SPRING 2023 FINC B 9325 Financial Econometrics – Time Series
Homework 5 Sample Solution
Geert Bekaert

Question 1:

The distribution of stock returns has fat tails (see lecture 1). As a result, we need models that deliver fat-tailed distributions (e.g., to manage risks, to price options on stocks). One way of doing this is to introduce jumps or crashes in the model. This problem describes the building blocks of how to do that. Consider a Bernoulli-distributed variable:

$$B_t = \begin{cases} 0 & \text{with probability } 1-p \\ 1 & \text{with probability } p \end{cases}$$

and two independent Standard Normal variables ε_t and δ_t. Define the jump as

$$J_t = B_t \times (\mu_J + \sigma_J \delta_t)$$

and let log returns follow the process:

$$r_t = \mu + \sigma\varepsilon_t + J_t.$$

Assume B_t is independent of ϵ_t and δ_t.

1. Derive the population mean, variance, skewness and excess kurtosis of this distribution.

   Solution:

   Mean

$$\begin{aligned}\mathbb{E}[r_t] &= \mathbb{E}[\mu + \sigma\epsilon_t + J_t] \\ &= \mu + \mathbb{E}[B_t(\mu_J + \sigma_J\delta_t)] \\ &= \mu + p\mu_J\end{aligned}$$

   Variance

$$\begin{aligned}\mathbb{V}[r_t] &= \mathbb{V}[\mu + \sigma\epsilon_t + J_t] \\ &= \mathbb{V}[\sigma\epsilon_t] + \mathbb{V}[B_t(\mu_J + \sigma_J\delta_t)] \\ &= \sigma^2 + \mathbb{V}[B_t\mu_J] + \mathbb{V}[B_t\sigma_J\delta_t] \\ &= \sigma^2 + \mu_J^2 p(1-p) + \sigma_J^2(p(1-p) + p^2) \\ &= \sigma^2 + \mu_J^2 p(1-p) + \sigma_J^2 p\end{aligned}$$

Skewness

First,

$$\begin{aligned}
\mathbb{E}[(r_t - \mathbb{E}[r_t])^3] &= \mathbb{E}[(\sigma\epsilon_t + (B_t - p)\mu_J + \sigma_J B_t \delta_t)^3] \\
&= \mu_J^3 \mathbb{E}[(B_t - p)^3] + 3\mu_J \sigma_J^2 \mathbb{E}[(B_t - p)B_t^2 \delta_t^2] \\
&= \mu_J^3 \mathbb{E}[B_t^3 - 3pB_t^2 + 3p^2 B_t - p^3] + 3\mu_J \sigma_J^2 p(1-p) \\
&= \mu_J^3 p(1-p)(1-2p) + 3\mu_J \sigma_J^2 p(1-p) \\
&= p(1-p)\mu_J((1-2p)\mu_J^2 + 3\sigma_J^2)
\end{aligned}$$

Using this result,

$$\begin{aligned}
\text{skewness} &= \frac{\mathbb{E}[(r_t - \mathbb{E}[r_t])^3}{\mathbb{V}[r_t]^{3/2}} \\
&= \frac{p(1-p)\mu_J((1-2p)\mu_J^2 + 3\sigma_J^2)}{(\sigma^2 + \mu_J^2 p(1-p) + \sigma_J^2 p)^{3/2}}
\end{aligned}$$

