Problem 1.

Provide a proof of the Roll estimator for the bid-ask spread based on the return auto-correlation. Using the original paper (Roll, 1984), it should be straightforward to pin down assumptions that will allow for a proof of the results.

Solution.

Following Roll (1984), assume that:

1) The asset is traded in a informativelly efficient market;

2) The probability distribution of the asset return is stationary;

3) There are no trading costs;

4) There is a market maker providing liquidity which knows the efficient value of the asset.

The market maker operates by choosing a bid-ask spread s symmetric around the efficient price p. Under these assumptions, the price conveil all the information available about the market and any changes to the price must be due to unantecipated information. Therefore, the price path can be described by Figure 1 where each possible path has equal probability.
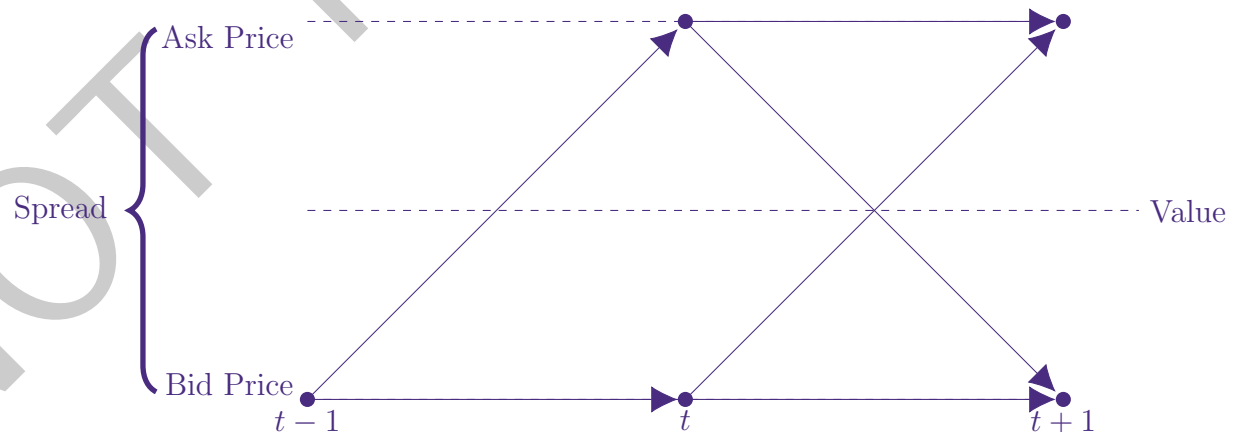

Figure 1: Price path in the Roll model

In other words, the joint distribution of the price change (Δp) at t and $t+1$ conditional on p_{t-1} being at the ask (A) or bid (B) price is given by:

$$p_{t-1} = B \qquad\qquad\qquad p_{t-1} = A$$

$p_{t-1} = B$:

		Δp_t	
		0	$+s$
	$-s$	0	$1/4$
Δp_{t+1}	0	$1/4$	$1/4$
	$+s$	$1/4$	0

$p_{t-1} = A$:

		Δp_t		
		$-s$	0	
	$-s$	0	$1/4$	
	0	$1/4$	$1/4$	Δp_{t+1}
	$+s$	$1/4$	0	

Since p_{t-1} is equally likely to be either in the bid or the ask and the expected value of Δp_t is zero, we can calculate the covariance:

$$\mathbb{C}\left(\Delta p_t, \Delta p_{t+1}\right) = \mathbb{P}\left\{p_{t-1} = B\right\}\mathbb{E}\left[\Delta p_{t+1}\Delta p_t | p_{t-1} = B\right] + \mathbb{P}\left\{p_{t-1} = A\right\}\mathbb{E}\left[\Delta p_{t+1}\Delta p_t | p_{t-1} = A\right]$$

$$= \frac{1}{2} \times \frac{1}{4}(-s) \times (+s) + \frac{1}{2} \times \frac{1}{4}(-s) \times (+s)$$

$$= -\frac{1}{4}s^2$$

and construct an estimator of the spread s as:

$$\widehat{s} = \sqrt{-\frac{1}{4}\widehat{\mathbb{C}}\left(\Delta p_t, \Delta p_{t+1}\right)}$$

For this estimation to be consistent, we need to impose the assumptions from Theorem 7.2.2 in Brockwell and Davis (2009) that the returns are stationary, i.e.,

$$\Delta p_t = \sum_{j \in \mathbb{Z}} \psi_j Z_{t-j} \qquad Z_t \sim I.I.D\left(0, \sigma^2\right)$$

with $\sum_{j \in \mathbb{Z}} |\psi_j| < \infty$ and $\sum_{j \in \mathbb{Z}} \psi_j^2 |j| < \infty$. $\qquad\qquad\qquad\square$

Problem 2.

We now take an initial look at predictive return regressions, where macro-finance variables are used to forecast the future equity premium. The task is to gauge whether there is (statistically significant) evidence of predictability in the S&P equity-index returns minus the risk-free rate (use 3-month T-bill rates – available from FRED at the St. Louis Fed).

In formal notation, $\mathbf{r^e(t+1) = r(t+1) - i(t)}$, where $r(t+1)$ is the monthly continuously compounded nominal return on the S&P index from the end of month t to end of month $t+1$, and $i(t)$ is the 3-month T-bill rate at the end of month t.

