NOVANTAS
PERSPECTIVE

Forecasting Percentage Growth: Do Your Dependent Variable Transformations for PPNR Stress Testing Make Sense?

WES WEST, SEAN WALLACE AND RYAN SCHULZ

We welcome your feedback and are happy to continue the conversation about this article or other Treasury and Risk viewpoints.
Pete Gilchrist at pgilchrist@novantas.com or Jonathan "Wes" West at jwest@novantas.com.

In PPNR Stress Testing, one of the most common transformations of the dependent variable (DV) is to predict growth in balances, revenue, and expenses from period to period. Implementing this correctly, however, comes with a catch: the preferred transformation is a logarithmic one, i.e., the difference in each period's log-transformed balance ("diff-log" or $\Delta\ln$), as opposed to percentage change, i.e., balances today divided by balances yesterday ("percent change" or $\Delta\%$). In this Perspective we outline why non-statisticians and statisticians alike should prefer a $\Delta\ln$ over a $\Delta\%$ transformation for balance variables.

PPNR Stress Testing modelers know that nearly all PPNR balance, revenue and expense series ("balances") are non-stationary, which generates problems in model development if not resolved[1]. For this discussion, a working definition of a non-stationary time series is that yesterday's value has a strong influence on today's value[2]. Mathematically, the equation is:

$$y_t = y_{t-1} + BX_t + u_t$$

Where y_t is the current value of the series being predicted, y_{t-1} is the predictor's value in the prior period, BX is the cumulative influence of the intercept and exogenous factors, and u_t is the error term (which may or may not be serially correlated).

When a series exhibits unit root/non-stationarity and therefore "yesterday is a good predictor of today", the influence of "yesterday" frequently overpowers any exogenous driver, defaulting to the non-useful outcome of:

$$y_t = y_{t-1} + u_t$$

This equation cannot inform how a bank's balance sheet would change under different macroeconomic conditions[3].

Transforming the time series to make it stationary requires elimination of the "yesterday" term from the right-hand side of the

1 *Rates also frequently exhibit unit root but should be treated differently for a variety of reasons discussed in greater depth in an earlier Perspective*

2 *More precisely: "A stationary time series is one whose statistical properties such as mean, variance, autocorrelation, etc. are all constant over time." From http://people.duke.edu/~rnau/411diff.htm*

3 *There is an exception when a cointegrating relationship is identified. This is a less common — but not impossible — PPNR modeling approach which we will address more completely in our next Perspective*

equation, which permits the predictive power of the exogenous variables to drive the regression. One approach is to difference the series, subtracting y_{t-1}, which translates the model from predicting the total balance to predicting the dollar change in balance.

However, in many cases when predicting balances — especially where portfolios have big changes between the maximum and minimum observed values — predicting growth will be more informative and more accurate than predicting the dollar change (and may better reflect what modelers believe the underlying model specification to be).

We often see model development teams come to this conclusion and transform the DV into a percent change:

$$PctChgDV_t = \Delta\%DV_t = \frac{DV_t}{DV_{t-1}} - 1$$

While this is often the *a priori* intuitive growth transformation, this fails to remove the y_{t-1} term from the right side of the equation, and therefore does not eliminate the unit root (the supporting math is in a later section). The more appropriate transformation is to take the natural log of the balance variable and then take the difference of the current and prior period natural log (which is the same as log-transforming the ratio) as follows:

$$DiffLogDV_t = \Delta\ln(DV_t) = \ln(DV_t) - \ln(DV_{t-1}) = \ln\left(\frac{DV_t}{DV_{t-1}}\right)$$

From a modeling perspective, the $\Delta\ln$ transformation has several properties that make it more useful than Δ%. From a validation perspective, the $\Delta\ln$ transformation adheres to key underlying principles of time series modeling that Δ% models do not.

CONCEPTUAL RATIONALE FOR ΔLN

Similar Ease of Interpretation: A positive characteristic of Δ% modeling is ease of interpreting the results. For example, assume we have the following model:

$$\Delta\%Bal_t = 0.01 + 1.5 * \Delta\%GDP_t + \varepsilon_t$$

The interpretation is transparent: the intercept (0.01) implies that — all else being equal — the DV will grow by 1% per period and the GDP coefficient (1.5) means that if GDP grew by 2% then the DV would grow by an incremental 1.5 x 2% = 3%.
A natural log transformation, while less familiar to some, will have the same interpretation. In this example, a $\Delta\ln$ transformation would yield an equation that looks similar:

$$\Delta\ln(Bal_t) = 0.01 + 1.5 * \Delta\ln(GDP_t) + \varepsilon_t$$

Importantly, the interpretation is *effectively identical*[4], with the 0.01 intercept translating to baseline 1% growth and the 1.5 GDP coefficient translating to approximately 1.5 x GDP's percentage growth. So ease of interpretation is not a bar to use of $\Delta\ln$ models.

Benefits of $\Delta\ln$ Models. There are practical benefits that make the $\Delta\ln$ transformation a better representation of growth than Δ% in time series regression models. The strongest benefit for $\Delta\ln$ transformations is that balance forecasts cannot go negative[5].

Using the same example model above without any truncation criteria, if GDP shrinks by 70% (an unlikely — and terrifying — scenario), then the forecast for bank balance growth is 1% + (1.5 x –70%) = –104%! Most validation units react negatively

4 $\ln(1+r) \approx r$ *when the growth rate is between —10% and +10%, and values are still similar when period growth rates range between —50% and +50%. For examples, see http://people.duke.edu/~rnau/411log.htm.*

5 *As noted earlier, modeling rates requires adjustment to the modeling techniques used for balances, revenues and expenses, in part because rates can go negative*

(pun intended) to any functional form that can produce these illogical results, regardless of how unlikely the generating scenario is. Using the $\Delta\ln$ equation keeps the predicted balance positive, with a more logical result that balances shrink by 83%.

