Intro — Introduction to financial statistics+

+These features are part of StataNow.

Description   Remarks and examples   References   Also see

Description

Financial statistics provides methods for statistical analysis of financial data. Financial data typically include time series on the returns of financial assets, such as stock price returns, foreign exchange returns, and bond yields. Statistical analysis of financial data often consists of fitting regression models to multiple time series, to cross-sectional data, or to panel data.

This introduction provides an overview of aspects of financial statistics. See [FIN] **fin** for Stata commands and a workflow for the analysis of financial data.

Remarks and examples

Remarks are presented under the following headings:

Introduction

Financial statistics is the branch of statistics concerned with financial data. Financial statistics blends financial theory and statistical methods. Financial theory provides guidance on which financial variables are important to investigate and which relationships among variables are likely to be meaningful and generates empirical implications of theory that can be tested with data. Statistical methods provide the empirical framework for testing theoretical models and provide quantitative measures of qualitative theoretical predictions.

Financial markets are characterized by an abundance of data. We are often interested in asset prices and returns, including equity returns, foreign exchange returns, and bond yields. Financial models propose relationships among these returns, both within and across time, which can be tested. Questions of interest involve both the cross-sectional and the time-series behavior of asset returns. Over any given time frame, some assets have higher returns than others; cross-sectional models attempt to capture this variation. Similarly, returns vary over time, and time-series methods are used to capture this behavior. A perennial question in asset pricing is whether returns can be forecasted using past data, a question that naturally fits into a time-series framework.

Thus, financial statistics blends theory and data, incorporating statistical methods already in use while also designing new techniques to address the particularities of financial data.

For an introduction to financial statistics, see Hurn et al. (2020). Campbell, Lo, and MacKinlay (1997) provide an advanced introduction to the econometrics of financial markets. Cochrane (2005) and Campbell (2018) provide introductions to asset pricing and combine theoretical models and empirical estimation. Boffelli and Urga (2016) provide an introduction to time-series methods in Stata with special attention given to financial data.

Characteristics of financial data

The most basic form of financial data is the price of a financial asset. Asset prices can be recorded at various frequencies, including annually, quarterly, monthly, weekly, daily, and high-frequency intraday. Many assets are traded continuously throughout trading days, so lower-frequency information is an aggregate of higher-frequency observations. This can introduce subtleties in the recording of financial data. For example, a monthly series on an asset price could record the beginning of month value, the middle of month value, the end of month value, or the average of daily observations. In turn, these daily observations could be beginning of day, end of day, or midpoint of daily range. Each of these aggregation methods incorporates different levels of information about the behavior of the asset price and combines that information in different ways. For the most part, in this manual, we will work with daily or monthly asset prices at their end-of-period values.

The return on an asset price is a measure of how that price changes over time. Several types of returns can be defined, and each has its uses. The most basic return is the change in the price, for example, the end-of-day value minus the previous day's end-of-day value. The percentage return is the percentage change in the price, which normalizes for the level of the price and makes comparison over time or across assets more straightforward. Percentage returns are of direct interest to financial market participants but have some properties that make statistical procedures challenging to apply directly. The logarithmic return is the difference in the log asset price, which has similar properties to the percentage return but has the advantage of having properties more suitable for statistical analysis.

The basic form of the data, then, is a collection of observations on one or more asset prices or asset price returns over time. An individual asset price measured over time is a "time series". A collection of asset prices measured over time is a "multiple time series".

Financial statistics is typically concerned with both the cross-sectional and the time-series behavior of asset returns. The cross-section of returns captures the notion that different assets pay different returns at a single point in time, and the researcher would like to understand drivers of this variation in returns. Over time, a researcher would like to know whether returns are forecastable from past returns or other characteristics. Combining the two ideas, a researcher might wish to know whether the variation in returns is persistent over time and which characteristics drive the cross-section of average returns.

