Key Issues in Developing Commercial Credit Default Models

Jack Rutner

Office of the Comptroller of the Currency

Economics Working Paper 2015-2
July 2015

Keywords: DFAST, default, probability of default, upgrades and downgrades, economic factors, data limitations, sparsity, obligors, segments, granular, risk ratings, risk categories, industry, temporal frequency, economic factors, selection, model, search procedures, overfitting, validation

JEL classifications: G21, G28

Jack Rutner is a senior financial economist at the Office of the Comptroller of the Currency. To comment, please contact Jack Rutner at the Office of the Comptroller of the Currency, 400 7th St. SW, Washington, DC 20219, or call (202) 649-5463; fax (202) 649-5743; or e-mail Jack.Rutner@occ.treas.gov.

The views expressed in this paper are those of the author alone and do not necessarily reflect those of the Office of the Comptroller of the Currency or the U.S. Department of the Treasury. The author would like to thank Richard Nisenson and John Culbertson for helpful comments, and Jessica Scully for editorial assistance. The author takes responsibility for any errors.

Key Issues in Developing Commercial Credit Default Models

Jack Rutner
July 2015

Abstract: The stress testing mandated by the Dodd–Frank Wall Street Reform and Consumer Protection Act has reinforced the ongoing development of quantitative models banks use to project expected losses. Banks have faced challenges in developing those models because of limitations on the data available, including incomplete information about obligors and the limited duration of the obligor data. Within the context of such limitations, this paper discusses key issues that can arise while developing commercial credit default models from obligor loan data. Such issues include organizing the data, selecting economic factors for the models, and then validating the models. This paper discusses why it is good practice to organize obligors by industry and rating categories. Organizing obligor data that way, though, requires deciding how granular to make industries and rate categories. The paper illustrates one way of reducing the granularity of risk rating states. Another challenge in data organization discussed here is selecting the temporal frequency. After the data are organized, model development follows, and, in particular, selecting the appropriate economic factors for the models. The paper discusses the current approach some banks use as well as an alternative that is data dependent. Regardless of approach, the models developed need validation. That is because models developed for projecting defaults can fit the data used to develop them very well, but projections by them may not fit new realizations of the data very well. A good approach to model validation is to see how well the models perform "out-of-sample" on data realizations from a time period not used in their development.

1. Introduction

Over the past several years, the larger commercial banks have moved away from using expert judgment and towards quantitative models to project expected loan losses in their commercial credit portfolios. That move has been reinforced by Dodd–Frank, which became public law in July 2010. Dodd–Frank requires that larger banks model loan losses conditioned on economic forces. Currently, the large banks maintain quantitative models for their commercial credit portfolios. Those portfolios include, but are not limited to, commercial and industrial loans as well as commercial real estate loans.[1] As required by Dodd–Frank, the commercial credit models developed by the banks generally must project loan losses over a multi-year period. The models project losses by decomposing the loss calculation into three components: the probability of default (PD); exposure at default; and loss given default. That decomposition follows the Board of Governors of the Federal Reserve System's outline in its document on stress testing and capital planning for large bank holding companies (Federal Reserve Board [2013]). This paper's focus is exclusively on the PD component.

In developing their commercial credit models to project PD, banks have faced some basic challenges. Many of the challenges arise because of limitations on the datasets banks use for modeling PD. One challenge is deciding what trade-offs to make in the level of granularity at which the models are estimated. The datasets' limitations include the following:

- **Incomplete obligor information.** The datasets used for Dodd–Frank Act stress tests (DFAST) provide no information about obligor financials. Instead they provide the risk ratings assigned to the obligors by the banks. The risk ratings, although based on those financials, are at best proxies for them. Moreover, modelers generally use the ratings to group together obligors with similar financial characteristics. It is from the common behavior of the obligors in each group that models are developed to forecast PD. That, though, is an inexact way of developing models to forecast PD.

[1] In addition, the commercial credit portfolio includes loans to depository institutions and to government and government-type entities.

- **Datasets of limited duration or coverage.** Some banks have only recently begun to retain historical datasets of obligors. Consequently, their datasets may go back only one or two economic cycles. Also, at some banks, the early history in the datasets is not representative of all obligors. Apparently, those banks originally recorded information only for obligors above a certain size. Finally, some banks, while large enough to fall within the purview of DFAST, are not large enough to have a sufficient number of obligors to estimate models at a very granular level for industries and credit risk states.

