

## Solutions for Practice Problems

1. Consider a 3-period model with  $t = 0, 1, 2, 3$ . There are a stock and a risk-free asset. The initial stock price is \$4 and the stock price doubles with probability  $2/3$  and drops to one-half with probability  $1/3$  each period. The risk-free rate is  $1/4$ .

(a) Compute the risk-neutral probability at each node.

**Solution:** Let  $q$  denote the risk-neutral probability of up-node and  $1 - q$  denote risk-neutral probability of the down-node. Then by the definition of the risk-neutral probabilities,

$$\begin{aligned} S_t &= E^Q \left[ \frac{1}{1+r} S_{t+1} \right] \\ &= \frac{1}{1+r} \left( q(2S_t) + (1-q) \left( \frac{1}{2} S_t \right) \right) \end{aligned}$$

with  $r = \frac{1}{4}$ . Solving this equation gives  $q = \frac{1}{2}$ . Notice that this calculation holds true at every non-terminal node. We conclude that the risk-neutral probability at each node is given by  $\frac{1}{2}$  probability of up-node and  $\frac{1}{2}$  probability of down-node.

(b) Compute the Radon-Nikodym derivative ( $d\mathbf{Q}/d\mathbf{P}$ ) of the risk-neutral measure with respect to the physical measure at each node.

**Solution:** The original (physical) measure assigns probabilities  $p_u = \frac{2}{3}$  and  $p_d = \frac{1}{3}$  to the up- and down-node, respectively. The risk-neutral measure assigns probabilities  $q_u = \frac{1}{2}$  and  $q_d = \frac{1}{2}$ . The Radon-Nikodym derivative at each node is a random variable that takes on the value

$$\left( \frac{dQ}{dP} \right) (u) = \frac{q_u}{p_u} = \frac{1/2}{2/3} = \frac{3}{4}$$

in the up-node and the value

$$\left( \frac{dQ}{dP} \right) (d) = \frac{q_d}{p_d} = \frac{1/2}{1/3} = \frac{3}{2}$$

in the down-node. Again, notice that this calculation is valid at each and every node.

(c) Compute the state-price density at each node.

**Solution:** Fix the current node at time  $t$  and let the state-price density at this node be denoted by  $\pi_t$ . Denoting the values of state-price densities at the children nodes by

$\pi_{t+1}(u)$  and  $\pi_{t+1}(d)$ , we have the following pricing equations:

$$\begin{aligned} S_t &= \frac{2}{3} \frac{\pi_{t+1}(u)}{\pi_t} \cdot 2S_t + \frac{1}{3} \frac{\pi_{t+1}(d)}{\pi_t} \cdot \frac{1}{2} S_t \\ &= \left( \frac{4}{3} \frac{\pi_{t+1}(u)}{\pi_t} + \frac{1}{6} \frac{\pi_{t+1}(d)}{\pi_t} \right) S_t \\ 1 &= \frac{2}{3} \frac{\pi_{t+1}(u)}{\pi_t} \cdot \frac{5}{4} + \frac{1}{3} \frac{\pi_{t+1}(d)}{\pi_t} \cdot \frac{5}{4} \\ &= \left( \frac{5}{6} \frac{\pi_{t+1}(u)}{\pi_t} + \frac{5}{12} \frac{\pi_{t+1}(d)}{\pi_t} \right) \end{aligned}$$

Solving this system of equations in the unknowns  $\frac{\pi_{t+1}(u)}{\pi_t}$  and  $\frac{\pi_{t+1}(d)}{\pi_t}$ , we get the solution

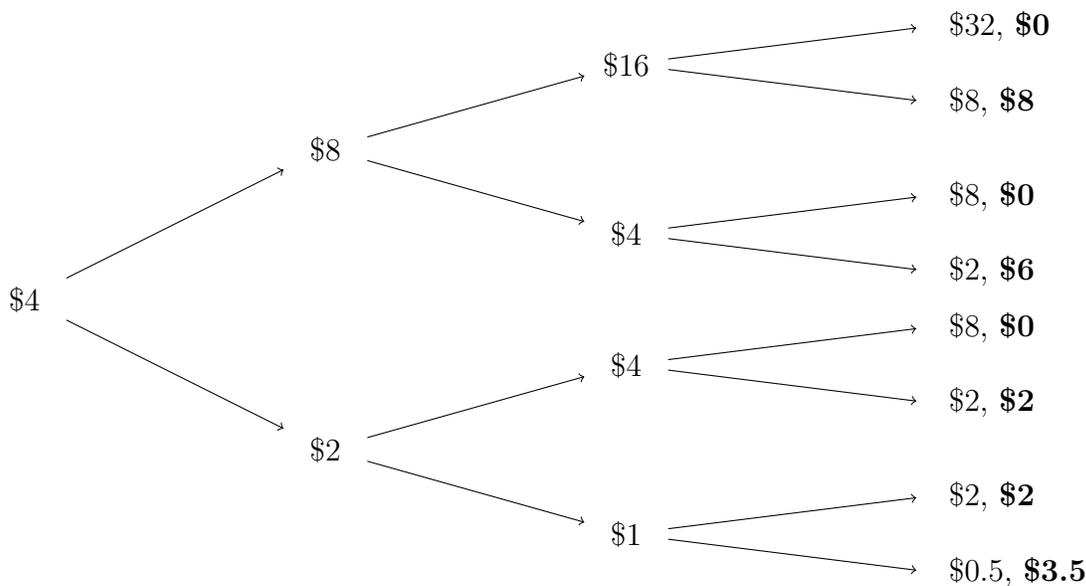
$$\begin{aligned} \frac{\pi_{t+1}(u)}{\pi_t} &= \frac{3}{5} \\ \frac{\pi_{t+1}(d)}{\pi_t} &= \frac{6}{5} \end{aligned}$$

Now, starting at the initial node at time  $t = 0$  and setting  $\pi_0 = 1$  allows us to solve for the state-price density at every node recursively. This calculation leads us to conclude that at a node  $\omega$  whose history consists of  $i$  up-movements and  $j$  down-movements, the state-price density is given by

$$\pi_t(\omega) = \left(\frac{3}{5}\right)^i \left(\frac{6}{5}\right)^j$$

(d) Price a lookback option with payoff at  $t = 3$  equal to  $(\max_{0 \leq t \leq 3} S_t) - S_3$  using risk-neutral probability.

**Solution:** The following binomial tree describes the evolution of stock price and the bold-face numbers next to the final stock price are the payoffs from the lookback option:



Recalling from part (a) that the risk-neutral probabilities of an up-movement and a down-movement are both  $\frac{1}{2}$ , all terminal nodes have the same Q-probability of  $(\frac{1}{2})^3 = \frac{1}{8}$ . Therefore the price of the lookback option is given by

$$\begin{aligned}
 C &= \left(\frac{1}{1+r}\right)^3 E^Q[D_3] \\
 &= \left(\frac{4}{5}\right)^3 \cdot \frac{1}{8} \left(8 + 6 + 2 + 2 + \frac{7}{2}\right) \\
 &= \frac{64}{125} \cdot \frac{1}{8} \cdot \frac{43}{2} \\
 &= \frac{172}{125}
 \end{aligned}$$

(e) Price the lookback option using state-price density and compare your answer to (d).

**Solution:** Using the state-price density we computed in part (c), we can calculate the price of the lookback option as

$$\begin{aligned}
 C &= E[\pi_3 D_3] \\
 &= \sum_{\omega} P(\omega) \pi_3(\omega) D_3(\omega)
 \end{aligned}$$

where the sum is across all the possible time-3 nodes  $\omega$ . But note that for any  $\omega$ , denoting the number of up-movements by  $i$  and the number of down-movements by  $j$ ,

we have

$$\begin{aligned}
 P(\omega) \pi_3(\omega) &= \left[ \left(\frac{2}{3}\right)^i \left(\frac{1}{3}\right)^j \right] \left[ \left(\frac{3}{5}\right)^i \left(\frac{6}{5}\right)^j \right] \\
 &= \left(\frac{2}{5}\right)^i \left(\frac{2}{5}\right)^j \\
 &= \left(\frac{2}{5}\right)^{i+j} \\
 &= \left(\frac{2}{5}\right)^3
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 C &= \sum_{\omega} P(\omega) \pi_3(\omega) D_3(\omega) \\
 &= \left(\frac{2}{5}\right)^3 \sum_{\omega} D_3(\omega) \\
 &= \left(\frac{2}{5}\right)^3 \left(8 + 6 + 2 + 2 + \frac{7}{2}\right) \\
 &= \frac{172}{125}
 \end{aligned}$$

Of course, we get the same answer as we did in part (d) using the risk-neutral probabilities.

2. Show that, under the risk-neutral measure, the discounted gain process

$$\hat{G}_t = \frac{P_t}{B_t} + \sum_{s=1}^t \frac{D_s}{B_s}$$

is a martingale (i.e.  $E_t^Q[\hat{G}_{t+1}] = \hat{G}_t$ ) from the definition of risk-neutral measure in lecture notes

$$P_t = E_t^Q \left[ \sum_{u=t+1}^T \frac{B_t}{B_u} D_u \right]$$

That is the reason why the risk-neutral measure is also called the "equivalent martingale measure" (EMM).

