1) If the price series of a stock is stationary, we can easily trade it profitably. Can you suggest how? Does that mean that if a price series is white noise, it would be easy to profit from it?

A price series that is white noise is stationary, thus easy to profit from it. Just buy when price is accidentally low compared to mean, and wait and sell until it is accidentally high compared to mean! (Or vice versa).

Quote from a former student Tom: "If the stock price is stationary then we can determine a mean price that we expect to be. It would also be easy to determine bounds where the price has strayed far from the mean in either direction. As the price drifted further from the mean the expected profit would increase because it would be expected that in the future the price will return to the mean.

If the series was white noise then theoretically it should be easy to profit from. First you could determine how often you wished to execute and then determine the variance to find out the noise level that is required to execute at the frequency you wished to trade. For instance since it is a white noise series we know that 68% of the time the price will be within 1 standard deviation of the mean. Maybe a simple strategy would be to sell if the price reached above the 1 standard deviation level, or buy if the price was below the 1 standard deviation level.

This may be possible from a theoretical standpoint, but if the amplitude of the noise was low, it [may] not be possible to implement because of fees and bid ask spreads."

2) If you difference the price series of any stock, you will get the returns series. The returns series of any stock is stationary, but this won't make it easy to trade it profitably. Can you see why? In particular, if a returns series is white noise, do you think you can trade it profitably? Do you remember (we have discussed it before) what possible property of a returns series will enable us to trade profitably?

As many of you pointed out, only auto-correlated returns can be traded profitably. If returns are uncorrelated white noise instead, we can't predict whether price will go up or down in the next period, so naturally we can't profit from it.

3) Dickey-Fuller test is my favorite test for stationarity of a time series. All it is testing is whether the change in price depends (linearly) on the current price, dP=k*P+noise. If the time series is a random walk, clearly there is no dependence. What should the sign of k be for a stationary price series?

$k < 0$ indicates mean-reversion (change in price is always opposite in direction to deviation from mean), which is virtually the same as stationarity.

4) The author said he prefers MA models over AR models because of parsimony. Do you remember why MA models are more parsimonious than AR models? Parsimony is important because the fewer parameters a model has, the better it can forecast out-of-sample data as opposed to just fitting in-sample data. Algorithmic trading is all about forecasting out-of-sample returns, so parsimony of models is especially important there. That's also why I seldom use any non-linear model in forecasting returns.

Ruppert p. 222: "...there is a potential need for large values of p when fitting AR processs."

Ruppert p. 223: "The idea behind AR processes is to feed past data back into the current value of the process …. The effect is to have at least some correlation at all lags. Sometimes data show correlation at only short lags, for example, only at lag 1 …. AR processes do not behave this way, and, …. Do not provide a parsimonious fit….. an MA(1) model has zero correlation at all lags except lag 1…."

5) Is AR(1) always stationary? Can AR(1) represent a random walk? Can AR(1) represent a momentum (trending) price series? How would you profitably trade a momentum price series?

Money quote from a former student Duncan: "The AR(1) is not always stationary. It can represent a random walk. It can represent a momentum/trending price/time-series, as well. Ruppert says, "If $|\phi|\geq 1$, then the AR(1) process is nonstationary, and the mean, variance and correlation are not constant" (Ruppert, 211). If $|\phi|=1$ the time series is considered a "random walk" and if $|\phi|>1$ the time series is considered "explosive" and trends (Ruppert, 211-212). To profitably trade a time series exhibiting momentum, you'd just need to first identify the trend, second, buy if the time series were trending higher / sell short if it were trending lower and third, exit before the trend ended or right after the trend ended."

6) Is MA(1) always stationary? Is it easy to see why or why not?

MA(1) is always stationary; as the white noise term is stationary.

7) How do you tell if AR(p) isn't a good fit for a time series?

Ruppert p.220: "Any significant residual autocorrelation is a sign that the AR(p) model does not fit well."

