Problem Set 3: Suggested Solutions

Cleaning the Data. A vital part of empirical analysis is preparing and cleaning the data. Often, the way the data is cleaned can have more impact than the choice of statistical test. Good papers detail the steps used to clean the data; you must describe in detail how you clean the data. However, this is not a class in data cleaning, so you should not spend hours cleaning the data, just use a reasonable approach. But if your reasonable approach is not ideal, you should also mention how you would clean the data if you were going to spend more time doing so.

The first issue is how to connect the BE/ME data—which is annual—to the return data—which is monthly. When testing market efficiency, we want to make sure that the variables we are using to decide whether to trade are available at the time of trading. In the case of book equity, for firms with December fiscal year-ends, the book equity (and all financial reports) are usually not reported for about 2-3 months historically—and in unusual cases it could be even longer. Accordingly, Fama & French assign the BE/ME data from calendar year t to the returns starting in July of year $t + 1$ to June of year $t + 2$.

A second issue is how to handle the size data. Market size mechanically reflects the previous month's returns. To be sure they are testing the effect of size, rather than the effect of lagged returns, Fama & French assign the market size from December of year t to the returns starting in July of year $t + 1$ to June of year $t + 2$. We will follow these same conventions.

The third and trickiest issue is how to handle the missing data, which in the data file is coded as "-99.99." Unsurprisingly, failure to account for this missing data will mess up our results, in particular since -99.99 is an extreme number to mistakenly use instead. It might be tempting to replace the -99.99s with zero; this is an error. Then we will act as if that portfolio earned a return of 0 when in fact we don't know the return. Admittedly, assuming a return of 0 will mess up our results less than assuming a return of -99.99%, but nonetheless, it will still mess up our results. To correct for the missing data, we are going to use all of the time-periods for which data exist, but omit the missing months from any individual regressions. This is ideal from the perspective of maximizing the use of our available information. But it means that some Betas are estimated with fewer months than the full time period, and some cross-sectional regressions are run with fewer than all 49 industries or 25 size/value portfolios. We will thus omit observations from the cross-sectional regressions for which either the monthly asset return does not exist in regression (b), or for which either the monthly asset return or corresponding size and BE/ME data does not exist in regression (f).

Part I: 49 Industry Portfolios

Note: In the following solution, the γ coefficients in the math equations are called "lambda" in the table printouts. Lambda as the symbol for risk premium is more standard in the literature.

a) **Cross-Sectional Regression.** Consider the following cross-sectional regression,

$$R_i = \gamma_0 + \gamma_M \beta_{iM} + \eta_i$$

where γ_0 and γ_M are regression parameters and $\beta_{iM} = cov(R_i, R_M)/\sigma^2(R_M)$. Taking expectations, we find

$$E[R_i] = \gamma_0 + \gamma_M \beta_{iM}.$$

The Sharpe-Lintner CAPM says that $E[R_i] = R_f + \beta_{iM}\left(E[R_M] - R_f\right)$ so $\gamma_0 = R_f$ and $\gamma_M = E[R_M] - R_f$. The Black CAPM says $E[R_i] = E[R_{0M}] + \beta_{iM}(E[R_M] - E[R_{0M}])$ so $\gamma_0 = E[R_{0M}]$ and $\gamma_M = E[R_M] - E[R_{0M}]$.

An additional testable hypothesis of both versions of the CAPM is that the market portfolio is mean-variance efficient (the relationships above depend only on the market portfolio being minimum variance). This implies that $E[R_M] - R_f > 0$ (Sharpe-Lintner) or that $E[R_M] - E[R_{0M}] > 0$ (Black), so we have that $\gamma_M > 0$.

b) **Fama-MacBeth with the Market Portfolio.** In Table 1, we report betas for the 49 portfolios estimated using all the data available for each industry. We estimate betas as a regression of R_i on R_M; alternatively, since the risk-free rate is changing over time, a regression of $R_i - R_M$ on $R_M - R_f$ will give (very slightly) different results.

