1. Introduction

The goal of this course is to develop a quantitative approach to portfolio management. Such an approach is especially useful when managing a systematic trading strategy, such as statistical arbitrage or automatic market making. However, many of the mathematical tools we will develop are useful for analyzing any portfolio, including portfolios built by fundamental managers.

This course is dedicated to the study of *portfolio theory*, interpreted quite broadly as the study of any collection of more than one asset. An important subfield of portfolio theory is the study of optimization: how to form portfolios that are optimal with respect to the investor's utility function, but we will also develop tools to study portfolios which weren't necessarily constructed by optimizing utility.

Changes in portfolio holdings are called *trades*. Although it's not always emphasized in studies of portfolio theory, (eg. all of the foundational work on "Modern Portfolio Theory (MPT)" ignores trading cost completely!) trading itself is a complex process with many decision variables that need to be considered and potentially optimized. As we'll discuss later, it isn't strictly necessary to separate studies of optimal trading and optimal portfolios if we are willing to increase the dimension of the space of portfolios by one, to include time as an additional dimension.

Let all of the assets we are going to consider for our portfolio be indexed by $i = 1, \dots, n$. The asset prices at time t will be denoted p_t with p_t^i the price for the i-th asset. The *return* of the i-th asset over the interval $[t, t+1]$ will be denoted

$$r_{t+1}^i = p_{t+1}^i / p_t^i - 1. \tag{1.1}$$

If there are stock splits or dividends between t and $t+1$, the effect of those should be included in the return variable, for example by adjusting the price before the split to be in the same units as price after the split before calculating (1.1). Once this adjustment has been made, the quantity (1.1) is called *total return*.

We will also sometimes work with *log-relative returns* or "logrels" for short, which are

$$\log(1 + r_{t+1}^i) = \log p_{t+1}^i - \log p_t^i.$$

The notation r_{t+1} without a superscript denotes the entire $n \times 1$ column vector of asset returns. When $t+1$ is in the future, r_{t+1} is a random variable. Any

structure inherent in this random process potentially gives rise to analogous structure in the portfolio's return. For example, let h_t^i denote our holdings in the i-th security at time t. Then $h_t \cdot r_{t+1}$ is the portfolio's return over the interval $[t, t+1]$. Information (eg. from a statistical model's forecast) about the variance-covariance matrix $\text{var}(r_{t+1})$ translates into information about $\sigma^2(h)$, the portfolio variance.

Whether or not it is explicitly framed in these terms, the goal of any rational investor is to maximize the utility of final wealth. We can conceptualize utility in terms of *decisions* and *outcomes*. Conditional on a decision d, the probability of outcome w is $p(w \mid d)$. A decision d is chosen by maximizing $\mathbb{E}[U(w) \mid d]$ where $U(w)$ quantifies the agent's utility associated to outcome w.

In trading problems, decisions are typically modeled as the decision to hold a specific portfolio sequence

$$\mathbf{x} = (x_1, x_2, \ldots, x_T),$$

where x_t is the portfolio the agent plans to hold at time t in the future. Often the relevant outcome is the trading profit. If r_{t+1} is the vector of asset returns over $[t, t+1]$, then the profit associated to decision $d = \mathbf{x} = (x_1, x_2, \ldots, x_T)$ is

$$\pi(\mathbf{x}) = \sum_t [x_t \cdot r_{t+1} - \mathscr{C}_t(x_{t-1}, x_t)] \tag{1.2}$$

where $\mathscr{C}_t(x_{t-1}, x_t)$ is the total cost (including but not limited to market impact, spread pay, borrow costs, ticket charges, financing, etc.) associated with holding portfolio x_{t-1} at time $t-1$ and ending up with x_t at time t.

Trading profit $\pi(\mathbf{x})$ is a random variable, since many of its components are future quantities unknowable at time $t = 0$. What is its distribution? Here we pause for a digression on central limit theorems. Theorem 1.1 is the classical version, while Theorem 1.2 gives an extension which applies to variables that aren't identically distributed.

Theorem 1.1. *Suppose $\{X_1, X_2, \ldots\}$ is a sequence of i.i.d. random variables with $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}[X_i] = \sigma^2 < \infty$. Then as n approaches infinity,*

$$\sqrt{n}\left(\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \mu\right) \xrightarrow{d} N(0,\, \sigma^2).$$

where $\xrightarrow{d}$ denotes convergence in distribution.