**Solution:** We want to show  $E_t^Q[\hat{G}_{t+1}] = \hat{G}_t$ .

$$E_t^Q [\hat{G}_{t+1}] = E_t^Q \left[ \frac{P_{t+1}}{B_{t+1}} + \sum_{s=1}^{t+1} \frac{D_s}{B_s} \right]$$

Now recall that

$$P_{t+1} = E_{t+1}^Q \left[ \sum_{u=t+2}^T \frac{B_{t+1}}{B_u} D_u \right]$$

Substituting this into the first expression, we have

$$\begin{aligned} E_t^Q [\hat{G}_{t+1}] &= E_t^Q \left[ \frac{P_{t+1}}{B_{t+1}} + \sum_{s=1}^{t+1} \frac{D_s}{B_s} \right] \\ &= E_t^Q \left[ \frac{1}{B_{t+1}} E_{t+1}^Q \left[ \sum_{u=t+2}^T \frac{B_{t+1}}{B_u} D_u \right] + \sum_{s=1}^{t+1} \frac{D_s}{B_s} \right] \\ &= E_t^Q \left[ E_{t+1}^Q \left[ \sum_{u=t+2}^T \frac{D_u}{B_u} \right] + \sum_{s=1}^{t+1} \frac{D_s}{B_s} \right] \\ &= E_t^Q \left[ \sum_{s=1}^T \frac{D_s}{B_s} \right] \\ &= E_t^Q \left[ \sum_{u=t+1}^T \frac{D_u}{B_u} \right] + \sum_{s=1}^t \frac{D_s}{B_s} \\ &= \frac{1}{B_t} E_t^Q \left[ \sum_{u=t+1}^T \frac{B_t}{B_u} D_u \right] + \sum_{s=1}^t \frac{D_s}{B_s} \\ &= \frac{P_t}{B_t} + \sum_{s=1}^t \frac{D_s}{B_s} \\ &= \hat{G}_t \end{aligned}$$

This shows that  $\hat{G}_t$  is a martingale under the risk-neutral measure.

3. Consider the following model of interest rates. Under the physical probability measure  $\mathbf{P}$ , the short-term interest rate is  $\exp(r_t)$ , where  $r_t$  follows

$$dr_t = -\theta(r_t - \bar{r}) dt + \sigma_r dZ_t,$$

where  $Z_t$  is a Brownian motion.

Assume that the SPD is given by

$$\pi_t = \exp \left( - \int_0^t r_u - \int_0^t \frac{1}{2} \eta_u^2 du - \int_0^t \eta_u dZ_u \right)$$

where  $\eta_t$  is stochastic, and follows

$$d\eta_t = -\kappa(\eta_t - \bar{\eta}) dt + \sigma_\eta dZ_t^\eta$$

where  $Z_t^\eta$  is a Brownian motion independent of  $Z_t$ .

- (a) Derive the dynamics of the interest rate under the risk-neutral probability  $\mathbf{Q}$ .

**Solution:** From the form of SPD  $\pi_t$ , the price of risk at time  $t$  is  $\eta_t$ . Therefore under the risk-neutral measure  $Q$ ,

$$dZ_t^P = dZ_t^Q - \eta_t dt$$

and the shock that drives the price of risk process,  $dZ_t^\eta$  remains unchanged under  $Q$ . Thus, the dynamics of the risk-free rate  $r_t$  can be written as

$$\begin{aligned} dr_t &= -\theta(r_t - \bar{r}) dt + \sigma_r dZ_t \\ &= -\theta(r_t - \bar{r}) dt + \sigma_r (dZ_t^Q - \eta_t dt) \\ &= (-\theta(r_t - \bar{r}) - \sigma_r \eta_t) dt + \sigma_r dZ_t^Q \end{aligned}$$

- (b) Compute the spot interest rates for all maturities. (Hint: look for bond prices in the form  $P(t, T) = \exp(a(T - t) + b(T - t)r_t + c(T - t)\eta_t)$ ).

**Solution:** Suppose the current time is  $t$  and assume that the bond matures at time  $T$ . Denote its price  $P(t, T)$ . By the definition of risk-neutral measure  $Q$ ,

$$P(t, T) = E_t^Q \left[ \exp \left( - \int_t^T r_s ds \right) \right]$$

Now let us guess the bond price  $P(t, T)$  in the functional form

$$P(t, T) = \exp(a(T - t) + b(T - t)r_t + c(T - t)\eta_t)$$

On the one hand, by the definition of the risk-neutral measure, we know that

$$E_t^Q \left[ \frac{dP(t, T)}{P(t, T)} \right] = r_t dt$$

On the other hand, we can calculate the same expression using Ito's Lemma:

$$\begin{aligned} \frac{dP(t, T)}{P(t, T)} &= -(a'(T - t) + b'(T - t)r_t + c'(T - t)\eta_t) dt \\ &\quad + b(T - t) \left( (-\theta(r_t - \bar{r}) - \sigma_r \eta_t) dt + \sigma_r dZ_t^Q \right) \\ &\quad + c(T - t) \left( -\kappa(\eta_t - \bar{\eta}) dt + \sigma_\eta dZ_t^{\eta, Q} \right) \\ &\quad + (b(T - t))^2 \cdot \frac{1}{2} \sigma_r^2 dt + (c(T - t))^2 \cdot \frac{1}{2} \sigma_\eta^2 dt \end{aligned}$$

and therefore,

$$\begin{aligned}
E_t^Q \left[ \frac{dP(t, T)}{P(t, T)} \right] &= -(a'(T-t) + b'(T-t)r_t + c'(T-t)\eta_t) dt \\
&\quad + b(T-t)(-\theta(r_t - \bar{r}) - \sigma_r \eta_t) dt + c(T-t)(-\kappa(\eta_t - \bar{\eta})) dt \\
&\quad + \frac{1}{2} ((b(T-t))^2 \sigma_r^2 + (c(T-t))^2 \sigma_\eta^2) dt
\end{aligned}$$

Equating this expression to  $r_t dt$ , we get a system of three ODEs (match the coefficients on constant,  $r_t$ , and  $\eta_t$  terms):

$$\begin{aligned}
0 &= -a'(\tau) + \theta \bar{r} b(\tau) + \frac{1}{2} \sigma_r^2 (b(\tau))^2 + \frac{1}{2} \sigma_\eta^2 (c(\tau))^2 \\
0 &= -b'(\tau) - 1 - \theta b(\tau) \\
0 &= -c'(\tau) - \sigma_r b(\tau) - \kappa c(\tau)
\end{aligned}$$

Also, we have a terminal condition that  $P(T, T) = 1$ , so that  $a(0) = 0$ ,  $b(0) = 0$ , and  $c(0) = 0$ .

Note that the second equation is an autonomous equation in  $b(\tau)$  and it is straightforward to solve. Given the solution for  $b(\tau)$ , we can then solve the third equation for  $c(\tau)$ . Then, finally, we can solve for  $a(\tau)$  from the first equation. The solutions are

$$\begin{aligned}
b(\tau) &= -\frac{1}{\theta} (1 - e^{-\theta\tau}) \\
c(\tau) &= \frac{\sigma_r}{\theta} \left( \frac{1}{\kappa} (1 - e^{-\kappa\tau}) - \frac{1}{\kappa - \theta} (e^{-\theta\tau} - e^{-\kappa\tau}) \right)
\end{aligned}$$

and  $a(\tau)$  is not reported here for simplicity (it's not difficult to compute, but the resulting expression is long and messy). The solutions for  $a(\tau)$ ,  $b(\tau)$ , and  $c(\tau)$  complete the characterization of bond price

$$P(t, T) = \exp(a(T-t) + b(T-t)r_t + c(T-t)\eta_t)$$

- (c) Compute the instantaneous expected rate of return on a zero-coupon bond with time to maturity  $\tau$ .

**Solution:** Suppose  $T - t = \tau$ , i.e., time to maturity is equal to  $\tau$ . Then,

$$\begin{aligned}
E_t^P \left[ \frac{dP(t, T)}{P(t, T)} \right] - r_t dt &= E_t^P \left[ \frac{dP(t, T)}{P(t, T)} \right] - E_t^Q \left[ \frac{dP(t, T)}{P(t, T)} \right] \\
&= b(\tau) \sigma_r \eta_t dt
\end{aligned}$$

To understand this calculation, recall that the only difference between the two instantaneous drifts is

$$dZ_t = dZ_t^Q - \eta_t dt$$

and the coefficient in front of  $dZ_t$  in  $\frac{dP(t, T)}{P(t, T)}$  is  $b(T-t)\sigma_r$ .

- (d) Show that the slope of the term structure of interest rates predicts the excess returns on long-term bonds. Discuss the intuition. Show that more volatility in the price of risk,  $\eta$ , means more predictability in bond returns.