Excess Kurtosis

First,

$$\begin{aligned}
\mathbb{E}[(r_t - \mathbb{E}[r_t])^4] &= \mathbb{E}[(\sigma\epsilon_t + (B_t - p)\mu_J + \sigma_J B_t \delta_t)^4] \\
&= \sigma^4 \mathbb{E}[\epsilon_t^4] + 6\sigma^2 \mu_J^2 \mathbb{E}[\epsilon_t^2 (B_t - p)^2] + 6\sigma^2 \sigma_J^2 \mathbb{E}[\epsilon_t^2 B_t^2 \delta_t^2] \\
&\quad + \mu_J^4 \mathbb{E}[(B_t - p)^4] + 6\mu_J^2 \sigma_J^2 \mathbb{E}[(B_t - p)^2 B_t^2 \delta_t^2] + \sigma_J^4 \mathbb{E}[B_t^4 \delta_t^4] \\
&= 3\sigma^4 + 6\sigma^2 p(1-p)\mu_J^2 + 6\sigma^2 \sigma_J^2 p \\
&\quad + \mu_J^4 \underbrace{(p - 4p^2 + 6p^3 - 3p^4)}_{p(1-p)(1-3(1-p)p)} + 6\mu_J^2 \sigma_J^2 p(1-p)^2 + 3p\sigma_J^4
\end{aligned}$$

Using this result,

$$\begin{aligned}
\text{excess kurtosis} &= \frac{\mathbb{E}[(r_t - \mathbb{E}[r_t])^4 - 3\mathbb{V}[r_t]^2}{\mathbb{V}[r_t]^2} \\
&= \frac{\mu_J^4 p(1-p)(1-6(1-p)p) + 6\mu_J^2 \sigma_J^2 p(1-p)(1-2p) + 3p(1-p)\sigma_J^4}{(\sigma^2 + \mu_J^2 p(1-p) + \sigma_J^2 p)^2} \\
&= p(1-p)\frac{\mu_J^4(1-6(1-p)p) + 6\mu_J^2 \sigma_J^2 (1-2p) + 3\sigma_J^4}{(\sigma^2 + \mu_J^2 p(1-p) + \sigma_J^2 p)^2}
\end{aligned}$$

2. Explain carefully why this Bernoulli-normal mixture is potentially a better model of returns than the log-normal model.

Solution:

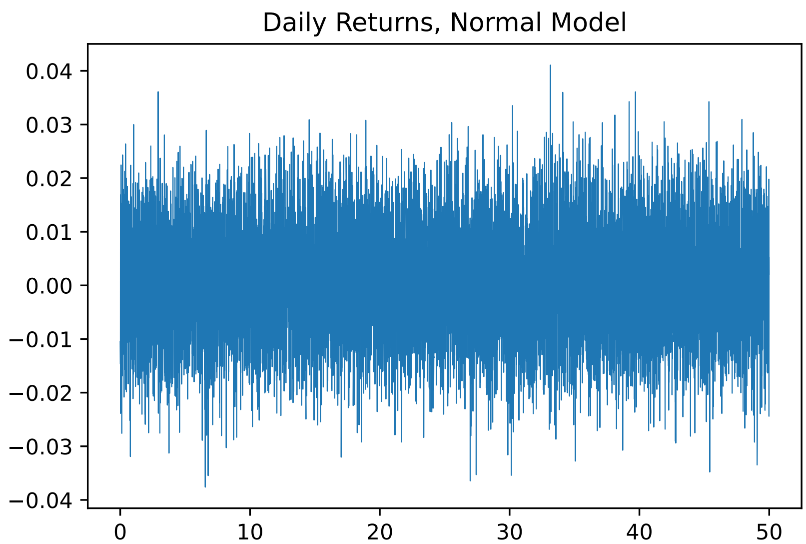
In the log-normal model, log returns are normally distributed, and thus have 0 skewness and 0 excess kurtosis. The Bernoulli-normal mixture model allows us to model skewness and excess kurtosis.

3. Suppose that log returns r_t are simply i.i.d. distributed $N\left(0.0004, 0.01^2\right)$. Plot a simulated series of 12,500 daily observations (50 years). Does it look like the data? What is missing? Next, suppose that log returns r_t are given by the above jump model with the following 5 parameters:

