

POLICY PERSPECTIVES



ABILITY-TO-REPAY BENCHMARK PROOF OF CONCEPT FOR AN INDUSTRY STANDARD SETTING ORGANIZATION (ISSO)

by Richard Cooperstein

Prepared for: **Structured Finance Association**

JULY 2021

ANDREW
DAVIDSON
& CO





03	OVERVIEW
05	EXECUTIVE SUMMARY
05	A. FRAMEWORK: METRICS AND DRIVERS → COHORTS AND BENCHMARK
06	B. SUMMARY OF RESULTS
07	DATA ANALYSIS
07	A. DATA SELECTION AND ANALYSIS
08	B. METRICS: DEFINING ABILITY-TO-REPAY
09	C. DRIVERS: SELECTING PERFORMANCE INDICATORS
11	D. MEASURING PERFORMANCE OF COHORTS
16	ESTABLISHING A BENCHMARK
19	APPLYING THE PROPOSED BENCHMARK TO THE NON-QM MARKET
20	CONCLUSION

Mortgages are complex long-term financial instruments crucial to buying homes and accumulating wealth for the working and middle class. In a dynamic economy like that of the United States, sources of income often change for prospective homeowners and borrowers in a constantly evolving job landscape. This dynamism is a mainstay of American economic growth. However, with such a dynamic economy coupled with a strong focus on homeownership as a social goal, the mortgage market is susceptible to damaging boom and bust credit cycles, as well as varying and imperfect access to credit for underserved populations.

The most notable such event was the lending environment leading up to the 2008 financial crisis. Efforts to make loan terms more flexible for borrowers quickly devolved into risky and damaging practices that harmed America's most vulnerable communities. Additionally, rampant and unchecked investment speculation—often fueled by loans with little to no income verification requirements—led to the worst financial crisis in 100 years. Once the dust settled, Congress passed the Dodd-Frank Act, creating enhanced oversight for the mortgage market as well as the Consumer Financial Protection Bureau to enforce the new rules. Ever since, the challenge facing both regulators and the market has been how to find a way to effectively serve all communities in a dynamic economy without running afoul of protections for consumers. Unfortunately, little progress has been made on this front in the past thirteen years. Instead, algorithms at Fannie Mae and Freddie Mac (the “GSEs”) serve as the de facto credit box given the favorable regulatory treatment these enterprises receive, while the remainder of the market has seen only limited innovation in finding ways to safely serve all creditworthy borrowers due to the less favorable regulatory framework for non-GSE loans.

A mainstay of the new rules was the concept that lending must not be based solely on the value of the underlying collateral, but on the borrower's capacity to successfully repay the loan. Such clear but flexible Ability-to-Repay (ATR) requirements were a key regulatory addition after the financial crisis, and today they are an important component of providing stability to the market while allowing for the extension of credit to underserved markets. While innovation has stagnated, much progress has been made since the financial crisis towards the goal of ensuring maximum safety in lending. For example, two of the largest drivers of the mortgage crisis have been addressed: risky mortgage contracts have largely disappeared, and the market generally requires conscientious documentation of income.

The regulatory framework has established much clearer standards on what originators should not do, but work remains to level the playing field for all origination and funding channels and to clarify what degree of innovation in underwriting practices could be utilized to expand access to credit. For example, greater clarity about the use of compensating factors to increase access to and use of the additional consumer credit data that has become available in the last 15 years would be valuable. Finally, originators, investors, and borrowers would all benefit if it were clear at origination whether a loan met the ATR standard.

For its part, the Consumer Finance Protection Bureau (CFPB) has outlined objectives for an ATR standard which include ensuring fair access, preventing consumer exploitation, leveling the playing field for lenders, and encouraging innovation. These objectives are appropriately timeless, but defining and implementing policy that

is both effective and practical is a great challenge. The GSEs have spent billions on their automated underwriting systems (AUS) and fulfillment infrastructure; their underwriting guides are hundreds of pages long. However, continuing to rely on GSE underwriting systems through the so-called “GSE patch” is not desirable, as it could suppress innovation, decrease competition, and hinder a vibrant private market. Equally important, as steps are taken to eliminate reliance on the GSEs, careful steps must be taken to avoid the risk of ATR policies becoming either complicated and intractable or overly simplistic.

