Rise and Fall: An Autoregressive Approach to Pairs Trading

Bora Uyumazturk (yuyumaz@stanford.edu) and Vasco Portilheiro (vascop@stanford.edu)

***Abstract*—We pursue a pairs trading strategy using an autoregressive model, and compare it to a standard baseline strategy. We find that estimating the shifting mean of price spread while trading greatly improves return on investment, and conclude by proposing improvements to be explored in future work.**

I. Introduction

The idea behind pairs trading is simple: find a pair of stocks which track each other, and when they diverge, buy the lower one and short the high. If they converge again, you pocket the difference. It is perhaps this simplicity that has attracted the attention of so many of Wall Street's quantitative minds. First explored in the 1980s by a group of applied mathematicians and scientists led by Nunzio Tartaglia, automated pairs trading was initially a great success, helping the team generate about $50 million of revenue for Morgan Stanley in 1987. The group disbanded in 1989 after a few years of less than satisfactory returns, but many high profile quantitative firms, such as D.E. Shaw and Long Term Capital, have since tried their hand at the strategy, hoping that, with a minor tweak here or there, they could capitalize on the algorithm's early promise.

Though pairs trading sounds intuitive, in applying the strategy one finds that it is more nuanced than it seems. There are three main problems:

1) How does one identify profitable pairs, that is, pairs which will continue to track each other out of sample?
2) Assuming you can close your position exactly when prices cross again, it is easy to see that your profit is the spread between the stocks when you initialize your long-short position. It follows that to maximize profits, one must initialize trades at peak divergence. Without seeing into the future, how does one predict when this will be?
3) In the event that a pair does diverge, how can one detect it and react accordingly?

In this paper, we explore and attempt to answer these questions. Precisely, we try to devise an algorithm which, given 6 months of stock price data, identifies pairs suitable for trading along with a sequence of profitable trades for each pair which take place over the course of the next 3 months.

A. Related Work

Despite the notoriety of pairs trading on Wall Street, the literature on the subject reamins sparse. The first rigorous analysis of the strategy takes place in Evan Gatev's [1] 2006 paper, in which he implements a "distance" approach model, which we take as our baseline. This approach chooses pairs which minimizes the square distance between the normalized prices of stocks, expressed as

$$\sum_t (p_t^A - p_t^B)^2,$$

where p_t^A and p_t^A are the normalized price paths of stocks A and B respectively. He paper also establishes a theoretical basis for pairs trading using the "Law of One Price", or LOP, which states that markets should "assign to similar payoffs prices that are close", which basically states that in a reasonable market, if two stocks offer the same returns, they should have similar prices. Here they also set the precedent for looking for pairs only within industries, arguing that stocks in the same industry are likely to be affected by the same non-stationary factors, so by looking at the spread between them one might expect that the non-stationary factors in effect cancel each other out. This amounts to reducing the risk of divergence between pairs during the trading period.

Another, more recent attempt at pairs trading can be found in Qu, et. al [2], which tackles pairs trading using more modern methods. They present an overview of methods such as the distance approach mentioned above, stochastic spread, and cointegration. The stochastic spread approach models a spread X_t as an Ornstein-Uhlenbeck Process:

$$X_t = \rho(\mu - X_{t-1}) + \sigma W_t,$$

where ρ, μ, and σ are parameters to be estimated using a Kalman filter, and W_t is standard Brownian motion. This inspired us to use the the discrete time equivalent of this process, which is the autoregressive model. In this paper they also introduce model based trading rules, such as estimating the probability of convergence given two stocks which have been fitted to the above process, and using that to decide whether or not to enter a trade.

As a general reference on mathematical finance and stochastic processes, we used [3], and [4], [5] provided background on linear discrete time random processes and parameter estimation for time series.

B. Data

We pulled data from Yahoo Finance, which provided us with open and close prices at each day, as well as stock meta-information such as company name, symbol, and sector, and we restricted our scope to only stocks listed on the S&P 500. This is for two reasons. The first is that two stocks more are likely to be influenced by the same fundamental factors if they are listed on the same exchange, since most exchanges have some common theme. This way we avoid identifying stock pairs which are correlated coincidentally. The second reason is practical. In order to find optimal pairs, our models require computing parameters or statistics for each distinct pair of stocks, so clearly exploring all stocks on multiple exchanges quickly becomes computationally unfeasible.