8) Money quote on overfitting, parsimony, and out-of-sample testing in general:

Former student Joseph: "For me parsimony is critical due to the potential for data mining bias. A problem I see many hedge fund analysts get into is that they fit too many models. If you spend your day fitting models, eventually one will fit by pure chance. If you try 20 models, for example, you need to be cognizant that 1 will fit well by pure chance. This stresses the need for parsimony such that out of sample tests work well. In my experience if the model can't be tied back to some theoretical foundation, even if it is a complex mathematical theory, it is likely a result of data mining bias."

While former student Steve responded: "The way around that problem is to split the time series data into training and Out-Of-Sample testing datasets, and minimize the RMSE of the OOS testing dataset. The key is not to use a random split of the data but instead split the data across a point in time, i.e. the training dataset is earlier and the testing dataset is later. An even more advanced form is known as Walk-Forward Analysis, which splits the data into multiple temporal OOS testing datasets. These algorithms are more realistic than using theoretical Information Criteria such as AIC and BIC."

9) I have attached a concise summary of ARMA modeling written by Myron.

10) Former student Eric wrote: "cross-validation is a more robust test of a model's goodess of fit. While I didn't know about it until after I turned in this week's assignment, the claret package provides some useful functions for splitting time series for cross-validation. See the bottom of this page for an overview: http://caret.r-forge.r-project.org/splitting.html. I'm definitely going to use these for time series model validation going forward. They are definitely better than my one of sample slices that I did this week."

11) Eric also mentioned: "Have you tested hierarchical time series models? A relatively new R package rts implements the idea. Because of the tension between explanable parsimony and living with a complex world where "true" coefficients are rarely really zero, hierarchical models seem to allow a degree of flexiblity that varies with sub-group sizes.

12) Former student Andy asked

a) "I'm not sure what this means: Corr(Yi, Yj) =p(|i – j|) *for all i and j for some function p(h). (Ruppert p. 202)*

Corr is the correlation between Yi and Yj, where Yi is the price at time i and Yj is the price at time j. Imagine a vector containing the prices in ascending order in time. Call this Y. Then shift this vector i-j time steps backward or forward. Call this Y'. Corr(Yi, Yj) is the correlation between these two vectors Y and Y'. The notation p(|i-j|) just means that this correlation depends only on the absolute difference between time i and time j, not on i and j separately.

b) Ruppert p. 202 "The function p is called the autocorrelation function of the process*. Note that p(h) = p(-h). Why? …"*

p is called autocorrleation because it is the correlation between the vector Y and the time-shifted vector Y' that I described in a). Since we only use the autocorrelation function with a positive input, it won't matter whether p(h)=p(-h).

Correlation is by definition the covariance divided by the standard deviations of 2 input vectors. Since both input vectors are the same in an autocorrelation, the product of standard deviations is equal to the variance of the vector, hence autocorrelation is the autocovariance divided by the variance of the vector. If gamma is an auto-covariance and p is the autocorrelation, then gamma(0) is the variance and p(0) = 1.

c) Ruppert p. 206 "To estimate p(.), we use the sample autocorrelation *function (sample ACF)…"*

As stated in b), autocorrelation is autocovariance divided by variance. But we don't really know the true autocorrelation, autocovariance, or variance, since we only have a finite sample of data. To estimate these quantities from the finite data, we plug in the data points into equation 9.2. This is similar to estimating the mean of a vector by adding all the elements and divide by the number of elements, or estimating the variance by adding all the squared deviations from the mean and dividing by the number of elements n (or sometimes n-1).

13) Former student William asked "

a) Lag is the idea that an observation has value that may be repeated over time and this time difference would be called lag?