Portfolio construction

An individual investing their wealth into a collection of assets is said to be building a "portfolio" of assets. Portfolios are linear combinations of assets with weights indicating shares in a particular asset. If there are, for example, K assets, then the portfolio construction problem is to choose weights $(w_1, \ldots, w_K)$ that satisfy some criterion. An individual might choose to hold a portfolio that achieves a goal that a single asset cannot achieve, such as reducing investment risk. When combined, two assets that tend to move against each other can generate a less variable stream of returns than either asset would in isolation.

The simplest criterion might be to maximize the return of the portfolio. On its own, this criterion would reduce to placing all one's investments into the single best-performing asset. However, if one is willing to sacrifice some average return, then portfolios can typically be constructed that imply lower variance in returns than any single asset. Formally, a minimum variance portfolio is the linear combination of assets that attains the lowest variance possible while still generating a desired target return. The necessary ingredients are the historical average returns of each asset and the covariance matrix of the returns. Collect the average historical returns of the K assets into an $K \times 1$ vector $\mathbf{r} = (r_1, r_2, \ldots, r_K)'$, and collect the historical covariance among the assets into an $K \times K$ covariance matrix $\mathbf{\Sigma}$. Then, for a portfolio with weights $\mathbf{w}$, the mean return and variance of the portfolio are

$$r_p = \mathbf{w}'\mathbf{r}$$
$$\text{Var}(r_p) = \mathbf{w}'\mathbf{\Sigma}\mathbf{w}$$

Suppose the investor wishes to hit a target return $\overline{r}_p$ with a portfolio that has as low a variance as possible. Then the variance-minimizing weights $\widehat{\mathbf{w}}$ that obtain a target return $\overline{r}_p$ are the weights that solve

$$\widehat{\mathbf{w}} = \text{arg min}_{\mathbf{w}} \text{Var}(r_p) = \mathbf{w}'\Sigma b\mathbf{w}$$
$$\text{s.t.} \quad \mathbf{w}'\mathbf{r} = \overline{r}_p$$
$$\sum_{i=1}^{K} w_i = 1$$

The first constraint states that the desired return must be $\overline{r}_p$, and the second constraint states that the weights must sum to one.

Another approach maximizes the return-to-risk ratio, also known as the Sharpe ratio. Given a risk-free asset r_f and the standard deviation σ_p of the portfolio, the risk-adjusted return on a portfolio is its Sharpe ratio,

$$S_p = \frac{r_p - r_f}{\sigma_p}$$

The numerator here is the so-called excess return: the return generated by the portfolio in excess of return that can be obtained with negligible risk, say, through investment in secured government debt. The maximum Sharpe ratio strategy chooses weights $\mathbf{w}$ to maximize the Sharpe ratio.

In both the variance minimization and Sharp ratio maximization cases, the weights must sum to one, but there is no requirement for the weights to be positive. A negative weight on an asset indicates that the asset is being "sold short". A short position is the act of borrowing an asset, selling it, and using the proceeds to finance long (positive weight) positions. Then the short position must eventually be closed, at which time the borrowed asset is paid back. Restricting the weights to be positive is thus a restriction on short selling. The above optimization problems can be modified to handle no-short-sales restrictions.

In addition to these optimizing strategies, a researcher might wish to set custom weights or, in the simplest case, to set equal weights on all assets in the portfolio. These settings may be used to assess alternative strategies or provide a baseline against which optimizing strategies can be assessed.

Stata can compute portfolios based on minimum variance or maximum Sharpe ratio, for a given target return or globally, with or without short sales. Stata also provides tools to create custom (nonoptimizing) portfolios and portfolios with equal weights. See [FIN] **finportfolio**.

Portfolio summaries

Individual assets and portfolios can be compared across a number of dimensions. Most of these comparisons involve the mean of returns, their "volatility" (that is, variance), and functions thereof.

Portfolio mean. The portfolio mean is the average return on the portfolio, given the weights and the average historical return on the assets in the portfolio. The units of the mean are the same as the units of the data: with monthly data, we have an average monthly return; with daily data, the average daily return; and so on.