**Solution:** First, let us discuss what we mean by predictability. In the case that returns are serially i.i.d and independent from any other random variables, there is no predictability because there is no other piece of information that allows us to “predict” the returns. In this question, we have shown in part (b) that the bond prices are of the form

$$P(t, T) = \exp(a(T - t) + b(T - t)r_t + c(T - t)\eta_t)$$

and in part (c) that the expected excess return on a bond with time to maturity  $\tau = T - t$  is given by

$$E_t \left[ \frac{dP(t, T)}{P(t, T)} \right] - r_t dt = b(T - t) \sigma_r \eta_t dt$$

When  $\eta_t$  is high, we can see from the first equation that the term structure of interest rates is steep and from the second equation that excess returns on long-term bonds are high. Therefore, when the term structure is steeper, and excess returns on long-term bonds are higher. The intuition is as follows: times of high price of risk  $\eta_t$  are typically thought of as recessions when people are more risk-averse. High price of risk drives down the price of bonds, resulting in high excess returns (the terminal value of the bond is constant at \$1 at maturity).

Another observation we can make is that more dispersion in the distribution of  $\eta_t$  means more predictability in excess bond returns. Note that in the extreme case when  $\eta_t$  is a constant, our problem reduces to a setting where returns are i.i.d over time and we do not have any predictability (because expected excess bond returns would equal  $b(T - t) \sigma_r \bar{\eta}$ , and excess bond returns would be random noise plus this mean). As  $\eta_t$  becomes more variable, the fraction of variation in the excess bond returns due to variation in the mean of the excess returns,  $b(T - t) \sigma_r \eta_t$ , becomes larger and we have more predictability. We can show that the variance of the stationary distribution of  $\eta_t$  is  $\frac{\sigma_\eta^2}{2\kappa}$ , so more variability in  $\eta_t$  could be due to higher volatility of shocks (higher  $\sigma_\eta$ ) or lower rate of mean-reversion (lower  $\kappa$ ).

4. Suppose that uncertainty in the model is described by two independent Brownian motions,  $Z_{1,t}$  and  $Z_{2,t}$ . Assume that there exists one risky asset, paying no dividends, following the process

$$\frac{dS_t}{S_t} = \mu(X_t) dt + \sigma dZ_{1,t}$$

where

$$dX_t = -\theta X_t dt + dZ_{2,t}$$

The risk-free interest rate is constant at  $r$ .

- (a) What is the price of risk of the Brownian motion  $Z_{1,t}$ ?

**Solution:** Let the price of risk for  $Z_{1,t}$  and  $Z_{2,t}$  be  $\eta_t = [\eta_{1,t} \quad \eta_{2,t}]^T$ . Then we must have

$$\mu(X_t) S_t - [\sigma S_t \quad 0] \eta = r S_t$$

This gives

$$\mu(X_t) S_t - \sigma \eta_{1,t} S_t = r S_t$$

whereas there is no constraint for  $\eta_{2,t}$ . Hence, the price of risk of  $Z_{1,t}$  is

$$\eta_{1,t} = \frac{\mu(X_t) S_t - r S_t}{\sigma S_t} = \frac{\mu(X_t) - r}{\sigma}$$

- (b) Give an example of a valid SPD in this model.

**Solution:**

$$\frac{d\pi_t}{\pi_t} = -r dt - \eta'_t dZ_t, \pi_0 = 1$$

Any  $\eta_{2,t}$  such that  $\eta_t$  satisfies the Novikov's condition will be allowed. The simplest example would be to let  $\eta_{2,t} = 0$ .

- (c) Suppose that the price of risk of the second Brownian motion,  $Z_{2,t}$ , is zero. Characterize the SPD in this model.

**Solution:**

$$\frac{d\pi_t}{\pi_t} = -r dt - \eta_{1,t} dZ_{1,t}$$

Let  $y_t = \ln \pi_t$ , by Ito's lemma,

$$\begin{aligned} dy_t &= \frac{1}{\pi_t} d\pi_t - \frac{1}{2\pi_t^2} (d\pi_t)^2 \\ &= -r dt - \eta_{1,t} dZ_{1,t} - \frac{1}{2} \eta_{1,t}^2 dt \end{aligned}$$

So

$$\begin{aligned} y_t &= -rt - \frac{1}{2} \int_0^t \eta_{1,t}^2 dt - \int_0^t \eta_{1,t} dZ_{1,t} \\ \pi_t &= e^{-rt - \frac{1}{2} \int_0^t \eta_{1,t}^2 dt - \int_0^t \eta_{1,t} dZ_{1,t}} \end{aligned}$$

- (d) Derive the price of a European Call option on the risky asset in this model, with maturity  $T$  and strike price  $K$ .

**Solution:** Existence of SPD implies existence of risk-neutral measure. Under risk-neutral measure, all traded assets must have drift  $r$  and the volatility is unchanged. So under risk-neutral measure,

$$\frac{dS_t}{S_t} = r dt + \sigma dZ_{1,t}^Q$$

The physical drift does not matter at all.

$$C_t = E_t^Q [e^{-r(T-t)} [S_T - K]^+]$$

Standard Black-Scholes formula applies. So

$$C_t = S_t N(d_1) - K e^{-r(T-t)} N(d_2)$$

where

$$d_{1,2} = \frac{\ln(S/K) + (r \pm \frac{1}{2}\sigma^2) T}{\sigma\sqrt{T}}$$

5. Consider a European call option on a stock. The stock pays no dividends and the stock price follows an Ito process. Is it possible that, while the stock price declines between  $t_1$  and  $t_2 > t_1$ , the price of the Call increases? Justify your answer.

**Solution:** Yes, it is possible. For instance, if volatility becomes extremely large once price hits a low level, we could have the situation described in the question. The underlying reason is that call option is an increasing function of both volatility and stock price. So if price drop is accompanied by large increase in volatility, we may have a higher call price.

The easiest way to see this is through a 3-period binomial tree example: Suppose price at time-0 is 100. At time-1, it may go up to 120 or go down to 10. If price goes up to 120, it may go further up to 140 or down to 100 at time-2. If price goes down to 10, the volatility becomes huge. As a result, price may go up to 10000 or down to 1 at time-2. Let interest rate be 0 and consider an European call with strike at 150 and maturity of two periods.

Let's find the call option price at each node. If price goes up to 120 at time-1, then the call is worth 0 since, under no circumstances at time-2, the call will be in the money (both 140 and 100 are less than strike 150). If price goes down to 10, let's find the call price by solving the contingent claim prices. Let the time-1 price of 1 unit of payoff at time-2 and state  $S_2 = 10000$  be  $q_{10000}$  and the time-1 price of 1 unit of payoff at time-2 and state  $S_2 = 1$  be  $q_1$ . Since interest rate is 0, we have

$$q_{10000} + q_1 = 1$$

Since the time-1 stock price is 10, we also have

$$10000q_{10000} + q_1 = 10$$

Solving the simultaneous equation, we have

$$q_{10000} = \frac{1}{1111} \text{ and } q_1 = \frac{1110}{1111}$$

Since the call pays  $(10000 - 150) = 9850$  dollars if price goes up to 10000 at time-2 and 0 dollars if price goes down to 1. We have the call price, at time-1 and state  $S_1 = 10$ ,  $C_{10} = 9850q_{10000} = \frac{9850}{1111} \approx 9$ .

Let's solve for time-0 contingent claim prices  $q_{120}$  and  $q_{10}$  of time-1 payoffs. Zero interest rate implies

$$q_{120} + q_{10} = 1$$

Stock price movements implies

$$120q_{120} + 10q_{10} = 100$$

This gives

$$q_{120} = \frac{9}{11} \text{ and } q_{10} = \frac{2}{11}$$

Hence, the call price at time-0 is  $C_{100} = C_{10}q_{10} \approx \frac{18}{11} \ll 9 = C_{10}$ . So stock price declines at time-1 but call price increases.

The same principle applies in continuous-time as long as we allow volatility to become large once stock price hits a low level. This is certainly possible since the stock price process follows a general Ito process.