The challenges arising from the data limitations then become how to (1) organize the data and (2) model the data. The purpose of this paper is to discuss those basic challenges. Section 2 focuses on how to organize the data. In particular, it discusses the reasons for modeling (1) the commercial portfolio by industrial segment rather than in its entirety, (2) PD in a framework of upgrades and downgrades[2] from initial credit risk categories, and (3) at an annual frequency rather than a quarterly frequency. Section 3 then focuses on selecting which economic factors to include as well as model validation. It also discusses an alternate approach to the one that many banks use when deciding which economic factors to select for their models. The approach of at least some banks has been to develop PD models from a small set of economy-wide economic factors. Section 3 provides an alternative approach that develops models from a larger set of factors. The larger set also includes industry-specific factors that align with the industrial segments being modeled. No matter the economic factors selected, or their selection process, the models then have to be validated. Section 3 discusses that as well. Section 4 concludes by observing that a great many decisions have to be made in designing models of upgrades and downgrades, that the process is an iterative one, and that the decisions made must be articulated to management.

[2] More general terms for upgrades and downgrades are credit migrations or credit transitions.

2. Organizing Data by Industry and Risk States

2.1 Segmenting Obligors by Industry

A PD model for commercial credit can be constructed for the entire commercial credit portfolio without segmenting obligors by industry. This section shows abstractly and factually why industrial segments matter, and thus, why it is good practice to segment obligors by industry. Data sparsity, however, acts to limit the level of segmentation achievable.

A primary reason for using industrial segments is to group together obligors sharing similar characteristics into the same segment. Obligors in the same segment can be expected to respond in about the same way to the same economic factors as compared to obligors of other segments. The models developed by industry segment should then be relatively more stable and easily explainable. An added benefit of segmenting obligors by industry is to inform management about in which segments of the commercial credit portfolio risk is concentrated in. It thus becomes good practice[3] to model a commercial credit portfolio segmented by industry.[4]

To understand abstractly why industry segments matter, suppose a bank's entire portfolio is composed of only two segments: segment A and segment B. Suppose further, in response to prevailing economic factors, segment A's PD is 1 percent, while segment B's PD is 0.5 percent, and that the two segments have equal weights in the bank's commercial credit portfolio. The entire portfolio then has a PD of 0.75 percent. Now, suppose management decides to increase segment B's share of the portfolio to 75 percent. For the modeler who has developed a PD model for the entire portfolio, the model appears to be unstable insofar as the portfolio's PD has changed, even though no economic factor has. By segmenting obligors by industry, the modeler should expect to achieve, at least potentially, greater stability for each segment's model. In turn, the model's forecasts should be more reliable.

[3] The Federal Reserve Board considers banks using only one segment (i.e., no industry breakout) to be using "weaker practices" (see p. 22).

[4] For banks participating in DFAST, it is easy to split the portfolio into industrial segments. That is because the DFAST datasets they have typically assign a Standard Industrial Classification code or a North American Industrial Classification System code to each obligor.

To show factually why segments matter, figure 1 plots default frequencies of eight industrial segments from the credit analytics data collected by the OCC. The figure shows the percentage of defaulted dollars in each segment. The two recessions in that period are shaded red. Figure 1 shows that the default frequencies of different segments do not proceed in lockstep. That is true for both the magnitude and the timing of peaks in each of the industrial segments' default frequencies. Sometimes, default frequencies peak at the same time; other times they do not. At most times, the magnitudes are different.

Figure 2 shows that by isolating two segments: distribution and services. It is apparent from the figure that the magnitude of the default frequency in the services segment generally differs from that of the distribution segment, being almost always greater. Also, the figure shows that the timing of peaks in default frequencies differs: during and after the tech bust, no essential change occurred in the default frequency of the distribution segment. By contrast, the default frequency of the services segment increased sharply a year after the bust. During the Great Recession, the services segment's default frequency peaked one quarter after the recession ended. The distribution segment's default frequency, though, did not peak until three quarters later.