Excess return. The excess return of a portfolio is its return minus the return on a safe asset. Safe assets are those that provide guaranteed or near-guaranteed returns, serving as a benchmark for riskier portfolios. Some securities, like certain short-term bonds, can serve as a risk-free asset. Excess returns represent returns in excess of the risk-free benchmark.

Portfolio standard deviation. The portfolio standard deviation (square root of the variance) captures the variability of the portfolio around its mean and is the simplest measure of portfolio risk.

Market beta. A portfolio's market beta is the regression coefficient in a regression of portfolio returns on market returns. It captures volatility, not in itself, but with respect to the market as a whole. A beta of 1 indicates that when the market is up 1%, the asset or portfolio is also up 1%. A beta of 0 indicates that the portfolio does not move with the market. A negative beta indicates that the portfolio is a "hedge": it tends to rise when the market falls. Betas can be constructed for other characteristics as well, by including them as independent variables in the financial regression.

A portfolio's standard deviation and market beta capture different kinds of risk. Comparing portfolios by their standard deviations penalizes portfolios that are more volatile, regardless of the timing of that volatility. On the other hand, comparing portfolios by their market beta provides more information about risk. Portfolios with high beta tend to produce low returns at the same time as the market as a whole. They also tend to produce high returns when the market overall is also producing high returns. Portfolios with high beta are said to be risky because they fail to pay off when other assets also fail to pay off. By the same token, they can be enticing for those who have an appetite for risk. Assets or portfolios with zero or negative market beta pay off precisely when the market does not, which may be valuable for investors seeking more safety in their portfolio.

Sharpe ratio. The Sharpe ratio, mentioned earlier, divides the excess return of a portfolio by its standard deviation, thus providing a measure of return to risk. A Sharpe ratio of 1, for example, indicates a tradeoff of 1% higher volatility for each 1% in excess returns. A Sharpe ratio of 0.3 indicates a 0.3% excess return for every 1% of higher volatility. A Sharpe ratio of 3 indicates 3% higher excess return for every 1% of higher volatility.

Treynor ratio. The Treynor ratio divides the excess return by the market beta of the portfolio.

Jensen's alpha. Jensen's alpha is the intercept in a regression of the portfolio return on the market return. It captures average return of the portfolio after accounting for the portfolio's covariance with the market.

Additional information about financial regressions, whose outputs are ingredients in market beta, the Treynor ratio, and Jensen's alpha, is presented below. Stata can compute many of these summary statistics automatically after creating a portfolio; see [FIN] **finsummarize**.

Financial regressions

Risk and return

A basic principle in asset pricing is that assets with higher risk ought to compensate with higher returns or, put another way, that any asset exhibiting unusually high returns only does so because it takes on higher risk.

The classic measure of risk is the variance of an asset's return (or its standard deviation). In one direction, assets that have higher returns should have higher risk, or there would be no incentive to hold lower-return assets. In the other direction, assets that are riskier should have higher returns to entice investors to hold such risky assets. Asset-pricing models characterize risk in terms of observable variables. These variables, often called factors or risk factors, are then used as independent variables in models of asset returns.

Work by Sharpe (1964), Lintner (1965), Jensen (1968), and Black (1972) used the mean-variance optimization theory described in *Portfolio construction* to derive the capital asset pricing model (CAPM), in which the appropriate measure of risk is not the variance of an asset's return itself but the covariance of the asset's return with "the market", a measure of the return to all assets. Intuitively, assets that pay off when the market is doing poorly serve as a hedge and can thus trade with lower expected average return; but assets that pay off when the market is doing well require a higher expected return to entice investors to hold them. Financial regressions characterize the exposure of assets to market risk.

Later extensions to the CAPM added additional risk factors and are known as multifactor CAPMs.