6. Suppose that the stock price  $S_t$  follows a Geometric Brownian motion with parameters  $\mu$  and  $\sigma$ . Compute

$$E_0 [(S_T)^\lambda].$$

**Solution:** Using Ito's lemma, we have

$$\begin{aligned} dS_t^\lambda &= \lambda S_t^{\lambda-1} dS_t + \frac{1}{2} \lambda (\lambda - 1) S_t^{\lambda-2} (dS_t)^2 \\ &= \lambda S_t^{\lambda-1} (\mu S_t dt + \sigma S_t dZ_t) + \frac{1}{2} \lambda (\lambda - 1) \sigma^2 S_t^\lambda dt \\ &= S_t^\lambda \left[ \lambda \left( \mu + \frac{1}{2} (\lambda - 1) \sigma^2 \right) dt + \lambda \sigma dZ_t \right] \end{aligned}$$

Let  $X_t = S_t^\lambda$ . Then we have

$$\frac{dX_t}{X_t} = \mu_X dt + \sigma_X dZ_t$$

where

$$\begin{aligned} \mu_X &= \lambda \left( \mu + \frac{1}{2} (\lambda - 1) \sigma^2 \right) \\ \sigma_X &= \lambda \sigma \end{aligned}$$

Hence,  $X_t$  also follows geometric Brownian motion. As a result,

$$X_T = X_0 e^{(\mu_X - \frac{1}{2}\sigma_X^2)T + \sigma_X Z_T}$$

Therefore,

$$\begin{aligned} E[S_T^\lambda] &= E[X_T] \\ &= X_0 e^{\mu_X T} \\ &= S_0^\lambda e^{\lambda(\mu + \frac{1}{2}(\lambda-1)\sigma^2)T} \end{aligned}$$

7. Suppose that, under  $\mathbf{P}$ , the price of a stock paying no dividends follows

$$\frac{dS_t}{S_t} = \mu(S_t) dt + \sigma(S_t) dZ_t$$

Assume that the SPD in this market satisfies

$$\frac{d\pi_t}{\pi_t} = -r dt - \eta_t dZ_t$$

(a) How does  $\eta_t$  relate to  $r$ ,  $\mu_t$ , and  $\sigma_t$ ?

**Solution:** By definition of SPD, we must have  $\pi_t S_t$  as a martingale. By Ito's lemma,

$$\begin{aligned} d(\pi_t S_t) &= \pi_t dS_t + S_t d\pi_t + dS_t d\pi_t \\ &= \mu(S_t) \pi_t S_t dt - r \pi_t S_t dt - \sigma(S_t) \eta_t \pi_t S_t + [\dots] dZ_t \\ &= \pi_t S_t (\mu(S_t) - r - \sigma(S_t) \eta_t) dt + [\dots] dZ_t \end{aligned}$$

To qualify for a martingale process, the drift of  $\pi_t S_t$  must be 0. So

$$\begin{aligned} \mu(S_t) - r - \sigma(S_t) \eta_t &= 0 \\ \eta_t &= \frac{\mu(S_t) - r}{\sigma(S_t)} \end{aligned}$$

(b) Suppose that there exists a derivative asset with price  $C(t, S_t)$ . Derive the instantaneous expected return on this derivative as a function of  $t$  and  $S_t$ .

**Solution:** Recall that in general, if  $X_t$  is the price of a traded asset with the dynamics

$$\frac{dX_t}{X_t} = \mu_X dt + \sigma_X dZ_t$$

then the expected excess return is given by

$$\begin{aligned} (\mu_X - r) dt &= E^P \left[ \frac{dX_t}{X_t} \right] - E^Q \left[ \frac{dX_t}{X_t} \right] \\ &= \sigma_X \cdot \eta_t \end{aligned}$$

Now note that the price of our derivative asset  $C(t, S_t)$  satisfies

$$\begin{aligned} dC(t, S_t) &= \frac{\partial C(t, S_t)}{\partial t} dt + \frac{\partial C(t, S_t)}{\partial S_t} dS_t + \frac{1}{2} \frac{\partial^2 C(t, S_t)}{\partial S_t^2} (dS_t)^2 \\ &= \left( \frac{\partial C(t, S_t)}{\partial t} + \frac{\partial C(t, S_t)}{\partial S_t} (\mu(S_t) S_t) + \frac{1}{2} \frac{\partial^2 C(t, S_t)}{\partial S_t^2} (\sigma(S_t) S_t)^2 \right) dt \\ &\quad + \left( \frac{\partial C(t, S_t)}{\partial S_t} (\sigma(S_t) S_t) \right) dZ_t \end{aligned}$$

and thus

$$\begin{aligned} \frac{dC(t, S_t)}{C(t, S_t)} &= \frac{1}{C(t, S_t)} \left( \frac{\partial C(t, S_t)}{\partial t} + \frac{\partial C(t, S_t)}{\partial S_t} (\mu(S_t) S_t) + \frac{1}{2} \frac{\partial^2 C(t, S_t)}{\partial S_t^2} (\sigma(S_t) S_t)^2 \right) dt \\ &\quad + \frac{1}{C(t, S_t)} \left( \frac{\partial C(t, S_t)}{\partial S_t} (\sigma(S_t) S_t) \right) dZ_t \\ &= \mu_C(t, S_t) dt + \sigma_C(t, S_t) dZ_t \end{aligned}$$

So we can express the expected excess return on this derivative as

$$\mu_C(t, S_t) - r = \frac{1}{C(t, S_t)} \left( \frac{\partial C(t, S_t)}{\partial t} + \frac{\partial C(t, S_t)}{\partial S_t} (\mu(S_t) S_t) + \frac{1}{2} \frac{\partial^2 C(t, S_t)}{\partial S_t^2} (\sigma(S_t) S_t)^2 \right) - r$$

or

$$\sigma_C(t, S_t) \eta_t = \frac{1}{C(t, S_t)} \left( \frac{\partial C(t, S_t)}{\partial S_t} (\sigma(S_t) S_t) \right) \cdot \frac{\mu(S_t) - r}{\sigma(S_t)}$$

- (c) Derive the PDE on the price of the derivative  $C(t, S)$ , assuming that its payoff is given by  $H(S_T)$  at time  $T$ .

**Solution:** Under risk-neutral measure, all traded securities have instantaneous expected return  $r$ . So under risk-neutral measure, the stock has drift  $rS_t$  and the derivative

$$E_t^Q \left[ \frac{dC}{C} \right] = \left[ \frac{\partial C}{\partial t} + \frac{\partial C}{\partial S} rS_t + \frac{1}{2} \frac{\partial^2 C}{\partial S^2} \sigma(S_t)^2 S_t^2 \right] dt / C = r dt$$

We have a PDE

$$\frac{\partial C}{\partial t} + \frac{\partial C}{\partial S} rS_t + \frac{1}{2} \frac{\partial^2 C}{\partial S^2} \sigma(S_t)^2 S_t^2 - rC = 0$$

- (d) Suppose that there is another derivative trading, with a price  $D(t, S_t)$  which does not satisfy the PDE you have derived above. Construct a trading strategy generating arbitrage profits using this derivative, the risk-free asset and the stock.

**Solution:** Suppose for derivative  $D(t, S_t)$ , we have

$$\frac{\partial D}{\partial t} + \frac{\partial D}{\partial S} rS_t + \frac{1}{2} \frac{\partial^2 D}{\partial S^2} \sigma(S_t)^2 S_t^2 - rD > 0$$

at some  $(t, S_t)$ . Then consider holding the derivative, shorting  $\frac{\partial D}{\partial S}$  shares and financing the above position  $(D - \frac{\partial D}{\partial S} S_t)$  by borrowing at short rate. The instantaneous gain is

$$\begin{aligned} & \underbrace{\frac{\partial D}{\partial t} dt + \frac{\partial D}{\partial S} dS_t + \frac{1}{2} \frac{\partial^2 D}{\partial S^2} (dS_t)^2}_{\text{Gain from derivative}} \underbrace{- \frac{\partial D}{\partial S} dS_t}_{\text{Gain from stock}} \underbrace{- \left( D - \frac{\partial D}{\partial S} S_t \right) r dt}_{\text{Interest payment}} \\ &= \frac{\partial D}{\partial t} dt + \frac{1}{2} \frac{\partial^2 D}{\partial S^2} \sigma(S_t)^2 S_t^2 dt - \left( D - \frac{\partial D}{\partial S} S_t \right) r dt \\ &= \left( \frac{\partial D}{\partial t} + \frac{\partial D}{\partial S} r S_t + \frac{1}{2} \frac{\partial^2 D}{\partial S^2} \sigma(S_t)^2 S_t^2 - r D \right) dt > 0 \end{aligned}$$

This is a riskless arbitrage. We can form a similar riskless arbitrage if

$$\frac{\partial D}{\partial t} + \frac{\partial D}{\partial S} r S_t + \frac{1}{2} \frac{\partial^2 D}{\partial S^2} \sigma(S_t)^2 S_t^2 - r D < 0$$

8. Consider a futures contract with price changing according to

$$\begin{aligned} F_{t+1} &= F_t + \lambda + \mu_t + \sigma_F \varepsilon_t, \\ \mu_{t+1} &= \rho \mu_t + \sigma_\mu u_t \end{aligned}$$

where  $\varepsilon_t$  and  $u_t$  are independent IID  $\mathcal{N}(0, 1)$  random variables. Assume that the interest rate is constant at  $r$ . Your objective is to construct an optimal strategy of trading futures between  $t = 0$  and  $T$  to maximize the terminal objective

$$\mathbb{E} \left[ -e^{-\alpha W_T} \right]$$

where  $W_T$  is the terminal value of the portfolio. Assume the initial portfolio value of  $W_0$ .

(a) Formulate the problem as a dynamic program. Describe the state vector, verify that it follows a controlled Markov process.

