

Preventing Another Crisis: Quality Data for MBS Markets

Andrew Kumiega

Illinois Institute of Technology

565 W. Adams, Chicago, IL 60661, USA

Tel: 1-312 906-6513 E-mail: akumiega@iit.edu

Ben Van Vliet (Corresponding author)

Stuart School of Business, Illinois Institute of Technology

565 W. Adams, Chicago, IL 60661, USA

Tel: 1-312 906-6513 E-mail: bvanvliet@stuart.iit.edu

Apostolos Xanthopoulos

American Intercontinental University

5550 Prairie Stone Parkway, Hoffman Estates, IL 60192, USA

Tel: 1-630 606-7830 E-mail: toli_xan@yahoo.com

Received: March 3, 2012

Accepted: April 4, 2012

Published: May 15, 2012

doi:10.5430/afr.v1n1p162

URL: <http://dx.doi.org/10.5430/afr.v1n1p162>

Abstract

Mortgage-backed securities and derivatives pricing and risk models often assume static input distributions. As real-world uncertainty increases, the need for real-time data updates becomes imperative. Quality standards for pool level data would ensure the orderly re-pricing of risk. Many industries abide by government mandated quality data standards. We argue that what the financial industry needs is what the NIST already provides to manufacturing and the NASS provides to agriculture. The financial industry has evolved and now needs continuous monitoring framework for the securitization process to control the complex mathematical models and technological systems that enable disintermediation in the mortgage markets.

Keywords: Mortgage crisis, Quality data standards

1. Introduction

The finance industry relies upon the mortgage securitization process for fixed income investing. This process enables the transfer of debt from the original issuer—local banks and brokers—to institutional investors—mutual funds, pension funds and insurance companies. The securitization process is supported by complex mathematical models and extensive data for valuation. This process has created a secondary market for these products, allowing firms to tailor their portfolios to meet cash flow targets using evermore complex financial instruments.

The securitization process and its models worked well in a stable economy, when inputs into the models were (in effect) statistically stationary and loan risks were homogeneous across pools. The assumptions of static credit scores, consistent prepayment and default rates and stable asset prices may have seemed reasonable. Providing real-time data updates to these models throughout the chain of the origination-to-distribution model was thought to be unnecessary, since the securitization process created homogeneous loans that allow for slow moving global input values.

Another assumption was that securitization provided risk mitigation through diversification. During the financial crisis of 2008, however, the lack of a process for updating key input data for mortgages became evident, as the ability to sort pools of performing loans from underperforming loans became critical. The lack of data only postponed the inevitable re-pricing of pools with deteriorated credit. This resulted in a panic that effectively froze the mortgage-backed markets, in contrast to the orderly real-time re-pricing of risk in listed markets.

Lo and Mueller (2010) have challenged the stationarity assumptions so prevalent in MBS models (Note 1). They defined a five-level taxonomy as an alternative to that mathematical precision:

Level 1: Complete Certainty. Outcomes are deterministic. Uncertainty does not exist.

Level 2: Risk without Uncertainty. Known outcomes come from a known probability distribution. Through *a priori* deduction, no empirical data and no statistical inference are necessary. The probabilities have been derived.

Level 3: Fully Reducible Uncertainty. Uncertainty is due to unknown probability distributions. Given large amounts of empirical data, *a posteriori* statistical inference can arrive at outcomes and probabilities close to those derived in Level 2.

Level 4: Partially Reducible Uncertainty. Data generation processes are ambiguous. Distributions are unstable. While significant empirical inference may explain some of the outcomes and their probabilities *ab ambiguitate*, a significant amount of unexplainable uncertainty exists. This leads to model uncertainty.

Level 5: Irreducible Uncertainty. A state of total ignorance beyond reasoning prevails.

By disambiguating uncertainty in this way, we can address methods of justification for each level. The relevant question is: on what basis do we justify belief in the repeatability of a model's outputs? The problem is not unique to finance.

