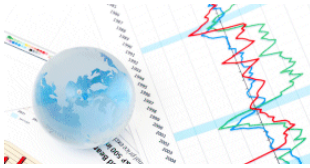


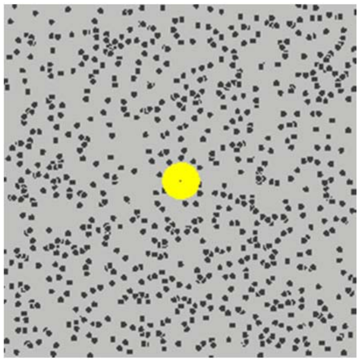
Lecture notes on forecasting

Robert Nau
Fuqua School of Business
Duke University
http://people.duke.edu/~rnau/forecasting.htm

The random walk model

Some history

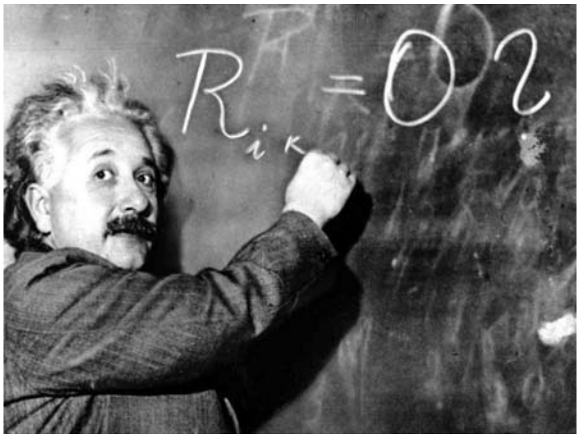
"Brownian motion" is a **random walk** in continuous time**.** It is named after the Scottish botanist **Robert Brown**, who observed in 1827 that microscopic particles in water undergo spontaneous random movements, which (we now know) are due to statistical fluctuations in collisions with the surrounding molecules, which are in constant motion due to heat.



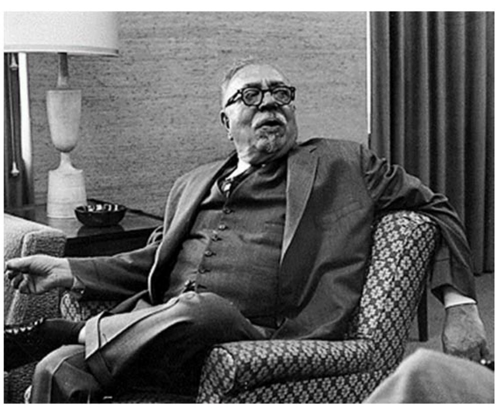
http://en.wikipedia.org/wiki/Brownian_motion

Louis Jean-Baptiste Alphonse Bachelier, a French mathematician, was the first person to model Brownian motion, as part of his PhD thesis, *The Theory of Speculation*, published 1900.

His thesis, which discussed the use of Brownian motion to evaluate stock options, is historically the first paper to use advanced mathematics in the study of finance.



Five years later, **Albert Einstein** (1905) independently discovered the same stochastic process and applied it in thermodynamics.

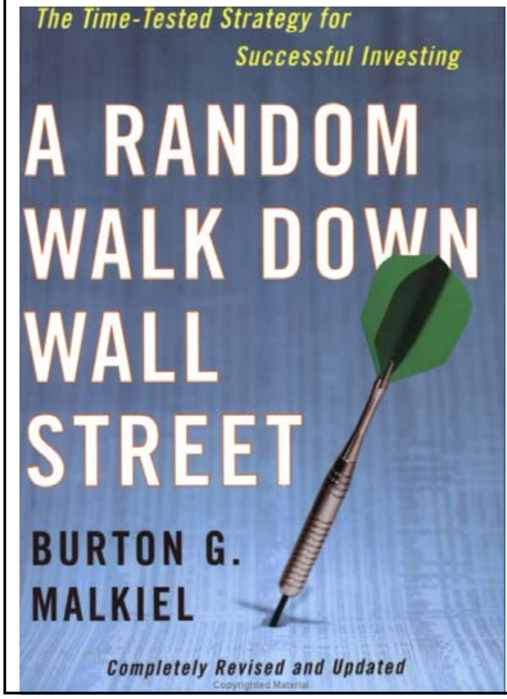


The mathematical properties of a continuous-time random walk were later worked out by **Norbert Weiner**, hence such a process is known as a **Wiener process**,