In our models, we looked at the normalized time series rather than the raw data. This simply means dividing the entire time series by the initial value, so that every stock begins at value 1. This way each stock is viewed on the scale of percent returns so that metrics such as square distance are more meaningful.

II. Models

Each model takes as input about 6 months of historical stock data, identifies pairs and estimates parameters using that data, and then uses that information to make real-time trades during the following 3 months. For the rest of the paper, let

$$S_t^{A,B} = p_t^A - p_t^B$$

denote the spread between two stocks, A, and B (we will omit A and B when we are speaking in generality).

A. Baseline

As alluded to above, we adopted the model from [1] as our baseline. Pair selection amounts to choosing the 20 stocks pairs from each sector of the S&P 500 which minimize the square distance expression above. A trade is entered only if the absolute value current spread, $|S_t|$, is greater than two times the standard deviation of the spread over the history provided. There is no exit strategy, and the trade only closes when the spread crosses 0 or when the 3 month trading period is over.

B. Autoregressive Model

An autoregressive process of order p, denoted $AR(p)$, is a random walk of the form:

$$X_t = \phi_0 + \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t,$$

Where ϕ are parameters and ε_t is independent white noise, distributed as a $\mathcal{N}(0, \sigma^2)$ random variable. Colloquially, this model assumes that X_t depends on the p previous values of the process in a linear way. Looking at the $p = 1$ case, we find that the process can be rewritten as:

$$X_t = X_{t-1} + \rho(\mu - X_{t-1}) + \varepsilon_t,$$

where $\mu = \frac{\phi_0}{1-\phi_1}$ and $\rho = (1 - \phi_1)$. This makes clear the mean-reverting nature of the process, casting it as the discrete time analogue of the well known Ornstein-Uhlenbeck process. One limitation of the model is that it assumes stationary mean (assuming $||(\phi_1, ..., \phi_p)||_2 < 1$), which is given by

$$E[X_n] = \frac{\phi_0}{1 - (\sum_{i=1}^{p} \phi_i)}.$$

We attempted to model spread between two stocks as an $AR(p)$ process.

Parameter Estimation

Let $\Phi = (\phi_0, \phi_1, ..., \phi_p)$. We choose to estimate parameters using Maximum Likelihood. Taking the first p values of the process as fixed, we find the log-likelihood to be

$$l(\Phi, \sigma^2) = (n - 1 - p)\log(\frac{1}{\sqrt{2\pi\sigma^2}}) + \frac{1}{\sigma^2} \sum_{i=p+1}^{n} (X_i - \phi_0 - (\sum_{j=1}^{p} \phi_j X_{i-j}))^2.$$

Unsurprisingly, this greatly resembles the loss function from least squares regression, and we find that, defining our design matrix and labels as:

$$X = \begin{pmatrix} 1 & X_1 & X_2 & \dots & X_p \\ 1 & X_2 & X_3 & \dots & X_{p+1} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & X_{n-p-1} & X_{n-p} & \dots & X_{n-1} \end{pmatrix}, Y = \begin{pmatrix} X_{p+1} \\ X_{p+1} \\ \vdots \\ X_n \end{pmatrix}$$

our MLE estimates are given by the following familiar expressions:

$$\hat{\Phi} = (X^T X)^{-1} X^T Y$$
$$\hat{\sigma}^2 = \frac{1}{n-1-p}(\sum_{i=p+1}^{n} (X_i - \phi_0 - (\sum_{j=1}^{p} \phi_j X_{i-j}))^2)$$

Pair Selection

Realizing the importance of the stationary mean assumption, we wanted to pick pairs whose means seemed as consistent as possible. To do this, we used the parameters estimated above, choosing in each sector the 20 pairs which minimized:

$$||(\phi_1, \phi_2, ..., \phi_p)||_2,$$

which corresponds to choosing pairs whose spread returns to the mean the quickest. One can view this as analogous to the square distance metric from the baseline, but where distance is computed in the ϕ parameter space, instead of the historical time series space.

