Mgmt 237e: Homework 3
ARMA Models

Drew Creal
TA: Patrick Kiefer

January 18, 2016

Please use Matlab to solve (2) and (3). The homework is due on January 26. Use the electronic drop box to submit your answers. Submit the Matlab file and the file with a short write-up of your answers. The points for each question are allocated equally to each of the sub-questions.

Problem 1: Fundamentals of AR(p) models

(20 pts) Consider the second-order autoregressive process:

$$y_t = \phi_0 + \phi_2 y_{t-2} + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{N}(0, \sigma^2)$$

where $|\phi_2| < 1$.

1. Find the first and second-order auto-covariances.
2. Find the partial autocorrelations $\phi_{1,1}$ and $\phi_{2,2}$.
3. Find the impulse response function.
4. Determine the forecast function $\widehat{y}_t(h) = \mathrm{E}_t(y_{t+h})$ and compute the forecast errors $v_t(h)$. Report the variance of the forecast errors.

Problem 2: Applying the Box-Jenkins methodology[1]

(60 pts) In PPIFGS.xls you will find quarterly data for the Producer Price Index. Our goal is to develop a quarterly model for the PPI, so we can come up with forecasts. Our boss needs forecasts of inflation, because she wants to hedge inflation exposure. There is not a single 'correct' answer to this problem. Well-trained econometricians can end up choosing different specifications even though they are confronted with the same sample. However, there definitely are some wrong answers.

1. We look for a covariance-stationary version of this series. Using the entire sample, make a graph with four subplots:

   (a) Plot the PPI in levels.

   (b) Plot ΔPPI

   (c) Plot $\log PPI$

   (d) Plot $\Delta \log PPI$.

2. Which version of the series looks covariance-stationary to you and why? Let's call the covariance stationary version $y_t = f(PPI_t)$.

3. Plot the ACF of y_t for 12 quarters. What do you conclude? If the ACF converges very slowly, re-think whether y_t really is covariance stationary.

4. Plot the PACF of y_t for 12 quarters. What do you conclude?

5. On the basis of the ACF and PACF, select four different ARMA model specifications.

   (a) Using the entire sample, estimate each one of these. Report the coefficient estimates and standard errors. Check for stationarity of the parameter estimates.

   (b) Plot the residuals. (Note: the residuals will have conditional heteroskedasticity or 'GARCH effects'. We will talk about this in Lecture #5. However, in well-specified models, the residuals should not be autocorrelated.)

   (c) Report the Q-statistic for the residuals for 8 and 12 quarters, as well as the AIC and BIC. Select a preferred model on the basis of these diagnostics. Explain your choice.

[1]In Matlab, there is an **Econometrics Toolbox** and a series of functions : 'arima, estimate, forecast, infer, simulate, lbqtest' that can help you solve this problem. Alternatively, you can download Kevin Sheppard's **MFE toolbox**, which is freely available. You can just Google this and find it.

6. Re-estimate the 4 models using only data up to the end of 2005 and compute the MSPE (mean squared prediction error) on the remainder of the sample for one-quarter ahead forecasts:

$$\frac{1}{H}\sum_{t=1}^{H} v_t^2$$

where H is the length of the hold-out sample, and v_i is the one-step ahead prediction error. Also report the MSPE assuming there is no predictability in y_t, i.e. assuming y_t follows a random walk. What do you conclude?

Problem 3: AR(1) model of the yield curve

(40 pts) Use the Fama-Bliss zero coupon bond yields available in 'fama_bliss_data.xlsx'. These are monthly data from June 1952 to December 2014 on yields of maturity 1m, 3m, 12m, 24m, 36m, 48m, 60m. Zero coupon yields are artificially constructed securities that are calculated based on the trades of coupon bounds observed in practice.[2] The goal of this problem is to familiarize you with how bond pricing models work in practice by working with the simplest model: the Vasicek (1977) model.

1. Using the Box-Jenkins toolbox, estimate different AR(p) models for the 1-month yield. Which model do you find fits the data the best? (NOTE: If you desire, you may drop the data from 2009 to 2014 when the short-term interest rate is near zero. This period of time is very difficult to model with an AR model.)

2. Estimate an AR(1) model on monthly data for the 1-month yield (this is the Vasicek (1977) model). How do the residuals look? What features of the 1 month yield does the Vasicek (1977) model miss capturing?