Lag is the time difference between 2 observations, it is not necessarily the time it takes for a pattern to repeat. For a random but stationary system, a random pattern still takes a random amount of time to repeat. The lag is not the average time this takes.

b) ***Autocorrelation is the idea that some future value is connected with some past or present value. Hence it may be predictable?***

Autocorrelation is predictable if the time series is stationary.

c) Autoregression is the idea that a certain pattern may be repeated*. Some pattern may have short durations and others larger. They may have different frequencies too. Lag is a measure of this difference in time for a certain pattern to repeat. Autoregression refers to the fact that a certain pattern can be regressed over and over in different intervals of time.*

Autoregression means that both the dependent and the independent variables in a regression are from the same time series, albeit with a lag between them.

d) White noise is just the pops you might get on a record player *(if you remeber them). It is the background and hopefully is random.*

e*) In an ACF plot if an autocorrelation event happens (exceeds our field goal posts) in a short period of time and decays then it indicates ...? stationarity*

Any statistically significant correlation at non-zero lag indicates time series is not white noise. It can still be stationary though.

f) What is phi? *I know it is the parameter. The examples of phi = .98 or phi = -.6 or phi = 1 or phi =1.01 don't register for me. Is phi the variation, slope or what? Or is this interpretation dependent upon something else?*

You can think of phi as a regression coefficient in the autoregressive relationship AR(1). It relates the future value of Y to a past value.

g) Lastly, interpret this statement: The ACF plot shows no no short-term autocorrelation, which is another sign that ARMA(1,1) is satisfactory. *Does this mean that our model has modeled the autocorrelations (removed them?) at least in the short term and therefore are working for us?*

It is important to note that the ACF plot in Fig. 9.13 is that of the **residuals** of an ARMA(1, 1) model. It does not mean that the time series is not autocorrelated. It means that the ARMA(1, 1) is a good model of the time series.

14) Using an AR(p) model for forecasting/prediction is intuitive. We want to predict future value y(t+1) using the most recent p past values of y with a linear equation. So if we want to predict y(t+1), we only need to input y(t), y(t-1), …, y(t-p+1) and nothing else into the model. See the unnumbered equation in Ruppert between 9.41 and 9.42.

However, using an MA(p) model for forecasting is a bit more tricky. Let's just consider MA(1), i.e. q=1. You might think this should be simple: the predicted value at t+1 should be just y(t+1)=mu, where mu is the estimated mean of the time series. Thus you might think if the mean is fixed using a training set, there is no input to the equation for the prediction! But that is incorrect. y(t+1) is actually mu – theta*epsilon(t), as Ruppert shows in the unnumbered equation immediately after Equation 9.44. How are we to determine epsilon(t), assuming mu and theta are again estimated values determined in the training set? Well, epsilon(t) is just the difference between y(t) and mu: epsilon(t)=y(t)-mu. So once again we are using the past q values of y to predict the next value of y, just like AR.

I have never understood why no textbooks ever explicitly show how MA(q) actually uses past values of y to predict future values as I demonstrated above. They always show how to do that using epsilon, but epsilon is not directly observed!

15) Former student William pointed out that MA models are easier (more accurate) to estimate than AR models because means are easier to estimate than linear regression coefficients.

16) Ashwini asks "When we do the time series modeling, do we always have to convert non-stationary data (trend/seasonal) into stationary process (may be using the differencing method or some othe transformation)? May I know the consequences of not doing transformation into stationary? Once we do the stationary transformation and have the forecast, how do we convert them back into their original scale that is non-stationary? Do we always use ARIMA to model non-stationary data?

Do we consider all stationary processes as white noise process or vice versa?"

My response: "It isn't always necessary or even possible to transform non-stationary data to stationary data. I generally prefer to work with log price series instead of returns series. In fact, I wrote in my latest book "Machine Trading" that

"Would it be advantageous to model log returns ΔY instead of Y using ARMA(p, q)? It would be if we can further reduce the lag p and q from the ones obtained when modeling prices (or log prices) using ARMA(p, q). Unfortunately, I have never found that to be true. For example, modeling the log of AUD.USD time series using ARIMA(p, 1, q) gives p = 1, and q = 9. The equivalence of an ARIMA(p, 1, q) model on log prices to an ARIMA(p, 0, q) model on log returns should not be confused with the statement that an ARMA(p, q) = ARIMA(p, 0, q) model on log prices is equivalent to some ARMA(p′, q′) model on log returns. The latter statement is false. An ARMA model in ΔY's can always be transformed into an ARMA model in Y's. But an ARMA model for Y cannot always be transformed into an ARMA model for ΔY. This is because an ARMA model for ΔY can only have ΔY as independent variables, whereas an ARMA model for Y can have both ΔY (which is just the difference of two Ys) and Y as independent variables. Hence, a model for Y is more flexible and gives better results. If we want to have a model for ΔY that have both ΔYs and Ys as independent variables, we have to use a VEC(p) model, to be discussed at the end of the next section on VAR(p)."