The difficulty with pair selection is that there is no canonical accuracy measure for it. For any set of price data, there is no set of agreed upon "best pairs." This makes it hard to either validate the first step of any given pairs trading model, or to compare different models' pair selections. Both of these tasks are key to shedding light on which path to follow in order to improve results. In light of this, we devised a method to:

1) given a full data set (that is, including both parameters estimation and trading periods), rank all possible pairs

2) based on this ranking, score any given model's selection of pairs

In order to accomplish 1, for each pair, we calculated the range of the moving average of the spread over 20 day windows. We then ranked these pairs in increasing order of range of average spread, the intuition being that minimizing this value is a proxy for stabilizing the spread over the whole data set. For 2, we devised a conservative scoring strategy, which rewarded models for choosing pairs from top 4% of the ranked $\binom{500}{2} = 124750$ pairs. Specifically, we assigned a full point for any pair chosen from the top 100 ranked pairs, 0.5 for a pair from the top 1000, and 0.25 for a pair from the top 5000. The total point count is then divided by the number of considered pairs. In order to get a maximum score of 1, a model would then have to select pairs only form the top 100. Figure 1 displays the results of using this pair selection scoring strategy.

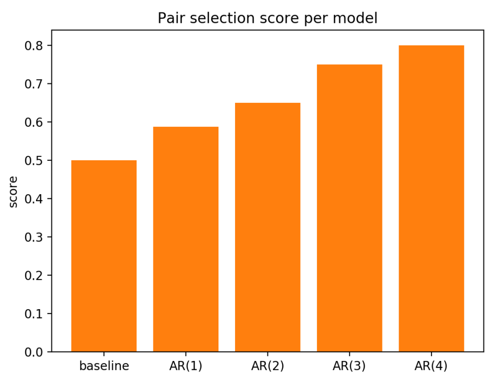

Fig. 1. Scores for model pair selection, including baseline and autoregressive models up to order 4.

It is clear that using an autoregressive model improves pair selection over the baseline strategy. Furthermore, with each increase in model order, there is about a further 0.1 increase in score. However, in further development and testing, we chose to focus on models of order 1 and 2. One unavoidable reason for this is pronounced slow-down in testing with each increase in model order. There are, however, independent motivations for this choice. First, we suspect, without testing, that increased model order risks overfitting during parameter selection. Second, as the scoring strategy is conservative, the scores for order 1 and order 2 models are in fact more than satisfactory, and our results during testing suggested that a more impactful area of improvement would be the trading strategy applied after pairs are selected.

Trading Strategy

The advantage of this model-based approach is that it allowed us to estimate various probabilities which would assist in making trading decisions. We list our trading rules below.

Convergence Probability (CP): Let S_t denote the spread between two stocks at time t. When considering entering a trade, we first estimated the probability of the convergence of the spread to the mean in the time remaining in the trading period using a Monte-Carlo simulation. Let us denote this quantity $P(convergence \mid S_t)$. In our first iteration of the trading strategy, we entered a trade at time t if

$$P(convergence \mid S_t) \geq 0.5.$$

Intuitively, this simply means that we should only trade if the spread is likely to converge before the end of the trading period. We hypothesized that this would minimize loss by preventing trades towards the end of the trading period that clearly did not have time to turn a profit.

Growth Threshold (GT): While the last rule was intended to minimize loss, our next rule attempts to maximize profit. As stated earlier, a trade made at time t which does reconverge to the mean makes exactly $|S_t - \mu_S|$, where S is the spread and μ_S is the historical mean of the spread. Therefore in order to maximize return, one would like to trade when the spread is at a local extrema, i.e. right before it begins to reconverge. Though it is impossible to know exactly when this will be, our model-based approach allows us to compute the probability the spread will increase at the next time step. It is easy to see that given fitted model parameters,

$$S_t \mid S_{t-1}, ..., S_{t-p} \sim \mathcal{N}(\phi_0 + \sum_{i=1}^{p} \phi_i S_{t-i}, \sigma^2),$$

which makes it relatively easy to compute $P(S_t > S_{t-1})$ or $P(S_t < S_{t-1})$ analytically. Our trading rule, then, was to only enter a position if

$$P(|S_{t+1} - \mu_S| > |S_t - \mu_S|) < \tau,$$

where in practice we found $\tau = 0.2$ to be effective. In English, this means that we only trade if waiting another time step will likely decrease our returns.