Sometimes the i.i.d. restriction may be too restrictive. It turns out that it isn't really needed as long as certain analytic regularity conditions are satisfied.

Theorem 1.2. *Suppose* $\{X_1, X_2, \ldots\}$ *is a sequence of independent random variables, each with finite expected value* μ_i *and variance* σ_i^2*. Define* $s_n^2 = \sum_{i=1}^n \sigma_i^2$*. Suppose that for every* $\epsilon > 0$*,*

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{i=1}^{n} \mathbb{E}\big[(X_i - \mu_i)^2 \cdot \mathbb{1}_{\{|X_i-\mu_i|>\epsilon s_n\}}\big] = 0\,. \tag{1.3}$$

where $\mathbb{1}_S$ *denotes the indicator function of a set S, i.e.* $\mathbb{1}_S(x) = 1$*, if* $x \in S$ *and 0 otherwise. Then the random variables*

$$Z_n := \frac{\sum_{k=1}^n (X_k - \mu_k)}{s_n}$$

converge in distribution to a standard normal random variable as $n \to \infty$*.*

The condition (1.3) was introduced by Lindeberg (1922) and is known as "Lindeberg's condition." We will not use this condition, but you should know that the central limit theorem can apply to independent random variables that are not identically distributed.

Note that the variable $\pi(\mathbf{x})$ can be viewed as a sum of contributions from n assets, where n could conceivably be quite large: there are as many as 10,000 liquid equities traded in the world as of the time of writing. The components r_t^i and r_t^j for $i \neq j$ are not independent, but consider what might happen if we could replace them with residuals from some well-specified model with independent residuals. Then the central limit theorem (CLT) would apply to $\pi(\mathbf{x})$.

Final wealth w must equal initial wealth plus the trading profit,

$$w = w_0 + \pi(\mathbf{x}).$$

Consider the standard CARA utility function

$$U(w) = -e^{-\gamma w} = -e^{-\gamma(w_0+\pi(\mathbf{x}))} = -e^{-\gamma w_0} e^{-\gamma\pi(\mathbf{x})}$$

where $\gamma > 0$ is the Arrow-Pratt index of absolute risk aversion. If $\pi(\mathbf{x})$ is normally distributed, then $U(w)$ is negative-lognormal. It follows that maximization of $\mathbb{E}[U(w)]$ is accomplished by maximizing $u(\mathbf{x})$ defined by

$$u(\mathbf{x}) := \mathbb{E}[\pi(\mathbf{x})] - (\gamma/2)\mathbb{V}[\pi(\mathbf{x})] \tag{1.4}$$

Here, $\mathbb{E}$ and $\mathbb{V}$ denote mean and variance as forecast at time $t = 0$. We now substitute (1.2) into (1.4), to find

$$u(\mathbf{x}) = \sum_t \Big[x_t^\top \mathbb{E}[r_{t+1}] - \frac{\gamma}{2} x_t^\top \mathbb{V}[r_{t+1}] x_t - \mathscr{C}_t(x_{t-1}, x_t) \Big] \tag{1.5}$$

Much of portfolio theory amounts to understanding these three terms in greater detail. One can go a long way towards understanding the first two terms by means of linear factor models, which we now discuss.

A typical example of a structural model that can be imposed upon returns is to assume they are linear functions of a set of unobservable stochastic processes f_t, called *factor returns*, i.e.

$$\underset{[n\times 1]}{r_{t+1}} - \underset{[n\times 1]}{r^{\mathrm{rf}}_{t+1}} = \underset{[n\times p]}{X_t}\,\underset{[p\times 1]}{f_{t+1}} + \underset{[n\times 1]}{\epsilon_{t+1}}, \tag{1.6}$$

where for concreteness, we have indicated the matrix dimensions of each element below it. In useful models of this type, usually $p \ll n$ so that the model provides a very meaningful reduction in the number of parameters.

The vector r^{rf}_{t+1} is the relevant vector of *risk-free rates*; if the returns are local currency returns, one must be careful to have the correct risk-free rate for each currency! It is reasonable to suppose that whatever factors drive equity returns do not also drive risk-free rates, since such rates are set by central banks. The same assumption is made in the CAPM; market beta drives the *risk-adjusted return*. The f_{t+1} cannot be observed directly, so information about them must be inferred from a statistical model, of which (1.6) is only a partial specification.

