Global TAA

Approaches to Asset Allocation

The theme is to detail the various quantitative approaches to tactical global asset management.

The Common Ingredients

Research team

The first step is to put together a dedicated research team. It is critical that the project leader have some knowledge of regression econometrics. In particular, knowledge is needed on the following topics:

- model building
- general specification tests
- overfitting and cross validation
- residual diagnostics, including heteroskedasticity and serial correlation
- nonlinear regression
- model choice techniques

It is also critical that a research protocol is developed. Too often, members of the research team get sidetracked into searching for the best "R^2". It is important to realize that the success of the Global TAA effort will not be measured by the R^2. Success will be judged relative to a benchmark return.

Data

The research team must have easy access to a variety of data. It is often appropriate to designate one member of the research team to data. The collection and maintenance of the database is very important. Tactical decisions need to be made quickly after new data arrives. It is best to invest in a database system that takes the new data and automatically runs the quantitative programs.

Computing

This is not really a major issue. While workstations are desirable, most calculations can be performed quickly on PCs - given that the focus is on monthly or quarterly returns. Database systems are also desirable. While most top-down data management exercises can be handled within Excel, the bottom up projects are not feasible within a spreadsheet. The bottom-up projects may include up to 10,000 securities along with vectors of attributes for each security.

Top Down - X-Opt

I call this "Top Down X-Opt" because the predicted stock returns are not used in any type of optimizer, i.e. a mean-variance optimizer which chooses portfolios which have the highest expected returns per unit of variance. The steps are as following:

- Build country-by-country forecasting models based on benchmark return from MSCI or IFC.
- After validation or models, forecast out of sample returns.
- Sort country returns
• Invest in portfolio of highest expected return countries.

This approach does not tell us how many countries to invest in, nor does it tell us how much to invest in each country. Some possibilities are:

- Equal weight countries selected by judgement
- Value weight countries selected by judgement
- Use judgement to both weight and select the highest expected returns countries.

"Hedge" strategies are also possible. This involves taking long positions in the highest expected returns countries and short positions in the lowest expected returns countries. A number of caveats are in order for this style of strategy.

- Some countries may not be shortable.
- Some countries, while shortable, may be prohibitively expensive to short.
- Short positions almost always involve positive investments, hence the hedge strategy is not a zero net investment strategy.
- The "hedge" strategy may not really be a hedge. That is, the long and the shorts are not necessarily offsetting.

Let me elaborate on this last point. Usually, a hedge strategy is measured with respect to some benchmark. For example, consider a U.S. equity hedge strategy. High expected returns securities are purchased and low expected returns securities are sold. The "beta" or sensitivity of the high expected returns portfolio is calculated with respect to a benchmark, like the S&P 500 stock indexd return. The "hedge" portfolio is constructed by selling the portfolio of low expected returns securities that has the same beta as the high expected returns portfolio. This produces a zero beta or hedged portfolio. The expected return on the hedge portfolio is the "alpha".

Importantly, this portfolio does not have zero volatility. The beta measures the average movement in the portfolio given a movement in the benchmark. The higher the R^2 in the beta regression, the more closely the portfolio and the benchmark move. In the U.S. hedge strategy, the volatility would be much lower than holding either side of the hedge. Nevertheless, volatility could be on the order of one half the S&P 500 volatility.

With the international portfolio, it is more difficult to achieve this hedge. First, a benchmark needs to be designated. This depends on your performance criteria. A global manager may choose the Morgan Stanley Capital International (MSCI) world index. An international manager may choose the MSCI EAFE. A U.S. manager may choose the S&P 500.

Three issues arise (which we will detail in the International Risk Management). First, the betas of some of the countries could be zero. This is especially the case with some of the emerging stock markets. Harvey (1995, Pxx) shows that many of the emerging markets have zero betas and some have negative betas. The problem arises in the following way. If you are shorting India, which has a negative beta, you are increasing the risk of the portfolio.

The second issue has to do with the low R^2 in the beta regression. Harvey (1991, Pxx) shows that the average correlation among developed markets is 41% (which implies an R^2 of only 16%). The correlations among emerging markets are lower. Hence, the hedge portfolio, even if it has zero beta risk, will have substantial volatility.
The third issue has to do with the stability of the betas. As we will see, the betas could be very unstable. The usual method of obtaining betas relies on a regression analysis which essentially averages the historical comovements between the country and the benchmark. There is no guarantee that the future comovements will look like the past. We will tackle this problem later by proposing dynamic risk models which explicitly forecast the future comovements between the country and the benchmark.

