$$\gamma_y(1)(1-\phi^2) = (\theta-\theta\phi+1+\theta^2+2\theta\phi)\sigma_\eta^2 = \phi\lambda^2,$$

and divide the second equation here by ϕ and subtracted it from the first equation to obtain

$$\sigma_\eta^2\theta(1-\phi^2)/\phi = -(1-\phi^2)$$
$$\sigma_\eta^2 = -\phi/\theta.$$

Now substitute this into the first equation:

$$-\phi(1+2\theta\phi+\theta^2)/\theta = 1-\phi^2+\lambda^2$$
$$-\phi(1+2\theta\phi+\theta^2) = \theta-\theta\phi^2+\theta\lambda^2$$

or

$$\phi\theta^2+\theta(\phi^2+\lambda^2+1)+\phi = 0$$

This quadratic equation in θ has two solutions given by

$$\theta = \frac{-(\phi^2+\lambda^2+1) \pm \sqrt{(\phi^2+\lambda^2+1)^2-4\phi^2}}{2\phi}$$

To obtain a single solution, we can require that the ARMA(1,1) process is invertible (i.e., $|\theta| < 1$):

$$\theta = \frac{-(\phi^2+\lambda^2+1) + \sqrt{(\phi^2+\lambda^2+1)^2-4\phi^2}}{2\phi}$$
$$\sigma_\eta^2 = -\frac{\phi}{\theta} = \frac{2}{(\phi^2+\lambda^2+1) - \sqrt{(\phi^2+\lambda^2+1)-4\phi^2}}.$$

□

(iv) Now suppose that υ_t and ε_t are jointly normally distributed and have a correlation ρ. Re-derive the autocovariance function for y_t. Are the coefficients (λ, ϕ, ρ) identified? Find an observationally equivalent ARMA representation.

Solution. To find autocovariance function for y_t when there is a non-zero correlation ρ between υ_t and ε_t, we simply add the cross product to our earlier calculations and obtain

$$\begin{aligned}\gamma_y(0) &= \frac{\lambda^2}{1-\phi^2} + 2\lambda\rho \\ \gamma_y(h) &= \left(\frac{\lambda^2}{1-\rho^2} + \lambda\rho\right)\phi^{|h|}, \qquad h \neq 0,\end{aligned}$$

which imply that although $\lambda\rho$ is sign-identified, neither λ nor ϕ nor ρ is uniquely identified.

Write the observationally equivalent ARMA(1,1) process as

$$y_t = \rho y_{t-1} - \eta_1 + \theta\eta_{t-1}, \qquad \eta_t \sim \text{iid } N(0, \eta_t^2).$$

Matching autocovariance functions as in (iii) gives

$$\sigma_\eta^2 = -\phi(1+\lambda\rho)/\theta,$$

with solution

$$\begin{aligned}\theta &= \frac{-(\phi^2+\lambda^2+1+2\lambda\rho) + \sqrt{(\phi^2+\lambda^2+1+2\lambda\rho)^2 - 4\phi^2(1+\lambda\rho)^2}}{2\phi(1+\lambda\rho)} \\ \sigma_\eta^2 &= -\frac{\phi}{\theta} = \frac{2(1+\lambda\rho)}{(\phi^2+\lambda^2+1+2\lambda\rho) - \sqrt{(\phi^2+\lambda^2+1+2\lambda\rho)^2 - 4\phi^2(1+\lambda\rho)^2}}.\end{aligned}$$

□

(v) We derived the Kalman filter iterations under the assumption that the errors in the measurement equation and the state-transition equation are independent. Generalize the Kalman filter iterations for the above state-space model to allow for a non-zero correlation ρ between υ_t and ε_t.

Solution. Initialize the recursion by assuming that state vector has the unconditional distribution distribution $s_0 \sim N(\bar{s}_0, p_0)$, with

$$\begin{aligned}\bar{s}_0 &= Es_0 = 0 \\ p_0 &= \text{var}(s_0) = \frac{1}{1-\phi^2}.\end{aligned}$$

Assume that at time t the previous iteration yields

$$s_{t-1} \mid y^{t-1} \sim N(\bar{s}_{t-1|t-1}, p_{t-1|t-1}).$$

(1) Forecast: Our assumptions together with $\varepsilon_t \sim$ iid $N(0,1)$ imply that

$$s_t \mid y^{t-1} \sim N(\bar{s}_{t|t-1}, p_{t|t-1}),$$

where

$$\begin{aligned}\bar{s}_{t+1|t} &= E(s_t \mid y^{t-1}) = \phi\bar{s}_{t-1|t-1} \\ p_{t|t-1} &= \text{var}(s_t \mid y^{t-1}) = \phi^2 p_{t-1|t-1} + 1.\end{aligned}$$

(2) Likelihood: The marginal distribution of y_t conditional on y^{t-1} is of the form

$$y_t \mid y^{t-1} \sim N(\bar{y}_{t|t-1}, f_{t|t-1}),$$

where

$$\begin{aligned}\bar{y}_{t|t-1} &= E(y_t \mid y^{t-1}) = \lambda\bar{s}_{t|t-1} \\ f_{t|t-1} &= \text{var}(y_t \mid y^{t-1}) = \lambda^2 p_{t|t-1} + 1 + 2\lambda^2\rho.\end{aligned}$$

(3) Update: The joint distribution of the s_t and y_t is

$$\begin{pmatrix} s_t \\ y_t \end{pmatrix} \mid y^{t-1} \sim N\left(\begin{pmatrix} \bar{s}_{t|t-1} \\ \bar{y}_{t|t-1} \end{pmatrix}, \begin{pmatrix} p_{t|t-1} & \lambda p_{t|t-1} + \rho \\ \lambda p_{t|t-1} + \rho & f_{t|t-1} \end{pmatrix} \right),$$

where the covariance captures the correlation between υ_t and ε_t and is obtained from

$$\begin{aligned} \operatorname{cov}(s_t, y_t \mid y^{t-1}) &= \operatorname{cov}(s_t, \lambda s_t + \upsilon_t \mid y^{t-1}) \\ &= \lambda \operatorname{var}(s_t \mid y^{t-1}) + \operatorname{cov}(\phi s_{t-1} + \varepsilon_t, \upsilon_t \mid y^{t-1}) \\ &= \lambda p_{t|t-1} + \rho \end{aligned}$$

Applying Bayes' Theorem gives

$$s_t \mid (y_t, y^{t-1}) \sim N(\bar{s}_{t|t}, p_{t|t}),$$

where

$$\begin{aligned} \bar{s}_{t|t} &= \bar{s}_{t|t-1} + (\lambda p_{t|t-1} + \rho)(y_t - \bar{y}_{t|t-1})/f_{t|t-1} \\ p_{t|t} &= p_{t|t-1} - (\lambda p_{t|t-1} + \rho)^2 / f_{t|t-1} \end{aligned}$$

This completes one iteration of the Kalman filter. $\square$