**Solution:** The state vector for this problem is  $(t, W_t, \mu_t)$ . Let  $\theta_t$  be the control variable that represents the number of futures contracts at time  $t$ . Clearly,  $\mu_{t+1}$  follows a Markov process given the state variable  $\mu_t$ . Furthermore, to verify that  $W_t$  is a controlled Markov process, note that

$$\begin{aligned} W_{t+1} &= \theta_t (F_{t+1} - F_t) + (1 + r) W_t \\ &= \theta_t (\lambda + \mu_t + \varepsilon_{t+1}) + (1 + r) W_t \end{aligned}$$

and as such, the conditional distribution of  $W_{t+1}$  only depends on the state variables  $W_t$  and  $\mu_t$ .

- (b) Derive the value function at  $T$  and  $T - 1$  and optimal trading strategy at  $T - 1$  and  $T - 2$ .

**Solution:** Let us start from  $t = T$ . The value function is simply given by

$$J(T, W_T, \mu_T) = -\exp(-\alpha W_T)$$

Now, for  $t = T - 1$ , the Bellman equation says

$$\begin{aligned} J(T - 1, W_{T-1}, \mu_{T-1}) &= \max_{\theta_{T-1}} E_{T-1} [J(T, W_T, \mu_T)] \\ &= \max_{\theta_{T-1}} E_{T-1} [-\exp(-\alpha W_T)] \\ &= \max_{\theta_{T-1}} E_{T-1} [-\exp(-\alpha (\theta_{T-1} (\lambda + \mu_{T-1} + \epsilon_T) + (1 + r) W_{T-1}))] \end{aligned}$$

Simple algebra leads to the first order condition

$$\theta_{T-1}^* = \frac{\lambda + \mu_{T-1}}{\alpha \sigma_F^2}$$

Plugging this into the Bellman equation, we get the value function at time  $t = T - 1$ :

$$J(T - 1, W_{T-1}, \mu_{T-1}) = -\exp\left(-\frac{1}{2\sigma_F^2} (\lambda + \mu_{T-1})^2 - \alpha (1 + r) W_{T-1}\right)$$

For  $t = T - 2$ , the Bellman equation is given by

$$\begin{aligned} J(T - 2, W_{T-2}, \mu_{T-2}) &= \max_{\theta_{T-2}} E_{T-2} [J(T - 1, W_{T-1}, \mu_{T-1})] \\ &= \max_{\theta_{T-2}} E_{T-2} \left[ -\exp\left(-\frac{1}{2\sigma_F^2} (\lambda + \mu_{T-1})^2 - \alpha (1 + r) W_{T-1}\right) \right] \end{aligned}$$

where

$$\begin{aligned} F_{T-1} &= F_{T-2} + \lambda + \mu_{T-2} + \sigma_F \epsilon_{T-1} \\ \mu_{T-1} &= \rho \mu_{T-2} + \sigma_u u_{T-1} \\ W_{T-1} &= \theta_{T-2} (\lambda + \mu_{T-2} + \epsilon_{T-1}) + (1 + r) W_{T-2} \end{aligned}$$

We can repeat a similar calculation as before and arrive at the optimal control

$$\theta_{T-2}^* = \frac{\lambda + \mu_{T-2}}{\alpha (1 + r) \sigma_F^2}$$

9. Suppose you can trade two assets, a risk-free bond with interest rate  $r$  and a risky stock, paying no dividends, with price  $S_t$ . Assume  $S_{t+1} = S_t \times \exp(\mu + \sigma \epsilon_t)$  where  $\epsilon_t$  are IID  $\mathcal{N}(0, 1)$  random variables.

Assume that whenever you buy the stock you must pay transaction costs, but you can sell stock without costs. Specifically, when you buy  $X$  dollars worth of stock, you must pay  $(1 + \tau)X$ , so the fee is proportional, given by  $\tau$ . Your objective is to figure out how to trade optimally to maximize the objective

$$E[-e^{-\alpha W_T}]$$

where  $W_T$  is the terminal value of the portfolio.

- (a) What should be the state vector for this problem? Formulate the problem as a dynamic program, verify the assumptions on the state vector and the payoff function.

**Solution:** The state vector for this problem is  $(t, X_t, B_t)$ , where  $X_t$  and  $B_t$  represent the dollar value of stock holdings and bond holdings at the beginning of period  $t$ . Let  $\lambda_t$  be the control variable that represents the net increase in the dollar value of the stock holdings as a result of rebalancing at time  $t$ . So, for example, if  $\lambda_t > 0$ , then we invest more in the stock at time  $t$ . Note that we incur a transaction cost of  $\tau\lambda_t$  in period  $t$  if  $\lambda_t > 0$  and none otherwise. Under this definition, the dollar value of stock holdings at the end of the period  $t$  is  $X_t + \lambda_t$ . Then the dynamics of  $X_t$  and  $B_t$  are given by

$$\begin{aligned} X_{t+1} &= \exp(\mu + \sigma\epsilon_{t+1})(X_t + \lambda_t) \\ B_{t+1} &= (1 + r)(B_t - \lambda_t - \tau \cdot \max(\lambda_t, 0)) \end{aligned}$$

It is clear that  $X_t$  and  $B_t$  are controlled Markov processes because their conditional distributions of  $X_{t+1}$  and  $B_{t+1}$  only depend on the state variables  $X_t$  and  $B_t$  and the control variable  $\lambda_t$ . Furthermore, as such, these state variables capture all the relevant information regarding the dynamics of the variable of ultimate interest,  $W_t = X_t + B_t$ .

- (b) Write down the Bellman equation.

**Solution:** The Bellman equation is given by

$$J(t, X_t, B_t) = \max_{\lambda_t} E_t[J(t + 1, X_{t+1}, B_{t+1})]$$

where

$$\begin{aligned} X_{t+1} &= \exp(\mu + \sigma\epsilon_{t+1})(X_t + \lambda_t) \\ B_{t+1} &= (1 + r)(B_t - \lambda_t - \tau \cdot \max(\lambda_t, 0)) \end{aligned}$$

The Bellman equation in the terminal time period is simply

$$J(T, X_T, B_T) = -\exp(-\alpha(X_T + B_T))$$

10. Suppose we observe returns on  $N$  independent trading strategies,  $r_t^n$ ,  $n = 1, 2$ ,  $t = 1, \dots, T$ . Assume that returns are IID over time, and each strategy has normal distribution:

$$r_t^n \sim \mathcal{N}(\mu_n, \sigma^2)$$

Assume  $\mu_1 > \mu_2$ .

- (a) Estimate the mean return on each strategy by maximum likelihood. Express  $\hat{\mu}_n$  as a function of observed returns on strategy  $n$ .

**Solution:** It is a standard calculation (see, for example, lecture notes) to show that the maximum likelihood estimate of the mean of a normal distribution is the sample mean. Hence

$$\hat{\mu}_n = \frac{1}{T} \sum_{t=1}^T r_t^n$$

- (b) Since returns are normally distributed,  $\hat{\mu}_n$  is also normally distributed. Describe its distribution. (In general, for arbitrary return distribution,  $\hat{\mu}_n$  is only approximately normal).

**Solution:** Since  $r_t^n$ ,  $t = 1, \dots, T$ , are drawn from IID normal distribution, the sample mean  $\hat{\mu}_n$  is a sum of independent normal variables and is again normally distributed. Since  $r_t^n \sim N(\mu_n, \sigma^2)$ , the sample mean  $\hat{\mu}_n$  is distributed  $N\left(\mu_n, \frac{\sigma^2}{T}\right)$ .

- (c) What is the distribution of  $\max_n(\hat{\mu}_n)$ ? characterize it using the CDF function.

**Solution:** Note that  $\hat{\mu}_1$  and  $\hat{\mu}_2$  are independent so that the cumulative distribution function (CDF) of  $\max(\hat{\mu}_1, \hat{\mu}_2)$ , denoted  $F_{max}(\cdot)$ , is given by

$$\begin{aligned} F_{max}(x) &= P(\max(\hat{\mu}_1, \hat{\mu}_2) \leq x) \\ &= P(\hat{\mu}_1 \leq x) \cdot P(\hat{\mu}_2 \leq x) \\ &= F_1(x) \cdot F_2(x) \\ &= \Phi\left(\frac{\sqrt{T}(x - \mu_1)}{\sigma}\right) \cdot \Phi\left(\frac{\sqrt{T}(x - \mu_2)}{\sigma}\right) \end{aligned}$$

where we use the fact that  $\hat{\mu}_n \sim N\left(\mu_n, \frac{\sigma^2}{T}\right)$ .

- (d) Suppose you are interested in identifying the strategy with the higher mean return. You pick the strategy with the higher estimated mean. What is the probability that you have made a mistake?