Table 1: 49 Industries Estimated Betas with Market

	Agric	Food	Soda	Beer	Smoke	Toys	Fun
Beta	0.92	0.73	0.83	0.94	0.63	1.21	1.42
	Books	Hshld	Clths	Hlth	MedEq	Drugs	Chems
Beta	1.12	0.90	0.81	1.13	0.84	0.84	1.04
	Rubbr	Txtls	BldMt	Cnstr	Steel	FabPr	Mach
Beta	1.21	1.14	1.16	1.35	1.36	1.11	1.24
	ElcEq	Autos	Aero	Ships	Guns	Gold	Mines
Beta	1.29	1.25	1.30	1.17	0.84	0.60	0.97
	Coal	Oil	Util	Telcm	PerSv	BusSv	Hardw
Beta	1.30	0.87	0.78	0.66	1.09	0.89	1.11
	Softw	Chips	LabEq	Paper	Boxes	Trans	Whlsl
Beta	1.64	1.34	0.99	1.67	0.95	1.14	1.09
	Rtail	Meals	Banks	Insur	RlEst	Fin	Other
Beta	0.97	0.94	1.04	1.12	1.28	1.31	1.05

In Table 2, we report the results of the Fama-MacBeth regressions. The proxy used for the market portfolio is mean-variance efficient if and only if $\gamma_M > 0$, so we can test for this as our null. Since $t(\gamma_M) = 0.877$, we cannot reject the hypothesis that $\gamma_M > 0$. Thus, we cannot reject the mean-variance efficiency of the market portfolio at a reasonable significance level (say, 5%). In order to reject the efficiency of the market portfolio we should be able to rule out $\gamma_M > 0$, which is possible at the 5% confidence level only if we had found $t(\gamma_M) < -1.65$ (this is a one-tailed test). While we fail to reject that $\gamma_M > 0$, it is remarkable that there is so little relationship between the expected return of a portfolio and its β. In fact, these results show that the Security Market Line is close to flat.