**Top-Down Opt**

This approach follows the first initial steps of Top Down X-Opt. Country forecasting models are built, validated and out-of-sample forecasts are formed. The difference is that the information in both the volatilities and correlations is used in determining optimal portfolio weights. Top Down Opt is usually performed within the context of optimal portfolio control techniques. These techniques minimize variance for target levels of expected returns and maximize expected returns for target levels of variance. That is, Opt almost always refers to a portfolio strategy conducted within the mean-variance paradigm.

So the extra steps involved are:

- Forecast variance out of sample
- Forecast covariance (or correlation) out of sample.

There are a number of different possibilities which are detailed in the section on volatility. Here are some options:

- Use historical estimates of volatility and correlation (perhaps based on the last five years of data). This is essentially an equally weighted moving average. These are called *unconditional* variances and covariances.
- Use modified historical estimate that places more weight on recent information. An example of this is an exponentially weighted moving average (EWMA). This is what J.P. Morgan calls RiskMetrics.
- Use average conditional variance and covariance.
- Use conditional variance and covariance.

**Average conditional variance, covariance and correlation**

Let me elaborate on the last two possibilities (the first two are simple). The average conditional variance is defined as:

$$ACV(r_i) = \text{AVERAGE}((r - E[r|Z])^2)$$

Note the difference between this and the usual variance:

$$V(r_i) = \text{AVERAGE}((r - E[r])^2)$$

In the usual estimator, we square the returns minus the average return and take the average value. In the average conditional variance estimator, we square the returns minus their *predicted* values.

The intuition behind this estimator is that it measures the average dispersion of the realized returns from their predicted values.
The same type of formula can be applied to covariance:

\[ \text{ACC}(r_i, r_j) = \text{AVERAGE}\{ (r_i - E[r_i|Z])(r_j - E[r_j|Z]) \} \]

Note the difference between this and the usual covariance:

\[ \text{C}(r_i, r_j) = \text{AVERAGE}\{ (r_i - E[r_i])(r_j - E[r_j]) \} \]

The intuition is identical to the variance estimator. The average conditional covariance tells us how two securities move relative to their predicted values. For example, a positive average conditional covariance tells us that when one security is above (below) its predicted value the other security is, on average, above (below) its predicted value.

Importantly, the average conditional variance and covariance estimators are easy to implement. No special econometric techniques are required. Once the forecasting models are built for the returns, it is simply some averages that need to be calculated. Indeed, another way to expressing the average conditional variance is just:

\[ \text{ACV}(r) = \text{AVERAGE}(\text{residuals}^2) \]

where residuals are the regression residuals from the forecasting model for the returns. Indeed, some regression programs report this statistic as the Mean Squared Error. Note that the residual is just defined as:

\[ \text{residual} = r - E[r|Z] \]

Similarly, the average conditional covariance is just

\[ \text{ACC}(r_i, r_j) = \text{AVERAGE}(\text{residual}_i \times \text{residual}_j) \]

Also, note that the mean value of the residuals is always zero (by construction in regression). Hence, one can just feed the residuals into any statistical routine and calculate the variance, covariance and correlation. The output will give the average conditional variance, average conditional covariance and the average conditional correlation.

However, note that an average of past observations is not necessarily the best forecast of the future. Indeed, this average places equal weights on all historical observations. One possibility is to use an exponentially weighted moving averages on squared residuals and the residual cross-products. This would give more weight to recent observations. But even with this modification, it is not clear that this is the best method to forecast future volatility and correlation. However, it could be a significant improvement over naive historical averages of returns from their mean values (the unconditional measures).

**Conditional variance, covariance, and correlation**

For these measures, see the section on volatility models. The intuition for the conditional variance is that we want to provide the best forecast of the squared deviation from the predicted return. That is, we are predicting the dispersion from what we predicted - we are not predicting whether the dispersion will be above or below the predicted value. Indeed, we can't predict the residual itself. If we can, then our forecasting model is missing something (i.e. if you can predict the model mistakes, then the model needs to be rebuilt).
Investment horizon is important here. If your rebalancing period is one quarter, you should look at the quarterly dispersion and covement measures. You may not care about some of the intra-quarter movements that you might be measuring with the conditional variance fit on monthly data.

**Bottom-Up - X-Opt**

I will spend a considerable amount of time discussing this style of asset management in a later lecture. For now, the idea is to select individual securities. From a variety of methods, forecasted winners are purchased and forecasted losers are sold. Briefly, the two most popular methods are portfolio attribute classification and cross-sectional regression.