The vector ϵ_{t+1} is called the *residual vector*. It is usually appropriate to model the components of ϵ_t corresponding to different assets as independent. If there were a dependence that we can understand the source of, then it's reasonable to think we could introduce additional factors into the model so that the residuals from the new model would be independent, even if the original ones weren't. The residuals of different assets may differ widely in their distributions. For example, a small cap tech stock might have a large residual variance, while a large company with very broadly-diversified business lines might find that its fortunes rise and fall with the market itself, and thus it has a low residual variance. For the latter company, most of its variance comes from its market beta and the variance of the market. This already suggests that variance could potentially be decomposed into contributions from various sources.

The matrix X_t is called, alternatively, the *design matrix* or the matrix of *factor loadings* or the matrix of *independents*. By contrast, r_{t+1} (or anything appearing on the left side of an equation such as (1.6)) are referred to as *dependents*. Importantly, if the returns pertain to the interval $[t, t+k]$ for some $k > 0$, then the independents X_t must be composed of quantities that are available to us as of time t. Any violation of this requirement is called *lookahead bias* and typically renders the model useless.

Before continuing, let us consider some simple examples of the kinds of factors that might end up as columns in the design matrix. The simplest is a column of 1's, also known as the *intercept*. More generally, suppose we have a *classification* or a *labeling* which maps each stock to exactly one of a set of categories or *classes*. We can create a column for each class, and place 1's in the rows which belong to that class, and 0's in the remaining rows. An industry classification in which each stock belongs to exactly one industry is an example.

If X_t consists of a single column containing each stock's market beta, then the model (1.6) is structurally compatible with the CAPM, with the important difference that the return to the "market-beta factor" is now being viewed as an unobservable hidden parameter which must be estimated via statistical inference. We will see later that all factor returns are returns to suitable portfolios whose weights can be calculated from the Moore-Penrose pseudoinverse of X_t, but the return to the market-beta factor will usually not be the same as the return to a capitalization-weighted basket such as the S&P 500.

As a further example, suppose we wanted a factor to capture the *size effect* for US equities, where "size" refers to market capitalization. Let c^i denote the capitalization of the i-th company. The empirical distribution of $\{c^i\}$ will have a fat right tail, but $\{\ln(c^i)\}$ might be closer to normal. One model is that companies in the middle of this distribution have no size effect, while companies two sigmas above the mean have twice the effect of companies one sigma above the mean. One could thus compute a score as

$$s^i = (\ln(c^i) - \mu_c)/\sigma_c \tag{1.7}$$

where μ_c and σ_c are (robust) estimates of the location and scale of the distribution of $\ln(c)$. The most recent values of s^i available at time t then become a column in the design matrix X_t, implying a relationship of the form

$$r_{t+1} = f_s s + \{\text{other effects}\} + \epsilon_{t+1}.$$

Later, we will learn to interpret the fitted coefficients such as f_s as portfolio returns.

More generally, we could think about applying the procedure we just described for market capitalization to almost any continuous data $v \in \mathbb{R}^N$. Similar ideas have been applied to continuous variables such as earnings-to-price,

book-value-to-price, liquidity, financial leverage, foreign-currency beta, price-momentum, dividend-yield, and others. These are sometimes called *style factors* because it could be a portfolio manager's "style" to buy stocks that look cheap based on book value, or stocks with high dividend yields.

Since the model (1.6) is linear, in each case, careful consideration should be given to how the data should be transformed in order that a linear relationship to returns is a reasonable hypothesis. For example, the data may come in the form of a ranking, for which there is no reason to believe it is related to expected return by a simple scaling.

In eq. (1.7), and indeed in many applications, robust estimators of location and scale are desirable. The construction and properties of such estimators forms a whole subfield within statistics which we do not have time to cover exhaustively. However, it is worth noting that for univariate data, a surprisingly efficient (82% in the normal case) estimator of scale can be formed as

$$Q_n(\mathbf{x}) := d\,\{|x_i - x_j| : i < j\}_{(k)}, \quad \text{where} \quad k = \binom{\lfloor n/2 \rfloor + 1}{2} \approx \frac{1}{4}\binom{n}{2} \tag{1.8}$$

so $Q_n(\mathbf{x})$ is approximately the first quartile of the $\binom{n}{2}$ inter-point spacings. The constant d is a scale factor which is set to ensure Fisher consistency in a location-scale family, which need not be normal. In the normal case, $d \approx 2.2219$. Note that $Q_n(\mathbf{x})$ can be computed in $O(N \log N)$ time, as explained in a famous paper of Rousseeuw and Croux (1993).

