# Developing robust PPNR estimates

**Risk** September 2015

**Authored by:**
Abhishek Anand
Jeff Schonert

# Contents

# Developing robust PPNR estimates

Supervisory expectations for analytical rigor have increased over the successive annual cycles of the Comprehensive Capital Analysis and Review (CCAR). Beginning with the predominantly qualitative approach used for the Supervisory Capital Assessment Program in 2009, both bank holding companies (BHCs) and the Federal Reserve have been moving to more analytically based approaches, with a particular recent emphasis on pre-provision net revenue (PPNR). In fact, CCAR 2014 was the first time the Federal Reserve used detailed statistical models to evaluate PPNR submissions. These models were further updated and refined for CCAR 2015.

With the rising expectations around analytics, BHCs have struggled to keep pace with regulatory standards. While this has been a challenge for both loss forecasting and PPNR, loss estimation analytics have been less challenging thanks to the existence of solid analytical capabilities prior to CCAR. However, because of a lack of both analytical infrastructure and capability, PPNR has been much more difficult. BHCs have traditionally forecast revenue through a bottom-up, expert judgment approach, as opposed to the top-down, statistical model approach consistent with CCAR. This practice partly accounts for much of the negative feedback many BHCs have received to their capital plans on qualitative grounds.

Moreover, since publication of SR 11-7[1], the quality bar for analytical approaches has been raised significantly. Where in the past any quantitative model was considered a big step in the right direction, PPNR models are now reviewed carefully and judged critically across multiple dimensions, including conceptual soundness, statistical rigor, and comprehensiveness of documentation. BHCs have struggled to strike the right balance in model governance, resulting too often in unreasonable expectations at one end of the spectrum or lax controls at the other.
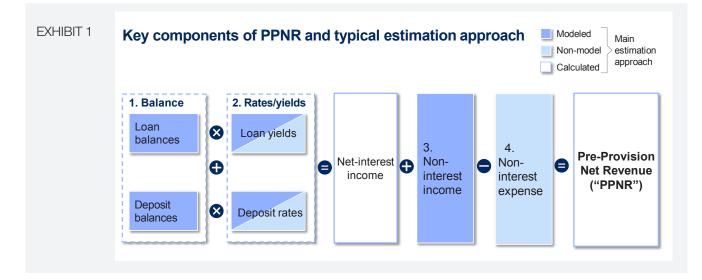
Based on assisting a number of BHCs in navigating this PPNR journey, we believe BHCs need to focus on four key dimensions:

- Understanding key PPNR challenges
- Building strong PPNR models
- Using management overlays
- Employing "non-model models"

[1] Refers to Supervisory Letter SR 11-7 on guidance on Model Risk, here: http://www.federalreserve.gov/bankinforeg/srletters/sr1107.htm

# 1. Understanding key PPNR modeling challenges

PPNR estimation typically entails estimation of four key components, as shown in Exhibit 1 below:

EXHIBIT 1



**Key components of PPNR and typical estimation approach**

Typically BHCs have struggled in developing statistically robust PPNR models in four major areas:

- Segmentation
- Lack of historical data
- Conceptual soundness
- Statistical robustness

## 1.1 SEGMENTATION

Determining appropriate segmentation for PPNR models has been a challenge for most BHCs. Issues have been seen at both ends of the spectrum, with BHCs developing models that are either too high level or too granular. An excessively high or low level of granularity makes statistical modeling inaccurate because of "confounding"[2] of macro relationships or the introduction of significant noise.

## 1.2 LACK OF HISTORICAL DATA

Historical data for developing robust stress-test models must cover at least one full economic cycle. Typically for net-interest income models this implies covering an interest rate cycle and for non-interest income models covering one business cycle. The last rising-rate period in the United States was in 2004, implying that net interest income models require data going back at least to that year. Non-interest income models that are more dependent on the business cycle should have data going back at least to 2006–07. Most BHCs have struggled to obtain consistent data going back eight to ten years due to changes in organization structure, business strategy or lack of robust data systems, leading to significant limitations on the data needed to build robust statistical models.