**Portfolio attribute classification**

In this method, portfolios are formed based on particular attributes. For example, at the end of the year, suppose we have 2000 securities and the price to book (PB) ratio for each security. Sort by PB and form quintile portfolios. Track the return over the next year. Re-form the quintiles at the end of the year. Depending on the results, we can implement a simple "value"-based strategy. One can, say, purchase the low PB portfolio and sell the high PB portfolio. Usually, the portfolios are value weighted.

The portfolios can be sorted again. Suppose we sort each PB quintile based on firm market capitalization (SIZE). Then there are 25 portfolios. Again, the performance of these portfolios and a strategy can be implement. For example, we may choose to buy low PB and low SIZE portfolios and sell high PB and high SIZE portfolios.

An important limitation of this analysis is that we can't sort on too many attributes at the same time. With three attributes, we would have 125 portfolios. With 2000 securities, there would be times when very few securities fell into a particular portfolio.

Overall, the appeal of this method is its simplicity. However, the method has severe limitations in practice. In addition, note that no optimization has been performed.

**Cross-sectional regression**

With this method, returns in quarter \( t \) are regressed on attributes that are available at \( t-1 \). Using the estimated coefficients, an out-of-sample forecast of the returns in quarter is formed by multiplying the regression coefficients times the attributes that are available today, at time \( t \).

Forecasts are obtained and some portfolio strategy is implemented, such as purchasing the high expected return securities and selling the low expected return securities. It is even possible with this method to implement a Bottom-Up Opt. However, some special problems arise. We cannot go ahead and use the standard mean-variance tools. It is not feasible to put 2,000 securities into a mean-variance optimizer. However, there are some ways around this which will be discussed later.
Global TAA

Forecasting International Returns

The Traditional Approach

The traditional approach to asset management input assumes that asset prices follow random walks (potentially with some drift). In these models, this month's price is the best forecast of next month's price. In the model with drift, the forecasted return is this month's price plus a constant. In both of these models, returns (differences in the log prices) are unforecastable.

\[ P_t = \text{drift} + P_{t-1} + \epsilon_t \]

implies

\[ R_t = \text{drift} + \epsilon_t \]

Notice that \( \epsilon_t \) is so-called white noise, which implies that the returns are not forecastable. The expected return is just the drift which is often measured as an average.

Traditional asset management will calculate historical average returns, historical average volatility and historical average covariance and apply these inputs in a mean-variance framework following Markovitz (1959).

This framework imposes the assumption that in a more general model:

\[ R_t = d_0 + d_1(Z_{t-1}) + \epsilon_t \]

where \( Z_{t-1} \) represents lagged information, the coefficient \( d_1 \) is exactly zero. With this constraint, the R-square of the model is exactly zero. Hence, the traditional approach starts with the constraint that the R-square is zero.

The traditional approach provides a lower bound for our exercise. If we fail to forecast returns, we will also get a zero R-square. Hence, in failing we will fall back on the traditional approach. If we get a zero R-square, the forecasted return is a constant.

The General Setup

We will be interested in building models which explicitly incorporate conditioning information. The research protocol gives us a set of guidelines for avoiding specification problems and data mining traps. We will develop a parsimonious model which has been tested on an quasi-out-of-sample basis, i.e. we hold out some data for out-sample-testing (it is quasi because the data are know to us).

After developing the model, we implement using the following steps. Suppose the date is October 31.

1. Fit regression with returns through October 31 and information variables through September 30.
2. Save coefficients, $d_0, d_1, \ldots$
3. Use coefficients and information variables available October 31 to form an out-of-sample forecast for the return in November.

**Information Variables**

**An Economic Model**

In specifying information variables, we must be keenly aware of the data mining problem. Variables must make economic sense to be included in the regression model. Hence, we must start with an economic model. Usually, we use some variant of the Lucas (Econometrica, 1978) asset pricing model.

\[
\text{Forecasted Cash Flows into Future} \\
\text{Price}_t = \sum \text{---------------------------} \\
\text{Expected discount rates (for each period)}
\]

This is just the present value formula. However, it has some unique features compared to the Finance 101 version.

- First, cash flows are being explicitly forecast. The forecasts are based on information today.
- The expected discount rate is forecasted. It can be different for each period.

One could future decompose the expected discount rate into a part due to:

- Risk free rate
- Risk exposure of firm
- Risk premium for risk exposure

Each of these could change through time and may be different depending on the horizon of the forecasted cash flow.