As predictors, we will choose the PE-ratio (price/earnings ratio), the DP-ratio (dividend/price ratio), and the relative interest rate (current 3-month T-bill rate minus the average 3-month T-bill rate over the prior last 12 month). This is labeled **RREL**.

The data are available from Bob Shiller's website and are posted on the Canvas site – Shiller has a pre-calculated **PE** ratio and provides both çdividend and **price level** data for the S&P index. The variable construction can be seen from the Excel formulas generating the scaled data from the raw observations. As indicated, **T-bill** data are available from many sources including FRED.

The relevant predictive OLS regression take the form:

$$r^e(t+1) = a + b'X(t) + u(t), \qquad t = 0, 1, \ldots, T$$

where t refers to the (end of) month for the observation.

a) Run the regression above with only the PE-ratio as the explanatory (X) variable. Use the sample period from January 1963 - December 2022. Obtain OLS standard errors, assuming no heteroskedasticity or autocorrelation in the innovations, and assess the significance of the relevant regression coefficient and compute the R^2 statistic.

b) Repeat the exercise from a), but with only the DP-ratio as the explanatory variable.

c) Repeat the exercise from a), but with only the RREL as the explanatory variable.

d) Repeat the exercise from a), but use both the PE-ratio and the RREL as regressors.

e) Now, assess the significance of the regression coefficient(s) in the above regressions using heteroscedasticity robust standard errors (White standard errors).

f) Repeat 2.e) with heteroscedasticity and autocorrelation consistent (Newey-West) st. errors.

g) Does the change of method for computing standard errors make any difference to the size of the estimated standard errors? Does it change your conclusions?

Solution.

The results are presented in Tables 1 and 2. The first table presents the results for the regressions for the Value Weighted Index, while the second table presents the results for

the Equal Weighted Index. The numbers in parenthesis correspond to the t-statistics using homoskedastic standard errors. The numbers in brackets, using White heteroskedastic std. errors. Finally, the numbers in curly brackets show the t-statistics using the Newey-West standard errors. As we can see, there is not much predictability in the returns when it comes to these factors. Both the Price-to-Earnings and the Dividend-to-Price ratios are not statistically significant in columns (a), (b) and (d) in both tables. The Relative Interest Rate (RREL) is significant at 5% under all std. errors but the R^2 it is able to generate is really low (usually around 0.9%).

Comparing the t-statistics from the tables we also conclude that the method to estimate the standard errors increases as we add more robustness (Homoskedastic $\rightarrow$ White $\rightarrow$ HAC). However, for every variable in our model, this change in the standard errors is not enought to change the significance of the coefficients.

Table 1: Testing for Returns Predictability (Value Weighted)

| | Dependent variable: vwretd | | | |
	(a)	(b)	(c)	(d)
Constant	0.003	0.003	0.005***	0.005
	(1.063)	(0.631)	(3.317)	(1.424)
	[0.959]	[0.591]	[3.323]	[1.268]
	{0.892}	{0.583}	{3.206}	{1.110}
P/E	0.000			0.000
	(0.752)			(0.307)
	[0.624]			[0.250]
	{0.570}			{0.211}
D/P		0.093		
		(0.666)		
		[0.604]		
		{0.622}		
RREL			-4.619**	-4.512**
			(-2.470)	(-2.370)
			[-2.375]	[-2.316]
			{-2.311}	{-2.193}
Observations	719	719	719	719
R^2	0.001	0.001	0.008	0.009
Adjusted R^2	-0.001	-0.001	0.007	0.006
Residual Std. Error	0.043 (df=717)	0.043 (df=717)	0.043 (df=717)	0.043 (df=716)
F Statistic	0.565 (df=1; 717)	0.443 (df=1; 717)	6.100** (df=1; 717)	3.093** (df=2; 716)

Note: *p<0.1; **p<0.05; ***p<0.01

Stars w.r.t homoskedastic std. errors

( ): Homoskedastic t-stat; [ ]: White t-stat; { }: HAC t-stat

Table 2: Testing for Returns Predictability (Equal Weighted)

	Dependent variable: ewretd			
	(a)	(b)	(c)	(d)
Constant	0.004	0.001	0.008***	0.005
	(1.036)	(0.286)	(4.201)	(1.395)
	[0.814]	[0.277]	[4.208]	[1.083]
	{0.805}	{0.302}	{4.469}	{0.994}
P/E	0.000			0.000
	(1.315)			(0.863)
	[0.916]			[0.593]
	{0.925}			{0.546}
D/P		0.224		
		(1.397)		
		[1.290]		
		{1.420}		
RREL			-5.518**	-5.171**
			(-2.566)	(-2.363)
			[-2.405]	[-2.239]
			{-2.522}	{-2.147}
Observations	719	719	719	719
R^2	0.002	0.003	0.009	0.010
Adjusted R^2	0.001	0.001	0.008	0.007
Residual Std. Error	0.050 (df=717)	0.050 (df=717)	0.050 (df=717)	0.050 (df=716)
F Statistic	1.730 (df=1; 717)	1.951 (df=1; 717)	6.583** (df=1; 717)	3.663** (df=2; 716)

Note: *p<0.1; **p<0.05; ***p<0.01
Stars w.r.t homoskedastic std. errors
( ): Homoskedastic t-stat; [ ]: White t-stat; { }: HAC t-stat

References

P. J. Brockwell and R. A. Davis. *Time series: theory and methods.* Springer science & business media, 2009.

R. Roll. A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of finance*, 39(4):1127–1139, 1984.