C. Adjusted Mean Model

In testing, we found that is was not uncommon during trading for the spread of a given pair to abruptly shift away from the mean it displayed during the parameter approximation period. This would often cause losses from ineffective trading, as the parameters, and therefore the trading rules described above, were no longer accurate for the given pair. We regard such a shift as reflecting some essential change in the market. Such a change would in essence shift the balance between to stocks, and therefore what the mean of their spread should be. In order to mitigate this effect, we conceived a model which would be resilient to shocks in the market that change the mean for any given pair of stocks, which we call the adjusted-mean model, $AM(p, q)$ (p being order and q being window size, explained in the next paragraph), which provides a way to change the mean around which trades are made, and makes consequent changes to the trading strategy.

	Overall	Consumer Discretionary	Energy	Financials	Health Care	Industrials	IT	Materials	Real Estate	Telecom
Baseline	-0.0063	-0.0121	-0.0006	0.0003	-0.0014	-0.0068	-0.0115	0.0164	0.0008	-0.0129
AR(1), CP	-0.0081	-0.0077	-0.0179	-0.0068	-0.0192	-0.0040	-0.0061	-0.010	-0.0018	-0.0021
AR(1), CP, GT	-0.0076	-0.0076	-0.0164	-0.0079	-0.0190	-0.0033	-0.0096	-0.0098	-0.0008	-0.0004
AR(2), CP	-0.0080	-0.0062	-0.0269	-0.0045	-0.0084	-0.0021	-0.0058	-0.0126	-0.0027	-0.0073
AR(2), CP, GT	-0.0067	-0.0065	-0.0243	-0.0033	-0.0055	0.0002	-0.0071	-0.0127	-0.0008	-0.0058
AM(2, 5)	0.0094	0.0105	0.0088	0.0088	0.0108	0.0120	0.0106	0.0089	0.0072	0.0074

TABLE I. TEST SET (11/25/2013-11/25/2015) AVERAGE PAIR ROI

	Overall	Consumer Discretionary	Energy	Financials	Health Care	Industrials	IT	Materials	Real Estate	Telecom
Baseline	-0.0097	-0.0082	-0.0050	-0.0038	-0.0063	-0.0046	-0.0031	-0.0085	-0.0053	-0.0492
AR(1), CP	-0.0079	-0.0056	-0.0087	-0.0005	-0.0060	-0.0058	-0.0134	-0.0067	-0.0083	-0.0185
AR(1), CP, GT	-0.0076	-0.0049	-0.0067	0.0024	-0.0070	-0.0064	-0.0141	-0.0052	-0.0074	-0.0234
AR(2), CP	-0.0106	-0.0075	-0.0165	-0.0001	-0.0042	-0.0151	-0.0142	-0.0046	-0.0134	-0.0273
AR(2), CP, GT	-0.0072	-0.0074	-0.0116	0.0017	0.0033	-0.013	-0.0084	-0.0009	-0.0096	-0.0248
AM(2, 5)	0.0090	0.0102	0.0124	0.0096	0.0112	0.0080	0.0097	0.0099	0.0063	0.0039

TABLE II. TRAINING SET (11/25/2015 - 11/25/2017) AVERAGE PAIR ROI

Mean Adjusting Strategy

The model builds on top of the autoregressive model, using the same initial parameter estimation method. The parameter p is therefore the same as the order of the autoregressive model used for parameter estimation. The difference is in how the adjusted-mean model behaves during trading. We observe the moving average μ_t of the spread over a window w_t of size q

$$w_t = \begin{pmatrix} S_{t-q+1} \\ \vdots \\ S_t \end{pmatrix}, \ \mu_t = \frac{\mathbf{1} \cdot w_t}{q} = \frac{1}{q} \sum_{i=0}^{q-1} S_{t-i}$$

which provides the insight necessary to determine when the spread has deviated from the mean approximated by the autoregressive process. When such a deviation occurs, intuitively, we want to wait until the mean stabilizes to a new value, and perform pairs trading around that new mean. In particular, we determine that such a deviation occurs when the sample variance of the running mean,

$$\sigma^2_{\mu_t} = \frac{(w_t - \mu_t)^2}{q-1}$$

becomes significantly larger than the predicted variance of the autoregressive process, given by

$$\sigma^2_{S_t} = \frac{\sigma^2}{1 - (\sum_{i=1}^{p} \phi_i)^2},$$

where σ^2 is the estimated variance of the white noise in the autoregressive process. In practice, we set a tolerance of $\pm\sigma^2_{S_t}$ for the difference between the sample variance of the running mean $\sigma^2_{\mu_t}$ and the predicted process variance $\sigma^2_{S_t}$. When this tolerance is surpassed, the adjusted-mean model makes no trades, until the samples variance $\sigma^2_{\mu_t}$ falls back within the parameters. Once the running mean stabilizes at μ_t, we perform the following parameter update:

$$\phi_0 = \mu_t(1 - \sum_{i=1}^{p} \phi_i).$$

Recalling that the mean of the autoregressive process S_t is given by

$$E[S_n] = \frac{\phi_0}{1 - (\sum_{i=1}^{p} \phi_i)},$$

the update above in effect changes the mean around which we now trade to μ_t.