The behavior of the industrial segments' default frequencies in figures 1 and 2 leads to the conclusion that industrial segments matter, and hence, obligors should be segmented by industry. That, though, may not be so for every bank. A particular bank's dataset may show smaller differences across industries than do the data collected by OCC Credit Analytics. In such instances, the need for the bank to segment obligors by industry would be lessened.

The challenge when modeling by industrial segments is to determine the degree of granularity achievable without compromising confidence in the model. Each level of granularity has its advantages and disadvantages. The advantage of more granular segments is that obligors in them are much more alike than in less granular segments. Consequently, models based on very granular segments reflect obligor behavior in each segment with much less error. The disadvantage of very granular segments is the sparsity of low probability events with the data available because of the dataset's limitations. The consequence of sparsity is the inability to model such events, in particular the downgrade to default, with a high degree of confidence or to

forecast them well. The advantage of less granular segments is the greater number of low probability events observed in each segment because less granular segments have more obligors. With more low probability events observed, risk rating changes can then be modeled and forecasted with greater confidence. The disadvantage is the increased dissimilarity of obligors in each of the sectors.

To conclude, data limitations thus constrain the degree of industry granularity that can be achieved. Models of obligor behavior with greater granularity have less error because of the greater similarity of obligors in them. At the same time, because of data limitations, excessive granularity leads to greater data sparsity, and the models can be expected to have more error and to be less stable.

2.2 Risk Categories

2.2.1 Granularity of Risk Categories

As with industry, it is possible to model the PD of an entire portfolio without considering initial risk states. The reason to consider initial risk states is analogous to segmenting obligors by industry: within the same risk state obligors are more similar to one another than to obligors of other risk states, and modeling groups of similar obligors should improve model performance. The multi-period time horizon of DFAST is the reason for modeling PD within a framework of upgrades and downgrades. The multi-period time horizon entails knowing how many obligors are in each risk state at the beginning of each period following the first. That can be determined only by modeling upgrades and downgrades from the previous period's starting risk states.[5] Of course, categorizing obligors by risk state raises the same challenge as with industry: how granular to make the risk states. The more granular the risk states, the sparser the observations within certain risk states and the less reliable the model.

[5] A side benefit of modeling upgrades and downgrades is being able to project changes in the quality of the portfolio.

As indicated, DFAST's multi-year time horizon necessitates modeling risk rating upgrades and downgrades. Section 2.1 shows that industrial segments matter. Together, those conditions lead to the conclusion that the modeling of upgrades and downgrades should be by industry segment. Two characteristics in the obligor data make it difficult for a bank to maintain the usual granularity of its risk ratings categories when modeling upgrades and downgrades. The two, infrequent rating changes and a high concentration of obligors in one risk rating, result in a sparsity of observed upgrades and downgrades. When banks segment obligors into more granular industries, observations get even more sparse. Such sparsity reduces confidence in a model of upgrades and downgrades developed from such data.[6] Combined with the existing data limitations, the challenge is to decide how to trade off the level of granularity of risk states for that of industry segments.

One way of mitigating sparsity is to reduce the granularity of risk ratings by consolidating some of the ratings. Table 1 illustrates one way of doing that. In it, 10 risk ratings are assumed[7] along with a high concentration of obligors in risk rating 6. Given the assumed high concentration in risk rating 6, the consolidation of the risk ratings is almost predetermined by data sparsity. It is into three non-defaulted categories and one default category.[8] The three non-defaulted categories are here titled high pass; medium pass, being those obligors in risk rating 6 (sometimes with 7[9]); and weak pass. When the concentration in risk rating 6 is not high, risk ratings can be more granular than presented in table 1.

Table 1 also illustrates how upgrades and downgrades are connected to the consolidated risk categories. The *q*s of the matrix are the percentage of upgrades or downgrades of obligors in each of the non-defaulted rate categories, with the *q*s of each row summing 100. The object of

[6] In some cases, sparsity can lead to unexpected results. For example, ordinarily the PD is expected to be monotonically related to the quality of the risk rating. With data sparsity, the observed frequency of default by happenstance may not be monotonically related.

[7] Many banks usually have more ratings, generally about 21. Those 21 can, however, be made to conform to the 10 categories of figure 1.

[8] More granular risk categories are, of course, possible. Creating such categories, however, depends on the characteristics outlined earlier.