**Solution:** We are interested in the probability that  $\hat{\mu}_2$  is greater than  $\hat{\mu}_1$  (so that we mistakenly infer that the second trading strategy has the higher mean). Recalling that  $\hat{\mu}_1$  and  $\hat{\mu}_2$  are independent and  $\hat{\mu}_n \sim N\left(\mu_n, \frac{\sigma^2}{T}\right)$ , we know that

$$\hat{\mu}_1 - \hat{\mu}_2 \sim N\left(\mu_1 - \mu_2, \frac{2\sigma^2}{T}\right)$$

Furthermore,

$$\begin{aligned} P(\hat{\mu}_1 \leq \hat{\mu}_2) &= P(\hat{\mu}_1 - \hat{\mu}_2 \leq 0) \\ &= 1 - \Phi\left(\frac{T(\mu_1 - \mu_2)}{2\sigma^2}\right) \end{aligned}$$

We see that this probability is decreasing in the distance between the true means  $\mu_1 - \mu_2$  and decreasing in the number of observations  $T$ . On the other hand, this probability is increasing in  $\sigma^2$ , reflecting the difficulty to estimate the mean when the distribution has large variance.

11. Suppose interest rate follows an AR(1) process

$$r_t - \bar{r} = \theta(r_{t-1} - \bar{r}) + \varepsilon_t$$

where  $\varepsilon_t$  are IID  $\mathcal{N}(0, \sigma^2)$  random variables. You want to estimate the average rate,  $\bar{r}$ , based on the sample  $r_t, t = 0, 1, \dots, T$ . Assume that we know the true value of  $\theta$ .

(a) Derive the estimate of  $\bar{r}$  by maximum likelihood.

**Solution:** It is easily seen that

$$L(\bar{r}, \theta, \sigma^2 | r_1, \dots, r_T) = \prod_{t=1}^T f(r_t | \bar{r}, \theta, \sigma^2; r_0, \dots, r_{t-1})$$

so that

$$\begin{aligned} \mathcal{L}(\bar{r}, \theta, \sigma^2 | r_0, \dots, r_T) &= \sum_{t=1}^T \ln f(r_t | \bar{r}, \theta, \sigma^2; r_1, \dots, r_{t-1}) \\ &= \sum_{t=1}^T \ln \left( \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} ((r_t - \bar{r}) - \theta(r_{t-1} - \bar{r}))^2\right) \right) \end{aligned}$$

In particular, the maximum likelihood estimate of  $\bar{r}$ , call it  $\hat{\bar{r}}$ , satisfies the first order condition

$$\begin{aligned} 0 &= \frac{\partial \mathcal{L}}{\partial \bar{r}} \\ &= \frac{1 - \theta}{\sigma^2} \sum_{t=1}^T ((r_t - \bar{r}) - \theta(r_{t-1} - \bar{r})) \end{aligned}$$

It follows that

$$\hat{\bar{r}} = \frac{1}{T(1 - \theta)} \sum_{t=1}^T (r_t - \theta r_{t-1})$$

- (b) Show that this estimate is valid even if the shocks  $\varepsilon_t$  are not normally distributed, as long as the mean of  $\varepsilon_t$  is zero.

**Solution:** Now we assume that  $\varepsilon_t$  are independent over time and  $E[\varepsilon_t] = 0$ . Note that

$$\begin{aligned}\hat{r} &= \frac{1}{T(1-\theta)} \sum_{t=1}^T (r_t - \theta r_{t-1}) \\ &= \frac{1}{T(1-\theta)} \sum_{t=1}^T ((1-\theta)\bar{r} + \varepsilon_t) \\ &= \bar{r} + \frac{1}{T(1-\theta)} \sum_{t=1}^T \varepsilon_t\end{aligned}$$

By the Law of Large Numbers,

$$\frac{1}{T(1-\theta)} \sum_{t=1}^T \varepsilon_t \rightarrow 0$$

in probability, and hence we establish that  $\hat{r}$  converges in probability to the true value  $\bar{r}$  and therefore is consistent.

- (c) Treating  $\varepsilon_t$  as IID, derive the asymptotic variance of your estimator of  $\bar{r}$ . Do not use Newey-West, derive the result from first principles. How does the answer depend on  $\theta$ ?

**Solution:** For this part, we maintain that  $\varepsilon_t$  are IID over time,  $E[\varepsilon_t] = 0$ , and  $Var[\varepsilon_t] = \sigma^2$ . In the last part, we saw that

$$\hat{r} = \bar{r} + \frac{1}{T(1-\theta)} \sum_{t=1}^T \varepsilon_t$$

and hence

$$\sqrt{T}(\hat{r} - \bar{r}) = \frac{1}{1-\theta} \frac{1}{\sqrt{T}} \sum_{t=1}^T \varepsilon_t$$

By the Central Limit Theorem,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \varepsilon_t \Rightarrow N(0, \sigma^2)$$

and we conclude that

$$\sqrt{T}(\hat{r} - \bar{r}) \Rightarrow N\left(0, \frac{\sigma^2}{(1-\theta)^2}\right)$$

Note that the asymptotic variance of our estimator is increasing in  $\theta$ . This makes sense, because in the limiting case where  $\theta = 0$ ,  $r_t$  are actually IID over time and asymptotic variance of the sample mean is equal to the variance of the shocks. In the other limiting case where  $\theta \rightarrow 1$ , we are very close to the unit-root case (random walk) where we do not have mean-reversion and hence estimation of the long-run mean becomes increasingly difficult and imprecise.

Moreover, calculating theoretical asymptotic variance of the estimator as in this question provides an alternative to constructing standard errors using the Newey-West method. There are advantages and disadvantages to each method. Theoretical asymptotic variance is a convenient way to see inner workings of the estimator (in our example, the dependence of asymptotic variance on the persistence of AR(1) process) with excellent finite sample properties, assuming correct model specification. However, derivation of asymptotic variance of an estimator may not be so straightforward in more complex situations and standard errors obtained this way are more sensitive to model misspecification compared to the more model-independent Newey-West standard errors.

12. Suppose you observe two time series,  $X_t$  and  $Y_t$ . You have a model for  $Y_t$ :

$$Y_{t+1} = \rho Y_t + (a_0 + a_1 X_t) \varepsilon_{t+1}, \quad t = 0, 1, \dots, T$$

where  $\varepsilon_{t+1} \sim \mathcal{N}(0, 1)$ , IID. Assume that the shocks  $\varepsilon_t$  are independent of the process  $X_t$  and the lagged values of  $Y_t$ . There is no model for  $X_t$ .

(a) Using the GMM framework, which moment condition can be used to estimate  $\rho$ ?

**Solution:** We first want to establish that

$$E[(a_0 + a_1 X_t) \varepsilon_{t+1} | Y_t] = 0$$

To see why this is true, note that

$$\begin{aligned} E[(a_0 + a_1 X_t) \varepsilon_{t+1} | Y_t] &= E[a_0 + a_1 X_t | Y_t] \cdot E[\varepsilon_{t+1} | Y_t] \\ &= E[a_0 + a_1 X_t | Y_t] \cdot E[\varepsilon_{t+1}] \\ &= E[a_0 + a_1 X_t | Y_t] \cdot 0 \\ &= 0 \end{aligned}$$

where the first equality follows from independence of the processes  $X_t$  and  $\varepsilon_t$ , the second equality follows from independence of  $\varepsilon_{t+1}$  with the lagged values of  $X_t$ , and the third equality follows from the assumption that  $\varepsilon_{t+1} \sim N(0, 1)$ .

Now having derived this condition, our usual arguments now allow us to derive the moment conditions

$$E[g(Y_t) (a_0 + a_1 X_t) \varepsilon_{t+1}] = 0$$

for any function  $g$ .

(b) Argue why it is valid to estimate  $\rho$  using an OLS regression of  $Y_{t+1}$  on  $Y_t$ .

**Solution:** In particular, if we pick  $g(Y_t) = Y_t$ , then the moment condition becomes

$$E[Y_t \cdot (a_0 + a_1 X_t) \epsilon_{t+1}] = 0$$

and our sample analogue is

$$E_T[Y_t \cdot (a_0 + a_1 X_t) \epsilon_{t+1}] = 0$$

or

$$E_T[Y_t \cdot (Y_{t+1} - \rho Y_t)] = 0$$

Note that the GMM estimate  $\hat{\rho}$  that solves this sample moment condition is also the OLS estimate from the regression

$$Y_{t+1} = \rho Y_t + u_t$$

where  $u_t$  represents unspecified error terms. The reason why the GMM estimate of  $\rho$  coincides with the OLS estimate is that the sample moment condition for  $\hat{\rho}$  simply says that the residual term  $Y_{t+1} - \hat{\rho}Y_t$  is orthogonal to the regressor  $Y_t$ , and this orthogonality between the fitted residuals and the regressors is the first order condition of the OLS estimate. Hence it is valid to estimate  $\hat{\rho}$  by simply running the regression

$$Y_{t+1} = \rho Y_t + u_t$$

to find the OLS estimate of  $\rho$ .

(c) Suppose that the variance of the estimator  $\hat{\rho}$  is  $(1/T)\sigma_\rho^2$ . Describe how you would test the hypothesis that  $\rho = 0$ .