$$(\mu, \sigma, p, \mu_J, \sigma_J) = (0.0004, 0.01, 0.01, -0.03, 0.04)$$

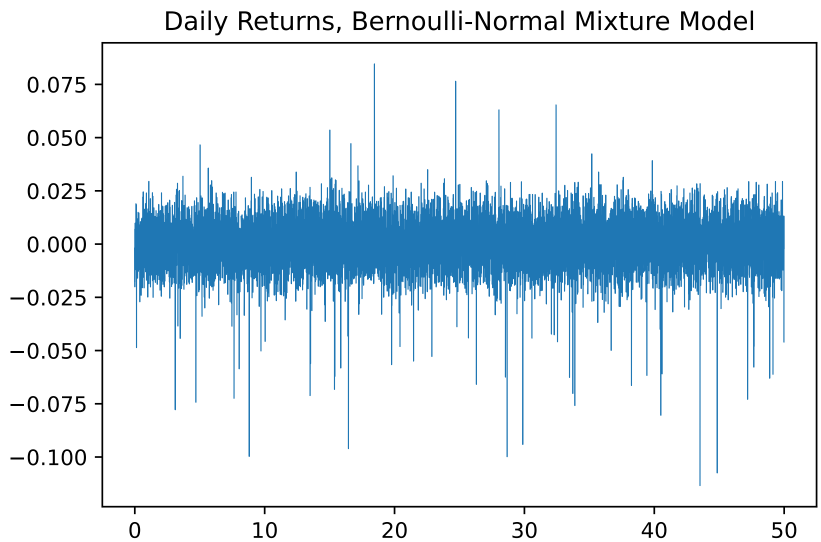
Using these parameter estimates, what are the unconditional mean, standard deviation, skewness and kurtosis of the simulated log stock returns? Again, plot a simulated series of 2500 observations (T = 10 years). Does it look like the data? What is missing?

Solution:



In the data, there are large outliers (see Lecture 1), which are missing in this simulated data. Moments for the mixture model:

	Normal	Analytical	Empirical
Mean	0.0004	0.00010	0.00006
Variance	0.0001	0.00012	0.00013
Skewness	0	-1.20882	-1.21055
Excess Kurtosis	0	10.72904	9.28821



This model is clearly better fitting the unconditional moments of daily stock returns well. However, the fit is only unconditional and happens through having regularly quite big negative shocks. In the actual data, stress times have persistence and there are periods with higher volatility and more negative skewness, which this i.i.d jump model cannot capture.

4. Here, we will assess the properties of the asymptotic White standard errors (no autocorrelation, but heteroskedasticity and non-normality robust standard errors). Simulate N samples from the return process with jumps in subquestion 3 with T = 1 (250 observations), where T indicates years (and we assume 250 trading days per year). Within each of the N samples, calculate the sample mean and construct the "t-stat" of the sample mean, under the null that the sample mean equals zero, using the central limit theorem from the notes and the (correct in this case) assumption that there is no autocorrelation in returns to construct the standard error.

   Give the mean, standard deviation, skewness and kurtosis of the N "t-stats" and sample means you have constructed. Do the sample means and "t-stats" look normally distributed? Compare the average standard error across the N samples to the standard deviation of the N sample mean estimates. Compare and discuss.

   Repeat this analysis for T $= 0.10, 0.50, 2,$ and 4 years worth of daily data and note any patterns of interest.

Solution:

$$\text{t-stat} = \frac{\bar{r}}{s/\sqrt{T}}$$

where $\bar{r}$ is the sample mean and s is the sample standard deviation. We calculate the Jarque-Bera test-statistic, given by

$$JB = \frac{S^2}{6/N} + \frac{(K-3)^2}{24/N}$$

where S is the sample skew, and K is the sample kurtosis. The JB test statistic is distributed $\chi(2)$.