**Economic Categories**

The following are a list of general categories:

- Inflation
- Business Cycle
- Fundamental
- Default Risk
- Microstructure of Market
- World Integration
- Political Risk
- Momentum
- Sentiment

Each of these categories has some relation to the general model. They are not mutually exclusive. That is, information variable may fit into more than one of these categories. Indeed, there can be correlation among the categories.
We will focus the discussion on the building of a model for a particular country. It will be generalized to an international setting.

**Inflation**

Inflation itself is not that useful of a forecasting attribute because it measures past inflation. We are really interested in future inflation. This will affect both the numerator and the denominator of our valuation formula. Some candidate variables include:

- Interest rates. Contain expected inflation.
- Term structure. Contain term structure of expected inflation.
- Survey expectations, e.g. Money Market Services.
- The change in survey expectations.

**Business Cycle**

The business cycle affects the expected cash flows in the numerator. We need leading indicators of the business cycle. One might immediately go to index of leading indicators (LEI). However, this index has produced a remarkable number of false signals. It is better to go to the components and variants of the components. Among the macro economic information:

- Housing starts
- Building permits
- New orders for capital equipment
- Moving average of new claims for unemployment insurance
- Corporate earnings
- Money supply aggregates
- Help wanted ads
- Capacity utilization
- Expected GDP growth

It is probably not wise to use last quarter's GDP growth. That is backward looking and we want something that promises to forecast the future.

There are also some financial variables that could be useful.

- Dividend yield of aggregate market index
- Term structure of interest rates
- Interest rates relative to 6 or 12 month moving average

The dividend yield is very cyclical [see Fama and French (Journal of Financial Economics, 1989)] and is a traditional forecasting variable. Harvey (Journal of Financial Economics, 1988) shows that the term structure forecasts real economic activity. A number of authors, including Campbell (Journal of Financial Economics, 1987) show that the term structure also forecasts stock returns.
Fundamental Valuation

There is a long history of using fundamental - or accounting - based information in stock selection. Recently, Ferson and Harvey (1993, 1995) show the importance of certain country based fundamental valuation metrics in the cross-sectional stock selection problem. The variables of interest are:

- Book to price ratios
- Price to earnings ratios
- Price to cash ratios
- Dividend yields
- Expected price to earnings ratios

The first four ratios are available through MSCI for developed countries and the IFC has coverage of all but the price to cash ratio. IBES has recently made available global aggregates of their survey data for a number of accounting aggregates. I have included the expected price to earnings ratio. However, there are many additional useful variables in the IBES database.

Default Risk

This category is correlated with the business cycle category. Default risk will change the discount rate in the denominator of the valuation equation. Default risk is highly correlated with the business cycle. A popular specification includes:

- Moody's Baa minus Aaa bond yields
- Low grade bond returns minus high grade bond returns


Microstructure of Market

The microstructure of the market could affect the discount rate applied to the firm in that market. That is, if the market is thinly traded, investors will demand a premium for transacting in that market. Some possible measures include:

- Trading volume or change in volume
- Volume divided by number of shares issued
- Turnover (value of trading divided by value of market capitalization)
- Number of margin transactions
- Short interest and change in short interest
- Volatility (both time series and cross-sectionally measured among the component stocks in the index, see Bekaert and Harvey (1995))
- Asset concentration ratios, on scale of 0 (all firms equal size) to 1 (one dominant firm)
- Industry concentration ratios, on scale of 0 (equal dispersion of industry) to 1 (all firms in one industry).
World Integration

The degree of market integration will affect the expected discount rate. Integration means that the same risk project will require the same discount rate no matter where the project is located. In reality, some markets are segmented. As a result, a larger discount rate is required. Integration works on the denominator of the valuation formula. See Bekaert and Harvey (Journal of Finance, 1995). They propose two proxies for openness:

- Trade sector, which is exports plus imports divided by GDP
- Growth in exports
- Size of equity market, which is market capitalization divided by GDP

Bekaert and Harvey (1995) specify a functional form for the degree of integration which relies on these variables.

Bekaert and Harvey (1995) pursue the intuition that a market that appears segmented because of numerous investment restrictions could be integrated if investors can access the market in different ways. They examine country funds and American Depository Receipts (ADRs) as proxies for this accessibility.

- Index of countries funds
- Index of ADRs
- Composite index

Political Risk

Much has been written and talked about political risk but not much mentioned. Obviously, political risk could influence the discount rate in the denominator of the valuation formula. A negative political event could adversely affect the numerator (cash flows).

Erb, Harvey and Viskanta (1994, 1995) have examined a proxy for political risk using Institutional Investors' Country Credit Rating. More recently, they are pursuing a project based on Political Risk Services'(PRS) indices.