**Solution:** We are assuming that

$$Var(\hat{\rho}) = \frac{1}{T}\sigma_\rho^2$$

To test the hypothesis that  $\rho = 0$ , we make use of the  $\chi^2$  test (refer to page 32 of Lecture Notes 8). We construct the test statistic as

$$\xi = \frac{\hat{\rho}^2}{Var(\hat{\rho})} = T \frac{\hat{\rho}^2}{\sigma_\rho^2}$$

We reject the null hypothesis of  $\rho = 0$  if the test statistic is sufficiently large, i.e., if

$$\xi \geq \bar{\xi}$$

where the cutoff point  $\bar{\xi}$  is such that

$$CDF_{\chi^2(1)}(\bar{\xi}) = 1 - \alpha$$

and  $\alpha$  is the size of our hypothesis test.

(d) Write down the conditional log-likelihood function  $\mathcal{L}(\rho, a_0, a_1)$ .

**Solution:** Given the model

$$Y_{t+1} = \rho Y_t + (a_0 + a_1 X_t) \epsilon_{t+1}$$

where  $\epsilon_{t+1}$  are IID  $N(0, 1)$ , the log-likelihood function is

$$\begin{aligned} \mathcal{L}(\rho, a_0, a_1) &= \sum_{t=1}^T \ln f(Y_t | X_0, \dots, X_{t-1}, Y_0, \dots, Y_{t-1}; \rho, a_0, a_1) \\ &= \sum_{t=1}^T \ln \left( \frac{1}{\sqrt{2\pi\sigma_{t-1}^2}} \exp \left( -\frac{(Y_t - \rho Y_{t-1})^2}{2\sigma_{t-1}^2} \right) \right) \\ &= \sum_{t=1}^T \left( -\frac{1}{2} \ln(2\pi\sigma_{t-1}^2) - \frac{(Y_t - \rho Y_{t-1})^2}{2\sigma_{t-1}^2} \right) \end{aligned}$$

where the conditional standard deviation of  $Y_{t+1}$ , denoted  $\sigma_t^2$ , is given by

$$\sigma_t^2 = (a_0 + a_1 X_t)^2$$

Therefore,

$$\mathcal{L}(\rho, a_0, a_1) = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^T \ln((a_0 + a_1 X_{t-1})^2) - \frac{1}{2} \sum_{t=1}^T \left( \frac{Y_t - \rho Y_{t-1}}{a_0 + a_1 X_{t-1}} \right)^2$$

(e) Suppose that the parameters  $a_0$  and  $a_1$  are known. Derive the maximum-likelihood estimate for  $\rho$ .

**Solution:** We now suppose that  $a_0$  and  $a_1$  are known. Then the maximum likelihood estimate of  $\rho$  satisfies the first order condition

$$\begin{aligned} 0 &= \frac{\partial \mathcal{L}}{\partial \rho} \\ &= \sum_{t=1}^T \left( \frac{Y_t - \rho Y_{t-1}}{a_0 + a_1 X_{t-1}} \right) \left( \frac{Y_{t-1}}{a_0 + a_1 X_{t-1}} \right) \\ &= \sum_{t=1}^T \left( \frac{Y_{t-1} (Y_t - \rho Y_{t-1})}{(a_0 + a_1 X_{t-1})^2} \right) \end{aligned}$$

and hence

$$\hat{\rho} = \frac{\sum_{t=1}^T \frac{Y_{t-1} Y_t}{(a_0 + a_1 X_{t-1})^2}}{\sum_{t=1}^T \frac{Y_{t-1}^2}{(a_0 + a_1 X_{t-1})^2}}$$

13. Suppose we observe a sequence of IID random variables  $X_t \geq 0$ ,  $t = 1, \dots, T$ , with probability density

$$pdf(X) = \lambda e^{-\lambda X}, \quad X \geq 0$$

- (a) Write down the log-likelihood function  $\mathcal{L}(\lambda)$ .

**Solution:** The log-likelihood function is given by

$$\begin{aligned} \mathcal{L}(\lambda|X_1, \dots, X_T) &= \sum_{t=1}^T \ln p(X_t|\lambda) \\ &= \sum_{t=1}^T (\ln \lambda - \lambda X_t) \\ &= T \ln \lambda - \lambda \sum_{t=1}^T X_t \end{aligned}$$

- (b) Compute the maximum likelihood estimate  $\hat{\lambda}$ .

**Solution:** The maximum likelihood estimate  $\hat{\lambda}$  satisfies the first order condition

$$\begin{aligned} 0 &= \frac{\partial \mathcal{L}}{\partial \lambda} \\ &= \frac{T}{\lambda} - \sum_{t=1}^T X_t \end{aligned}$$

and we get

$$\hat{\lambda} = \left( \frac{1}{T} \sum_{t=1}^T X_t \right)^{-1}$$

- (c) Derive the standard error for  $\hat{\lambda}$ .

**Solution:** To compute the standard error of the maximum likelihood estimator  $\hat{\lambda}$ , we resort to the general GMM standard errors (refer to page 28 of the Lecture Notes 8). We have

$$\begin{aligned} \hat{d} &= \hat{E} \left[ \frac{\partial^2 \ln p(X_t|\lambda)}{\partial \lambda^2} \Big|_{\lambda = \hat{\lambda}} \right] \\ &= \frac{1}{T} \sum_{t=1}^T \frac{\partial^2 \ln p(X_t|\lambda)}{\partial \lambda^2} \Big|_{\lambda = \hat{\lambda}} \\ &= \frac{1}{T} \sum_{t=1}^T \left( -\frac{1}{\hat{\lambda}^2} \right) \\ &= -\frac{1}{\hat{\lambda}^2} \end{aligned}$$

and

$$\begin{aligned}
 \hat{S} &= \hat{E} \left[ \left( \frac{\partial \ln p(X_t | \lambda)}{\partial \lambda} \right)^2 \middle| \lambda = \hat{\lambda} \right] \\
 &= \frac{1}{T} \sum_{t=1}^T \left( \frac{1}{\hat{\lambda}} - X_t \right)^2 \\
 &= \frac{1}{\hat{\lambda}^2} - \frac{2}{\hat{\lambda}} \left( \frac{1}{T} \sum_{t=1}^T X_t \right) + \left( \frac{1}{T} \sum_{t=1}^T X_t^2 \right)
 \end{aligned}$$

But we note that

$$\hat{\lambda} = \left( \frac{1}{T} \sum_{t=1}^T X_t \right)^{-1}$$

and we can simplify  $\hat{S}$  to

$$\hat{S} = -\frac{1}{\hat{\lambda}^2} + \left( \frac{1}{T} \sum_{t=1}^T X_t^2 \right)$$

Finally, the variance of our estimator is given by

$$\text{Var}(\hat{\lambda}) = \frac{\hat{S}}{T \cdot \hat{d}^2}$$

where all the matrix multiplications simplify a great deal since we are working in the scalar case. In particular, putting together expressions for  $\hat{d}$  and  $\hat{S}$ , we get

$$\begin{aligned}
 \text{Var}(\hat{\lambda}) &= \frac{1}{T} \hat{\lambda}^4 \left( -\frac{1}{\hat{\lambda}^2} + \left( \frac{1}{T} \sum_{t=1}^T X_t^2 \right) \right) \\
 &= \frac{1}{T} \left( -\hat{\lambda}^2 + \hat{\lambda}^4 \left( \frac{1}{T} \sum_{t=1}^T X_t^2 \right) \right)
 \end{aligned}$$

where

$$\hat{\lambda} = \left( \frac{1}{T} \sum_{t=1}^T X_t \right)^{-1}$$

14. Suppose you observe a series of observations  $X_t$ ,  $t = 1, \dots, T$ . You need to fit a model

$$X_{t+1} = f(X_t, X_{t-1}; \theta) + \varepsilon_{t+1}$$

where  $E[\varepsilon_{t+1} | X_t, X_{t-1}, \dots, X_1] = 0$ . Innovations  $\varepsilon_{t+1}$  have zero mean conditionally on  $X_t, X_{t-1}, \dots, X_1$ . You also know that innovations  $\varepsilon_{t+1}$  have constant conditional variance:

$$E[\varepsilon_{t+1}^2 | X_t, X_{t-1}, \dots, X_1] = \sigma^2$$

The parameter  $\sigma$  is not known.  $\theta$  is the scalar parameter affecting the shape of the function  $f(X_t, X_{t-1}; \theta)$ .

- (a) Describe how to estimate the parameter  $\theta$  using the quasi maximum likelihood approach. Derive the relevant equations.