Sample Mean

	T=0.1	T=0.5	T=1	T=2	T=4
mean	0.000226	0.000131	0.000103	0.000095	0.000107
standard deviation	0.002209	0.001003	0.000706	0.000504	0.000361
skewness	-0.234079	-0.122068	-0.111882	-0.004974	0.037630
excess kurtosis	0.569872	0.075092	-0.008048	0.128786	-0.100585
average se	0.002078	0.000981	0.000701	0.000499	0.000352
std of sample mean	0.002209	0.001003	0.000706	0.000504	0.000361
JB test stat	22.663552	2.718379	2.088975	0.695205	0.657561
JB p-value	0.000012	0.256869	0.351872	0.706380	0.719801

T-Stat

	T=0.1	T=0.5	T=1	T=2	T=4
mean	0.146513	0.181724	0.183145	0.219126	0.323797
standard deviation	1.083402	1.024793	1.013301	1.019074	1.032135
skewness	0.012487	0.132335	0.075751	0.164824	0.120872
excess kurtosis	0.470001	-0.054959	-0.009605	0.121771	-0.107417
JB test stat	9.230206	3.044609	0.960221	5.145677	2.915763
JB p-value	0.009901	0.218208	0.618715	0.076319	0.232729

Because the data are essentially i.i.d., convergence to a normal distribution is very swift, depsite the fact that data are drawn from a very non-Gaussian distribution.

Question 2:

Download a set of monthly data from Canvas. The variables are:

- `div_yld`: annual dividend yield
- `log_excess_ret`: log excess returns

1. Run a simple one step ahead predictability test:

$$12 \times r_{t+1} = \alpha + \beta dy_t + e_{t+1}$$

Use "robust" standard errors (controlling for heteroskedasticity, so -called "White" standard errors) and test the null that $\beta = 0$. You can express both in actual percent, so e.g. 0.03 for the dividend yield, and 0.15 for the returns. By multiplying the return with 12, the regression coefficient indicates how much percentage points a one percent (absolute) change in the dividend yield (e.g. from 3% to 4%), changes the annualized equity premium.

Solution:

Using the White (1980) estimator of the covariance matrix, we have

$\hat{\beta}$	se	p-value
2.545	2.269	0.262

We cannot reject the null of no predictability at the 5% level.

2. Create $r_{t+k,k} = r_{t+k} + r_{t+k-1} + \ldots + r_{t+1}$. Set $k = 12$. Run the same regression but with $r_{t+12,12}$ as the dependent variable, with monthly data. These data have an overlapping structure. What process does the residual follow? How would you adjust for that in your standard errors? Again, run the test that β is zero. Also, report the R^2.

Solution:

$$r_{t+k,k} = \alpha + \beta dy_t + e_{t+1}$$

Under the null hypothesis that $\beta = 0$, we have $r_{t+k,k} = k\alpha + \sum_{i=1}^{k} e_{t+i}$. The residual follows a $MA(k-1)$ process. We adjust for that with Newey West standard errors with 18 lags or Hansen-Hodrick (1980) standard errors with 11 lags.

	$\hat{\beta}$	R^2	se	p-value
with OLS se	2.781	0.029	0.601	0.000
with Robust se (White 1980)			0.587	0.000
with NW se (lags = 18)			1.703	0.102
with HH se (lags = 11)			1.921	0.148

With the NW standard error or HH standard error, we cannot reject the null at the 5% level. The other standard errors are, of course, incorrect.

3. Set up a bivariate VAR in r_t and dy_t. Compute the BIC criterion for lags 1 through 3. Run the VAR for the chosen lag order and report the VAR results.

   Solution:

   VAR model:

   $Z_t = \mu + \Phi_1 Z_{t-1} + \Phi_2 Z_{t-2} + \ldots + \Phi_p Z_{t-p} + \epsilon_t, \quad Z_t = [r_t \;\; dy_t]' \quad \epsilon \sim N(0, \Sigma)$

   BIC calculation:

   $$BIC(p) = \log(|\Sigma_p|) + N^2 p \frac{\ln T}{T}, \quad \Sigma_p = \frac{1}{T} \sum_{t=1}^{T} \hat{\epsilon}_t \hat{\epsilon}_t'$$

lag order	BIC
1	-20.84
2	-20.80
3	-20.77

   The lag order that minimizes the BIC is 1. The estimated VAR coefficients are estimated equation-by-equation using OLS.

   $$\hat{\mu} = \begin{bmatrix} -0.0030 \\ -0.0003 \end{bmatrix}, \hat{\Phi}_1 = \begin{bmatrix} 0.0851 & 0.2445 \\ -0.0031 & 0.9880 \end{bmatrix}$$

   The eigenvalues of $\hat{\Phi}_1$ are given by (0.086, 0.987).