---

[2] In statistics, a confounding variable is an extraneous variable in a statistical model that correlates (directly or inversely) with both the dependent variable and the independent variable.

## 1.3 CONCEPTUAL SOUNDNESS

An ongoing struggle for BHCs has been a lack of coordination and engagement between modeling teams and business unit personnel. The disconnect may vary between two extremes: one in which PPNR is treated as a "headquarters" exercise with no line-of-business involvement, and the other in which PPNR is driven by the lines of business with little central oversight and consistency. Within the first extreme, this has led to statistical models that may look adequate based on statistical validity tests but lack business sense. Within the second extreme, this has led to estimations that are overly reliant on judgment without sufficient quantitative support.

## 1.4 STATISTICAL ROBUSTNESS

Most BHCs have had difficulty in producing PPNR models that have an appropriate level of statistical rigor. With the statistical packages now available, building models is actually relatively easy. However, assessing the models' validity requires personnel who are trained in advanced time series statistical modeling, resources that few BHCs have had available. Some of the more common problems are the lack of "stationarity"[3] of model variables or auto-correlation in historical data that lead to spurious relationships. The stationarity issue is often overlooked, resulting in seemingly highly predictive models that are actually fundamentally flawed. Other commonly observed issues for PPNR models include low predictive power and the failure to test underlying statistical assumptions (such as normality).

# 2. Building strong PPNR models

As shown in Exhibit 2 below, building strong PPNR models requires focusing on eight key considerations.

EXHIBIT 2

**Key considerations for PPNR modeling**

**1 Segmentation**
- **How do we ensure optimal segmentation** to balance the conceptual soundness of segments but avoid going to overly granular levels where noise in data leads to model robustness issues?

**2 Estimation metric**
- **What is the optimal estimation metric** e.g., end-of-period balance vs. cash-flow components of balances, revenues vs. underlying activity volume driver?

**3 Historical data**
- **What are the minimum historical data requirements** required to develop robust stress testing models? What are the potential solutions for banks with historical data limitations?

**4 Model structure**
- **What is the appropriate model structure** considering the trade-off between statistical accuracy and conceptual soundness?

**5 Statistical robustness**
- **What are the expectations for statistical robustness** for stress-testing models and potential tests to demonstrate it?

**6 Output analysis**
- **How do we ensure that the outputs reflect an informed view of the business** and are consistent with actual results based on historical data?

**7 Sensitivity analysis**
- **How can sensitivity analysis be used to test model robustness** and identify potential mitigations for model limitations?

**8 Documentation**
- **What are the expectations for documentation** to ensure that the process is transparent and repeatable?

---

[3] Stationarity—an underlying assumption in time series techniques is that the data are stationary. A stationary process has the property that the mean, variance, and autocorrelation structure do not change over time.

## 2.1 SEGMENTATION

The development of a modeling plan includes making decisions on the right level of granularity. A key characteristic of BHCs with more successful PPNR modeling practices is a segmentation plan that is built around the following dimensions:

- **Materiality threshold** — It is important to establish a threshold below which statistical models will not be developed. Such a threshold is cru-cial for optimal allocation of scarce modeling resources, and for discouraging modelers and business unit personnel from evolving toward an overly granular approach.

- **Portfolio characteristics** — It is important to segment the portfolio by product or customer characteristics that influence the behavior of key metrics across scenarios (for example, interest bearing versus non-interest bearing for deposits).

- **Macroeconomic drivers** — Products with different macroeconomic drivers should be modeled as separate segments to avoid confounding of variables. The determination of distinct behavior is typically a combination of business judgment and simple statistical analyses (for example, correlation, and principal component analysis).

- **Model architecture alignment** — PPNR models typically use, and are used by, other model work streams (such as credit losses and risk-weighted assets). It is vital to keep close coordination in segmentation with these other modeling work streams to ensure smooth and accurate flow of inputs and outputs.