Table 2: 49 Industries Lambda estimated using Fama-Macbeth for Beta with Market

	Intercept	Lambda(Beta)
Mean	0.806	0.222
Stderr	0.189	0.253
t-stat	4.275	0.877

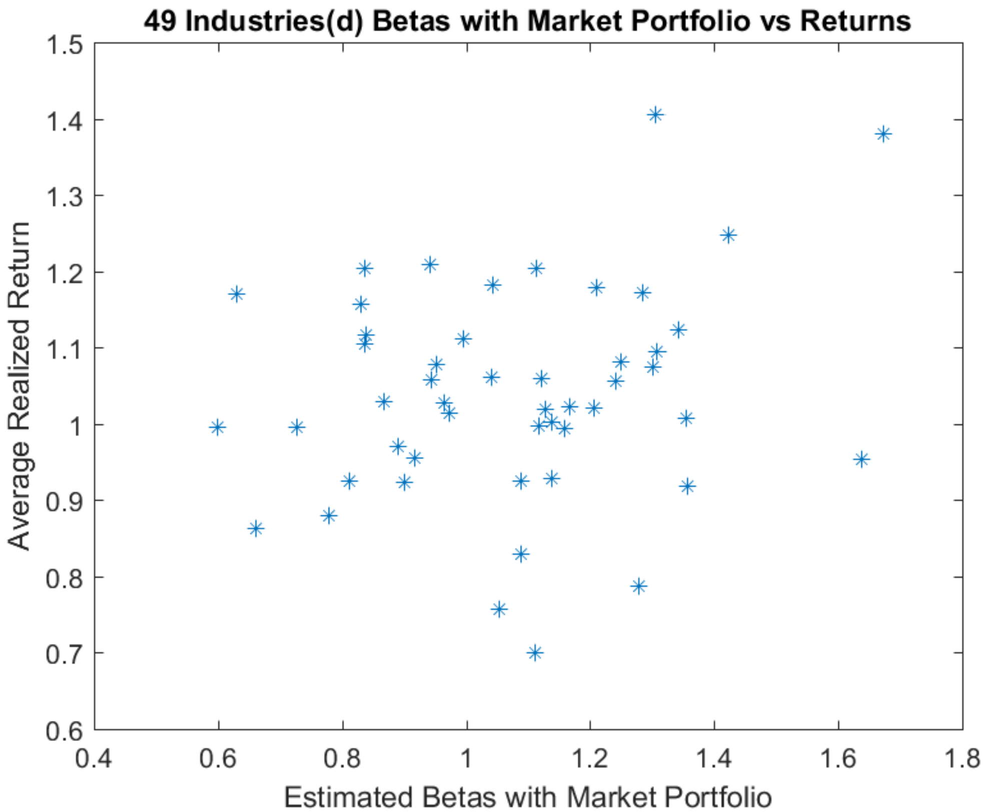
c) **Regression of Averages.** The γ_0 and γ_M computed by means of a cross-sectional regression of average returns are equal to the estimates of γ_0 and γ_M using the Fama-MacBeth procedure, but *only if all months have the same number of observations and the values of 'X' (i.e., estimated Betas) are constant*. Given that we have missing observations for some months, this does not hold, so our estimates will be slightly different. The standard errors are always (potentially) different. The standard errors estimated using OLS are consistent and unbiased only if the error terms are uncorrelated and homoscedastic. However, this assumption typically does not hold for returns data. In particular, the residuals in the cross-section considered above are correlated across assets within each month and heteroskedastic. It can be shown that the Fama-MacBeth procedure captures the cross-sectional correlation of errors through the variation in period-to-period slope estimates (γ_M here), and thus achieves a more consistent estimate of the standard errors and is therefore a superior choice relative to OLS. We could achieve the same end if we knew what the residual covariance matrix was in our regression of average returns on betas (GLS), but this matrix is difficult to estimate. Table 3 provides the cross-sectional regression results using averages.

Table 3: 49 Industries Lambda estimated using average returns & Beta with Market

	Intercept	Lambda(Beta)
Mean	0.908	0.123
Stderr	0.093	0.085
t-stat	9.767	1.454

d) **Plot.** Figure 1 provides a plot of the data points. The relationship does not look strongly positive. If the CAPM were true (or, alternatively phrased, if the market portfolio were the tangency portfolio) then all points should be on an upwards-sloping straight line.

Figure 1:



e) **Theory, γ_{size} and $\gamma_{B/M}$.** Since there is no size or book-to-market term in the CAPM and we are already controlling for beta, γ_{size} and $\gamma_{B/M}$ should both equal zero.

f) **Fama-MacBeth with the Market, Size, & Value.** In the following table we report the results for the Fama-MacBeth regressions using the market beta, size and book-to-market. Again, we cannot reject that the Beta premium is positive. Further, the point estimates for the premium on size and the book-to-market ratio are negative and positive respectively but neither are statistically different from zero. This implies that we cannot reject that the proxy for the market portfolio is on the minimum variance frontier, i.e. we cannot reject the CAPM. However, we might get more power by sorting on size and value directly.

Table 4: 49 Industries(f) Lambda estimated using Fama-Macbeth for Beta with Market, ln(Size) & ln(BE/ME)

	Intercept	Lambda(Beta)	Lambda(ln_Size)	Lambda(ln_BE_ME)
Mean	1.082	0.098	-0.048	0.008
Stderr	0.292	0.224	0.039	0.072
t-stat	3.699	0.436	-1.241	0.105

Part II: 25 Size & Value Portfolios

1)

b) **Fama-MacBeth.** In Table 5, we report the betas for the 25 portfolios.