Bachelier had derived the price of an option where the share price movement is modelled by a Wiener process and derived the price of what is now called a barrier option (namely the option which depends on whether the share price crosses a barrier). **Black and Scholes***, following the ideas of Osborne and Samuelson, modelled the share price as a stochastic process known as a **geometric Brownian motion (with drift).**

In **geometric** Brownian motion, the **natural log** of the variable is a continuous-time random walk with normally distributed steps**.**

*Fischer Black and Myron Scholes, together with Robert Merton, received the Nobel Memorial Prize in economics in 1997, a year before the spectacular collapse of the hedge fund Long Term Capital Management, on whose board both Scholes and Merton sat.



"The market is not a perfect random walk. But any systematic relationships that exist are so small that they are not useful for an investor…

The history of stock price movements contains no useful information that will enable an investor to outperform a buy-and-hold strategy in managing a portfolio."

(p. 151)

Don't take financial advice from babies!

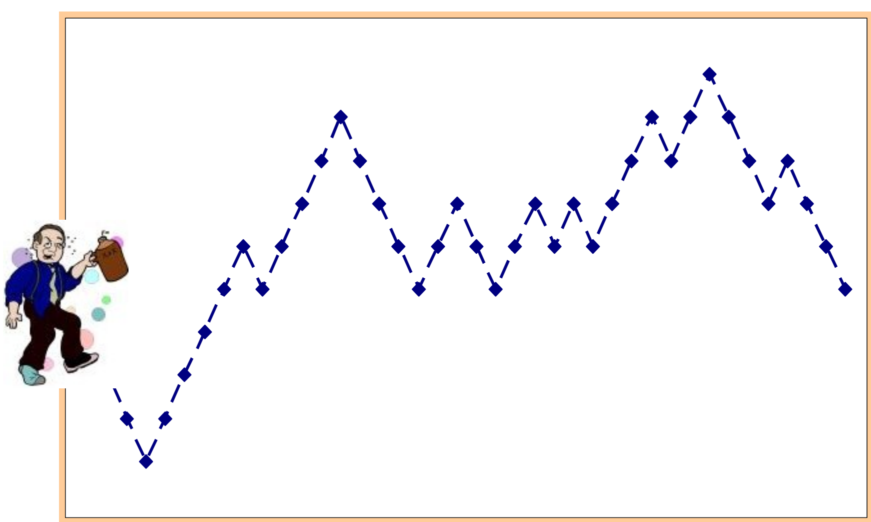
The random walk model

- A time series is a *random walk* if its period-to-period *changes* are statistically independent & identically distributed ("i.i.d.")
- In each period it takes an independent random "step" away from its last position
- If the *mean* step size is non-zero, it is a random walk "with drift" (i.e., trend)

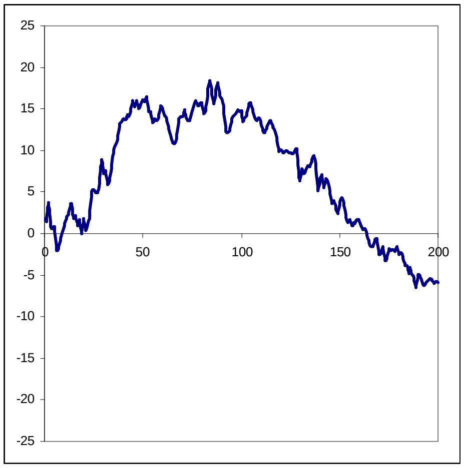
Analogies

- Random walk: a drunk person staggering left and right with equal probability while moving forward (this was Bachelier's own example)
- Random walk with drift: a drunk person with one shoe
- Continuous random walk ("Brownian motion"): a drunk snail

Walking the "drunkard's walk"



A random walk with little or no drift often does not look "random"! It may appear to have trends, cycles, "head and shoulders" patterns and other interesting features by *sheer chance*.