PPNR model segmentation is typically more aggregated and less nuanced than a bank's planning and budgeting process, which is driven by managerial considerations. Coordinating across separate business unit management teams is key to developing an integrated business perspective that aligns with the model segmentation.

## 2.2 MODELING METRIC

The simulation of loan cash-flows (for example, contractual payment, pre-payment, new originations, and credit losses) is critical to ensure accuracy of net-interest income estimation. While most banks have infrastructure to conduct such simulations, determining when to focus on aggregate balance models versus more detailed component models (such as separate models for originations and roll-offs) is a critical question. To address this question, banks typically consider three key criteria:

- **Macroeconomic impact on components** — BHCs should consider if there is sufficient business rationale for the different components behaving similarly or differently. Whatever the decision, it should be supported by sound business judgment.

- **Historical behavior** — BHCs should analyze the historical behavior of components to determine if it is suitable for statistical modeling (for example, in many instances BHCs' prepayment rates for certain portfolios may be highly stable or close to zero as a result of prepayment penalties; hence it may not be possible to model the prepayment rate and the BHCs should assume a constant prepayment rate).

- **Ability to develop robust models** — BHCs should assess their ability to develop component-based models before deciding to go down this path. This involves considerations of data availability of components, as well as considerations of modeling capacity. For BHCs that do not have robust existing models, a way ahead may be to develop models in generations, where first-generation models are balance models that are further refined by second-generation component model

For non-interest revenues the key question is whether the BHC should model revenues or the underlying drivers of revenues. In line with stated supervisory expectations, BHCs should try to estimate the underlying drivers of revenues (for instance, credit card sales for interchange fees, assets under management for asset management fees) to ensure that the macroeconomic impact on the volume of fee-generating activity and the fee-per-unit volume can be isolated. Typically BHCs develop statistical models for the volume component, while the fee-per-unit-volume is often estimated using non-model approaches. For some non-interest revenues where the underlying volume driver may be esoteric (for example, liquid secondary-market trading revenues) BHCs often develop models to directly estimate revenues.

For loan and deposit yields, BHCs typically estimate beta (sensitivity of customer rates to benchmark rates) or spreads (difference of customer rates to benchmark rate). Most BHCs have used non-model approaches to estimate these parameters due to data challenges resulting from flat rates in recent history and significant management judgment used in setting rates. However, many BHCs are now transitioning to using statistical modeling approaches for estimating rates and spreads.

## 2.3 MODELING DATA

The quality of data will affect the degree to which statistical models can be developed. In cases where no satisfactory data can be sourced, it is better to use expert judgment or an analytical approach rather than attempt to build a statistical model based on unsound data. Whether considering internal or external data, the following criteria are crucial:

- **Data quality**—Before beginning model development, modelers need to examine the dataset for potential outliers that may affect model results. This process will include looking for large "spikes" and structural breaks in the data. The most important aspect of this process is that modelers need to work closely with business unit personnel and the chief data officer's organization to understand the underlying causes of any data irregularities. The business rationale should be documented, and modelers should avoid making unjustified assumptions about how to treat irregularities.

- **Governance and controls**—All data (whether internal or external) should come from approved sources that have strong controls (for example, data are reconciled to Securities and Exchange Commission (SEC) filings on a regular basis, edit rights are tightly controlled). Modelers should work with the enterprise data organization to ensure they are sourcing data in a way that is considered acceptable by the validation team.

- **Representativeness**—In addition to covering a complete interest-rate cycle, modeling data needs to include at least one period of significant stress to be able to model different scenarios. If, for example, balances show only growth over the entire time period (for instance, due to an aggressive expansion strategy) then this data will likely not predict probable behavior in future stress scenarios. In such cases, alternative data sources (such as aggregate industry data, similar markets) should be considered.

Many banks struggle to source high-quality and sufficiently long internal data for PPNR. Instead, they are often able to source three to four years' worth of high-quality internal data. For such cases, a two-stage approach may be taken. This consists of developing a model using external data over a longer time period and then regressing available internal data on the external data in order to "rescale" to capture BHC-specific factors, thereby aligning the model's output to the bank's portfolio. While this methodology provides a viable alternative to using models solely based on three to four years of internal data, using long-term internal-only data is always the preferred approach, so BHCs must provide strong justification for any use of external data.