Capital asset pricing model

Financial analysts seek to understand the sources of variation in returns. One of their key tools is financial regression. Let the return of asset i at time t be r_{it}. Then a financial regression relates the asset's return to a risk-free rate r_t^f and independent variables.

$$r_{it} - r_t^f = \alpha_i + \beta_i(r_t^m - r_t^f) + \boldsymbol{\delta}'\mathbf{x}_t + e_{it}$$

The dependent variable is the time series of asset returns in excess of the risk-free rate, $r_{it} - r_t^f$. The key independent variable is the return on the market, also in excess of the risk-free rate, $(r_t^m - r_t^f)$. The parameter β_i is the coefficient on adjusted market returns, that is, the "market beta". The presence of additional independent variables $\mathbf{x}_t$ allows for other factors or controls with coefficients $\boldsymbol{\delta}$. The term α_i is an intercept in the regression and is of particular interest in financial statistics. Finally, e_{it} is a mean-zero error term. Such time-series regression models are known as CAPMs. These are time-series regressions because the estimate of β_i is driven by variation in a single stock over time.

In a financial regression, both the slope coefficient β_i and the intercept α_i are of interest. The slope coefficient measures the comovement of an asset with the market itself. An "aggressive" asset is one with $\beta_i > 1$, so that the asset moves more than one for one with market movements. A "conservative" asset has $0 < \beta_i < 1$; it moves in the same direction as the market but less than one for one. An asset with $\beta_i = 0$ has returns that are uncorrelated with market returns. Finally, an asset with $\beta_i < 0$ moves against the market, yielding positive returns during market downturns and vice versa; it serves as a hedge. The intercept α_i captures average returns that are not explained by the independent variables in the regression. Under the CAPM theory, the intercept should be zero. This forms the basis for a test of the theory.

So far, we have looked at a single asset. But we could fit one regression for each asset, allowing for a different intercept and slope coefficient for each asset. The resulting model is

$$\begin{aligned} r_{1t} - r_t^f &= \alpha_1 + \beta_1(r_t^m - r_t^f) + \boldsymbol{\delta}_1'\mathbf{x}_t + e_{1t} \\ \vdots \quad &= \quad \vdots \\ r_{Kt} - r_t^f &= \alpha_K + \beta_K(r_t^m - r_t^f) + \boldsymbol{\delta}_K'\mathbf{x}_t + e_{Kt} \end{aligned}$$

Notice that the adjustment variable r_t^f and the independent variables are common to all regressions. Letting $z_{it} = r_{it} - r_t^f$, $\mathbf{z}_t = (z_{1t}, \dots z_{Kt})'$, and $f_t = r_t^m - r_t^f$, the collection of time-series regressions is a model,

$$\mathbf{z}_t = \boldsymbol{\alpha} + \boldsymbol{\beta} f_t + \boldsymbol{\Gamma}\mathbf{x}_t + \mathbf{e}_t$$

where $\boldsymbol{\alpha}$ is the $K \times 1$ vector of intercepts, $\boldsymbol{\beta}$ is the $K \times 1$ vector of slope coefficients, $\boldsymbol{\Gamma}$ is an $K \times C$ matrix of coefficients on the C independent control variables $\mathbf{x}_t$, and $\mathbf{e}_t$ is the $K \times 1$ vector of disturbance terms.

The CAPM is based on the assumption that the factors f_t explain all the variation in expected returns. This assumption generates two testable implications. First, additional factors beyond those included in one's theory should enter the regressions with coefficients of zero. Second, the intercepts $\boldsymbol{\alpha}$ should all be jointly zero. The latter implication is frequently assessed using the Gibbons–Ross–Shanken test.

In Stata, you can use `finregress capm` to fit a CAPM; see [FIN] **finregress capm**.