One significant note to make is that the mean adjusting strategy relies on the assumption that the rate of reversion of a pair is intrinsic to the pair itself, and is independent of the mean of the spread at any given time. It interesting that this assumption is supported by the equations above, if we note that the variance $\sigma^2_{S_t}$ of the autoregressive process does not depend on ϕ_0. This captures the idea that similar stocks remain correlated in price due to the same market effects, while the spread in their prices may change.

Trading Exit Strategy

In addition to the CP and GT trade entering strategies inherited from the autoregressive model, the adjusted-mean model employs a more complex trading exit strategy. Namely, for each trade entered, an expected time of convergence $E[t_{convergence}]$ is calculated, during the same Monte-Carlo simulations used in the CP strategy to calculate $P(convergence \mid S_t)$. This value is used to close trades that may go awry due to a shifted mean. Regardless of whether the mean is moving, at each time step the adjusted-mean model checks whether the pair has converged, or, failing that, whether the expected time of convergence for the current trade has been surpassed. If this is the case, the model closes the trade as soon as the spread is larger than when the trade was entered, or when the spread converges. What this intuitively means is that if the expected convergence time passed, and we are in not in a better position than we were when we entered the trade, we exit.

III. EXPERIMENTAL RESULTS

Using the precedent set by [2], we report results in terms of average pair return on investment. Specifically, let A, B be a pair of stocks, and let $T^{A,B} = \{(q_i^A, q_i^B, o_i, c_i)\}$, be a series of trades, where q_i^A denote how many shares of A were bought, q_i^B denotes how many shares of B (negative shares indicate shorting), and o_i is opening time of trade, and c_i is closing time. Also let $p^A(t)$ be the price of stock A at time t (same for $p^B(t)$). Then the return on investment, or ROI of pair A, B is defined as:

$$ROI(A,B) = \sum_i \frac{q_i^A(p^A(c_i) - p^A(o_i)) + q_i^B(p^B(c_i) - p^B(o_i))}{|q_i^A|p^A(o_i) + |q_i^B|p^B(o_i)}.$$

Our test set came from 11/25/2013-11/25/2015, while our test set came from 11/25/2015 - 11/25/2017. For both train and test results, we averaged our returns over disjoint periods of 185 trading days, where we learn parameters for the first 125 days and trade for final 60.

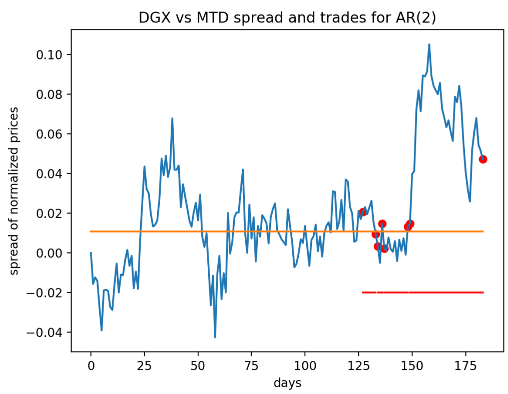

Fig. 2. Plot of spread for example pair over parameter estimation period for an $AR(2)$ model, with only the CP trading rule, and subsequent trading. Red dots mark trade times. Note that the final trade is a clear loss and a perfect example of spread divergence.