## 2.4 MODEL STRUCTURE

When defining the approach for PPNR models, statistical soundness must be balanced with implementation needs as well as clear business understanding. Once the variable being modeled has been chosen, the form of the model structure must be decided. The following are some strengths and weaknesses of the major different model structures:

- **Level model** — This is the simplest approach, in which the dependent and independent variables are used without any transformations. However, this simplistic approach often suffers from material statistical deficiencies, including the use of non-stationary variables and spurious correlation with unintuitive drivers.

- **First-difference models** — First-difference models transform key variables for use in modeling by subtracting the current period's value from a previous period. This first-difference approach often produces variables that are stationary and results that are relatively easy to interpret. Banks with leading practices tend to use first-difference models for most of their PPNR.

- **Error-correction models** — Similar to level models, error correction models use untransformed variables but allow for both short-term and long-term error-correction terms. Although these models are more statistically robust than level models, they may be more difficult to develop in comparison with first-difference models.

- **Second-difference models** — In rare cases it may be necessary to consider second differences. Caution must be used when developing second-difference models. There should be a strong business rationale for going to second differences. In particular, modeling first-differenced dependent variables against second-differenced independent variables may produce unintuitive forecast results.

## 2.5 STATISTICAL SOUNDNESS

Model development must reflect a balance between rigorous statistical methods and sound business judgment. Modeling must be neither a purely data mining exercise nor just a veneer on top of undocumented business judgment. Within this balance, several key statistical properties are consistent across strong PPNR models:

- **Stationary variables** — A variety of tests (including augmented Dickey–Fuller test, Phillips-Perron, and Kwiatkowski–Phillips–Schmidt–Shin) can be used to evaluate the stationarity of variables. The results of these statistical tests must be combined with modelers' judgment to determine if a variable is stationary. Supervisors generally recommend that BHCs use more than one stationarity test given its importance.

- **Residuals free from serial correlation** — Models with residuals that have serial correlation may produce coefficient estimates that are biased, leading to unreliable hypothesis-testing and interval estimation. Common tests to detect serial correlation include generalized Durbin-Watson and Breusch-Godfrey. One frequent shortcoming is that banks typically test only first-order serial correlation. Instead, up to fourth order should be tested when using quarterly data. These tests establish the extent to which the model suffers from serial correlation, which may then be corrected by a variety of means (for example, lagged structural residuals).

- **Strong out-of-sample performance** — Because PPNR models are used to make predictions over nine quarters, the stability of model predictions is crucial to the soundness of the model. One potential way of assessing this stability is to define an out-of-sample window that is comparable to the nine quarters, and then vary this window throughout the modeling dataset. For each iteration, the root mean square error should be computed for both the out-of-sample and in-sample windows. For the model to be considered robust, the out-of-sample error should be comparable to the in-sample error.

- **Verification of basic statistical assumptions** — Because linear regression is typically used for building PPNR models, the fundamental assumptions that are needed to make linear regression appropriate must be verified. These include that residuals are normally distributed and do not suffer from heteroskedasticity[4]. Additionally, the variables in the final model should show no evidence of multicollinearity[5].

Although the above statistical properties are highly desirable, they may not be uniformly passed in all sound models. Modelers should apply judgment in evaluating whether the failure of any of these properties is grounds for rejecting the overall model. There may be cases where some statistical properties fail, but the modeler documents it as a limitation and includes mitigating controls to address the shortcoming.

## 2.6 OUTPUT ANALYSIS

The purpose of PPNR models is to make predictions in a statistically valid way, meaning that one of the key metrics by which a PPNR model is judged is the quality of its predictions. There are two categories of predictions used for evaluating a model:

- **Historical back-testing** — Besides looking at the overall accuracy across all data points, modelers should examine the trend of the errors. Such examination allows one to assess the model's performance in different regimes. For example, a model may systematically over-predict balances during periods of economic stress. Such analysis informs the use of the model, including any overlays that are made to it.