- Institutional Investors' Country Credit Rating (available semiannually)
- PRS Political Risk Index (all PRS available monthly)
- PRS Credit Risk Index
- PRS Economic Risk Index
- PRS Composite Index

Momentum

There are various economic reasons why serial correlation in returns may arise. These reasons focus on the speed of adjusting production as a result of economic shocks. Momentum is often proxied by:

- Previous month's return
- Earnings growth
• Number of new highs or lows
• Up/Downs
• Moving average cross-over rules, e.g. MA12 minus MA3
• Flow of funds to mutual funds
• Various "technical indicators"

Sentiment

This factor will influence the expected cash flows and the discount rates. It is often measured by:

• Consumer confidence
• Put/call ratio
• Consumer debt (also credit card debt)
• Art index (or discretionary expenditures)
• Closed end fund discount or premiums

Model Building

For each country, the availability of data is determined. The next step is to conference with the research team. A small set of variables are pre-selected for each country. Small being 10-15 variables. The variables should have a theme. That is, the same type of variables should be chosen across different markets. The model building process then begins with strict adherence to the research protocol. Out-of-sample validation is critical to the success of this exercise.
Global TAA

Econometric Tools for Asset Management

Regression Model Basics

We will be interested in forecasting \( R_t \) as a function of lagged information \( Z_{t-1} \). It is logical to start with a linear regression model. Later we discuss the generalization of this linear model using nonparametric density estimation techniques.

The linear regression model is with a single explanatory variable:

\[
R_t = d_0(Z_0) + d_1(Z_{1,t-1}) + \text{residual}_t \quad [1]
\]

where \( d_0, d_1 \) are regression coefficients.

This is often presented as

\[
R_t = d_0 + d_1(Z_{1,t-1}) + \text{residual}_t \quad [2]
\]

The \( d_0 \) is interpreted as the intercept and the \( d_1 \) as the slope coefficient. Equation [1] and [2] are identical. Remember we have a single explanatory variable. It turns out, in the standard implementation, of regression, that the \( Z \) contains two variables: \( Z_1 \) might be an interest rate level and \( Z_0 \) is a constant vector of ones. In a spread sheet, one can think of the first column as the returns, say from January 1970 through December 1994, the second column has a "1" in every row, the third column is the interest rate from December 1969 through to November 1994 (it is lagged). Notice I have no time subscript on \( Z_0 \) because it is just a column of ones.

Suppose we ran the following regression:

\[
R_t = d_0(Z_0) + \text{residual}_t \quad [3]
\]

This is a regression on the column of ones. What is \( d_0 \) in this case? It is just the average return. It is also an equally weighted average return. According to regression theory, the coefficient is

\[
d_0 = \text{INV}(Z'Z)Z'R \quad [4]
\]

where \( Z \) is just a column of ones. This can be broken down into two parts.

\[
\text{INV}(Z'Z) = \text{INV}(\#\text{obs}) = \frac{1}{\#\text{obs}}
\]

\[
Z'R = \text{SUM}(\text{returns})
\]
Hence, it is obvious that the \( d_0 \) is the average return, i.e. the sum divided by the number of observations!

Why are we focussing on this trivial regression? Well, the traditional style of asset management uses average returns (as well as variances and covariances) the mean-variance optimization. Sometimes, moving-window averages (MA) are used, say the last five years. In this case, \( Z_0 \) would have zeros in the initial rows and "1"s in the last 60 rows (assuming monthly data is used). Sometimes, exponentially weighted moving averages (EWEMA) are used. Again, we can set the \( Z_0 \) to handle this.

What is the R-square of this regression in [3]. Remember, the definition of R-square is the variance of the regression fitted values divided by the variance of the dependent variable. An R-square of 1.0 or 100% implies that the fitted values exactly coincide with the realized returns.

\[
\text{R-square} = \frac{\text{Var(fitted)}}{\text{Var(R)}} = \frac{\text{Var}(d_0)}{\text{Var}(R)} = 0
\]

The R-square is zero. Why? The variance of a constant, \( d_0 \), is exactly zero. Remember definition of variance. It is the squared deviation of the variable from its average. Since the variable is always equal to one, there is no variance.

Another way of looking at this exercise is to note that those using this style of model are assuming that no other \( Z \) variable influences future returns. In fact, in running this special regression (and, indeed, you do not need to run a regression, you simply need to push the average button), they are assuming the \( d_1 \) and other coefficients are exactly equal to zero.