Other industries produce outputs which contain uncertainty (see Bilson, Kumiega, and Van Vliet, 2010). Manufacturing processes churn out physical parts that contain uncertain variations in size, but which are nevertheless within tolerance. The uncertainty in the size of machined parts is akin to Level 3. Manufacturing engineers use statistical controls to indicate when the machine is producing parts (or production lots) that no longer meet specification, that is, when the uncertainty has moved too far in the direction of Level 4. Unlike the clear boundaries between other levels, Level 3 is a continuum between Levels 2 and 4. MBS models assumed Level 2 stability and small production-lot-by-production-lot variation. Similar to machines, however, most financial models run in Level 3 due to uncontrollable environmental factors. The inputs and outputs of a given trading algorithm can at times be closer to Level 2 and at other times be closer to Level 4. The closer the model outputs are to Level 2 the more reliable its outputs; the closer to Level 4, the less reliable its outputs.

The agricultural industry runs much closer to Level 4. As weather, fertilizer costs, labor and financing costs, and demographics change, data must be updated almost continuously to make sense of the shifting distributions of random inputs and their impacts on commodity prices and farm values. The closer to Level 4, the more important real-time data updates become. As a previously stable economy deteriorates the uncertainty increases towards Level 4, prepayment (i.e. default and refinancing) probabilities become more ambiguous and heterogeneous by pool. In such a case, assumptions need to be continuously adjusted at the pool level to allow for a liquid market that continuously adjusts pool values.

Like airplane and automobile assembly lines, the risk and pricing systems that help estimate the value of mortgage backed securities and derivatives are production machines. Like agricultural crops, the valuations are dependent upon ambiguous environmental input data processes. Once we accept Lo and Mueller's taxonomy, then it is clear that the input data needs to be regularly updated so the assumption of stationarity can be validated. The root cause of the problem is not the models (nor are they the solution), but rather the lack of real-time data used as model inputs at the pool level (Note 2) after a mortgage or loan has been issued. The updating of input data at the pool level would be similar to continuously measuring production lots off an assembly line. The finance industry should adopt control mechanisms from other industries that have successfully implemented regulations and controls on stochastic processes. It is usually better to copy successful process controls that have worked for decades rather than re-inventing the wheel.

In this paper, we discuss uncertainty in MBS markets with applications of quality data standards to the financial industry in order to mitigate systemic risks. By analogy, we will discuss how other industries have used similar approaches to systemic crises. Applying quality and data standards is not new. We propose that the industry view the MBS markets from a quality framework and evolve to continuous monitoring in a fashion similar to the way manufacturing and agricultural industries have data standards and publicly available aggregated statistics.

Over the decades, quality techniques have progressed from the military standard (Mil-Spec) method of testing to continuous process monitoring. The same idea of continuously updating a process with new data is also applicable to the MBS market. Prior to the advent of securitization process, all financial risk was borne by the original issuer. After its advent, what third parties need is continuous monitoring and data to re-price securities daily. However, the continuous process to monitor existing pools using updated data and predict pool-by-pool prepayment and default was not created, even though the process had evolved from a single bank to a complex production facility churning out MBSs, CMOs, and CDOs.

Times have changed. Mortgage data systems (designed thirty years ago) were not built to handle the types and volumes of pool-level real estate and credit data available today. Credit rating agencies that continuously update individuals' ratings did not exist. Zillow did not yet provide real-time price estimates and comparables. Property taxes were not available electronically. The ability to calculate pool level statistics and distribute them via the internet was not possible. What the MBS system needs is to move forward to continuous quality monitoring systems, similar to other high-tech industries implementing ISO 9001 standards for stochastic output processes, such as chemical plants (see ASQ, 2002).

2. Background

Finance as a discipline is derived from the dynamics of postponing consumption (i.e. savings or investment) and receiving some monetary reward for it (i.e. interest rate). Throughout such inter-temporal reallocations of private consumption patterns, economic growth is achieved as funds get transferred from surplus-spending units into deficit-spending units in a well-functioning financial system. But unless all financial intermediation falls under the umbrella of direct financing (as was the case of the savings and loans associations in the 1970's and 80's), asymmetric information will plague this transfer of funds, as it occurs through financial markets. Simply, counterparties will not have identical information about each other.