- **Forecasting** — The behavior of a PPNR model under different scenarios (both Federal Reserve and BHC) will be one of the most critical factors businesses evaluate. One factor to consider is whether forecasts show correct ordering of results. If models show stress forecasts that are better than baseline there should be strong rationale for such behavior. Another factor is whether forecasts show clear separation between scenarios. If the model's forecasts under different scenarios are largely identical, then there may be questions about whether the model exhibits sufficient mac-roeconomic sensitivity.

## 2.7 SENSITIVITY ANALYSIS

A crucial but frequently overlooked aspect of PPNR model development is understanding model sensitivity. There has been increased regulatory emphasis on quantifying each model's sensitivity to three model inputs:

- **Individual data points** (how does the model change if data points are removed or added?)

- **Macroeconomic drivers** (how does the forecast change if underlying drivers are shocked by different multiples of their standard deviation?)

- **Modeling assumptions**

Sensitivity analysis can be conducted by modelers at the level of individual models, or for total PPNR aggregated across key dimensions (such as geography or major business unit). The latter approach may be performed in the results aggregation system by running shocked scenarios.

## 2.8 DOCUMENTATION

Rigorous and thorough documentation is one of the most crucial aspects of PPNR modeling. One of the more common shortfalls is that the documentation does not reflect all of the thought and considerations that went into the modeling. A lagging practice is to present the final model as the only possible model; instead, a "trail" of candidate models should be documented along with the rationale for their rejection. Such documentation should cover the end-to-end process, including data sourcing, segmentation, business engagement, statistical testing, and self-identified limitations.

---

[4] Heteroskedasticity refers to the circumstance in which the variability of a variable is unequal across the range of values of a second variable that predicts it.

[5] Multicollinearity is a phenomenon in which two or more predictor variables in a multiple regression model are highly correlated, meaning that one can be linearly predicted from the others with a non-trivial degree of accuracy.

One of the most crucial topics that regulators often scrutinize is the variable-selection process, in which modelers list all variables considered, the theoretical rationale behind their testing, and the rationale for including or excluding them in the final model. This type of detailed explanation often results in model documentation reaching into the hundreds of pages for a single model. The documentation for each model also needs a well-written executive summary that provides the business and statistical rationale for the model, the key outputs from it, and any important limitations and considerations in simple and easy-to-understand language.

# 3. Using management overlays

Management overlays are used by BHCs to a varying degree to adjust model results. BHCs typically make overlays to address the following concerns:

- **Identified model limitations** — BHCs make overlays to address model limitations (both statistical and non-statistical) that were identified during model development or validation. Some examples of such limitations are exponential growth projected by log-change models or negative balance or fees projected by absolute-change models. These include limitations observed in historical data (for example, past M&A activity, management decisions) that are not repeatable.

- **Forward-looking business considerations** — BHCs also make overlays to incorporate forward-looking management views of the business. Typical examples of such overlays include impact of aggressive marketing plans or declared M&A activities.

## 3.1 KEY CONSIDERATIONS FOR APPROPRIATENESS OF OVERLAYS

There are two important considerations for making management overlays:

- BHCs should make overlays only to address documented model limitations and should avoid making favorable overlays that are predicated on aggressive and quick corrective management actions.

- Overlays should be transparent and repeatable. The factors that were considered and the decisions made for both the rationale and magnitude of the overlays should be subject to stringent internal scrutiny because they will almost certainly be carefully and critically scrutinized by regulators.

## 3.2 GOVERNANCE OF MANAGEMENT OVERLAYS

BHCs should implement strong governance and control processes for management overlays, including the establishment of a model overlay forum that typically includes the head of capital planning, the head of the business, and the head of model validation to review the proposed overlays. In addition, BHCs should develop thorough documentation of the overlays including the rationale for the overlay, the derivation of the magnitude of the overlay, and considerations of the argument in historical periods. Typically, model overlay documentation is created for each overlay that is reviewed by the model overlay forum. Any discussion in the forum while reviewing the overlay should be documented.

