Key data gaps your data vendors may not be able to solve in physical risk modeling

Traditional credit-risk methods have yet to incorporate the effects of climate change on asset returns. Although banks are quickly adapting their risk strategies to integrate these climate risks into their calculations, one key challenge they face is securing accurate and sufficient data to use in their risk modeling.

KPMG LLP has developed a physical risk model to assess the incremental credit losses that traditional credit-risk models don't capture. In this blog, we'll take a closer look at the specific data challenges that KPMG has overcome and describe how you can improve your data and modeling infrastructures to help prepare your climate-risk strategy for a low-carbon future.

Geospatial data is often imprecise, inaccurate

One key challenge in evaluating the effects of climate change is the poor quality of geospatial data. Banks use this data to evaluate the potential environmental risks to the physical collateral securing their loans. But often it is too rudimentary, sparse, or even inaccurate.

Bank records typically contain basic geographical data (e.g., address data) for loans secured by real estate. But KPMG observed that street numbers, city names, and ZIP codes in these records often contain typos or point to a location that doesn't even exist. For loan portfolios that aren't secured by property, the geographic data is typically much sparser. For example, commercial and industrial loans may be partially or fully collateralized by the borrower's cash flows, so banks must know where the borrower's suppliers, customers, factories, and warehouses are located. Then, banks need to obtain specific data on the physical risks to each of these assets, such as the collateral's flood risk, elevation, and proximity to geological features. These data attributes help banks segment, and understand, the physical risks to which these assets are exposed.

This geographic and physical data can be found from various data providers, who are quickly augmenting their capabilities to capture geographical data attributes. But their comprehensiveness and accuracy remain sticking points. Some data providers can supplement geospatial data using CUSIP/ISIN as a key for public securities, but this data gap persists for private securities and loan portfolios. In addition, public data can partially satisfy this demand for geographical data. For example, the Federal Emergency Management Agency (FEMA) has performed decades of flood modeling and has mapped flood zones across the United States, based the flood risks to a specific area.

Although public data can be useful, banks should proceed with caution. For instance, when developing its flood-risk model, KPMG used data from FEMA and the United States Geological Survey (USGS) to capture some measurement of flood risk to each asset we modeled. However, we found that the FEMA flood zone classification alone is too broad to capture all the specific flood risk to an asset, and this data challenge afflicts private data vendors too. We improved the granularity of geospatial data in these public data sets by using interpolation techniques, but the resulting data was still too coarse and ultimately diminished our model's precision.

Finding collateral's physical characteristics

The banking industry still lacks granular and high-quality data related to the physical characteristics of collateral and its adaptations to climate change. Very few internal or external data sets that KPMG inspected contained detailed collateral information, even though academic research demonstrates how critical physical characteristics are to precise physical risk modeling. Case in point: FEMA's Hazards US (HAZUS)[1] tool, which estimates the flood risk for any property in the United States, requires the user to input collateral characteristics (e.g., occupancy type, foundation type, and number of floors) to estimate economic losses from flood damages.

When it comes to loan collateral, the physical attributes that banks require will depend on the collateral class and type of physical risk. Loans secured by non-real-estate collateral may share the same physical risks as real-estate loans but will exhibit distinct susceptibility to floods, wildfires, and other physical risks. Data providers that specialize in collecting data corresponding to specific asset classes or physical risks have begun to fill niches in the market. But comprehensive, high-quality data for all asset classes and physical risks remains elusive.

Banks should consider supplementing their existing data with open-source data sets from FEMA, the USGS, and various other government sources as an interim solution until better and more granular data becomes available. For example, we utilized building characteristic information from FEMA's National Flood Insurance Program (NFIP) Redacted Claims Insurance data set[2] to simulate property characteristics in our residential mortgage model.

In short, banks need to identify interim data-procurement strategies based on what data is currently available and a target operating strategy to procure this data in the future.

Better historical climate data needed

Another issue that's preventing banks from estimating the physical and transition risks to their portfolios is inadequate historical data related to climate events and environmental regulations. Most banks haven't tracked how various physical risks and evolving government policies have impaired asset returns, and very few external data sets exist today that would enable banks to quantify these relationships. The Network for Greening the Financial System (NGFS) published its "Progress report on bridging data gaps" in May 2021 to summarize current data challenges and to present results from NGFS surveys and two closed-door workshops with banks and buy-side firms.[3] The report broadly classifies data requirements into six categories—(1) exposure quantification, (2) investment and lending decisions, (3) macroeconomic modeling, (4) financial stability monitoring, (5) climate-related disclosures, and (6) scenario analysis and stress testing—but concludes that the industry still hasn't coalesced around a specific list of data attributes that are necessary for each type of analysis. Unsurprisingly, data vendors have been slow to provide the necessary climate event data given that the industry has only a vague and scattered understanding of the data needed to model physical risk.

KPMG tried to model the historical relationship between floods and property values by analyzing NFIP's Redacted Claims data set and USGS's data sets that contain measurements for peak flooding from extreme flood events. These public data sets, however, have deficiencies in their geospatial data that prevented us from discerning the true relationship between flood damage and economic loss. We used geospatial interpolation techniques such as kriging[4]; but, even with advanced modeling techniques, publicly available data was inadequate to provide a precise estimate.

To make up for the shortcomings of publicly available information, KPMG collaborated with third-party data providers, such as First Street Foundation, to obtain more granular and thorough data sets.

Banks need to understand the immense cost of developing models that use physical and hydrological dynamics and should strongly consider third-party vendors that specialize in modeling property-level climate event data.

[1] Source: Hazus, FEMA website

[2] Source: OpenFEMA Dataset: FIMA NFIP Redacted Claims, FEMA website

[3] Source: "Progress report on bridging data gaps," Network for Greening the Financial System, May 2021

[4] Source: "How Kriging Works," ArcGIS website

Contact us

Adam Levy
Advisory Principal
Modeling & Valuation
KPMG US
T: 312-665-2928
E: adamlevy@kpmg.com

Max Teng
Senior Associate
Modeling & Valuation
KPMG US
T: 212-758-9700
E: maxteng@kpmg.com

Some or all of the services described herein may not be permissible for KPMG audit clients and their affiliates or related entities.

kpmg.com/socialmedia