[9] The assumption made here is that risk rating 7 is very comparable to 6. If 7 is not, it can be combined with 8 and 9.

model development is to project those *q*s. The number of upgraded or downgraded obligors in the next period is then determined by multiplying the projected *q*s of each row by the number of non-defaulted obligors of the row in the current period.

The matrix has an additional category most banks do not include in their models: payoffs (F). The category comprises exits from the portfolio from the initial three non-defaulted risk categories through payoffs of loan obligations. The reason for its inclusion is discussed in section 2.2.2.

Table 1. Stylized Consolidation of Risk Ratings and Upgrade and Downgrade Matrix

Risk ratings	Consolidated risk categories
1	High pass (H)
2	
3	
4	
5	
6	Medium pass (P)
7	
8	Weak pass (W)
9	
Greater than 9	Default (D)

Upgrade and downgrade matrix

	H_1	P_1	W_1	D_1	F_1	Final risk ratings
H_0	q_{hh}	q_{hp}	q_{hw}	q_{hd}	q_{hf}	
P_0	q_{hp}	q_{pp}	q_{pw}	q_{pd}	q_{pf}	
W_0	q_{wh}	q_{wp}	q_{ww}	q_{wd}	q_{pf}	

Note: Risk rating categories are H: high pass; P: medium pass; W: weak pass; D: default; F: payoff. qij = The likelihood (or frequency) of an obligor being upgraded (or downgraded) from risk rating i to risk rating j.

From the matrix in table 1, the problem caused by sparsity becomes apparent when obligors are infrequently upgraded or downgraded. In that circumstance, the values of the diagonal cells of the matrix—q_{hh}, q_{pp}, and q_{ww}—are larger. Furthermore, because each row's *q*s sum to100, the off-diagonal elements then must be smaller.[10] When upgrades or downgrades are observed

[10] The changing quality of the industry portfolio in the matrix can be determined by the changing sizes of the cell values. Specifically, as cell values increase on the right-hand side of the matrix, except for column F_1, the quality of the industry portfolio can be said to be decreasing. That is because a movement toward larger cell values on the right-hand side of the matrix represents an increase in the likelihood of risk rating downgrades and defaults.

infrequently, it becomes difficult to model them with much confidence. Comparably, as P_0 gets larger, fewer observations are left for the other rows of the matrix, and those become difficult to model as well.

To conclude, a DFAST model has to take into account industry segments and risk states. Yet the dataset's limitations mean that neither segments nor risk states can be too granular. Anyone developing models from the kind of datasets used for DFAST will have to decide how to trade off the granularity of the industrial segments with that of risk states when organizing the data. As a practical matter, some banks appear to prefer trading off greater granularity of risk states for lesser granularity of industry segments. Ultimately, though, each bank has to achieve a kind of Goldilocks position for itself, and that depends on its own specific data. The position has to be one that is neither too granular nor too coarse for its purposes.

2.2.2 Adding Two Credit States

While banks may have to consolidate risk ratings when modeling by industrial segment, under certain conditions they may have to add two credit states: a payoff state described in section 2.2.1 and possibly an entry state. Entry comprises new obligors entering the portfolio. Consideration should be given to adding those states when their characteristics do not conform to expectations during stress periods because determining what happens to banks in stress periods is the entire purpose of DFAST.

Consideration should be given to the payoff state when payoffs are relatively large. At some banks, exits from the portfolio through payoffs dwarf exits through default. Moreover, the payoff state seems to get particularly large in stressful periods. The entry state should be considered as well should new obligors, when compared with existing obligors, be both distributed very differently across risk ratings and relatively large.

2.3 Temporal Frequency of the Data

This section discusses an alternative to the temporal frequency some banks use in modeling PD, which is the one-quarter frequency of the DFAST datasets. The alternative others use is an annual frequency. The first part of the section describes the advantages of using an annual frequency for risk ratings. The second part discusses aligning the frequency of the economic factors with the frequency of the risk ratings. The final part discusses the challenge posed when using an annual frequency, instead of a quarterly frequency. The consequence of that is to reduce the number of data points by about 75 percent, and thereby compound a situation of already sparse data points. A methodology is proposed to overcome the challenge.