"It Ain't the Things You Don't Know That Hurt You, It's the Things You Know... That Ain't So!"

A related statistical illusion: the "hot hand in basketball"

- Many basketball players are perceived as "streaky" shooters (e.g., Allen Iverson), but statistical analysis shows that the chance of making a given field goal or free throw is roughly independent of what happened on the last few shots:

  "Chance is a very powerful force in creating streaks"

- See the Hot Hand in Sports website at

  http://thehothand.blogspot.com/

How to identify a random walk

- Plot the *first difference,* i.e., the period-to-period changes, of the original time series (Y_DIFF1 or DIFF(Y))
- If the first difference has *constant variance* and *no significant autocorrelations*, the original series is a random walk (at least locally).
- If the series is logically bounded above or below or has a stable long-run average, then it is not a “true” random walk, but the random walk model may still be appropriate for short-term forecasting or as a benchmark for comparing fancier models.

Relation to mean model

- If a time series is a random walk, then its *first difference* is described by the *mean model*.
- Thus, you should predict that the next *change* will equal the *average* change.
- The average change may or may not be zero, depending on whether there is “drift”.

Random walk forecast equation

- In the random walk *without drift* model, the *k*-step ahead forecast is the last observed value:

$$\hat{Y}_{t+k} = Y_t$$

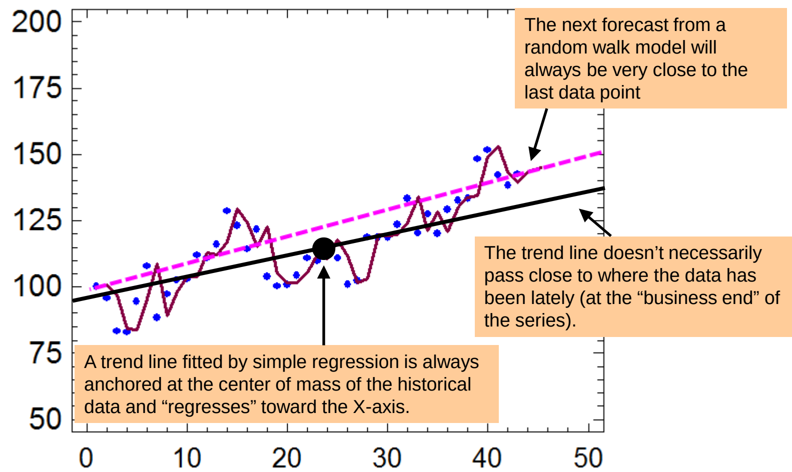
- In the random walk *with drift* model, the *k*-step ahead forecast is the last observed value plus *k* times the estimated drift (trend) per period:

$$\hat{Y}_{t+k} = Y_t + kd$$

Relation to linear trend model

- A linear trend model assumes that there is a trend line fixed "somewhere in space" around which the data varies in an i.i.d. manner.
- The *fitted* trend line always passes through the *center of mass* of the data and it regresses toward the X-axis
- A random-walk-with-drift model also assumes that there is a constant trend, but it continually "re-anchors" the trend line on the last observed data point.
- Confidence intervals for the RW model widen "parabolically" as the forecast horizon increases, but confidence intervals for the linear trend model hardly widen at all. Which is more realistic for your data set?

Linear trend vs. random walk with drift



Linear trend vs. random walk with drift

- If the growth pattern in the data is irregular or not perfectly linear, the linear trend model may fit badly near the *end* of the series—which is where the forecasting action occurs!
- Because the linear trend model anchors the trend line in the exact center of the data, its goodness of fit near the end of the series is very sensitive to the amount of history used.
- Plot of historical RWD forecasts looks like an exact copy of the data, but shifted right by 1 period (& perhaps moved up)

Autocorrelation & the random walk model

- In a true random walk, there is *strong* autocorrelation in the original series, but *no* significant autocorrelation in the *first difference* of the series at *any* lags.
- This means there is no pattern in the data except "the change next period will equal the average change"
- Hence the random walk model is sometimes called the "naïve" model…
- ..but it's not really naïve! It says you *can't do better than this*, and it has profound and financially important implications for widths of confidence intervals for forecasts at horizons of more than one period ahead.