Table 5: 25 Size-Value Estimated Betas with MARKET

	Small-Low	Small-2	Small-3	Small-4	Small-Hi
Beta	1.63	1.41	1.37	1.27	1.38

	2-Low	2-2	2-3	2-4	2-Hi
Beta	1.26	1.23	1.20	1.21	1.38

	3-Low	3-2	3-3	3-4	3-Hi
Beta	1.24	1.13	1.12	1.16	1.38

	4-Low	4-2	4-3	4-4	4-Hi
Beta	1.09	1.08	1.12	1.15	1.42

	Big-Low	Big-2	Big-3	Big-4	Big-Hi
Beta	0.96	0.95	0.97	1.11	1.31

In Table 6, we report the results of the Fama-MacBeth regressions.

Table 6: 25 Size-Value Lambda estimated using Fama-Macbeth for Beta with MARKET

	Intercept	Lambda(Beta)
Mean	0.611	0.456
Stderr	0.329	0.357
t-stat	1.859	1.277

The proxy used for the market portfolio is mean-variance efficient if and only if $\gamma_M > 0$, so we can test for this as our null. Since γ_M is positive, we definitely cannot reject the hypothesis that $\gamma_M > 0$. So we cannot reject the mean-variance efficiency of the market portfolio using this test.

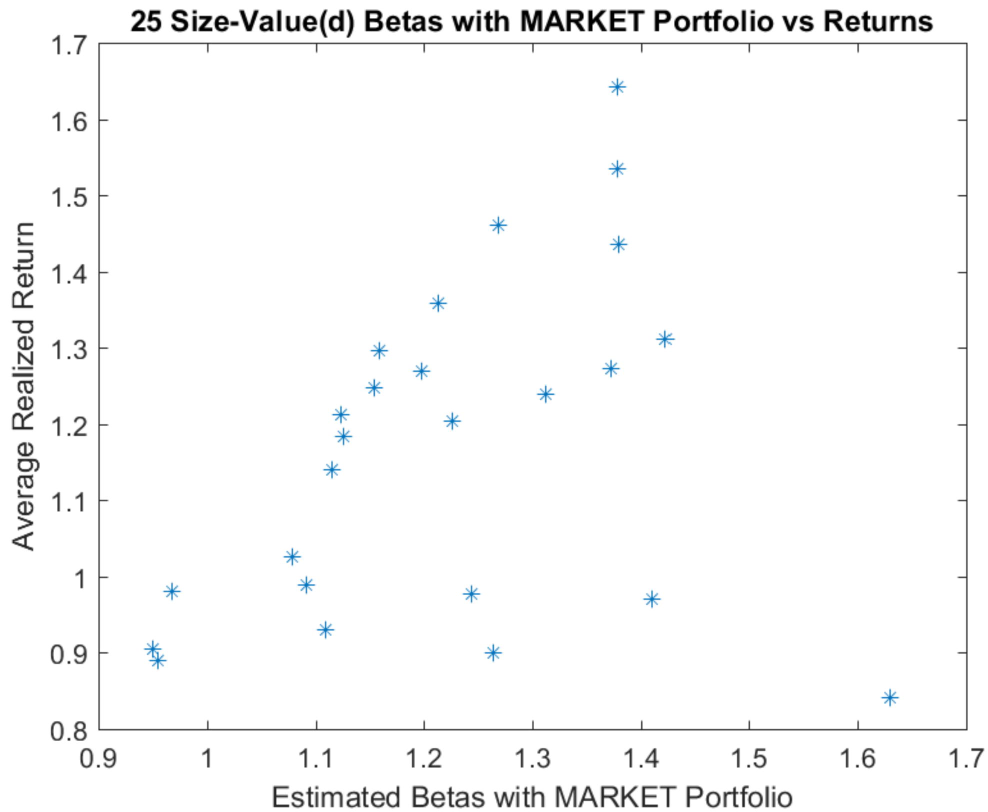
c) **Regression of Averages.** The γ_0 and γ_M computed by means of a cross-sectional regression of average returns are equal to the estimates of γ_0 and γ_M using the Fama-MacBeth procedure, but *only if all months have the same number of observations and the values of 'X' (i.e., estimated Betas) are constant*. Given that we have missing observations for some months, this does not hold, so our estimates will be a slightly different. However, since only very few observations are missing in this case, the estimated coefficients are essentially the same. However, the standard errors will still be different (per the previous discussion). Here, the standard error for the γ_M coefficient is most clearly different from that of the FM procedure. Table 7 provides the cross-sectional regression results using averages.