2.3.1 Advantages of an Annual Frequency

Several advantages arise from using an annual frequency rather than a quarterly one. The primary advantage, though, is to place the economic event affecting an obligor's risk rating in the same or close to the same time period as the obligor's reassignment to another rating. The reason an economic event may not be in the same period in a quarterly frequency is that the event can happen in one quarter while the reassignment does not occur until several quarters later. A second advantage of an annual frequency is that it mitigates seasonality effects from an obligor's reassignment to another risk rating. Such effects can arise when an obligor's risk status changes but the bank waits until some specified quarter of the year to reassign obligors' risk ratings. Such seasonal reassignments act like noise in the data. The use of annual frequencies is meant to mute that noise.

2.3.2 Aligning Economic Factors, Detrending, and Lagged Values

When an annual frequency is used for obligors' risk states, consideration should be given to having the temporal frequency of the economic factors on the same basis. Aligning those frequencies is analogous to having risk states and economic factors on a quarterly basis. For

economic factors, there are additional advantages of being at an annual frequency: the elimination of any seasonality in the factors and the reduction in any non-seasonal noise.[11]

After economic factors are recomputed at an annual frequency, it is good practice to detrend them. That is because they will be used to model the probability of an upgrade or downgrade which is trendless, except possibly in stress periods.[12] Moreover, even in stress periods, the trend is bounded. It is more sensible to model a trendless probability with trendless economic factors than with trending factors.[13]

Finally, when modeling upgrades and downgrades, current economic conditions may not immediately affect an obligor. Thus, the PD models should allow for both contemporaneous and lagged economic factors. Exactly how long the lag should be depends on the characteristics of the obligor data.[14]

2.3.3 Overlapping Data

While modeling at an annual frequency is generally advantageous, doing so has the significant disadvantage of shrinking the number of data points available to estimate the model. The shrinkage leaves only about 25 percent of the observations from the original dataset. The challenged posed by the shrunken dataset is that it may result in otherwise significant economic factors being omitted from the model.[15] One way to meet that challenge is to employ the oft-used

[11] Also, many factors that should be considered in modeling industries are unavailable on a seasonally adjusted basis.

[12] Clearly, the probability can change, but because it is bounded between zero and one, it cannot have a trend.

[13] Detrending for some factors means calculating annual average growth rates. For others, it means calculating a change in the annual average of the factor.

[14] In our own modeling, we used a two period lag.

[15] That is because with fewer observations, the error bands on the factors' coefficients are larger. Also, there is a reduction in the quality of the coefficient estimates themselves.

method of overlapping data[16] (see Brorsen and Harri [2001]), which are time series at annual frequencies displaced by one quarter. The use of overlapping data allows modelers to recapture most of the data points lost in aggregating up to an annual frequency.[17]

3. Economic Factors: Selection and Validation

3.1 Factor Selection

In constructing the upgrade/downgrade models for DFAST, another challenge is determining the appropriate economic factors to include in the models being developed. Yet, as Campos et al. put it: "We don't know what we don't know, and so we cannot know how best to find out what we don't know." So, how is that challenge to be met?

The challenge is often met by prejudging which factors to include in the models. That is, a judgment is made *ex ante* about the factors most likely to explain upgrades and downgrades. Frequently, the judgment is made that the appropriate factors are economy-wide. Examples are real gross domestic product (GDP), the GDP deflator, interest rates, and possibly one or two housing-related factors. Furthermore, some banks restrict such factors to just a few in any particular model. Most surprisingly, even when modeling by industrial segment, economy-wide factors are judged to be appropriate.

An alternative method should be considered in meeting the challenge of which factors to include. This alternative takes into account that upgrades are being modeled by industry segment. It also

[16] Overlapping data are not quite the complete answer because they introduce serial correlation that biases the statistics to accept factors that otherwise would be rejected. There is, though, a simple rule of thumb to correct the bias in the test statistic in many cases: double the standard error of each coefficient.

[17] Overlapping data can also help overcome a disadvantage of using an annual frequency: determining the quarter of the DFAST stress scenario in which a bank is most at risk. Some banks have associated an annual frequency with a particular quarter and have used one-fourth (or the fourth root) to represent the quarter. An alternative methodology is to start the forecast period in the year in which the first three quarters are known and back-out the unknown fourth quarter by subtracting from the forecast the known first three quarters. In the subsequent overlapping year, the first two quarters are known, and the third quarter is the backed-out value of the previous overlapping forecast. Then the fourth quarter is known by subtracting the first three quarters from the overlapping forecast. This practice can be continued until all the quarters of the stress scenario have been backed out. With each of the quarters of the scenario backed out, the quarter in which the bank is most at risk can be determined.

addresses the challenge of how to select the appropriate factors for each segment from a large set of factors under consideration.