4. The population coefficients of the β's in Questions 1) and 2) can be computed from the VAR dynamics, assuming it captures the true dynamics of r_t and dy_t. Report these coefficients with a standard error. How did you compute the standard error?

   Solution:

   Let Γ_0 be the unconditional variance of the VAR, and Γ_j be the lag j cross-covariance matrix for $j > 0$. The notation $A(i, \ell)$ refers to the i-th row and ℓ-th column element of the matrix A. Let β_1 and β_2 refer to the

β's in Questions 1) and 2), respectively. We have

$$\begin{aligned}\beta_1 &= \frac{\text{cov}(12 \times r_{t+1}, dy_t)}{\text{var}(dy_t)} \\ &= \frac{12 \times \text{cov}(r_{t+1}, dy_t)}{\text{var}(dy_{t+1})} \\ &= \frac{12 \times \Gamma_1(1,2)}{\Gamma_0(2,2)}\end{aligned}$$

$$\begin{aligned}\beta_2 &= \frac{\text{cov}(r_{t+k,k}, dy_t)}{\text{var}(dy_t)} \\ &= \frac{\text{cov}(r_{t+1} + \ldots + r_{t+k}, dy_t)}{\text{var}(dy_t)} \\ &= \frac{\text{cov}(r_{t+1}, dy_t) + \ldots + \text{cov}(r_{t+k}, dy_t)}{\text{var}(dy_t)} \\ &= \frac{\Gamma_1(1,2) + \ldots + \Gamma_k(1,2)}{\Gamma_0(2,2)}\end{aligned}$$

The estimates of the coefficients are given by

$$\hat{\beta}_1 = \frac{12 \times \hat{\Gamma}_1(1,2)}{\hat{\Gamma}_0(2,2)}, \quad \hat{\beta}_2 = \frac{\hat{\Gamma}_1(1,2) + \ldots + \hat{\Gamma}_k(1,2)}{\hat{\Gamma}_0(2,2)}$$

where,

$$\begin{aligned}\text{vec}(\hat{\Gamma}_0) &= [I_4 - \hat{\Phi}_1 \otimes \hat{\Phi}_1]^{-1}\text{vec}(\hat{\Sigma}) \\ \hat{\Gamma}_k &= \hat{\Phi}_1 \hat{\Gamma}_{k-1}, \quad k > 0\end{aligned}$$

For standard errors, first let $\theta = [\mu, \text{vec}(\Phi_1), \text{vech}(\Sigma)]$. We have

$$\sqrt{T}(\hat{\theta} - \theta) \sim N(0, V)$$

and the variance-covariance matrix V can be computed using GMM. The orthogonality conditions are given by the 6 OLS orthogonality conditions along with 3 orthogonality conditions from

$$\mathbb{E}[\text{vech}(\epsilon_t \epsilon_t') - \text{vech}(\Sigma)] = 0$$

Note that β_1 and β_2 are functions of θ. Call either function g. By the Delta Method,

$$\sqrt{T}(g(\hat{\theta}) - g(\theta)) \sim N(0, [\nabla h(\theta)]' V [\nabla h(\theta)])$$

The estimates of the coefficients and standard errors are given by

	estimate	se
β_1	2.56	1.89
β_2	2.94	1.94

Clearly, the direct regression estimates and the VAR-implied estimates are quite similar.