Geometric random walk

- If the *log* of a series is a random walk, the original series is a *geometric random walk.*
- Recall that the change in the natural log is (approximately) the *percentage change* between periods :

$$\mathrm{LN}(Y_t) - \mathrm{LN}(Y_{t-1}) = \mathrm{LN}(Y_t / Y_{t-1})$$

$$\approx (Y_t / Y_{t-1}) - 1 = (Y_t - Y_{t-1}) / Y_{t-1} \quad !$$

- Hence, in a geometric random walk, the series takes random steps in (roughly) *percentage* terms

Percentage change is a more familiar and easy-to-think-about concept, but *change-in-the-natural-log* is theoretically the right way to measure relative changes when exponential growth is occurring or when small changes are compounded over many periods.

Diff-log vs. percent change

% change in X	diff-log of X		diff-log of X	% change in X
-90%	-2.303		-0.700	50.3%
-70%	-1.204		-0.600	45.1%
-50%	-0.693		-0.500	39.3%
-40%	-0.511		-0.400	33.0%
-30%	-0.357		-0.300	25.9%
-20%	-0.223		-0.200	18.1%
-10%	-0.105		-0.100	9.5%
-5%	**-0.051**		**-0.050**	**4.9%**
-2%	**-0.020**		**-0.020**	**2.0%**
0%	**0.000**		**0.000**	**0.0%**
2%	**0.020**		**0.020**	**-2.0%**
5%	**0.049**		**0.050**	**-5.1%**
10%	0.095		0.100	-10.5%
20%	0.182		0.200	-22.1%
30%	0.262		0.300	-35.0%
40%	0.336		0.400	-49.2%
50%	0.405		0.500	-64.9%
90%	0.642		0.600	-82.2%
100%	0.693		0.700	-101.4%

Almost the same up to around ± 10%, and pretty close up to ± 20%,

Diff-log of X is X_LN_DIFF1 in RegressIt and DIFF(LOG(X)) in Statgraphics

% change of X from period t to period t+1 is $(X_t - X_{t-1})/X_{t-1}$

Diff-log of X from period t to period t+1 is $\mathrm{LN}(X_t) - \mathrm{LN}(X_{t-1})$

Geometric random walk forecast equation

- In log units:

$$LN(\hat{Y}_{t+k}) = LN(Y_t) + k\ LN(1+r)$$
$$\approx LN(Y_t) + kr$$

…i.e., $\mathrm{LN}(Y)$ is predicted to grow *linearly* with trend *r*

- In unlogged (original) units:

$$\hat{Y}_{t+k} = Y_t(1+r)^k$$

…i.e., Y is predicted to *compound* at rate *r*

Best example: stock prices

- The geometric random walk is the “default” model for stock prices and many other financial assets for which speculative markets exist
- First proposed by Bachelier in 1900, it became the basis for the Black-Scholes options pricing model 70 years later.
- This means it is hard to beat the market by technical analysis (“charting”)...
- … or by fitting regression models to monthly, weekly, or even daily data.

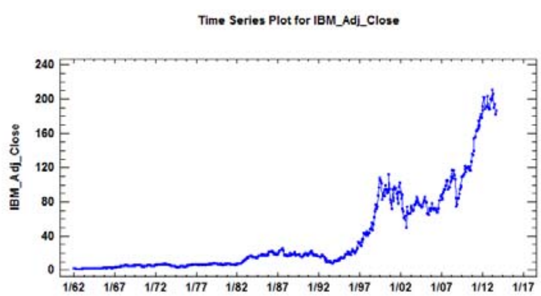
Why should stock prices behave like a geometric random walk?

- If everyone could predict that the stock market will go up more than average tomorrow, *it would have already gone up today*, hence future returns are independent of past returns (and other public information) *in an efficient market.*
- Investors generally think in terms of *percentage changes* in stock values when responding to informational events (earnings announcements, interest hikes, etc.), hence volatility is fairly constant in percentage terms.
- Remarkably, the average volatility has been consistent in percentage terms over the last century.