Using the average as a forecast forces the asset manager implement a strategy with a zero R-square. This is not necessarily a desirable strategy. Indeed, it implies that no other information affects expected returns. It implies that expected returns are constant (at least over the 60-month window of the MA).

Using a more general regression model, we can incorporate predictability. We can execute statistical tests to ensure that the predictability is genuine rather than an artifact of data snooping. The research protocol details procedures that avoid potential misspecification.

**Regression Diagnostics**

**Heteroskedasticity**

This problem occurs when the variance of the error term changes through time or across a cross-section of data. As a result, the least squares estimator will be unbiased but inefficient, i.e. you get the right point estimates for the parameters but the variances of the estimated parameters are not minimum variances.

The correction for heteroskedasticity is straight forward and involves weighted least squares. There are a number of approaches to this correction. Basically, each variable is transformed (often by dividing by a variance measure), and the regression is reestimated.

**Autocorrelation**
Autocorrelation or serial correlation is commonplace in time series regressions. Autocorrelation implies that the errors in previous periods carry over to the present period. Like heteroskedasticity, an autocorrelated regression will have unbiased but inefficient estimators. In fact, the variance of the regression coefficients will be underestimated leading one to falsely believe some parameters are statistically significant. Furthermore, if the model is used to forecast, the predictions will be inefficient (i.e. unnecessarily large sampling variances because we are not using important information -- in the previous error terms).

The solution strategy is to transform the regression variables. Suppose we have the following bivariate regression:

\[ Y_t = d_0 + d_1 X_t + \text{resid}_1t \]

and suppose the error follows a first-order autoregressive process:

\[ u_t = r_0 + r_1(u_{t-1}) + \text{resid}_2t \]

where \( r_1 \) is less than one and \( \text{resid}_2 \) is normally distributed and has constant variance. Transform the regression by multiplying the lag of \( Y \) by \( r_1 \) and taking the first difference:

\[ Y_t - r_1Y_{t-1} = d_0(1 - r_1) + d_1(X_t - r_1X_{t-1}) + (\text{resid}_1t - r_1\text{resid}_1_{t-1}) \]

or

\[ Y^*_t = d_0^* + d_1X^*_t + \text{resid}_1^*t \]

Now the model is properly specified and the estimation can proceed as usual.

**Multicollinearity**

One of the basic assumptions was that no linear dependence exists between any of the independent variables. The reason that this is important is that we need to invert the matrix, \( X'X \) to get the least squares estimator of the coefficients. The problem of multicollinearity does not arise when we have two independent variables that are exactly the same -- because the computer cannot estimate the coefficients. Multicollinearity arises when some of the independent variables are close to being the same.

The main consequence of multicollinearity is that the precision of the estimates deteriorate. It becomes very difficult to determine the relative influences of the independent variables. Investigators may be falsely led to drop variables that are insignificantly different from zero. Furthermore, the coefficient estimates could be sensitive to the block of data used, i.e. the first subperiod could deliver parameter estimates that are different from the second subperiod.

The usual solution strategy is to calculate the correlation matrix of all the independent variables. If two variables have a high degree of correlation, judgement should be used to determine which one to drop from the regression.

Another possibility is orthogonalization. Suppose \( Z_1 \) and \( Z_2 \) are correlated but you do not want to drop
one of the variables. One can regress $Z_2$ on $Z_1$ and save the residuals. A regression could then be run on $Z_1$ and these residuals. The interpretation of the residuals is that they are the part of $Z_2$ that is uncorrelated with $Z_1$.

**Omitted Variables**

This is a common specification error. In general, the parameter estimates will be biased as a result of omitting important variables. The only case where the bias disappears is if the omitted variable is uncorrelated with the included variables -- this case, however, is unlikely. If the omitted variable has a positive covariance with variable $X_i$, then the parameter estimate $d_i$ will be biased upward. The omitted variable problem will also affect efficiency. Inference about the coefficients will be wrong because the residual variance is biased upward.

**Errors in the Variables**

The errors in the variables problem arises when one or more of the independent variables are measured with error. In this case, the parameter estimates will be biased and the degree of bias depends on the variance of the measurement error.

The usual solution strategy is to opt for an *instrumental variables* estimator rather than ordinary least squares. The properties of this estimator are beyond the scope of this note.

**Detection and corrections for conditional heteroskedasticity**

Conditional heteroskedasticity occurs when the variance of the error term changes through time. Many financial time-series exhibit heteroskedasticity. Some examples are interest rates and volatilities of stock returns. Using ordinary least squares will deliver the correct estimates for the coefficients but *the standard errors and t-statistics will be incorrect*. If you are drawing inferences about the coefficients, then you must have the correct standard errors.