5. Estimate a simple restricted VAR. Project the return on just a constant; project the dividend yield onto a constant and its own lag. Save these coefficients. For each data point $t = 2$ through T, you have a vector of residuals. Compute its correlation matrix. Are the residuals correlated? Please run a bootstrap under the null of no predictability with these residuals. For each replication, draw T vectors of possible residuals from the set of observed ones (that is, all observed residual vectors have 1/T probability of occurring and draws are with replacement), reconstruct returns according to the restricted VAR and run the regressions in Questions 1 and 2 and create empirical distributions of the t -statistics. Do the test have good size properties for a 5% two sided test? [It is actually ok to square the t-statistics and investigate the 5% p-value for a $\chi^2(1)$] Also compute the empirical distribution of the R^2 for the regression in Question 2. Describe what you observe. Use 1,000 replications for this bootstrap.

   Solution:

   Denote $\epsilon = [\epsilon_1 \ \epsilon_2]'$. Correlation matrix of residuals:

	ϵ_1	ϵ_2
ϵ_1	1	-0.89
ϵ_2		1

   The residuals are negatively correlated. The restricted VAR is given by:

$$\begin{bmatrix} r_t \\ dy_t \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \alpha \end{bmatrix} \begin{bmatrix} r_{t-1} \\ dy_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

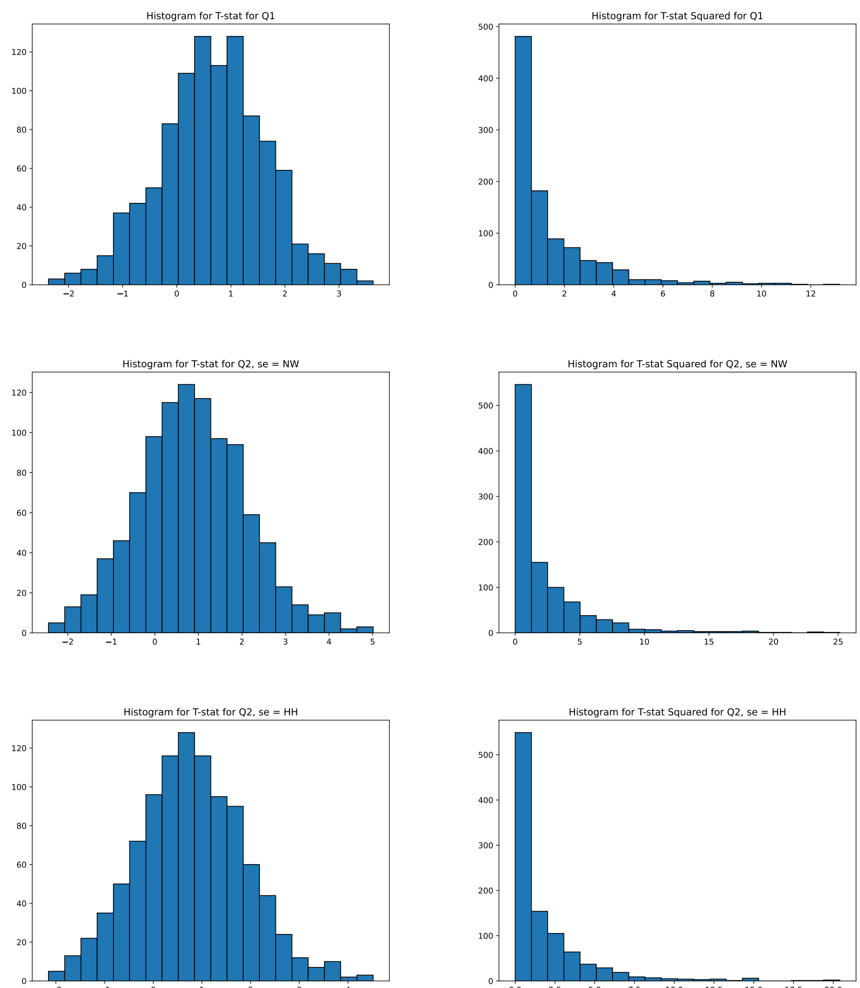
   The bootstrap procedure:

   - Estimate the coefficients $\hat{\mu_1}, \hat{\mu_2}, \hat{\alpha}$ and observed residuals $\{(\hat{\epsilon}_{1t}, \hat{\epsilon_{2t}})\}_{t=2}^{T}$.
   - Draw T vectors of possible residuals $\{(\hat{\epsilon}_{1t}^{boot}, \hat{\epsilon}_{2t}^{boot})\}_{t=1}^{T}$ with replacement from the set of observed residuals.
   - Reconstruct returns and dividend yield as:

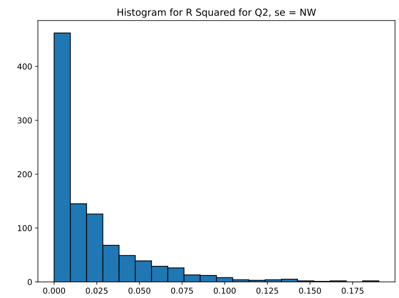
$$\begin{aligned} r_t^{boot} &= \hat{\mu_1} + \hat{\epsilon}_{1t}^{boot} \\ dy_t^{boot} &= \hat{\mu_2} + \hat{\alpha} dy_{t-1}^{boot} + \hat{\epsilon}_{2t}^{boot} \end{aligned}$$

     where dy_1^{boot} is randomly drawn from $\{dy_t\}_{t=1}^{T}$.

- Using the bootstrapped data $\{(r_t^{boot}, dy_t^{boot})\}_{t=1}^{T}$, run the regression either from Q1 or Q2, to obtain the t-stat of the β. Q1 uses robust SE and Q2 uses NW SE with 18 lags or HH SE with 11 lags.
- Repeat 1,000 times.



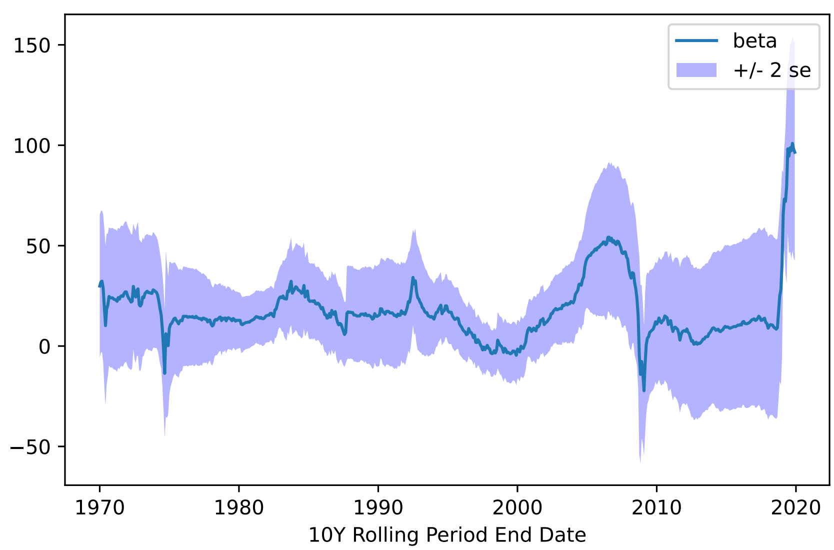
The 5% critical value for $\chi(1)$ is given by 3.84. For Q1, we have that 9% of the empirical distribution of the squared t-stat exceeds 5% critical value. For Q2 with NW se, we have that 18.2% of the empirical distribution of the squared t-stat exceeds the 5% critical value. For Q2 with HH se, we have that 13.8% of the empirical distribution of the squared t-stat exceeds the 5% critical value. Thus, for the multi-horizon regressions, we observe significant size bias and over-rejection for our standard tests.



We find that the mean (median) R^2-statistic is 0.02 (0.01).

6. Perform rolling regressions of the regression in Question 1 with 10 years of data. That is, the first regression uses the first 10 years of monthly data; and then you "roll" through the data set, adding one new data point and dropping the first. This allows you to potentially see any variation in parameter estimates. Plot estimates along with confidence intervals. What do you see? Why?

   Solution:



White (1980) standard errors are used. There is substantial time variation in the estimate of β, that may be due to structural breaks.