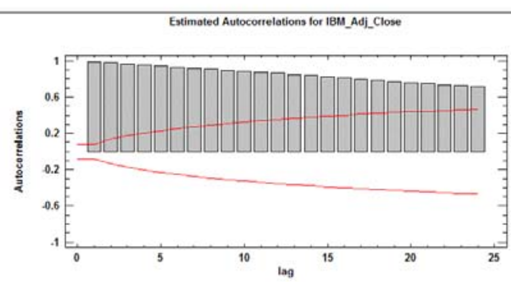
Forecasting from the GRW model

- First apply the standard RW model to the logged series to obtain forecasts and confidence intervals in logged units…
- Then "unlog" them (apply the "EXP" function) to obtain forecasts and confidence intervals in original units.
- The forecasting procedure in Statgraphics does all this automatically when you specify a random walk model in conjunction with with a log transformation.

Example

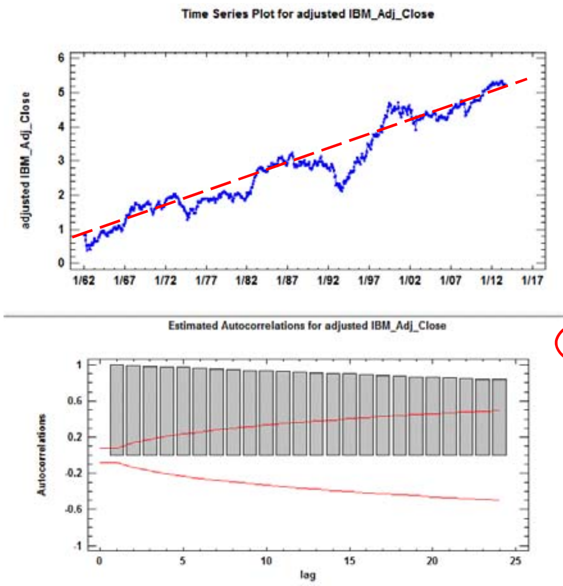


Original series shows erratic behavior, strong positive autocorrelation, peaks and valleys, exponential growth



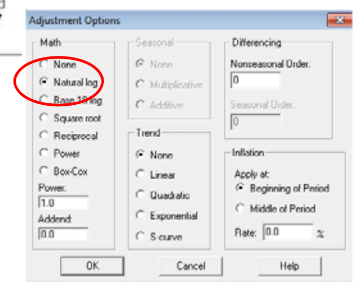
The heights of the bars are the autocorrelations. Here the autocorrelations are all close to 1 because the series tends to remain on the same side of its sample mean for long periods.

Example

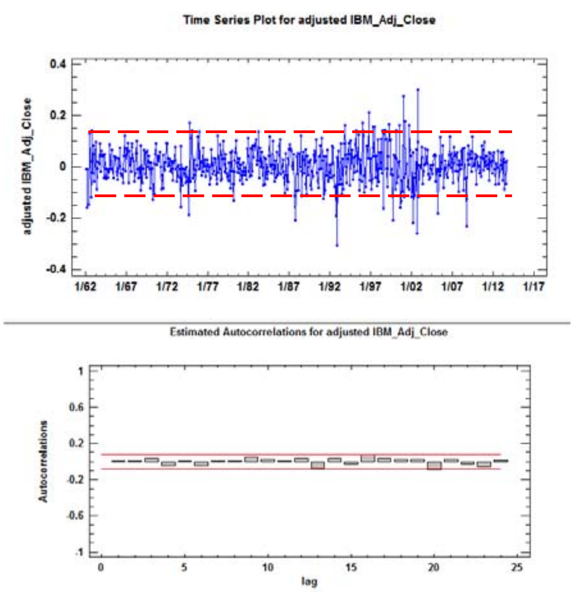


A natural log transformation linearizes the long-run growth.

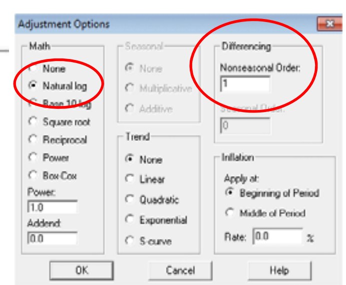
Slope of line drawn between 1st and last points is the average percent growth.