White noise process is certainly stationary, but not all stationary process is white noise. White noise means that there is no serial correlation in the prices (or returns), which isn't true for all stationary process."

17) Robert asked "what rules do you prefer for defining entry/exit points with mean reversion strategies?"

Often, I just apply the "Bollinger Band" trading strategy to a mean reverting price series. Those bands are defined by MovingAverage +/- N*MovingStandardDeviation, where N is typically 1 or 2.

However, much more sophisticated methods exist. See https://arxiv.org/abs/1408.1159 (Links to an external site.)Links to an external site. by Carr et al, or "Algorithmic and High-Frequency Trading (Links to an external site.)Links to an external site." by Cartea et al. Section 11.3 "Optimal Band Selection".

“*similar question as above, what signals do quants rely on to decide that momentum trends have broken/reversed?”*

Whether a momentum trend is broken is often determined by heuristic: if your position incurs a large enough loss, that means the trend is broken. So a stoploss will decide when you exit a momentum position. The magnitude of the stoploss is often determined by a multiple of the "Average True Range (Links to an external site.)Links to an external site." of that price series, or some other volatility measure. Alternatively, you can model the price series using ARIMA and apply the numerical method as outlined in the Carr paper cited above to find optimal exit rules.

Could we also discuss Figures 12.18 and 12.19 in the book: what is the real-world benefit of these forecasts…they are just horizontal lines? Can you realistically do anything with this sort of forecast?

Besides using a residual plot to determine whether a time series model is a good fit, one can check the AIC/BIC of the fit.This method is used by auto.arima to pick the best model - of course, even the best model may not be a good fit if AIC/BIC is too high.

Figures 12.18 and 12.19 are N-step-ahead forecasts. They may be of interest to economists who would like to have a view far into the future, but is not what traders typically use. Traders typically have much shorter term horizon: they want to know if price (or inflation rate) will rise tomorrow, or next month, not next year. So the ARIMA model they choose will typically provide a 1-period forecast, with the period equal to 1-day or 1-month, instead of making a N-period forecast far into the future. Why bother with the latter when one can adjust one's position every day in reaction to new information? Assignment 5 part B provides a glimpse of how a trader would use an ARIMA model for forecasting.

Could you provide us with some more thoughts/explanation on how to construct a portfolio of stocks that is stationary. And you use the term “value” in your question, which to me implies price, but thus far in this chapter we have focused on stationary returns series, so what do you mean by that?

An example portfolio of stocks that may be "stationary" is a long-short portfolio of 2 energy stocks. E.g. We will short Exxon and long Chevron. The net market value of this long-short portfolio will be approximately stationary, and therefore mean reverting. So short this portfolio if the market value is too high, and vice versa. Stationarity applies only to price series: returns series is always stationary but cannot be used for trading. (One cannot long or short returns: one can only long or short price, because price reflects the value of an asset but return does not.)

18) Scott said “Momentum investing began in the early 1990s and is a trend-following approach. According to Scott Bennett of Russell Investments, momentum investing presumes strong performance for a period of time can be a signal of continued outperformance for some future period. It capitalizes on behavioral biases and investor irrationality (Bennett, 2015) but is prone to severe downside volatility. Russell recommends careful monitoring of exposures and a value discipline with momentum investing. I’ve included the link below in my references for anyone wanting to red further.