The way the method takes into account that upgrades and downgrades are being modeled by industry segment is to include segment-specific economic factors in the model when warranted. The use of segment-specific factors should not, however, preclude the use of economy-wide ones. Segment-specific factors, generally speaking, include resource and labor costs of some segments as well as sources of revenue for them. In practice, such factors include producer prices by industry along with the corresponding consumer prices, gross product originating by industry, employee compensation, and a variety of GDP components. As an example, suppose the segment being modeled is the electric power industry. The potentially relevant segment-specific factors could be the producer price of electric power, the producer price of energy, the compensation of labor, the cost of transportation, etc.

When a bank specializes in lending to certain industries, then a strong argument can be made for using segment-specific factors even when modeling the entire commercial credit portfolio. In such an instance, the portfolio will be dominated by just a few industrial segments. The more appropriate factors for portfolio models would then be those most appropriate for the segments dominating the portfolio. Nonetheless, whether segment-specific factors should be used in place of economy-wide factors in such cases is a judgment each bank has to make. The best way to make it is to develop models with and without segment-specific factors, and then examine how well the models forecast with data not used in model development.

The use of segment-specific factors introduces many more economic factors than most banks typically use for DFAST models. How many more depends to some extent on how granular the segments are. Having many potential factors, however, is in the spirit of Campos et al.'s warning of not knowing what to include. Having many potential factors increases the probability of including at least some of the right ones in the model.

Once a list of potential factors has been selected, the next part of the alternative method is to address the challenge of selecting significant factors appropriate for the segments being modeled.

The way to do so is through a search procedure. That is, different factors can be hypothesized as being significant and then tested. If they are found to affect upgrades and downgrades significantly, they are retained. Otherwise they are discarded. The search can be manual or by machine (e.g., stepwise regression). The advantage of a machine search is that it is carried out in a very systematic manner. In either case, the data themselves determine the appropriate factors to be included in the models. Neither search method is statistically preferable, although the machine search is far faster.[18]

While the factor list can be lengthy, two constraints prevent too many factors from being included in a model. One is a data constraint, which occurs when there is a sparsity of upgrades and downgrades. That most often happens when modeling downgrades from high pass to default. In such cases, no more than one or two factors can be included.

A second constraint occurs when there are many upgrades and downgrades. That usually happens when modeling upgrades from medium pass or downgrades from it to weak pass. In such cases, the constraint is not from the data, which often allow a great many factors to be included in the model. The constraint is a "knowledge" constraint, which is that when many factors are included in the models, some are irrelevant. The disadvantage here is that the model can fit very well in-sample but often does not forecast very well out-of-sample. Modelers use the term *overfitting* to describe that situation. Several statistical tests have been developed to reduce the probability of including the irrelevant factors that lead to overfitting when developing a model with a great many data points (see Benjamini and Yekutieli [2001]).

3.2 Validation

After a model has been developed, it must be validated. Validation with in-sample data is not good practice because a good modeler can make a model fit the in-sample data very well through overfitting. Good practice for validating a model is to test it on out-of-sample data that are out-

[18] The reason neither is statistically preferable is that the statistic used for including a factor needs to be adjusted to account for the search process. That is so irrespective of whether the search process is done manually or by machine. As an example of why the adjustment is needed, suppose a 5 percent level of significance is chosen. Furthermore, suppose 100 searches are conducted. The likelihood is that in five of the searches, a factor will appear to be significant when it is not, and that will be so whether the search is manual or by machine.

of-time. That is, the modeler holds out observations after a certain time period when developing the model. The modeler then uses the model to forecast the held-out observations.[19] That way the ability of the model to forecast out-of-sample can be tested and different model set-ups compared. The held-out sample itself has to be long enough to determine how long the model stays on track. By determining how long that is, the modeler can ascertain the expected error in the forecasts over the DFAST horizon. Also, the modeler can ascertain how often the model has to be re-estimated.[20] Finally, testing against the held-out data can provide information about whether the model's forecasts are systematically biased. If they are, then the forecasts can be adjusted to account for any such biases.