The OLS estimator for the coefficients in (1.6) is

$$\hat{f}_{t+1} = (X_t' X_t)^{-1} X_t' r_{t+1}. \tag{1.9}$$

For the moment, let's assume that $X_t' X_t$ is invertible.

In terms of the factor model (1.6), the variance-covariance matrix of returns may be expressed

$$\Sigma_t = \operatorname{var}(r_{t+1}) = X_t \operatorname{var}(f_{t+1}) X_t' + \operatorname{var}(\epsilon_{t+1}) \tag{1.10}$$

Assume the residuals are uncorrelated across assets. Then the second term above, $\operatorname{var}(\epsilon_{t+1})$, is a diagonal matrix. Also, $\operatorname{var}(f_{t+1})$ is a $p \times p$ matrix where, recall, $p \ll n$. This essentially means that it's rarely necessary, in practice, to work with an $n \times n$ covariance matrix directly. This dramatically reduces the number of parameters that need to be estimated to something like $p(p+1)/2 + n \ll n^2$.

These quantities will appear again and again, so let's set some standard notation. Let

$$F_t = \text{var}(f_{t+1})$$

denote our forecast, as of time t, of the variance-covariance matrix of the p-vector f_{t+1}. Note that the $k \times k$ matrix F_t is a parameter of the model, and not the sampling variance that we used to compute standard errors and t-stats previously. Estimating this parameter matrix is somewhat subtle, and we will return to more sophisticated methods for doing so. A crude estimator can be obtained by computing sample variances and covariances from the time series of past factor returns $\{\hat{f}_s : s \le t\}$.

Similarly

$$\Delta_t = \text{var}(\epsilon_{t+1})$$

will denote our forecast, as of time t, of the (diagonal) variance matrix of ϵ_{t+1}. Then (1.10) is expressed as

$$\Sigma_t = \text{var}(r_{t+1}) = X_t F_t X_t' + \Delta_t.$$

Recall the form of the "standard linear regression model." The model is

$$y = X\beta + \epsilon \tag{1.11}$$

where y is an $n \times 1$ column vector of responses, X is an $n \times k$ matrix of predictors or explanatory variables, and the usual model has $\epsilon \sim N(0, \Omega)$, a multivariate Gaussian with variance-covariance matrix Ω. The simplest version has $\Omega = \sigma^2 I$.

As a warm-up let's derive the MLE for β. Note that

$$\frac{\partial}{\partial \beta_j}[X\beta]_i = \frac{\partial}{\partial \beta_j} \sum_k X_{ik}\beta_k = X_{ij}$$

Using this and the product rule,

$$\frac{\partial}{\partial \beta_j}(y - X\beta)'(y - X\beta) = 2(y - X\beta) \cdot (-\mathbf{x}_j)$$

where $\mathbf{x}_j$ is the j-th column of X. Setting this equal to zero, we have $\mathbf{x}_j \cdot y = \mathbf{x}_j \cdot X\beta$ for all j, or in linear algebra notation,

$$X'y = X'X\beta \tag{1.12}$$

There's always at least one minimizer of $-L(\beta) = \|y - X\beta\|^2$ because it's convex and bounded below. If $X'X$ is invertible, eqns. (1.12) have a unique solution which is the unique global minimum. If it isn't, eqns. (1.12) may have

lots of solutions which are all global minima, and later we'll see how to find all of them.

Recall the structural model for the cross-section of asset returns:

$$\underset{[n\times 1]}{r_{t+1} - r^{\text{rf}}_{t+1}} = \underset{[n\times p][p\times 1]}{X_t \; f_{t+1}} + \underset{[n\times 1]}{\epsilon_{t+1}}, \tag{1.13}$$

This has the form of the "standard model" (1.11). Suppressing subscripts for a moment, recall that if $X'X$ is invertible, the least-squares estimator is $\hat{f} = (X'X)^{-1}X'r$.