The purpose of this analysis is to provide an example approach for establishing a benchmark that can be used to analyze the performance of various origination and underwriting methods when determining a borrower’s ability to repay.

For example, while the benchmarks presented in this analysis are derived from fully documented loans using traditional documentation and underwriting, the benchmarks themselves can be used to evaluate any underwriting methodology.

Establishing a benchmark framework is the first step towards a durable structure that is fair and efficient. Among other issues, there is uncertainty about fulfilling the CFPB’s “document and verify” requirements fairly and without rebounding legal risk. The Structured Finance Association (SFA) is currently considering establishing a Mortgage Standards Setting Organization to help the industry analyze borrower ability to repay and help validate ATR requirements in both substance and spirit through a data- and performance-driven approach.

A. Framework: Metrics and Drivers → Cohorts and Benchmark

SFA's long-term objective is to work with participants across the housing finance industry to establish an efficient and verifiable means to validate compliance with the ATR rule that has the confidence of the CFPB and the market. As a framework for building a vibrant, efficient, and sustainable private market for loans that can be empirically shown to have effectively "considered and verified" borrower's ability to repay, we propose an **ATR Benchmark** that rests on a framework comprised of the **Metrics** and **Drivers** of mortgage performance.

Performance **Metrics** (delinquency) are how we measure household Ability-to-Repay over time. **Drivers** are the risk variables (credit scores, DTI, etc.) that are predictive of performance metrics and can reliably forecast a borrower's ability to repay.

We evaluate a set of metrics and drivers that is straightforward and relevant to (most) fully documented mortgage lending. Among other things, it enables assessment of the prior and existing ATR standards in relation to this objective, data-driven framework. Originators and policymakers can compare innovative underwriting approaches with the traditional underwriting-derived benchmark.

The underlying concept is as follows:

- First, fully documented, traditionally underwritten federal loans are carefully analyzed and aggregated into risk cohorts by drivers of performance to establish ATR performance benchmarks.
- Next, cohorts of loans that are not automatically deemed to be in compliance with the CFPB "safe harbor" are compared to the performance of loans that are within the safe harbor by using the established benchmarks.
- Finally, these performance metrics are compared across the business cycle to demonstrate that, if properly done, innovative underwriting practices can be objectively measured and compared to augment the determination that a borrower was adequately evaluated for his or her ability to repay.

B. Summary Results

Figure 1 summarizes the results from this investigation. We considered several metrics and concluded that 60 days or more delinquent at 24 months of loan age (“60+ in 24”) is the most effective and easily measurable metric of consumer Ability-to-Repay amongst the metrics we reviewed. We evaluated several drivers, and the empirical results show that FICO1 and DTI are most predictive of near-term consumer performance for traditionally underwritten loans.

FIGURE 1: METRICS AND DRIVERS			
Metrics		Drivers	
30+ DQ	Recovery rates are too high	FICO	Best short term DQ predictor
60+ DQ	About right, industry standard	DTI	Moderate DQ predictor
90+	Fewer observations, longer horizon	LTV	Poor short term DQ predictor
Ever DQ	Recovery rates are too high	Balance	Poor DQ predictor
DQ in 24	Well-targeted and commonly used	Coupon	Indirect effect, correlated

The paper proceeds by evaluating potential metrics and drivers, constructing cohorts, and finally recommending a potential benchmark for SFA member and mortgage industry consideration. Focusing on the most predictive measure(s) narrows down the bewildering array of potential choices to a small set of reasonable ones.

While the benchmark is based on traditional drivers of performance for fully documented loans, the benchmark can be used to evaluate loans originated under innovative approaches. In essence, the fundamental idea of this approach is that the new underwriting techniques should be able to produce ATR performance that is as good or better than the benchmark performance. In that way, the market is free to innovate while having a clear performance benchmark that is sensitive to market conditions to measure the success of various approaches.

1 The Proof-of-Concept datasets included FICO 5 which was released in 2004. However, FICO has released several updated versions since then and VantageScore also provides credit scores that are widely used.