Fama–MacBeth regression

One result of the above time-series procedure is an estimate of the risk measure β_i for each asset. Different assets have different levels of risk, and a key implication of the model is that assets with higher expected returns should have higher risk. Thus, while the CAPM regressions above are fit to time-series data, the single β_i estimate for each asset has a cross-sectional implication. Ignoring covariates, the expectation of returns is

$$E(r_i) - E(r^f) = \beta_i[E(r^m) - E(r^f)]$$

This implication can be tested in the cross-section of returns. The cross-sectional model is

$$E(r_i) - E(r^f) = \gamma + \lambda\beta_i$$

The dependent variable is the expected excess return on asset i. The independent variable is asset i's beta risk, β_i. The parameter λ captures how expected returns vary with beta risk; it is called the price of risk. If all sources of risk are "priced in" (that is, if they are all accounted for in setting the price of the asset), then $\gamma = 0$.

Fama and MacBeth (1973) devised a two-stage method for estimating the cross-sectional relationship between expected return and market risk. In the first stage, a collection of K time-series regressions of asset returns on the market portfolio is performed, as in the CAPM regressions above. This stage yields K estimated β_i coefficients. In the second stage, a cross-sectional regression is run at each time point t ($t = 1, 2, \dots, T$) with the cross-section of excess returns as the dependent variable and the market risk factors β_i as the independent variable. The result of the second stage is a time series of estimated intercepts and slopes, $\{\widehat{\gamma}_t\}$ and $\{\widehat{\lambda}_t\}$. The final parameter estimates are the time-series average of the intercept and slope,

$$\widehat{\gamma} = \frac{1}{T}\sum_{t=1}^{T}\widehat{\gamma}_t \qquad \widehat{\lambda} = \frac{1}{T}\sum_{t=1}^{T}\widehat{\lambda}_t$$

The single estimate $\hat{\lambda}$ measures the degree to which expected return rises for a unit increase in beta. Deviations from this line are the pricing errors for individual assets, the α_i in the time-series regression. And $\hat{\gamma}$ can be thought of as an average of the pricing errors; if the covariates in the model capture the mean of expected returns, then $\hat{\gamma}$ will equal zero. An estimate of $\hat{\gamma}$ substantially different from zero indicates the presence of residual variation in expected returns that is not captured by the model.

Fama and MacBeth (1973) estimate standard errors as the standard deviation of these time series of coefficients,

$$\sigma^2(\hat{\gamma}) = \frac{1}{T^2}\sum_{t=1}^{T}(\hat{\gamma}_t - \hat{\gamma})^2 \qquad \sigma^2(\hat{\lambda}) = \frac{1}{T^2}\sum_{t=1}^{T}(\hat{\lambda}_t - \hat{\lambda})^2$$

The Fama–MacBeth regression assumes the underlying returns are independently distributed through time but allows for correlation among returns within a time period. The regression does not account for the fact that the β_i are estimated. Shanken (1992) proposed a correction to the standard error calculation that accounts for this.

Cochrane (2005, 245–251) provides a detailed discussion of the Fama–MacBeth regression and its relationship to standard methods in panel-data estimation.

In Stata, you can use `finregress fmb` to fit a Fama–MacBeth regression; see [FIN] **finregress fmb**.

Time-series methods

Financial statisticians also use a number of time-series models to characterize, explain, and forecast financial data. This section provides an overview of models. This overview is not exhaustive. See [TS] **Time series** for a more comprehensive introduction to methods used in time-series analysis and Stata's tools for time-series data. This overview highlights some of the most widely used methods.

Autoregressive moving-average models

Autoregressive moving-average (ARMA) models are a flexible class of models for characterizing univariate time series. The arma model takes as its building block a noise process e_t to build a time-series y_t as the sum of its own past values and current and past values of the noise.

$$y_t = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + e_t + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q}$$

This is called an ARMA(p, q) model. The autoregressive terms are the terms involving lags of y_t itself. The moving-average terms are the terms involving the current and lagged values of the disturbances e_t. The parameters are $(\phi_0, \phi_1, \ldots, \phi_p, \theta_1, \theta_2, \ldots, \theta_q)$. The choice of p and q is called the lag order of the ARMA model. Various criteria can be used to select the lag order. Visual plots of autocorrelations can be used to assess dependence. Formal tests that use information criteria to tradeoff model complexity and explanatory power are also employed to estimate the proper lag order.