A second benefit of $\Delta\ln$ models is symmetry of results. For example, assume a $100 portfolio grows $20 in one period and immediately shrinks back to historic levels. The $\Delta\ln$ is symmetric with values of +0.18 and –0.18, while the percent change values are asymmetric with +20% growth and –17% shrinkage.

	PERIOD 0	PERIOD 1	PERIOD 2
Balance	$100	$120	$100
Δ%		20%	–17%
$\Delta\ln$		0.18	–0.18

While again $\Delta\ln$ results do not quite equate to percentage changes, the positive/negative symmetry makes the model development and forecasting process easier, and in other ways simplifies the interpretation of results. For example, in niche or large corporate portfolios, single client movements can double portfolio balances in one period (e.g., prior to a large M&A deal), only to return to normal the following period. Using Δ% models require two separate indicator variables, one for the 100% increase, and another for the 50% decrease (both variables are 1 in the appropriate period, but are assigned different coefficients). Using $\Delta\ln$, a single indicator variable with values of 1 and –1 can represent both the increase and decrease with the same coefficient estimate.

The positive/negative symmetry also results in more accurate forecasts when balance change and reversion happen over a longer timeframe. For example, if a bank expects CD balances to return to approximately historic levels if rates similarly revert (as we do), the modeling team will face an awkward situation if the DV is Δ%. A $\Delta\ln$ forecast of –0.80, +0.80 will end at the same value as the starting point (akin to a CD portfolio that shrank after the crisis but will grow when rates rise). Conversely, a Δ% forecast of –80%, +80% will end up at 36% of starting balances!

DIGGING INTO THE MATH: ΔLN WORKS AND Δ% DOES NOT

As discussed above, most series being modeled in PPNR Stress Testing can be represented as a standard textbook example of a "Random Walk" model, with the DV driven by the prior period DV and a set of exogenous drivers[6]. Using this model specification, the key goal of reordering terms — whether from a differences, $\Delta\ln$, or a Δ% model — is to eliminate the y_{t-1} term from the right-hand side. The resulting transformed DV will (usually) be stationary and thus suitable for time series modeling.

EQUATION	COMMENTARY
$y_t = y_{t-1} + BX_t + \varepsilon_t$	BX_t is function of exogenous factors
$y_t - y_{t-1} = y_{t-1} - y_{t-1} + BX_t + \varepsilon_t$	Subtract y_{t-1} from both sides
$\Delta y_t = BX_t + \varepsilon_t$	y_{t-1} term eliminated from right-hand side. Model proceeds where DV is change in y

6 *Extrapolated from "Time Series: Theory and Methods" 2nd Edition, Peter J. Brockwell and Richard A. Davis. Assumes y_t is $I(1)$. The core insight is applicable to any model with autoregressive error terms, e.g., an AR(1) model, by replacing ε_t with u_t where $u_t = \Theta u_{t-1} + \varepsilon_t$*

This objective of eliminating the y_{t-1} term from the right-hand side is possible when the growth rate is represented as the $\Delta\ln$:

EQUATION	COMMENTARY
$\ln(y_t) = Y_t$	By definition
$\Delta Y_t = \ln(y_t) - \ln(y_{t-1}) = \ln\left(\frac{y_t}{y_{t-1}}\right)$	Identity property of logarithms
$Y_t = Y_{t-1} + BX_t + \varepsilon_t$	As defined in the above example, BX_t is a function of exogenous factors and ε_t is the error term
$Y_t - Y_{t-1} = \Delta Y_t = \ln\left(\frac{y_t}{y_{t-1}}\right) = BX_t + \varepsilon_t$	As in the first example, the Y_{t-1} term eliminated from right-hand side

However, when performing a Δ% DV transformation of this specification of the DV relationship, it is not possible to eliminate the y_{t-1} term from the right-hand side: it is impossible to make a Δ% transformation stationary.

EQUATION	COMMENTARY
$y_t = y_{t-1} + BX_t + \varepsilon_t$	BX_t is function of exogenous factors
$\frac{y_t}{y_{t-1}} - 1 = \frac{y_{t-1}+BX_t+\varepsilon_t}{y_{t-1}} - 1 = \frac{BX_t+\varepsilon_t}{y_{t-1}}$	Calculate percentage change by dividing by y_{t-1} and subtracting 1
$\Delta\% y_t = \frac{BX_t + \varepsilon_t}{y_{t-1}}$	Note failure of meeting objective to eliminate y_{t-1} term from the right-hand side

ABOUT NOVANTAS **Novantas is the industry leader in analytic advisory and solution services for financial institutions.** Our Global Treasury & Risk unit partners with banks to advance their analytic capabilities — bringing to bear our thought leadership, advanced modeling techniques, and extensive experience. The Novantas PPNR Modeling and Forecasting team has worked with more than a third of CCAR banks on PPNR modeling engagements, and is routinely in contact with almost all CCAR banks, many DFAST banks, major international banks in 10+ countries, and U.S. and international regulators.

CONTACT US We welcome your feedback and are happy to continue the conversation about this article or other Treasury and Risk viewpoints. Please reach out to the head of Novantas Global Treasury and Risk, Pete Gilchrist at pgilchrist@novantas.com; or the head of Novantas PPNR Modeling and Forecasting, Jonathan "Wes" West at jwest@novantas.com.