Direct financing involves face-to-face transactions and gives the lender the opportunity to alleviate the presence of such asymmetries through the active collection of information. On the other hand, indirect financing, through securitization depends on the amount and quality of information in a passive way. Counterparties often substitute for this lack of direct control over information with quantitative models which, ironically, may require more robust sets of data (standardized, and systematically disseminated).

To be effective, information about consumer loans should have been collected regularly during the life of the loan, standardized and disseminated down to the pool-level for participants in all mortgage-related assets and the general public. The dissemination of the data to general public would result in academic research on the shift in the credit quality of pools with specific features which would have been a trigger to change the assumption that the world was in Level 2. Hopefully, the markets would have gone through a soft landing (as opposed to the current liquidity crisis) as the effective discount rate rose for the existing pools with poor credit. Sophisticated investors would have created relative value arbitrage strategies, which would have reduced the issuance of more loans of that had a high probability of default as identified by deteriorating credit scores of the average borrower.

3. Manufacturing Analogy

The federal government is already active in promoting quality data standards for many industries. The National Institute of Standards and Technology (NIST)'s "standards ensure that consumers are confident of the quantity and quality of the product purchased." "NIST supplies industry, academia, government, and other users with over 1,300 Standard Reference Materials (SRMs). These artifacts are certified as having specific characteristics or component content, used as calibration standards for measuring equipment and procedures, quality control benchmarks for industrial processes, and experimental control samples (NIST, 2011)."

In the early 1900's there arose a general understanding by the public of the dangers inherent in the designs of steamboat boiler, bridges and other structures. New regulations and standards were developed to prevent explosions (a common occurrence then), collapses and other failures. These are physical constructs that pose risks to physical health. The abstract and esoteric nature of MBSs has prevented recognition of the financial risks by the general public. Prior to 2008, most lay people likely assumed risk posed applied only to large financial institutions. But these firms' models did not account for the realities of bad input data; the designers had no fear of the systemic risks to societal financial health. Some understood the risks. Hedge fund manager John Paulson profited from the collapse.

For example, bolts are used in many industries, from mission-critical uses in bridges and airplanes to everyday uses in furniture and appliances. Correct manufacturing of bolts requires consistent-grade alloys as inputs and consistent quality of the finished product. The quality of the bolts is determined by the dimensions and metallurgical properties. Sampling and testing techniques confirm that bolts meet standards, and because of this we can trust that our planes and bridges won't fail. Two-inch, hex-head, quarter-inch diameter, 10/32-inch thread bolts, of course, are not homogeneous. Rather, their properties are statistically controlled to ensure each bolt falls within levels of tolerance. The composition of the bolts—the case hardness and depth, and metallurgical properties, such as 4130 carbon steel or 316 stainless steel—is certified and sent to buyers (Kumiega and Van Vliet, 2008).

Now, imagine what would happen if manufacturers stopped disclosing data on their bolts because it was too expensive, and all the bolts with similar external measurements were placed into a giant bin, with orders being filled by picking bolts out of the bin. What would happen is that a low-quality bolt would be used in a critical application, such as a

bridge. The risk of a bridge collapse in the future would be substantially higher than if today's rigid quality standards were used. Of course, no manufacturer would assume such risk, and any industry that used bolts would grind to a halt.

What holds true for bolts also holds true for many industries—pharmaceuticals, chemicals and food processing to name a few. It should also hold true for finance, where quality is based solely on measurement of external features of financial instruments. Firms in all of the aforementioned industries are required to disclose information on the internal composition of their products, except finance. The firms' testing and reporting methods must also meet national quality standards before they can sell their products to the general public. Again, this is not the case in finance, which is not required (either legally or ethically) to abide by any standard of quality whatsoever.

Due to the lack of standards in finance (particularly in the structured debt markets) there was a liquidity crisis because the financial machines stopped working properly when the input distributions became (Level 4) ambiguous.

Financial firms manufacture complex products—Collateralized Debt Obligations (CDO) and Credit Default Swaps (CDS)—to sell to public and private pension funds, with neither national standards for disclosure about the internal composition of those products nor performance metrics at issuance. Furthermore, the data collected is not updated at intervals over the life of the loan. This is counter-intuitive given the daily pricing and trading of these instruments.