The ARMA model is designed to characterize time series that are stationary. Stationary time series have a constant long-run mean and variance, meaning they do not exhibit trends. Some financial time series do exhibit trends; for example, the price of an asset might trend upward or downward over time. Even then, a time series could have an ARMA structure after differencing it one or more times to remove a trend. Such models are called autoregressive integrated moving-average (ARIMA) models and provide an additional layer of flexibility. In a financial context, although the price of an asset might exhibit trending behavior, the returns (percentage changes) of the price might be more stable. Then one would use an ARIMA model for the price but an ARMA model for the returns.

The ARIMA model is used to capture persistence and predictability in the conditional mean of a time-series process. An enduring topic in financial statistics is whether asset returns have any predictability at all; thus, test of an ARIMA structure against the null of no such structure takes on considerable importance.

Stata has a command to choose the lag order of an ARIMA model; see [TS] **arimasoc**. Stata also has a command to estimate the parameters of ARIMA models; see [TS] **arima**. For predictions and forecasting after ARIMA models, see [TS] **arima postestimation**.

Autoregressive conditional heteroskedasticity models

Stock returns show time-varying volatility: periods of high variability tend to cluster together, producing strings of periods of high volatility and strings of periods of low volatility. The ARMA model, however, cannot capture this feature of the data because, in an ARMA model, volatility is constant over time. The autoregressive conditional heteroskedasticity (ARCH) model extends the ARMA model by modeling time-varying dynamics in the conditional variance. Where an ARIMA model is designed to capture predictable movements in the mean of a time series (like the rate of return on an asset), an ARCH model is designed to capture predictable movements in the variance of the series (like the periods of high or low volatility in asset returns).

In an autoregressive model, the time-series process follows

$$
\begin{aligned}
y_t &= \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + e_t \\
e_t &\sim N(0, \sigma^2)
\end{aligned}
$$

where σ^2 is a constant. The ARCH model explains the variance process with autoregressive and moving-average terms:

$$
\begin{aligned}
e_t &\sim N(0, \sigma_t^2) \\
\sigma_t^2 &= \gamma_0 + \alpha_1 e_{t-1}^2 + \cdots + \beta_1 \sigma_{t-1}^2 + \cdots
\end{aligned}
$$

The three simplest models are the ARCH model of order 1,

$$\sigma_t^2 = \gamma_0 + \alpha_1 e_{t-1}^2$$

the generalized ARCH (GARCH) model of order 1,

$$\sigma_t^2 = \gamma_0 + \beta_1 \sigma_{t-1}^2$$

and the GARCH$(1, 1)$ model, which combines both,

$$\sigma_t^2 = \gamma_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

The ARCH terms are the analogue in the variance model to moving-average terms in an ARIMA model. The GARCH terms are the analogue in the variance model to the autoregressive terms in an ARIMA model. In an ARIMA model, a shock to e_t leads the mean of the series to change, with persistence in that change captured in the autoregressive and moving-average terms. Similarly, in a GARCH model, a shock to e_t^2 leads the variance of the series to change, with persistence in that change captured in the ARCH and GARCH terms. The result of a GARCH model is a series whose variance shows clustering of periods of high and low volatility.

The parameters estimated in a GARCH model are of direct interest in that they capture persistence in the conditional variance process. A fitted GARCH model can also be used to predict the conditional variance, both in sample and out of sample.

GARCH models were developed, in large part, to assist in characterizing financial time series. Engle (1982) introduced the ARCH model, and Bollerslev (1986) introduced the GARCH model. Since then, a large family of models has been developed to generalize ARCH models in various directions. The `arch` command estimates the parameters of many members of the ARCH family of models; see [TS] **arch**.