Table 7: 25 Size-Value Lambda estimated using average returns & Beta with MARKET

	Intercept	Lambda(Beta)
Mean	0.611	0.456
Stderr	0.325	0.264
t-stat	1.882	1.730

d) **Plot.** Figure 2 provides a plot of the data points. The relationship for most observations looks slightly positive, but there is a big outlier (the highest beta asset also has the lowest return). If the CAPM were true (or, alternatively phrased, if the market portfolio were the tangency portfolio) then all points should be on a positively sloped straight line.

Figure 2:



f) **Fama-MacBeth with the Market, Size, & Value.** In Table 8, we report the results for the Fama-MacBeth regressions using the market, size and book-to-market. The estimated premia on the size and book-to-market ratios are both statistically significant. This implies that the proxy for the market portfolio is not on the minimum variance frontier, and therefore is not a mean-variance efficient portfolio either. Furthermore, we can reject that the coefficient on Beta is positive. This test thus rejects the CAPM.

```
Table 8: 25 Size-Value Lambda estimated using Fama-Macbeth for Beta
with MARKET, ln(Size) & ln(BE/ME)
            Intercept   Lambda(Beta) Lambda(ln_Size) Lambda(ln_BE_ME)
Mean            2.208         -0.547          -0.081            0.262
Stderr          0.435          0.367           0.032            0.062
t-stat          5.073         -1.490          -2.544            4.240
```

2)

c) **Regression of Averages, Full Tangency.** Table 9 shows the results when we replace the market portfolio with the full-period tangency portfolio (estimated using the mvp.m Matlab function created for problem set 2). Now the coefficient on the Beta with the Tangency Portfolio is unequivocally significant! (See the Matlab results for tangency portfolio weights). Table 9 shows the Betas with the Full-Period Tangency Portfolio. Table 10 provides the Fama-MacBeth results.