3. Recall that if the Vasicek (1977) model was 'exactly correct' then the yields of all other maturities would be a deterministic function of $y_t^{(1)} = g_t$. In other words, given parameters of the model, we could calculate the yields $y_t^{(n)}$ at higher maturities $n > 1$ by iterating on the following set of difference equations:

$$\begin{aligned}
y_t^{(n)} &= a_n + b_n g_t \\
a_n &= -\frac{1}{n}\bar{a}_n \\
b_n &= -\frac{1}{n}\bar{b}_n
\end{aligned}$$

$$\begin{aligned}
\bar{a}_n &= \bar{a}_{n-1} - \delta_0 + \bar{b}_{n-1}\left[(1-\phi)\mu - \sigma\lambda_0\right] + \frac{1}{2}\sigma^2\bar{b}_{n-1}^2 \\
\bar{b}_n &= \bar{b}_{n-1}\left[\phi - \sigma\lambda_1\right] - \delta_1 \\
\bar{a}_1 &= 0 \\
\bar{b}_1 &= -1
\end{aligned}$$

Using the estimates of μ, ϕ, σ from part 2, calibrate the Vasicek (1977) model on monthly yields to get the best fit for the average yield curve in the sample. By calibrate, I mean choose the parameters λ_0 and λ_g to fit the data as closely as possible.[3] Hint: go to Lecture

[2] The Fama-Bliss data set uses the methodology from Fama and Bliss (1987) to construct this data. You should be aware that there are several alternative procedures for doing this in practice. But, we will not worry about this for this class.

[3] Recall from class, that we are assuming $\delta_0 = 0$ and $\delta_1 = 1$.

#3 and see what values I estimated for λ_0 and λ_g. Use this as your initial guess.

Plot the unconditional mean of the term structure from the data compared to the yields implied by the model. Discuss the results. Is this model a good fit for the data? In other words, does this model make good predictions for what yields of other maturities should theoretically be?

Some things you should learn from this problem: In the rest of this problem, I am giving you some background. You do not have to turn anything in. Recall the fundamental formula of asset pricing from class

$$P_t = \mathbb{E}_t\left[M_{t+1}X_{t+1}\right] \tag{1}$$

We used this formula to calculate an expression for bond prices. This formula is intuitive. It just says that the price of an asset today P_t is equal to the expected future cash flows (payoffs) X_{t+1} discounted back to today at the rate M_{t+1}.[4] The formula generalizes the deterministic (i.e. non-random) present value formulas you learn in undergrad finance classes. The key point is that for almost all assets both M_{t+1} and X_{t+1} are random variables. Importantly, the random variables M_{t+1} and X_{t+1} may be highly correlated. In fact, you will learn in your asset pricing class that how the random variables M_{t+1} and X_{t+1} covary is what determines an asset's risk premium.

Although the Vasicek (1977) model does not fit the data well, you can learn a lot about asset pricing from simple bond pricing models. There are two reasons why.

(a) First, the payoffs of zero-coupon bonds are known in advance and are *guaranteed*, i.e. by definition they have no default risk. This means that X_{t+1} is not a random variable. In

[4] A multi-period generalization of this formula is straightforward. Suppose an asset pays off at time $t+n$. We buy the asset today and hold it until time $t+n$. Today, the asset is worth P_t and tomorrow it is worth $X_{t+1} = P_{t+1}$ where $P_{t+1} = \mathbb{E}_{t+1}\left[M_{t+2}X_{t+2}\right]$. Now, you just need to repeatedly substitute this expression in

$$\begin{aligned}
P_t &= \mathbb{E}_t\left[M_{t+1}X_{t+1}\right] \\
&= \mathbb{E}_t\left[M_{t+1}P_{t+1}\right] \\
&= \mathbb{E}_t\left[M_{t+1}\mathbb{E}_{t+1}\left[M_{t+2}X_{t+2}\right]\right] \\
&= \mathbb{E}_t\left[\mathbb{E}_{t+1}\left[M_{t+1}M_{t+2}X_{t+2}\right]\right] \\
&= \mathbb{E}_t\left[M_{t+1}M_{t+2}X_{t+2}\right] \qquad \text{Law of iterated expectations} \\
&\vdots \quad \vdots \\
&= \mathbb{E}_t\left[M_{t+1}M_{t+2}\ldots M_{t+n}X_{t+n}\right]
\end{aligned}$$

The payoff at time $t+n$ represented by X_{t+n} is discounted back until today by the multiperiod SDF given by

$$M_{t:t+n} = M_{t+1}M_{t+2}\ldots M_{t+n}$$

.

zero coupon bond models, the only source of randomness is M_{t+1}, i.e. the rate at which you discount future cash flows back to today. This is why it is called the **stochastic discount factor**. Zero coupon bonds are special instruments because they allow us to see how cash flows are discounted without worrying about the additional uncertainty associated with what the payoffs will be.

(b) For a special class of models known as 'affine' models, the expectation in (1) can be solved in closed form, i.e. pencil and paper. The Vasicek (1977) model is an example of an affine model. The benefit of having closed form expressions for prices is that it provides intuition for how more complicated settings may work. In more complicated models, the expectation in (1) cannot be solved in closed form and you will have to use numerical techniques like Monte Carlo methods.

For your information, the Black and Scholes (1973) formula for option prices happens to be another example of an 'affine' model where the price of the asset can be calculated in closed-form. You will learn more about this in your derivatives class.

References

Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. *Journal of Political Economy 81*(3), 637–654.

Fama, E. F. and R. R. Bliss (1987). The information in long maturity forward rates. *American Economic Review 77*, 680–692.

Vasicek, O. A. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics 5*(2), 177–188.