4. Agricultural Analogy

The federal government is active in calculating and disseminating statistics for particular industries. The USDA's National Agricultural Statistics Service (NASS) "provides timely, accurate, and useful statistics in service to U.S. agriculture." According to the NASS website (USDA, 2011), the NASS conducts "hundreds of surveys every year and prepares reports covering virtually every aspect of U.S. agriculture. Production and supplies of food and fiber, prices paid and received by farmers, farm labor and wages, farm finances, chemical use, and changes in the demographics of U.S. producers are only a few examples."

The NASS's focus is "on the latest official statistics that help to minimize the uncertainties and risks associated with the production and marketing of commodities. NASS reports are often used directly and indirectly by farmers, producer organizations, agribusinesses, researchers, policymakers, and government agencies." With respect to the privacy of personal data, the NASS "safeguard[s] the privacy of farmers, ranchers, and other data providers, with a guarantee that confidentiality and data security continue to be [their] top priorities."

The NASS's weekly Crop Progress Report publishes data on corn, soybeans, wheat and other crops in the U.S. The report contains unbiased data for commodity investors as crop conditions evolve over the course of the season. The report also provides historical data for the prior year and five-year averages. This is all good data for industry observers to research and value commodity prices despite Level 4 uncertainty.

Particularly analogous is the farmland bubble of the 1970s.

For a number of years in the 1970s, American landowners saw the prices of farmland increase in value beyond the capacity of the land to produce adequate income to pay for the loans. However, through refinancing purchased land, producers were able to buy even more farmland. When inflation slowed, overleveraged producers and the rural banks that had made the loans faced a financial crisis. By 1987, farmland values bottomed out after a 6-year decline.

In light of the 1980s farm financial crisis, NASS responded with what is now known as the Agricultural Resource Management Survey (ARMS). It's the data source for annual estimates of the financial health of farming households and cost-of-production estimates for major agricultural commodities. The data are critical to the research and analysis mission of USDA's Economic Research Service, and are a key input for evaluating and revising Farm Bill programs (USDA, 2005).

Where over-leveraged land owners once pushed rural banks to the brink of collapse, the ARMS data now "provide a direct linkage between commodity production practices...and the financial status of the farm and its operator's household (USDA, 2005)."

The value of farmland is a derivative of the income produced by farming activity. The key indicators of income are other derivatives—the health of farming households and costs of production. Given the well-understood dynamics of farmland and commodity valuation, investors can update pricing models with real-time income forecasts, thereby preventing a bubble in land prices before it led to a crisis. We might ask: why didn't the farmland and farm mortgages blow up in the 2008 crisis? The answer appears to be that with updated data investors, loan officers and loan buyers were able to properly price land using mathematical models. Market participants knew the inputs to the farm cash flow model, and thus the industry was able to properly value farmland.

The analogy between overvalued farmland of the 1970's and the residential housing bubble of the 2000's presents a strong justification for better data dissemination in the MBS market. There is one difference that makes this need even stronger in the housing market. In the farmland crisis, a direct financing relationship between borrower and lender pushed rural banks into crisis before the government intervened. In the real estate crisis, the lender has been twice-removed (through securitization) from data about the borrower, making the need for ARMS-like services that much stronger, and its implementation all the more justified.

This is why the ARMS's direct linkage between commodity production and financial status of farms is so vital to farming industry sustainability. A similar implementation in the housing market is in the best interest of the government, the FDIC and taxpayers, who ultimately shoulder the burden of crises in both farmland and housing through subsidies or bailouts. In the same way that farmland value is based upon income from farming, residential home values is based upon household income. The debt-to-income on the mortgage pool is no different than average income per farmer in a county. Pool level income information should be disclosed.

In Bernanke (2004) a containment to the economic problem is prescribed in quality terms. Now, the industry needs to get to the root cause of the problem, which is data dissemination. Armed with real-time data from the Great Depression, the Fed was able to structure a soft landing and avoid a 1929-style collapse. If mortgage pool data were to be available to the private sector, the markets would re-value the optionality of waiting for more information in an orderly fashion, as irreversible investment decisions imply (see Bernanke, 1983).