Table 9: 25 Size-Value Estimated Betas with FULL-PERIOD TANGENCY

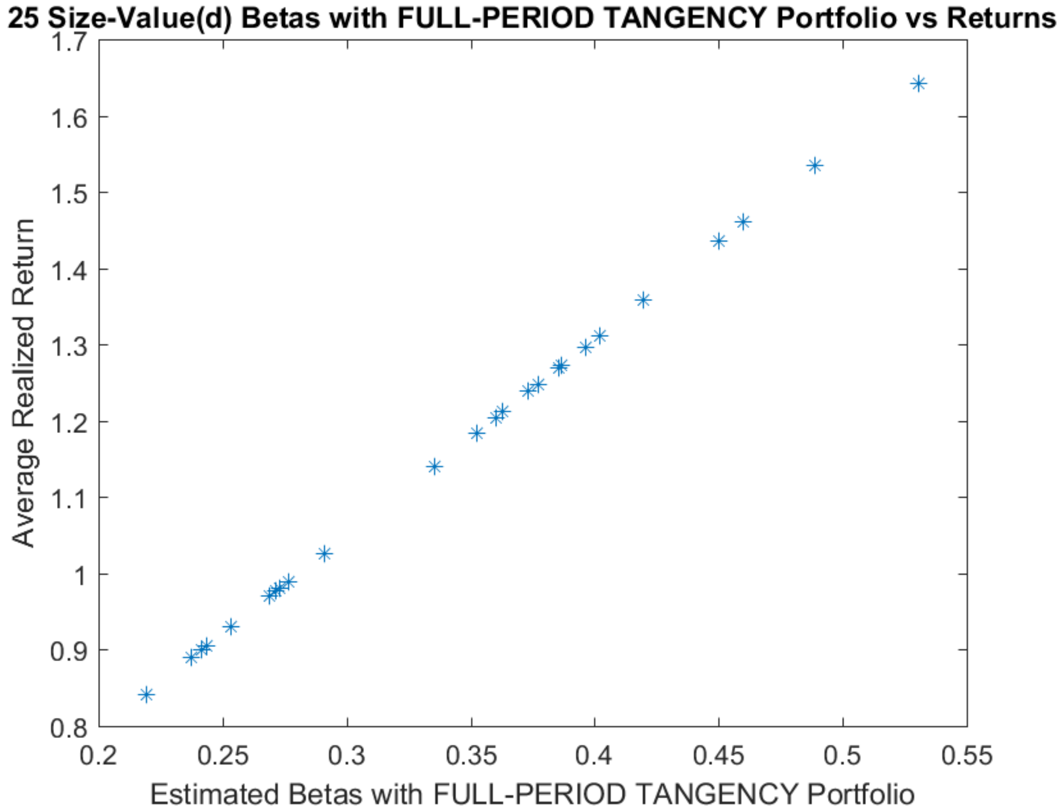
	Small-Low	Small-2	Small-3	Small-4	Small-Hi
Beta	0.22	0.27	0.39	0.46	0.53
	2-Low	2-2	2-3	2-4	2-Hi
Beta	0.24	0.36	0.39	0.42	0.49
	3-Low	3-2	3-3	3-4	3-Hi
Beta	0.27	0.35	0.36	0.40	0.45
	4-Low	4-2	4-3	4-4	4-Hi
Beta	0.28	0.29	0.34	0.38	0.40
	Big-Low	Big-2	Big-3	Big-4	Big-Hi
Beta	0.24	0.24	0.27	0.25	0.37

Table 10: 25 Size-Value Lambda estimated using Fama-Macbeth for Beta with FULL-PERIOD TANGENCY

	Intercept	Lambda(Beta)
Mean	0.278	2.572
Stderr	0.196	0.446
t-stat	1.418	5.770

d) **Plot.** Figure 3 provides a visual plot of the data points. All points (except for the asset with some missing data) are on a straight line! This plot beautifully illustrates Roll's critique: mathematically the relationship between the Beta with an in-sample tangency portfolio and returns is linear. This is why Roll says that what we really are testing is whether the market portfolio (for stocks) is the tangency portfolio of the test assets. But even if we show that the stock market portfolio is the tangency portfolio, we still don't know if the true market portfolio (i.e., including non-traded assets) is mean-variance efficient because we cannot measure the true market portfolio.

Figure 3:



3)

c) **Regression of Averages, Even-Odd Tangency.** Table 11 and 12 show the results when we replace the market portfolio with the portfolio of odd month returns in even years and even month returns in odd years using weights from the tangency portfolio constructed from even months in even years and odd months in odd years (and vice versa for even month returns in even years and odd month returns in odd years). (Estimated using the mvp.m Matlab function created for problem set 2.) Now the coefficient on the Beta with this alternative Tangency Portfolio is still very significant, although not quite as much as the full sample tangency portfolio. Table 11 shows the Betas with the Even-Odd Sample Tangency Portfolio. Table 12 provides the Fama-MacBeth results.