4. Conclusions

What should be apparent from this paper is that modelers have to make a great many decisions as they go from raw data to developed models because of the challenges posed by the data's own limitations. Such decisions range from the number of industrial segments to the number of risk categories after consolidating risk ratings, and finally what temporal frequency should be used when modeling the data. The next set of decisions is which and how many economic factors are to be included when estimating the models. After that is completed, the modelers have to test the model to validate the decisions they made in developing it.

The process for going from raw data to developed models may appear to be linear. It is not; it is iterative. Modelers have to posit alternate ways of categorizing obligors, develop models for each kind of categorization, and then see which categorization performs "best" in the validation stage of model development. Furthermore, if in validation the model does not produce acceptable results, modelers have to circle back and rethink whether they need to make the industrial segments finer or coarser, and correspondingly, whether they need to make the rate categories

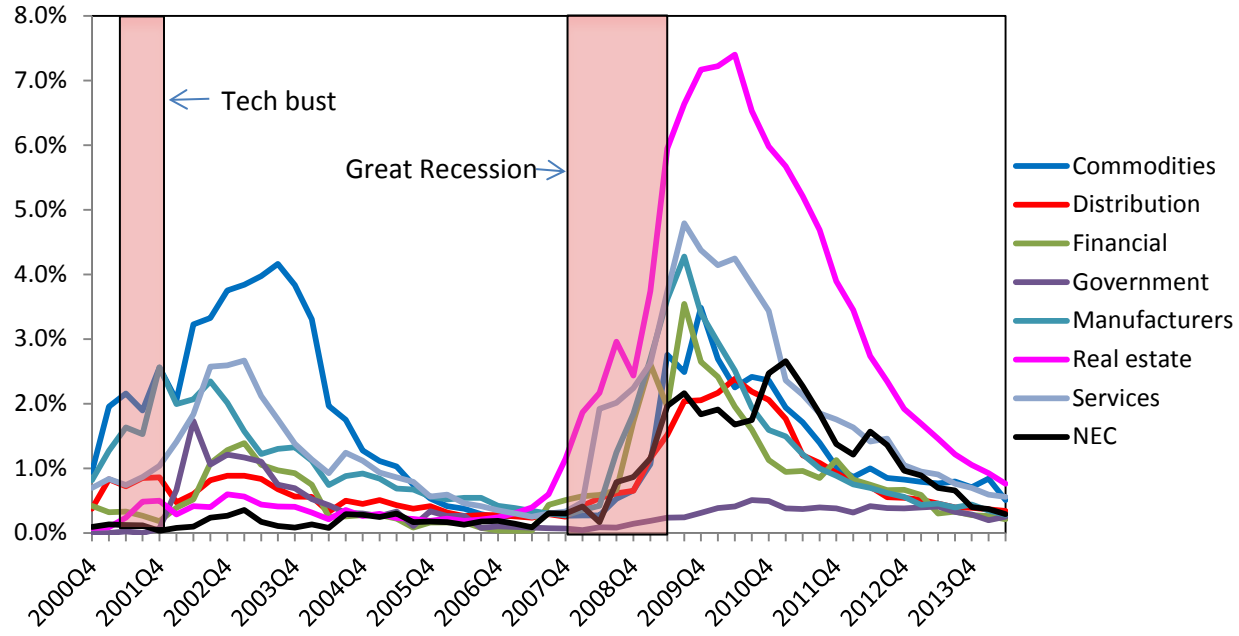
[19] An objection to out-of-time validation is that the model used for DFAST does not include the out-of-time held-out observations. The point, though, of validating the model out-of-time is to determine not so much the precise coefficients but whether the modeling approach itself is valid. If it is, then the model can be re-estimated to include the held-out observations.

[20] Implicit in this statement is that the nature of the underlying obligor data requires frequent re-estimations of the model.

finer or coarser. Modelers may also have to rethink whether the set of economic factors selected is appropriate for the industrial segments used for developing the models.