Vector autoregressive models

Where ARIMA models capture time dependence in a single variable, vector autoregressive (VAR) models jointly estimate dependence across a collection of variables. Returns to one asset can be affected by past returns to a different asset, reflecting cross-return dynamic dependence. Or returns to one asset can be affected by contemporaneous returns to other assets, reflecting contemporaneous or static dependence. VAR models allow the estimation of both kinds of dependence.

A VAR model has the structure

$$\mathbf{y}_t = \mathbf{A}_1\mathbf{y}_{t-1} + \cdots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{u}_t$$
$$E(\mathbf{u}_t\mathbf{u}_t') = \mathbf{\Sigma}$$

In this model, the parameters in the $(\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_p)$ matrices capture dynamic dependence across variables (returns). The residual for return i is u_i; cross-sectional dependence is captured in the off-diagonal elements of the covariance matrix of the residuals, $\mathbf{\Sigma}$. For example, if multiple assets are in the same economic sector, then sector-specific news is likely to cause correlated movements across all the asset returns in that sector.

The `var` command estimates parameters of a VAR model; see [TS] **var**.

Vector error-correction models

Vector error-correction (VEC) models add additional structure to the dynamics of variables in a VAR model. As in an ARMA model, a VAR model assumes the time series under consideration are stationary, meaning they do not show trends. The VEC model extends the VAR model to allow for a common, joint trend among the series. In macroeconomics and finance, such models arise naturally for variables that have trending behavior but whose ratio is stationary. In macroeconomics, gross domestic product and its components have trends, but the share of each component in gross domestic product can be modeled as stationary. In finance, asset prices and dividends each have trends, but the price-dividend ratio can be modeled as stationary.

In such situations, a VEC model arises. For two variables, say, log prices p_t and log dividends d_t, the VEC model takes the form

$$\Delta p_t = \alpha_1(d_{t-1} + \beta p_{t-1}) + a_{11}\Delta p_{t-1} + a_{12}\Delta d_{t-1} + u_{p,t}$$
$$\Delta d_t = \alpha_2(d_{t-1} + \beta p_{t-1}) + a_{21}\Delta p_{t-1} + a_{22}\Delta d_{t-1} + u_{d,t}$$

In this two-variable example, the variables are expressed in growth rates (log differences). The parameters $(a_{11}, a_{12}, a_{21}, a_{22})$ are typical VAR coefficients. New is the presence of the lagged linear combination of the levels of the variables, $d_{t-1} + \beta p_{t-1}$. This is the cointegrating vector, the relationship around which the variables fluctuate. The parameter β characterizes the cointegrating relationship. Easiest to interpret is $\beta = -1$, in which case the cointegrating vector is $d_t - p_t$, so that the growth rates of prices and dividends react to the lagged price-dividend ratio. The reaction coefficients are captured by (α_1, α_2), which can differ across equations. These parameters are interpretable: if $\alpha_1 > 0$, for instance, then an unusually high dividend-price ratio forecasts higher future price growth. The magnitude of the parameters (α_1, α_2) captures the speed of convergence back to the cointegrating trend.

The vec command estimates the parameters of VEC models; see [TS] **vec**. Campbell and Shiller (1987) use similar techniques to study two financial phenomena: the price-dividend relationship and the short-long interest-rate spread. VEC models can also be employed with more than two time series, for example, in characterizing exchange rate dynamics across a collection of countries.

Multivariate GARCH models

Multivariate GARCH (MGARCH) models capture time-varying heteroskedasticity in a multivariate setting. As in a VAR model, multiple variables are modeled jointly. As in a GARCH model, both the conditional mean and conditional variance are allowed to vary over time.