Table 11: 25 Size-Value Estimated Betas with MIXED RETURNS TANGENCY

	Small-Low	Small-2	Small-3	Small-4	Small-Hi
Beta	0.19	0.21	0.31	0.36	0.44
	2-Low	2-2	2-3	2-4	2-Hi
Beta	0.18	0.28	0.29	0.31	0.38
	3-Low	3-2	3-3	3-4	3-Hi
Beta	0.21	0.25	0.27	0.30	0.34
	4-Low	4-2	4-3	4-4	4-Hi
Beta	0.20	0.21	0.26	0.28	0.31
	Big-Low	Big-2	Big-3	Big-4	Big-Hi

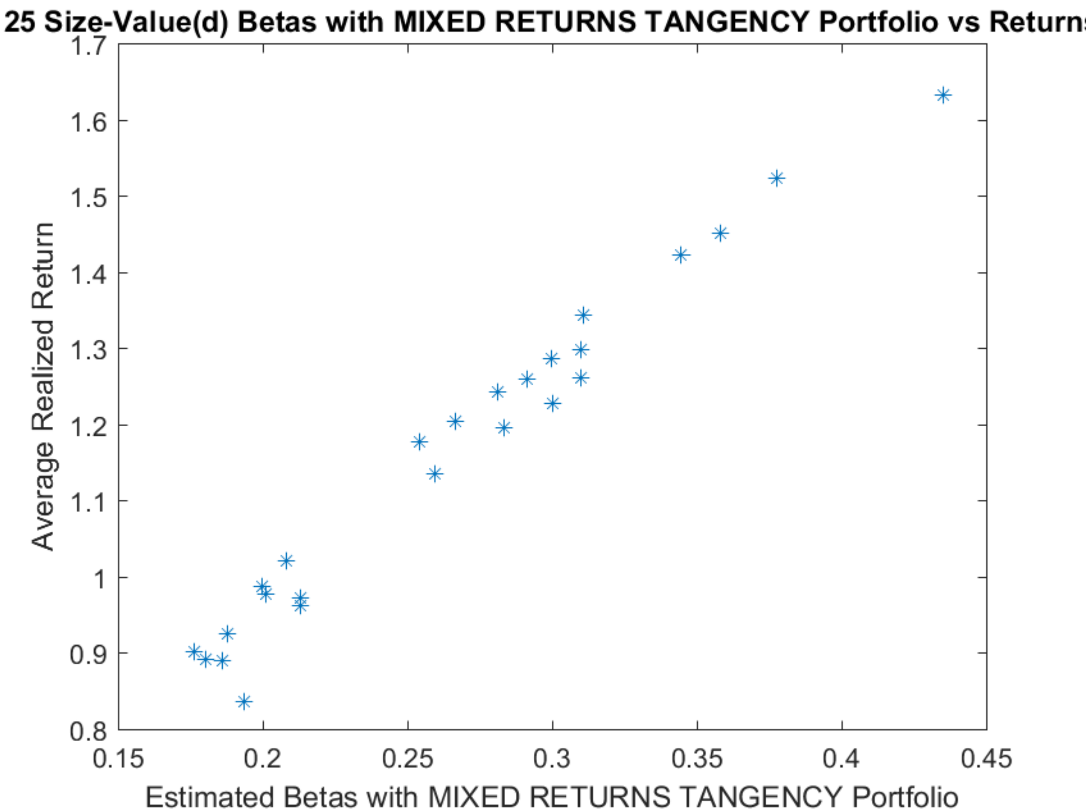
Beta	0.19	0.18	0.20	0.19	0.30

Table 12: 25 Size-Value Lambda estimated using Fama-Macbeth for Beta with MIXED RETURNS TANGENCY

	Intercept	Lambda(Beta)
Mean	0.339	3.094
Stderr	0.185	0.586
t-stat	1.830	5.281

d) **Plot.** Figure 4 provides a visual plot of the data points. If the CAPM were true (or, alternatively phrased, if the market portfolio were the tangency portfolio) then all points should be on a straight line.

Figure 4:



4) The main difference between (2) and (3) is that the tangency portfolio does not stay constant in part (3), so it is no longer the in-sample tangency portfolio in the second half of the sample, and therefore this portfolio does not have a perfectly linear relationship with the assets as the tangency portfolio we used in part (2) (although it is impressively close).