# 4. Employing "non-model models"

Although statistical models should be used where appropriate, it is equally important that statistical models are not developed when it is inappropriate to do so. There are four cases when statistical models should not be used:

- **Low macroeconomic sensitivity** — This situation applies for quantities that are not expected to have meaningful correlation with macroeconomic drivers. The most common examples are run-off portfolios, a new product in emerging markets (for example, a high-demand product in a high-growth country), or internally managed variables (most categories and types of non-interest expense). Any correlation with macroeconomic drivers that is modeled is likely to be spurious and misleading.

- **Insufficient data** — When internal data do not meet basic criteria (for instance, when data cover only some three to four years) and external data are unavailable, it is probably inappropriate to attempt a statistical model.

- **Modeling attempted but failed** — Some portfolios may have significant macroeconomic sensitivity and sufficient data, but all models developed fail basic criteria for sufficiency (for example, extremely low predictive power, unintuitive variables, poor results across multiple statistical validity tests). In these cases, it is difficult to build a model that both the business and the validation group will buy into, meaning that a non-model approach may be more appropriate.

- **Significant management judgment** — Some portfolios or metrics involve a significant level of management judgment that may or may not be directly macro-economy dependent. A typical example of such portfolios is risk treasury deposits for which the balance is based on the BHC's overall liquidity position.

## 4.1 KEYS TO BUILDING GOOD NON-MODEL MODELS

Because non-model models are often built on a combination of business analysis and statistical analysis, they are often executed by a team of FP&A (financial planning and analysis) specialists and model developers. FP&A typically takes more of a leading role, as the non-model approach is often closer to the projections they perform on a regular basis. Model developers often conduct statistical analyses that support FP&A estimates.

## 4.2 DIFFERENT APPROACHES TO TRIANGULATE RESULTS

In developing non-model estimates, three separate analyses should be attempted. Each analysis is independently useful, but the value in performing all three is that together they give a range of values that will inform the final decision. These analyses are:

- **Model of a similar portfolio** — This analysis applies a statistically sound model from a different portfolio that is similar in terms of geography, product, customer, or strategy. The growth rates predicted by the model can then be applied to the portfolio undergoing non-model estimation. In some cases, a model can be applied but the drivers can be customized based on the geography— for instance, a commercial real estate (CRE) model developed for a North America model and then applied to a European CRE portfolio.

- **Business budgeting tool** — This analysis takes the business's planned growth as the estimate for the base scenario, and then applies macro-economic sensitivity to derive the stress scenario behavior. Relevant macro-economic indicators are identified, and then the behavior of the indicators under the stress scenario is analyzed relative to their behavior under the base scenario.

- **Historical look back** — This analysis identifies relevant macroeconomic drivers for a portfolio and identifies stressed periods in the historical dataset. The balance or yield behavior during the stressed period is then extrapolated/rescaled as the stressed forecast. Similarly, the run rate over a recent time period is extrapolated as the base forecast.

Although non-model models do not have the same level of rigor as statistical models, they still require comparable levels of documentation and business syndication (despite common beliefs and practices to the contrary).

## Conclusions

Regulatory expectations around all areas of BHCs' stress-testing procedures have been increasing, in particular for PPNR. The expected rigor for PPNR has presented challenges for BHCs, with many struggling on business personnel engagement, data sufficiency, statistical robustness, and decisions about when and how to best use statistical models, management overlays, and non-model models.

As is generally the case with tough problems, there are no simple solutions/magic bullets. Getting better at PPNR models requires making a realistic assessment about where a BHC is today and where it needs to be in several years, and then mapping out a game plan to get there. BHCs are rightfully proud of their ability to understand and model credit losses. For most, modeling the entire balance sheet and income statement using an analytical and statistically rigorous approach is still relatively new, and it will require a concerted program to bring PPNR modeling up to where regulators believe it needs to be.

*Abhishek Anand and Jeff Schonert are consultants in McKinsey's New York office.*