**Solution:** Assume that  $\varepsilon_{t+1}|X_t, \dots, X_1$  follows Gaussian distribution  $N(0, \sigma^2)$ . Then the logged likelihood is

$$\begin{aligned} l(X_3, \dots, X_T; \theta, \sigma) &= \sum_{t=2}^{T-1} -\frac{1}{2} \ln 2\pi - \ln \sigma - \frac{\varepsilon_{t+1}^2}{2\sigma^2} \\ &= \sum_{t=2}^{T-1} -\frac{1}{2} \ln 2\pi - \ln \sigma - \frac{(X_{t+1} - f(X_t, X_{t-1}; \theta))^2}{2\sigma^2} \end{aligned}$$

To maximize the logged likelihood, we differentiate w.r.t  $\theta$  and F.O.C gives

$$\theta : \frac{1}{\sigma^2} \sum_{t=2}^{T-1} (X_{t+1} - f(X_t, X_{t-1}; \theta)) \frac{\partial f(X_t, X_{t-1}; \theta)}{\partial \theta} = 0$$

Upon simplifying, we have

$$0 = \sum_{t=2}^{T-1} \left( X_{t+1} - f(X_t, X_{t-1}; \hat{\theta}) \right) \frac{\partial f(X_t, X_{t-1}; \hat{\theta})}{\partial \theta} = \sum_{t=2}^{T-1} \hat{\varepsilon}_{t+1} \frac{\partial f(X_t, X_{t-1}; \hat{\theta})}{\partial \theta}$$

The equivalent moment condition is

$$E \left[ \varepsilon_{t+1} \frac{\partial f(X_t, X_{t-1}; \theta)}{\partial \theta} \right] = 0$$

which is valid.

- (b) Describe in detail how to use parametric bootstrap to estimate a 95% confidence interval for  $\theta$ .

**Solution:** Parametric bootstrap can be done through the following steps:

(1) Estimate  $\hat{\theta}$  using QMLE and obtain a sample of residuals  $\hat{\varepsilon}_{t+1} = X_{t+1} - f(X_t, X_{t-1}; \hat{\theta})$

(2) Fix  $\hat{\theta}$ . Generate  $i$ th sample (for  $i = 1, 2, \dots, N$ ) of  $(X_1^{(i)}, \dots, X_T^{(i)})$  by drawing randomly from the sample residuals with replacement to get  $\hat{\varepsilon}_{t+1}^{(i)}$  and  $X_{t+1}^{(i)} = f(X_t^{(i)}, X_{t-1}^{(i)}; \hat{\theta}) + \hat{\varepsilon}_{t+1}^{(i)}$ . Note that we need to exclude the "burn-in" sample and keep only last  $T$  observations. For  $i$ th sample, estimate  $\hat{\theta}^{(i)}$  with  $(X_1^{(i)}, \dots, X_T^{(i)})$ .

(3) Get  $\hat{\theta}_{2.5\%}$  and  $\hat{\theta}_{97.5\%}$  (5th and 95th percentile) from the sample of estimates  $\hat{\theta}^{(i)}$ ,  $i = 1, 2, \dots, N$ .

(4) The bootstrapped confidence interval is given by  $(\hat{\theta} - (\hat{\theta}_{97.5\%} - \hat{\theta}), \hat{\theta} - (\hat{\theta}_{2.5\%} - \hat{\theta}))$ .

- (c) Describe how to estimate the bias in your estimate of  $\theta$  using parametric bootstrap.

**Solution:** The bias can be estimated using

$$E \left[ \hat{\theta} - \theta \right] \approx \hat{E}_N \left[ \hat{\theta}^{(i)} - \hat{\theta} \right]$$

where  $\hat{E}_N [\cdot]$  represents the sample average of  $N$  bootstrapped estimates.

- (d) Derive the asymptotic standard error for  $\hat{\theta}$  (large  $T$ ) using GMM standard error formulas.

**Solution:** Recall that the moment condition for  $\theta$  is

$$E \left[ (X_{t+1} - f(X_t, X_{t-1}; \theta)) \frac{\partial f(X_t, X_{t-1}; \theta)}{\partial \theta} \right] = 0$$

Asymptotic variance of  $\hat{\theta}$  is  $\frac{1}{T} \left( \hat{d}' \hat{S}^{-1} \hat{d} \right)^{-1}$ .

$$\begin{aligned} \hat{d} &= \hat{E}_T \left[ (X_{t+1} - f(X_t, X_{t-1}; \theta)) \frac{\partial^2 f(X_t, X_{t-1}; \theta)}{\partial \theta^2} - \left( \frac{\partial f(X_t, X_{t-1}; \theta)}{\partial \theta} \right)^2 \right] \\ &\rightarrow E \left[ (X_{t+1} - f(X_t, X_{t-1}; \theta)) \frac{\partial^2 f(X_t, X_{t-1}; \theta)}{\partial \theta^2} - \left( \frac{\partial f(X_t, X_{t-1}; \theta)}{\partial \theta} \right)^2 \right] \\ &= E \left[ \varepsilon_{t+1} \frac{\partial^2 f(X_t, X_{t-1}; \theta)}{\partial \theta^2} \right] - E \left[ \left( \frac{\partial f(X_t, X_{t-1}; \theta)}{\partial \theta} \right)^2 \right] \\ &= 0 - E \left[ \left( \frac{\partial f(X_t, X_{t-1}; \theta)}{\partial \theta} \right)^2 \right] = -E \left[ \left( \frac{\partial f(X_t, X_{t-1}; \theta)}{\partial \theta} \right)^2 \right] \end{aligned}$$

Since for  $j > 0$ ,

$$\begin{aligned} &E \left[ \varepsilon_{t+1} \frac{\partial f(X_t, X_{t-1}; \theta)}{\partial \theta} \varepsilon_{t-j+1} \frac{\partial f(X_{t-j}, X_{t-j-1}; \theta)}{\partial \theta} \right] \\ &= E \left[ E[\varepsilon_{t+1} | X_t, X_{t-1}, \dots] \frac{\partial f(X_t, X_{t-1}; \theta)}{\partial \theta} \varepsilon_{t-j+1} \frac{\partial f(X_{t-j}, X_{t-j-1}; \theta)}{\partial \theta} \right] \\ &= 0 \end{aligned}$$

there is no autocorrelation. As a result,

$$\begin{aligned}
\hat{S} &= \hat{E}_T \left[ (X_{t+1} - f(X_t, X_{t-1}; \theta))^2 \left( \frac{\partial f(X_t, X_{t-1}; \theta)}{\partial \theta} \right)^2 \right] \\
&\rightarrow E \left[ (X_{t+1} - f(X_t, X_{t-1}; \theta))^2 \left( \frac{\partial f(X_t, X_{t-1}; \theta)}{\partial \theta} \right)^2 \right] \\
&= E \left[ \varepsilon_{t+1}^2 \left( \frac{\partial f(X_t, X_{t-1}; \theta)}{\partial \theta} \right)^2 \right] \\
&= E \left[ E[\varepsilon_{t+1}^2 | X_t, X_{t-1}] \left( \frac{\partial f(X_t, X_{t-1}; \theta)}{\partial \theta} \right)^2 \right] \\
&= \sigma^2 E \left[ \left( \frac{\partial f(X_t, X_{t-1}; \theta)}{\partial \theta} \right)^2 \right]
\end{aligned}$$

Thus, asymptotic variance is  $\frac{1}{T} \sigma^2 / E \left[ \left( \frac{\partial f(X_t, X_{t-1}; \theta)}{\partial \theta} \right)^2 \right]$ .

15. Consider an estimator  $\hat{\theta}$  for a scalar-valued parameter  $\theta$ . Suppose you know, as a function of the true parameter value  $\theta_0$ , the distribution function of the estimator, i.e., you know

$$CDF_{\hat{\theta} - \theta_0}(x)$$

(In practice, you may be able to estimate the above CDF using bootstrap). Note that this CDF does not depend on model parameters.

Based on the definition of the confidence interval, derive a formula for a confidence interval which covers the true parameter value with probability 95%.

**Solution:** Since the CDF of  $\hat{\theta} - \theta_0$  is independent of the parameter  $\theta_0$ , the 2.5 and 97.5 percentiles of the distribution, denoted as  $\alpha_{2.5\%}$  and  $\alpha_{97.5\%}$ , are fixed numbers independent of  $\theta_0$ . As a result,

$$\Pr(\alpha_{2.5\%} < \hat{\theta} - \theta_0 < \alpha_{97.5\%}) = 0.95$$

Rearranging the inequalities, we have

$$\Pr(\hat{\theta} - \alpha_{97.5\%} < \theta_0 < \hat{\theta} - \alpha_{2.5\%}) = 0.95$$

Hence, a 95% confidence interval is  $[\hat{\theta} - \alpha_{97.5\%}, \hat{\theta} - \alpha_{2.5\%}]$ , which illustrates why we have  $[\hat{\theta} - (\theta_{97.5\%}^* - \hat{\theta}), \hat{\theta} - (\theta_{2.5\%}^* - \hat{\theta})]$  as the bootstrapped confidence interval.  $\theta^* - \hat{\theta}$

has approximately the same distribution as  $\hat{\theta} - \theta_0$ . The 2.5 and 97.5 percentiles of the two distributions are also approximately the same. As a result,  $\alpha_{2.5\%}$  and  $\alpha_{97.5\%}$  can be approximated by  $(\theta_{2.5\%}^* - \hat{\theta})$  and  $(\theta_{97.5\%}^* - \hat{\theta})$ .

MIT OpenCourseWare  
<http://ocw.mit.edu>

15.450 Analytics of Finance  
Fall 2010

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.