Assuming this estimator is unbiased, and the true model is $r = Xf + \epsilon$, one has

$$\hat{f} = (X'X)^{-1}X'r = (X'X)^{-1}X'(Xf + \epsilon) = f + (X'X)^{-1}X'\epsilon\,.$$

This relation establishes consistency of the estimator $\hat{f}$. It also shows us that the sampling variance of the estimator is

$$\text{var}(\hat{f} - f) = \text{var}((X'X)^{-1}X'\epsilon) = (X'X)^{-1}X'\,\text{var}(\epsilon)X(X'X)^{-1}.$$

To go further, we'd have to specify a model for ϵ. The simplest model is $\text{var}(\epsilon) = \sigma^2 I$ in which case the above reduces to

$$\text{var}(\hat{f} - f) = \sigma^2(X'X)^{-1}$$

and we can estimate $\hat{\sigma}^2 = \frac{1}{n}\sum_i \hat{\epsilon}_i^2$. In this model, the standard error variance of the j-th component can be written

$$\text{var}(\hat{f}_j) = \sigma^2[(X'X)^{-1}]_{jj} = \frac{\sigma^2}{\text{SS}_j}\text{VIF}_j \tag{1.14}$$

where

$$\text{SS}_j = \sum_i (x_{ij} - \overline{x}_j)^2 \quad \text{and} \quad \text{VIF}_j = 1/(1 - R_j^2)$$

where R_j^2 is the unadjusted-R^2 (defined in eq. (1.15)) from a regression of X_j on all of the other columns $\{X_i : i \neq j\}$. Due to the role it plays in (1.14) as a multiplier for the ratio σ^2/SS_j, which would have been the value of $\text{var}(\hat{f}_j)$ if there were no relation among the columns, the quantity VIF_j is called the *variance inflation factor*. Finding a VIF larger than 10 is possibly a cause for concern, but one should always be suspicious of hard cutoffs; would 9.99 be a concern?

We now discuss the important R^2, or goodness-of-fit, statistic for multiple regression.

Relations of the form (1.13) can always be constructed, with or without underlying economic meaning. For any matrix X_t that is of full rank, one may

estimate f_{t+1} by least-squares and then (1.13) is satisfied trivially. If the matrix X_t were constructed in a completely arbitrary way, one would expect the cross-sectional variance of ϵ_{t+1} to be even larger than the variance of r_{t+1}.

To make this more precise, consider a generic linear regression $y = X\beta + \epsilon$ and let $\hat{\beta}$ denote the estimated or "fitted" parameters, and let $\hat{y} = X\hat{\beta}$ denote the vector of calculated values of y using the estimated parameters. The proper measure of goodness of fit to use is

$$R^2 = 1 - \frac{\sum(y-\hat{y})^2}{\sum(y-\overline{y})^2} \tag{1.15}$$

where $\overline{y}$ denotes the sample arithmetic mean of y. Eq. (1.15) is appropriate whether or not there is an intercept.

The intuition for (1.15) is as follows: if a prediction of the dependent variable y is to be made without using any knowledge about the independent variables x_j or their relationship to y, the best prediction would be a sample mean $\overline{y}$ ["best" in the sense that $\sum_{i=1}^{n}(y_i - v)^2$ is minimized for $v = \overline{y}$]. When a prediction (such as $\hat{y}$) is based on knowledge of the x_j and the model, then we expect the variance of $y - \hat{y}$ to be less than that of $y - \overline{y}$. Thus R^2 may be interpreted as the proportion of total variation of y about its mean $\overline{y}$ that is explained by the fitted model.

Later when we discuss robustness, we will note that when outliers are present, the median absolute deviation should be used in place of the sum of squares in (1.15).

Unfortunately, (1.15) always increases when new factors are added to the model, and hence isn't suitable as a general model-selection tool, but it can be used to compare different models having the same number of factors. We will return to the topic of model selection later in the course. For now, suffice it to say that there is a notion of *adjusted* R^2 in which we divide each of $\sum(y-\hat{y})^2$ and $\sum(y-\overline{y})^2$ by their respective degrees of freedom. This can be expressed as

$$R^2_{\text{adj}} = 1 - a(1-R^2), \quad a = \begin{cases} (n-1)/(n-k-1) & \text{intercept model} \\ n/(n-k) & \text{no-intercept model} \end{cases} \tag{1.16}$$

It follows from (1.16) that for $n \gg k$, one has $R^2 \approx R^2_{\text{adj}}$ so R^2_{adj} is of little use in selecting between, say, a model with 10 factors or a model with 14 factors when there are 5,000 assets.