Information will bring investors back to the table. If the assumption that a lack of transparency was a root cause, then a solution would be to have the Fed create a standard model and a standard set of data. This is similar to the Options Clearing Corp.'s role producing valuation and risk metrics for all listed equity options. These values are used to calculate firms' risk requirements. This method has worked well since 1972 and created a liquid derivatives market that is a model for all derivatives exchanges.

5. Quality Data Standards

The U.S. government has promoted quality since W. Edwards Deming first oversaw its implementation during World War II. Today, almost all products that are distributed to the general public must undergo rigorous quality inspections. History has shown that high-quality firms gain worldwide market share and become industry leaders. If strict quality standards for MBS data and its distribution were enacted, the U.S. financial industry would thrive like all other high quality industries (Kumiega and Van Vliet, 2008).

As mentioned, the manufacturing, chemical, pharmaceutical and food processing industries all provide detailed information on quality measurements and testing to the federal government. These industries employ mathematics and technology that are at least as complex as those found in the financial industry.

Data standardization needs to be undertaken by a governmental body similar to the NASS. The Federal Reserve, for example, could thus provide a function similar to the Options Clearing Corp. For manufacturing, the National Institute of Standards and Technology (NIST) performs these activities already. The mission of the NIST is to "promote U.S. innovation and industrial competitiveness by advancing measurement science, standards, and technology in ways that enhance economic security and improve quality of life (NIST, 2011)."

The step that follows standardization of information is data dissemination. Pool level data needs to be standardized in such a way that markets can extract signals of trouble from financial instruments independent of regulatory agencies or other intervening mechanisms. Prior to the crisis, information was primarily distributed to existing regulatory bodies, which lacked the mandate to provide updated input data. Furthermore, prior to the crisis, the average American would have objected to the release of their private information to for-profit private banks and insurance companies, even only for use in the calculation of aggregate statistics. Today, there is broader acceptance of the need to keep markets functioning through data transparency, much in the vein of the agriculture analogy.

6. Independent Validation

If two pools existed both with a 700 average credit score, and one pool went to 750, while the other went to 650, then MBSs based on these two pools should have different prepayment probabilities and prices. If this data was available, additional research as checks and balances would have reconciled the changes between credit states and risk, for example. Research would have found the fact that the pools were becoming heterogeneous. We would have seen the effects. The data is available, but the communication between the banks, the vendors, the investors and borrowers for the release of data to be aggregated never evolved.

In the absence of continuous mark-to-market validation of these models, the root cause of the credit crisis was not the models themselves, but rather the lack of reporting standards for the data these models use. This data could have been used by third parties for proxy validation. If the data was free and available, third parties and academics could research tail risk. In the future, data should be in the public domain similar to the agriculture reports published by NASS, but on a pool-by-pool or security-by-security basis. Financial firms would still be free to create derivative instruments that do not require data updating. As a form of disincentive those instruments could not be purchased, for example, by any public entity or government-backed pension fund.

Standardized current data would allow purchasers of MBSs to calculate their own valuations and implement specific risk measurements and forecasts. These activities are especially important as the compositional properties of the products change. As with so many other industries, the federal government should require national reporting standards on the composition of manufactured investment products. For example, for mortgage-related instruments, required aggregated statistics could include:

- Median credit score of the individuals that have been pool.
- Standard deviation of the current credit scores of the individuals that comprise the underlying components.
- Range between the 25th and 75th percentile credit score.
- Difference between the minimum and maximum credit score.
- Number of individuals with self-employed income.
- Difference between the appraised value and the property tax value.
- Credit statistics for prepayment groups.
- Income by pool in a fashion similar to the NASS reports.
- Change in tax values of properties.
- Ratio of debt to income.
- Ratio of Zillow price to income.
- Any statistic that can sort prepayment from foreclosure to refinancing.
- Ratio of cash profit per house or loan outstanding above or below some watermark versus income.

7. Conclusion

The financial industry has mathematical models. Those models have input data. Model risk rears its ugly head when *a priori* (Level 2) assumptions become inconsistent with empirical realities. Across the Level 3 continuum, the input data changes over time, approaching Level 4 and heterogeneity. Thus, the model assumptions are no longer supported by the ambiguous data. Of insufficient focus has been the contribution structural processes make to system-wide risk. The industry never built the infrastructure to provide input data as it changed over time.