Example



After a diff-log transformation, the variance is roughly constant (over the long run, at least) and autocorrelations are all insignificant, the signature of a geometric random walk



Forecasting from the random walk model

- The forecasts from the random walk model are extrapolated as *a straight line extending from the last observed data point*.
- In a random walk *without* drift, the line is *horizontal*.
- In a random walk *with* drift, it has a *non-zero slope.*
- *Drift estimation is hard* unless the sample size is huge and/or you have a theoretical basis for determining it from other data (as in the CAPM)
- If the drift is estimated from the sample, the forecasts are the extrapolation of a line through the 1st and last points!

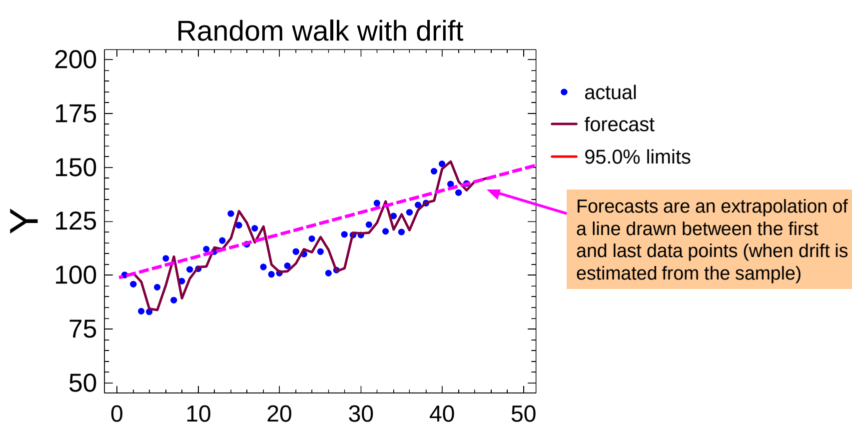
Standard errors of forecasts

- The standard error of a one-step-ahead RW forecast is *the sample standard deviation of the differenced series multiplied by SQRT(1+1/(n-1)) where n is the sample size.*
  - Same as the standard error of the forecast when applying the mean model to the differenced series, whose length is *n*-1 rather than *n*
- The standard error of a *k*-step-ahead RW forecast is equal to the standard error of a one-step-ahead RW forecast *multiplied by a growth factor of SQRT(k)*
  - Why….?

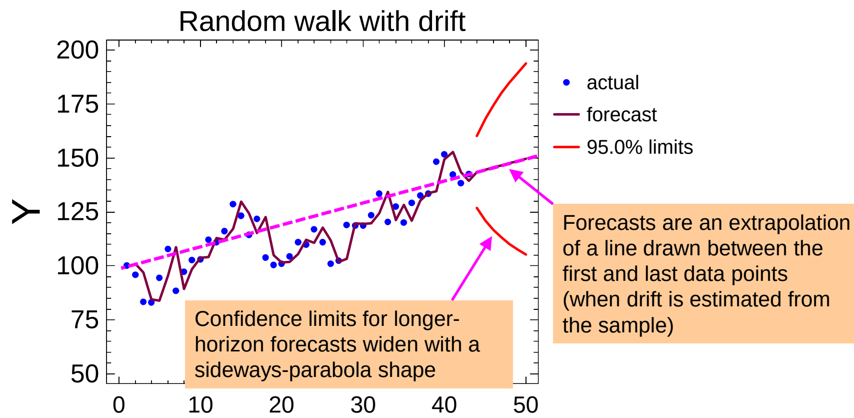
The square-root-of-time rule

- The one-step-ahead RW point forecast and its standard error are almost trivial to understand and to compute.
- What's interesting and profound is what the RW model implies for *standard errors of longer-horizon forecasts.*
- Because the steps are statistically independent, the *variance* of the sum of k steps is k times the variance of one step.
- The standard error of the k-step ahead forecast is the *standard deviation* of the sum of k steps, which is the square-root-of-k times the standard deviation of one step.
- Hence confidence intervals widen in proportion to the *square root of time* ("sideways parabola" shape).
- *This is the basis of options pricing models in finance.*