It is important to check for heteroskedasticity and to correct for it when it exists. Unfortunately, most computer packages do not have corrections for conditional heteroskedasticity. Most software packages (like Statgraphics) do not deliver corrected standard errors. However, the Fuqua version of Statgraphics has been modified to allow us to get heteroskedasticity consistent standard errors.

**Detection**

The best method of detection involves saving the residuals from a regression and plotting the residuals against time. If there is an obvious pattern, then it is likely that there is a conditional heteroskedasticity problem. Here are two more sophisticated tests.

**Breush-Pagan (1979) Test**

1. Run the regression in question and obtain coefficient estimates.
2. Save the residuals
3. Calculate the variance of the residuals, $vr$
4. Regress the squared residuals divided by the variance of the residuals on a set of information variables, $Z_i$
5. One half the regression sum of squares (i.e. the sum of the squares of the fitted values) is distributed Chi square with s degrees of freedom, where s represents the number of regressors in $Z$. 

http://www.duke.edu/~charvey/Classes/ba453/tools/tools.htm
(excluding the constant). A high Chi2 (low probability value) indicates evidence against the hypothesis that the coefficients are zero on the Z variables, which is evidence against the hypothesis of conditional homoskedasticity.

Engle (1982) Test

This is a test for Autoregressive Conditional Heteroskedasticity or ARCH. There is no statistics package available at this time that corrects for this type of heteroskedasticity. But you should be aware of this form -- since many financial time-series exhibit ARCH disturbances.

1. Run the regression in question and obtain coefficient estimates.
2. Save the residuals.
3. Regress the squared residuals on the lags of the squared residuals (you choose the number of lags).
4. The number of observations times the unadjusted R-square is distributed Chi-square with \( s \) degrees of freedom and \( s \) represents the number of lags used. A high Chi-square (low probability value) indicates evidence against the hypothesis of conditional homoskedasticity.

White (1980) Test

This is a popular test for Conditional Heteroskedasticity. The steps are as follows:

1. Run the regression in question and obtain coefficient estimates.
2. Save the residuals.
3. Regress the squared residuals on the original independent variables plus the cross-products and squares of the original independent variables.
4. The number of observations times the unadjusted R-square is distributed Chi-square with \( s \) degrees of freedom and \( s \) represents the number of regressors in the regression (excluding the constant). A high Chi-square (low probability value) indicates evidence against the hypothesis of conditional homoskedasticity.

Popular Corrections

There are three important references for corrections. White (1980, *Econometrica*) provides the most widely used correction (it is implemented in Statgraphics, SAS, and RATS). Hansen (1982, *Econometrica*) provides a more general correction whereby White is a special case. Newey and West (1987, *Econometrica*) provide an alternative to Hansen (in some situations's Hansen's correction will not work, however, you will know that it has not worked because the estimation routine will fail). Finally, Andrews (1991, *Econometrica*) provides the state-of-the-art correction. My recent research employs the Andrews' correction.
Tactical Global Asset Allocation

Research Protocol

In order to maximize the chance of successful model building, a strict research protocol should be set in place. This discipline involves seven steps.

1. Specifying the Problem

In this step, the problem is defined. In most cases, a forecasting problem is considered. A list of candidate explanatory variables is formulated. It is important that this list be developed in advance of the data gathering process (to minimize the impact of data snooping). Candidate variables should have economic meaning. For example, no forecasting model should contain variables that have no economic foundation (e.g. sunspots, Yankee batting averages and hemlines). In addition, a discipline should be developed in advance for the appropriate lag structure. No model should be presented, say, with the 5th and 17th lag of an explanatory variable. This smacks of data snooping. If a lag structure beyond 1 is needed, the justification must be developed in advance. Detailed notes of this process must be taken. An appendix to each research presentation must list the variables which were prespecified in stage one.

2. Data Collection

In this step, the data are obtained. It is important that the data are examined for potential errors. Even high profile data services like DATASTREAM can have data incorrectly keyed in. I recommend graphing each series. The eyeball is a powerful filter. In addition, when emerging markets are being considered extra attention needs to be paid to how the stock return is constructed (does it properly account for rights issues and dividends).