I can attest to the downside of these strategies in my first job out of school. We performed due diligence on a firm called [REDACTED] out of [REDACTED], CT for their international momentum strategy and their bull market performance was unreal but they would get hammered in downturns. The approach also had extremely high turnover, which can generate a lot of short-term gains. We ultimately concluded that the strategy was more appropriate for non-taxable, institutional-like clients such as foundations and kept the fund out of taxable accounts.

…They appear to do OK in market rebounds but have slowly bled excess return since June 2008. Their calendar year performance shows back in 2008 they were only down -4.11% to the benchmark and +0.78% 2009. So I may have over-exaggerated because it does look like they gave up a lot following

2000. Since inception, their upside market capture is 132% and downside capture is 96%, so over time they have protected on the downside relatively well. Interesting example, it looks like they might experience the momentum crash you mentioned in the early 2000s, correct?

In terms of your fund, are you able to shift the portfolio quickly after the month of tough performance or does it take longer to reposition the fund? The size of my companies' funds and liquidity of high yield and bank loans makes it relatively difficult to implement views quickly, especially in our long/short credit hedge fund which with leverage is about $1 billion AUM."

My response: "Interesting that you found Axiom's momentum strategies performed well in bull but not bear markets. Usually momentum strategies have the opposite behavior - there is a phenomenon called "momentum crash (Links to an external site.)Links to an external site." where these strategies actually suffer most during market rebounds. I used to work for a hedge fund that made huge returns during the financial crisis, putting them over $1B in AUM, but then suffered huge drawdown in 2009, forcing them to shut down. I observed this to happen in the momentum strategy in our fund as recently as last month!

…It appears that [REDACTED]'s strategy is mean reverting, short realized volatility. It exhibits the exact opposite behavior to a typical momentum strategy. The crash in early 2000 followed the burst of the dotcom bubble. A momentum strategy would have benefitted from that burst.

Some funds that advertise they are trend followers actually do the opposite. The opaqueness and duplicity in the fund management industry is staggering. (Think Bernie Madoff!) But numbers do not lie.

It is difficult to "reposition" a strategy quickly enough. A short drawdown does not necessarily indicate a regime shift. There is bound to be losses in the transition period. So the key, as the author of the momentum crash paper noted, is diversification across different trading styles and strategies, reducing exposure to any one factor."

19) Hua wrote "**This concept of profitting from stationarity instead of actually forecasting really opens up my thinking**. Intuitively, if I can find stocks with difference follow a stationary process with acceptable contant mean and variance; also afford to be able to hold it long enough for mean reversion, then should be not too hard to profit?

Do you have any additional recommendation on specific readings/paper for this strategy? I want to try to execute this strategy to see how profitable it can be."

My response: "My 1st and 2nd books Quantitative Trading and Algorithmic Trading discussed such mean reverting strategies extensively."

Hua wrote "**I see that you used Matlab in your books. For Algorithmic Trading, what was the reason you chose Matlab as personal choice, versus R or Python**?"

My response: "I also compared R, Python, and Matlab in my third book Machine Trading. Matlab is much more powerful than R, and faster than most Python implementations. But most importantly, Matlab is supported [by] a team of expert engineers and mathematicians, not by a network of anonymous users with uneven expertise, ….

There are well-known bugs in SciKit-learn (see new book by Dr. Marcos Lopez de Prado) that persist and no one is fixing them for various reasons. With Matlab, any known bug will be fixed no later than half a year when a new release comes out.

…In my third book, I quoted an academic study ((Aruoba, Bora˘gan, and Fernández-Villaverde,

2014, ‘‘A Comparison of Programming Languages in Economics.’’ NBER Working Paper No. 20263) which found that Matlab is about 5x faster than R, and 2x faster than Python. Besides, both Matlab and Python can be sped up further by compiling into C++ executables, but R can't.

Matlab costs $150 for a Home license, about the same as a Microsoft Word license, so cost is not really a factor.

In terms of sophisticated knowledge of math or statistics, Matlab, R, and Python are all just programming languages. You can use them in any way you like, simple or complicated. I find the documentation of Matlab quite superior, since it is professionally written by mathematicians and maintained and updated continuously.”