A general form of the MGARCH model is

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{u}_t$$
$$\mathbf{u}_t = \mathbf{H}_t^{1/2}\boldsymbol{\nu}_t$$

In the first equation, $\mathbf{y}_t$ is a collection of variables to be modeled, $\mathbf{x}_t$ represents a collection of independent variables that can include lags of $\mathbf{y}_t$, and $\mathbf{u}_t$ are disturbance terms. In the second equation, the disturbances are broken into two components: a time-varying component $\mathbf{H}_t$ and a collection of zero-mean, unit variance, independently and identically distributed shocks $\boldsymbol{\nu}_t$. The GARCH dynamics are embedded in the $\mathbf{H}_t$ matrix. The multivariate analogue of the GARCH$(1,1)$ model discussed previously is

$$\text{vech}(\mathbf{H}_t) = \mathbf{s} + \mathbf{A}\,\text{vech}(\mathbf{u}_{t-1}\mathbf{u}'_{t-1}) + \mathbf{B}\,\text{vech}(\mathbf{H}_{t-1})$$

MGARCH models are not identified in their full generality. Different submodels have been proposed that make different choices for the tradeoff between model complexity and parsimony.

The diagonal vech model fits the diagonal elements of $(\mathbf{A}, \mathbf{B})$ above, allowing every element of $\mathbf{H}_t$ to follow a univariate GARCH process. The family of conditional correlation MGARCH models fits a parsimonious model for the conditional correlations.

In all cases, the MGARCH model allows for comovements in volatility that cannot be captured by univariate GARCH models, because the VAR model allows for comovements in the conditional mean that cannot be captured by univariate ARIMA models. The mgarch command fits four types of MGARCH models; see [TS] **mgarch** for details.

This entry has described a collection of tools used by statisticians and econometricians to analyze financial time series. The next entry describes a workflow in Stata for analyzing such data, using tools from the fin suite of commands as well as Stata's general time-series commands.

References

Black, F. 1972. Capital market equilibrium with restricted borrowing. *Journal of Business* 45: 444–455. https://doi.org/10.1086/295472.

Boffelli, S., and G. Urga. 2016. *Financial Econometrics Using Stata*. College Station, TX: Stata Press.

Bollerslev, T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31: 307–327. https://doi.org/10.1016/0304-4076(86)90063-1.

Campbell, J. Y. 2018. *Financial Decisions and Markets: A Course in Asset Pricing*. Princeton, NJ: Princeton University Press.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay. 1997. *The Econometrics of Financial Markets*. Princeton, NJ: Princeton University Press. https://doi.org/10.2307/j.ctt7skm5.

Campbell, J. Y., and R. J. Shiller. 1987. Cointegration and tests of present value models. *Journal of Political Economy* 95: 1062–1088. https://doi.org/10.1086/261502.

Cochrane, J. H. 2005. *Asset Pricing*. Rev. ed. Princeton, NJ: Princeton University Press.

Engle, R. F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50: 987–1007. https://doi.org/10.2307/1912773.

Fama, E. F., and J. D. MacBeth. 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81: 607–636. https://doi.org/10.1086/260061.

Hurn, S., V. L. Martin, P. C. B. Phillips, and J. Yu. 2020. *Financial Econometric Modeling*. Oxford: Oxford University Press.

Jensen, M. C. 1968. The performance of mutual funds in the period 1945–1964. *Journal of Finance* 23: 389–416. https://doi.org/10.1111/j.1540-6261.1968.tb00815.x.

Lintner, J. 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics* 47: 13–37. https://doi.org/10.2307/1924119.

Shanken, J. 1992. On the estimation of beta-pricing models. *Review of Financial Studies* 5: 1–33. https://doi.org/10.1093/rfs/5.1.1.

Sharpe, W. F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19: 425–442. https://doi.org/10.1111/j.1540-6261.1964.tb02865.x.

Also see

[FIN] **fin** — Commands for analysis of financial data+

Stata, Stata Press, Mata, NetCourse, and NetCourseNow are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. StataNow is a trademark of StataCorp LLC. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2025 StataCorp LLC, College Station, TX, USA. All rights reserved.

For suggested citations, see the FAQ on citing Stata documentation.