Importantly, I usually require a two holdout sample methodology. For example, suppose we are forecasting the U.K. equity returns. We obtain data from say, 1970-2005. Our model would be fit over a shorter sample, say 1970-2001. Our model will be initially validated over the 2002-2004 sample (36 observations). After the validation process has been documented and the final model selected, the 2005-2006 data are obtained. A second validation is executed with these fresh data. An alternative strategy which has advantages when low frequency data are used is to randomly holdout 60 months. That is, use the data from 1970-2006 and randomly select 60 noncontiguous months for a holdout sample.

3. Estimation

In this step, the initial models are estimated. Linear and nonlinear models should be estimated with the most general econometric methodologies, such as the Generalized Method of Moments (GMM). [See, for example, Harvey P3 and Harvey and Kirby W15. The GMM is ideally suited as a forecasting method. The method handles linear and nonlinear problems. The method provides test statistics which are robust to departures from the traditional distributional assumptions. However, it is possible that some preliminary analysis could be done with the least squares (OLS) methodologies. It should be noted that, least squares is just a special case of the GMM.

The use of GMM allows for rigorous hypothesis testing of the model specifications. All tests should be heteroskedasticity-consistent and robust to potential moving-average errors induced in the data (this structure could be particularly important is raw transactions are being modeled which bounce between bid and ask and index situations which might suffer from infrequent trading).
"Stepwise" regression or GMM is not desirable. A small number of models should be specified in advance (step 1) and the final in-sample selection can be made using a variety of evaluation methodologies. The final selection should contain about three models. These models are chosen on the basis on traditional criteria such as adjusted R-square, Akaike Information Criterion, and Schwarz Criterion. All of these methods minimize squared errors and penalize models which have extra variables (reward parsimony). Parsimony is critical because it is an important determinant of out of sample performance.

The data snooping problem should be recognized at all stages of the research. With 20 randomly selected explanatory variables, one variable should enter the regression by chance with a t-statistic greater than 2.0. The prespecification of the variables helps minimize the snooping problem.

4. Validation

Each model must be validated on an out-of-sample basis. Using the holdout samples, the model performance is assessed using the same metrics. This helps identify the model which will be promoted.

The validation stage is clearly the most important. It potentially eliminates the data snooping problem. That is, if variables have been snooped to maximize, say the R-square, it is unlikely that these variables will perform on an out-of-sample basis. The validation will also eliminate forecasting models which contain too many variables. The overparameterized models will surely fail. Finally, the validation stage is important for identifying potential instability in the parameter estimates. One might have the right explanatory variables but the wrong functional form. While some of this problem should have been resolved in the estimation stage, any residual issues should be evident in the validation stage.

With the final model or models, the final part of the data is obtained for the second validation. This dataset is free of any possible snooping and provides a clean out of sample test.

5. Trading simulation

The statistical model building must next be transferred to a trading strategy. This stage is implemented on the two holdout samples. It may also be implemented on the in-sample data. There should be a close correspondence between model statistical fit and trading performance. However, when we are faced with three finalist models, it is unlikely that their predictive performances are statistically different. The trading simulation can help isolate the final model.

At this stage, levels of slippage and transaction costs need to be incorporated into the analysis. In addition, if large trades are being considered market impact must also be assessed.

6. Trading simulation benchmarking

A measure of the significance for the selected strategy is assessed by comparing the strategy's performance to two sets of benchmarks. First, a bootstrap distribution of random trading strategies is formed. The strategy in question should lie in the upper 10% tail of the bootstrapped distribution. Second, another distribution is formed with a set of 100 moving-average rules denoted (x,y,z). These moving-averages are applied to price (rather than return series). However, all evaluation of signals is done on the basis of returns. The moving average cross-over rule executes a buy (sell) when the short-term moving average, length x, goes below (above) the long term moving average, length y. The variable z denotes the band. For example, a band of 1% says that no trade is executed (neutral) when the short-term moving average is within 1% of the cross-over price. The selected strategy is evaluated
within the context of the moving average rules. This provides a tougher comparison to the randomly generated strategies.

7. Reporting

The paper trail is extremely important. A detailed appendix needs to be maintained which logs the models tested. Extensive model diagnostics are also presented in the appendix along with the graphical analysis. The diagnostics should include the standard battery of residual analysis tests, specification tests, details about the explanatory variables and their correlation structure, model comparisons, and model (parameter) stability tests.

The main body of reporting should detail: the statistical performance of the model of choice (in sample and out of sample), the next two closest competing models, the trading performance (correct hit rates, number of negative months in a row, maximum drawdown in any year, dollar profits, dollar standard deviations, Sharpe ratios, Graham-Harvey measures, and benchmark comparisons).

The report should explicitly detail the validation procedures employed.