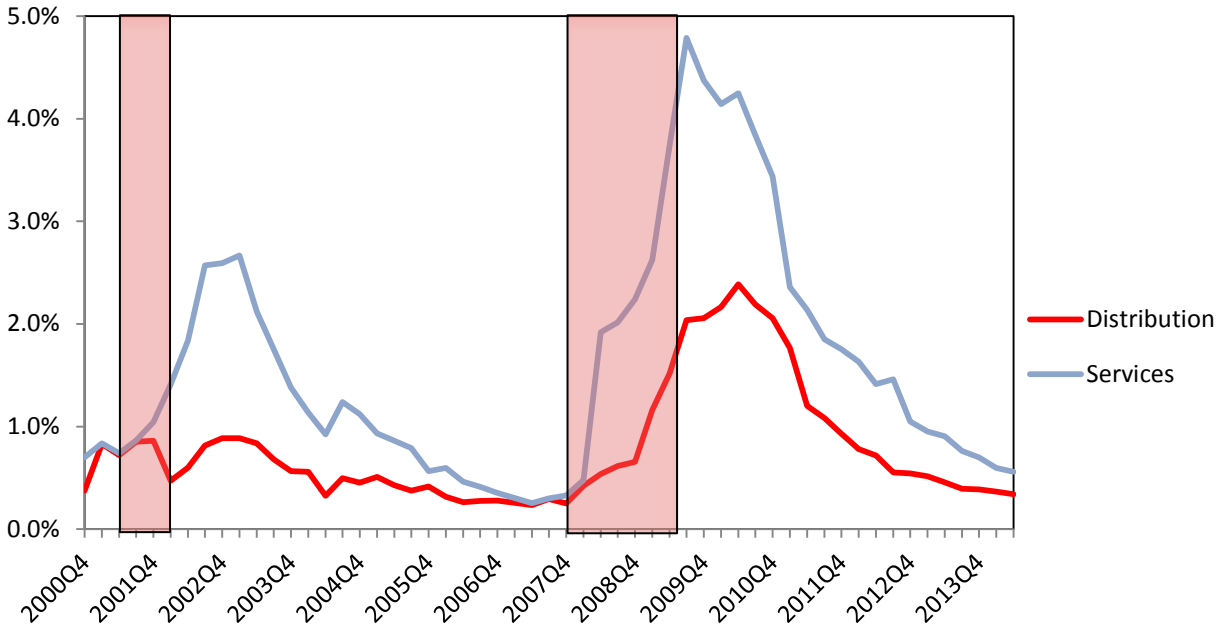
Whatever decisions modelers have made when developing the models, the modelers have to articulate those decisions to management and to others. For example, they need to articulate the options they considered, why they made the specific choices they did given those options, and the advantages and disadvantages of the choices made. The range of options includes the granularity of the industries and risk categories, the temporal frequency of the data, and the reason for the economic factors chosen for the models. Faced with various options, modelers may have to examine the results of competing models to determine which one provides better forecasts. Those results can then provide management with the rationale for the options ultimately chosen.

Figure 1. Eight Major Industrial Segments: Frequency of Default (2000Q4 Through 2014Q2)



Source: OCC Credit Analytics

Figure 2. Two Major Industrial Segments: Frequency of Default (2000Q4 Through 2014Q2)



Source: OCC Credit Analytics

Note: Currently, data for credit analytics are collected voluntarily from 35 large and midsize national banks covering almost 90 percent of commercial credits. The data begin in the last quarter of 2000 and go through the second quarter of 2014. Defaulted dollars are used here because obligor counts are not available from OCC credit analytics.

References

Benjamini, Yoav, and Daniel Yekutieli, 2001. “The Control of the False Discovery Rate in Multiple Testing Under Dependency.” *The Annals of Statistics*, 29:4, pp. 1165–1188.

Federal Reserve Board, August 2013. “Capital Planning at Large Bank Holding Companies: Supervisory Expectations and Range of Current Practice.” Washington, DC.

Brorsen, B. Wade, and Ardian Harri, 2001. “The Overlapping Data Problem,” *Quantitative and Qualitative Analysis in Social Sciences*, 3:3 pp. 78–115.

Campos, Julia, Neil R. Ericsson, and David F. Hendry, August 2005. “General-to-Specific Modeling: An Overview and Selected Bibliography.” Board of Governors of the Federal Reserve System: International Finance Discussion Papers, no. 838.