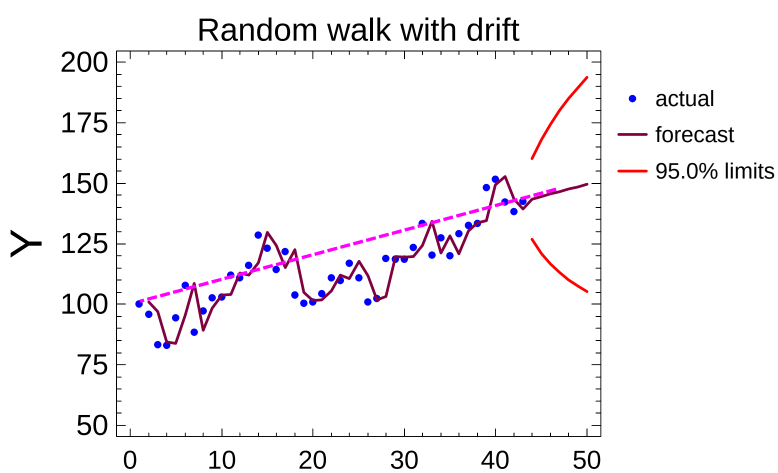
Forecasts and confidence limits for random-walk-with drift model



Forecasts and confidence limits for random-walk-with drift model

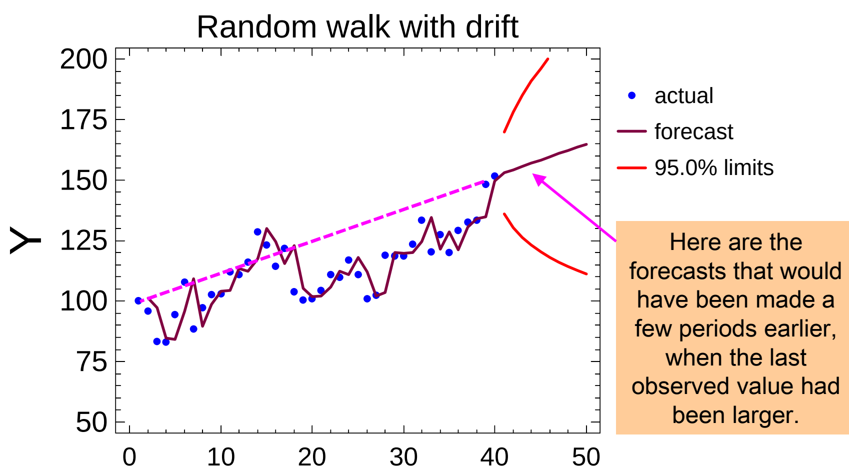


Updating of random walk forecasts



Forecasts into the future are a trend line re-anchored on the last observed data point. Past forecasts look like a plot of the data shifted to the right and slightly up.

Updating of random walk forecasts



Forecasts into the future are a trend line re-anchored on the last observed data point. Past forecasts look like a plot of the data shifted to the right and slightly up.

How to tell if drift (trend) is non-zero?

- The drift term often does not matter much to 1-step-ahead forecasts, but adds a possibly-important trend to longer-horizon forecasts.
- Ask whether it makes sense that the series should trend upward or downward indefinitely: if not, then assume no drift.
- It's difficult to test the hypothesis of zero drift by statistical methods (e.g., t-stat of the sample mean of diff-Y) unless the sample is very large, because finite samples of random walks often exhibit spurious trends ("hot hands") even with no drift.

How to estimate the drift?

- Usually it is NOT best to estimate drift from the average increase over the sample (i.e., the slope of a line between 1st and last data points) *unless* you have a very long history.
- You may have to rely on other assumptions and other information.
- In financial markets, the drift of an asset price (in % terms) theoretically ought to be determined by the *correlation of its returns with those of the market index* (which determines its “beta”), according to the Capital Asset Pricing Model.

Looking ahead

- Some of the more sophisticated forecasting models we will meet later (e.g., simple and linear exponential smoothing) are just fancied-up random walk models.
- The forecast line is extrapolated from the average position of *the last few points*.
- The trend in the forecasts is equal to the average trend of the *recent data*, not the whole sample.