It is rarely safe to assume independent homoskedastic errors. The Newey-West procedure is a straightforward method of calculating standard errors in more general situations. Rewrite the above as

$$\text{var}(\hat{f} - f) = \text{var}((X'X)^{-1}X'\epsilon) = (X'X)^{-1}\,\text{var}(X'\epsilon)(X'X)^{-1} \tag{1.17}$$

and if we suspect that the error terms may be heteroskedastic, but still independent, then a reasonable estimator of the required variance matrix is

$$\widehat{\text{var}}(X'\epsilon) = \sum_{i=1}^{n} \hat{\epsilon}_i^2 \cdot \mathbf{x}_i \mathbf{x}_i' \tag{1.18}$$

Plugging this into (1.17) gives the heteroskedasticity-consistent standard errors. Note that in practice, we may have other ways of estimating the variances $\text{var}(\epsilon_i)$, such as using time-series methods or information from the options markets. In such cases a GLS estimator may be more useful, which also removes the need for (1.18). We will return to these points.

Whatever the details of our estimation procedure for $\text{var}(\hat{f} - f)$, the vector of standard errors for the coefficients and the associated t-statistics are then readily computed:

$$\text{se}(\hat{f}) = \text{diag}(\text{var}(\hat{f} - f))^{1/2}, \quad t = \hat{f}./\text{se}(\hat{f}). \tag{1.19}$$

In common with matlab, we use "diag" to denote the vector that is the diagonal of a matrix, and a dot in front of an operation to denote that it is pointwise, not a matrix operation. Since the square root of an expression involving $(X_t^* X_t)^{-1}$ appears in the denominator of the formula (1.19) for the t-statistic, it becomes clear that near-colinearities in the design matrix can bias all of the individual factor t-stats towards zero. It is for this reason that a "symptom" of colinearity is sometimes said to be a significant F-statistic but no significant t-statistics.

Note that $\text{rank}(X_t' X_t) = \text{rank}(X_t)$, hence If $\text{rank}(X_t) < p$, then the OLS estimator is not uniquely defined. In other words, there is a continuous space of vectors $\hat{f}_{t+1}$ which all have the same value for the least-squares objective function. One says that the optimization has a *flat direction*. From the statistics standpoint, this is called an *identifiability problem* since one can't uniquely "identify" the parameter values. Such problems are fairly common in econometrics and we will encounter many ways of resolving them, including Bayesian methods and examination of the singular value decomposition.

Even if X_t is of full rank, OLS can be misleading if X_t is approximately rank-deficient. This is a deficiency in OLS and later we will discussed more advanced methods for handling identifiability problems. The first line of defense, in these

situations, should always be to understand why the model isn't identifiable. If there is a linear dependency between the columns where we didn't expect one, then we should understand why such a relation arises. There may be a problem with the model specification.

Whether or not $X'X$ is invertible, there will always be a *Moore-Penrose pseudoinverse* (Penrose, 1955; Moore, 1920) of X, denoted X^+, in terms of which the minimum-norm least-squares estimate is $\hat{\beta} = X^+ y$. Keep in mind that if one column is an approximate linear combination of the others, it's probably a much better idea to remove that variable from the model, rather than finding the minimum-norm coefficient vector that has the offending column included. A small-norm solution to a flawed model is still, well, flawed.

References

Lindeberg, Jarl Waldemar (1922). "Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung". In: *Mathematische Zeitschrift* 15.1, pp. 211–225.

Moore, E. H. (1920). "On the reciprocal of the general algebraic matrix". In: *Bulletin of the American Mathematical Society* 26, pp. 394–395.

Penrose, Roger (1955). "A generalized inverse for matrices". In: *Mathematical proceedings of the Cambridge philosophical society*. Vol. 51. 03. Cambridge Univ Press, pp. 406–413.

Rousseeuw, Peter J and Christophe Croux (1993). "Alternatives to the median absolute deviation". In: *Journal of the American Statistical Association* 88.424, pp. 1273–1283.