The never-built process to certify input data is very similar to using the wrong kind of steel to make bolts. If the right MBS data were available in a quality framework, the world would come again to the U.S. to trade these products. Data standards would increase interest rate costs due to the cost of building and operating the collection and aggregation systems. However, the end result of higher quality data would ultimately lower rates by removing part of the uncertainty involved in the continuing financial disintermediation of the mortgage-lending process.

Viewed through the lens of quality, MBS models are machines that have only produced out-of-specification products due to poor quality materials and standards. These products are now costing the American taxpayer billions of dollars in rework costs to make them into quality products again.

Quality data standards would ensure that the general public can have faith in their investments. Our hope is that the perspective of mathematical models as machines with Level 3 inputs will allow market participants and regulatory agencies to frame the current financial issues properly as crises of quality. We hope the financial community will adopt the quality mantle, as other industries have. The adoption of quality in the American financial industry will once again show it can be a leader and innovator in the arena of international finance.

What the American financial industry needs is what the NIST provides to manufacturing and the NASS provides to agriculture: quality data and documentation based on national standards. The financial industry needs a quality framework for the process of manufacturing and controlling complex mathematical models, technological systems and financial disintermediation.

References

- American Society for Quality (ASQ). (2002). *ISO 9001 Guidelines for the Chemical and Process Industries, Third Ed.* ASQ Chemical and Process Industries Division.
- Bernanke, Ben S. (1983). Irreversibility, Uncertainty, and Cyclical Investment. *The Quarterly Journal of Economics*. 98(1), pp. 85-106. <http://dx.doi.org/10.2307/1885568>
- Bernanke, Ben S. (2004). Monetary Policy Alternatives at the Zero Bound: An Empirical Assessment. *Brookings Papers on Economic Activity*. 2004(2), pp. 1-78. <http://dx.doi.org/10.1353/eca.2005.0002>
- Bhardwaj, Geetesh, and Rajdeep Sengupta. (2008). Did Prepayments Sustain the Subprime Market? *Research Division, Federal Reserve Bank of St. Louis, Working Paper 2008-039A*. pp. 1-44.
- Bilson, John F.O, Andrew Kumiega, and Ben Van Vliet. (2010). Trading Model Uncertainty and Statistical Process Control. *Journal of Trading*. 5(3), pp. 39-50. DOI: 10.3905/jot.2010.5.3.039. <http://dx.doi.org/10.3905/jot.2010.5.3.039>
- Brunnermeier, Markus K. (2009). Deciphering the Liquidity and Credit Crunch 2007 – 2008. *The Journal of Economic Perspectives*. 23(1), pp. 77-100. DOI:10.1257/jep.23.1.77. <http://dx.doi.org/10.1257/jep.23.1.77>
- Fabozzi, Frank J., and Franco Modigliani. (1992). *Mortgage and Mortgage-Backed Securities Markets*. Harvard Business School Press.
- Kumiega, Andrew, and Ben Van Vliet. (2008). Perspectives: In Crisis, Give Credit to Quality. *Quality Progress*. December.
- Lo, Andrew and Mark Mueller. (2010). WARNING: Physics Envy May Be Hazardous To Your Wealth! *Journal of Investment Management*. 8(2).
- National Institute of Standards and Technology (NIST). (2011). [Online] Available at: www.nist.gov/public_affairs/mission.cfm
- USDA National Agricultural Statistics Service (NASS). (2005). An Evolving Statistical Service for American Agriculture. Available at: [Online] www.nass.usda.gov/index.asp
- USDA National Agricultural Statistics Service (NASS). (2011). [Online] Available at: www.nass.usda.gov/index.asp

Note 1. We will use the term MBS to refer to mortgage TBA pools, securities based upon those pools, and securitized tranches and derivatives thereof (i.e. CMOs, CDOs, etc.). Our conclusions and recommendations are assumed to permeate the structure and process of mortgage securitization.

Note 2. Individual data should never be made available after the inception of a loan. Dissemination of aggregated mortgage pool statistics would be consistent with the U.S. Dept. of Agriculture's NASS Agricultural Resource Management Survey, which publishes aggregated statistics on the financial health of farming households.