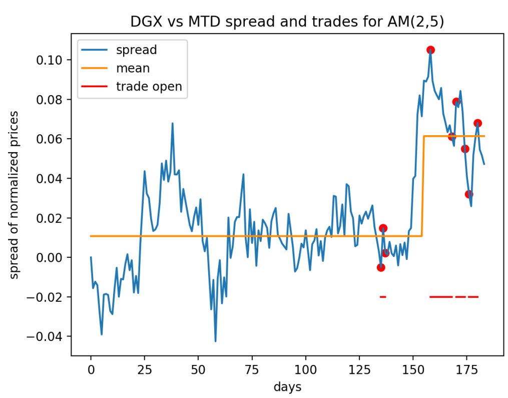

Fig. 3. Plot of spread for example pair over parameter estimation period for an $AM(2,5)$ model, with CP and GT trading rules, and subsequent trading. The shifting mean rules allow for deft handling of spread movement toward end of trading period.

IV. Discussion

From Tables I and II, we see that the predictive capabilities of the autoregressive approach does not make much of a difference (though the presence of a Growth Threshold rule consistently improves performance within models of the same order), except in the shifting mean model, which is the only one with positive returns across all industries. Although the pairs chosen are “good” in the sense of the metric mentioned in the Models section, too many pairs diverge for there to be net profit. Clearly the assumption that the mean is stationary is too strong, especially over a trading period of 3 months, during which economic conditions are bound to change. Looking at specific pairs, however, we find that, thought the mean may change, the mean reverting assumption of the autoregressive model is not completely inaccurate, and this is what accounts for the clear improvement in the adjusted means model. Figures 2 and 3 illustrate the difference in behavior between the standard $AR(2)$ model and the AM model.

Having observed this, we suggest an improvement to the adjusting means model, and try to place it on more stable theoretical footing. During trading, one can consider the parameters $\phi_1, ..., \phi_p$ and σ^2 as fixed, as they describe the dynamics of the stock pair, while attempting to optimize the bias, ϕ_0, since this is more susceptible to non-stationary market shocks. Considering only the last m weeks, say, one can then ask for a Maximum Likelihood estimate of ϕ_0 under these assumptions. Let t be the current time step. We then view the log-likelihood over the last m observations as a function of just ϕ_0 and take derivatives,

$$\begin{aligned}\frac{\partial}{\partial \phi_0} l(\phi_0) &= \frac{\partial}{\partial \phi_0} \sum_{i=1}^{m} (X_{t-i} - \phi_0 - (\sum_{j=1}^{p} \phi_j X_{t-i-j}))^2 \\ &= \sum_{i=1}^{m} (-X_{t-i} + \phi_0 + (\sum_{j=1}^{p} \phi_j X_{t-i-j})),\end{aligned}$$

and setting this equal to 0, we find that:

$$\hat{\phi}_0 = \frac{1}{m} \sum_{i=1}^{m} (X_{t-i} - (\sum_{j=1}^{p} \phi_j X_{t-i-j})).$$

Thus our new estimate of the mean, considering only the last m steps observations, is given by:

$$\begin{aligned}\hat{\mu}_t^m &= \frac{\hat{\phi}_0}{(1 - \sum_{j=1}^{p} \phi_j)} \\ &= \frac{\sum_{i=1}^{m} (X_{t-i} - (\sum_{j=1}^{p} \phi_j X_{t-i-j}))}{m(1 - \sum_{j=1}^{p} \phi_j)}.\end{aligned}$$

We suspect that the performance of the adjusting mean model would improve given these more refined update rules.

Acknowledgements

We would like to thank Andrew Ng and Dan Boneh for an informative and challenging course, as well as the TAs for all their hard work throughout the quarter.

Contributions

Bora Uyumazturk focused mainly on model selection, data extraction, testing, background research, and parameter estimation.

Vasco Portilheiro assisted in model selection and implementation, parameter estimation, as well as development of the pair selection metric and adjusted mean trading strategy.

References

[1] Gatev, E., W. Goetzmann and K. Rouwenhorst Pairs Trading: Performance of a Relative-Value Arbitrage Rule, The Review of Financial Studies (2006), 19, 798- 827.

[2] Qu, S., S. Mesmoeris, C. Davis et. al. ”Mean Reversion II: Pairs Trading Strategies”, Deutsche Bank Markets Research, (2016).

[3] Joshi, M.S. ”The Concepts and Practice of Mathematical Finance”, Cambridge University Press, (2008)

[4] Huerta, G. 2008 *AR, MA, and ARMA Models*, lecture notes, Time Series Analysis Stats 481, University of New Mexico, Week 3

[5] Stine, R. 2008 *Estimating an ARMA Process*, lecture notes, Time Series Analysis and Forecasting Stats 910, Wharton School of Finance