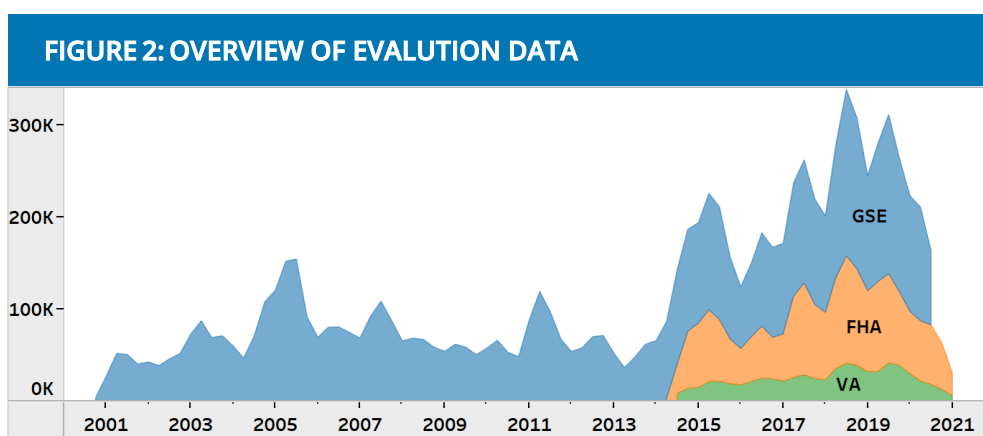
A. Data Selection

For this proof of concept, we used loan level data from the GSEs and from GNMA for FHA and VA loans.² The GSE data spans 2000–2020 and the FHA and VA data span 2014–2021. Financial benchmarks generally focus on basic assets to make comparisons easier for any permutation of related assets. One example is the S&P 500 index; it is simple, widely available, and straightforward to compare to individual stocks or esoteric ETFs for returns or risk. For mortgages, we propose that traditionally underwritten loans with extensive historical performance comprise the population to construct an appropriate benchmark. An example of how this benchmark can be used will be described below.

This data underlying our analysis is 30-year fixed-rate, fully documented loans excluding those with layered risk factors such as cash-out and investor loans to retain focus on traditional home mortgage lending.³ We used loans aged 22–26 months to limit the impact of

external factors, because our focus is consumer ability to repay, not ultimate losses. The cutoff doesn't have to be exactly 24 months, but this is consistent with the metrics used by the CFPB. The impact of external economics rises with longer time periods, and delinquencies sooner than 12 months are often considered defective originations.

Figure 2 shows observations over time for a 10% random sample of the qualifying loan pool, totaling about nine million loans. We combined Fannie Mae and Freddie Mac loans into one group because GSE loans perform similarly to one another from a credit perspective, based upon our research. FHA and VA loans are included but separately identified because we believe they may perform differently even with similar underwriting characteristics. We supplement our in-depth consideration of traditional mortgages with summary results of subprime mortgages from the mid-2000s.



² Fannie Mae Single-Family Loan Performance Data | Fannie Mae, http://www.freddiemac.com/research/datasets/sf_loanlevel_dataset.page, http://www.embs.com/public/html/gnm_loan_dailymon_v1.8_Feb2021.pdf

³ We also excluded extreme values such as DTI > 60, LTV > 105%, FICO < 600 and > 800, and missing values.

B. Metrics: Defining Ability-to-Repay

Our approach is consistent with CFPB's perspective⁴ as expressed in the 2019 Ability-to-Repay and Qualified Mortgage Rule Assessment Report, with a couple important refinements. The CFPB writes:

Because the affordability of a given mortgage will vary from consumer to consumer based upon a range of factors, there is no recognized metric that can directly measure whether the terms of mortgage loans made after the Rule's effective date reasonably reflect consumers' ability to repay.

This analysis instead measures a proxy for the lack of ability to repay across a wide pool of loans by considering the frequency of early borrower distress, measured as whether a borrower was ever 60 or more days past due within the first two years after origination.

This measure is referred to as the "early delinquency rate" in the analyses in this chapter. The focus on early delinquencies is intended to capture borrowers' difficulties in making payments soon after the origination of the loan, even if these delinquencies do not lead to a borrower potentially losing their home.

While the CFPB recommends **ever-60+** before 24 months, we recommend **60+ DQ at 24 months** as the ATR Metric because many early delinquencies cure. (Fannie Mae data⁵ shows that about 1/3 of early 60-day delinquent loans improve before they are 24 months old.)

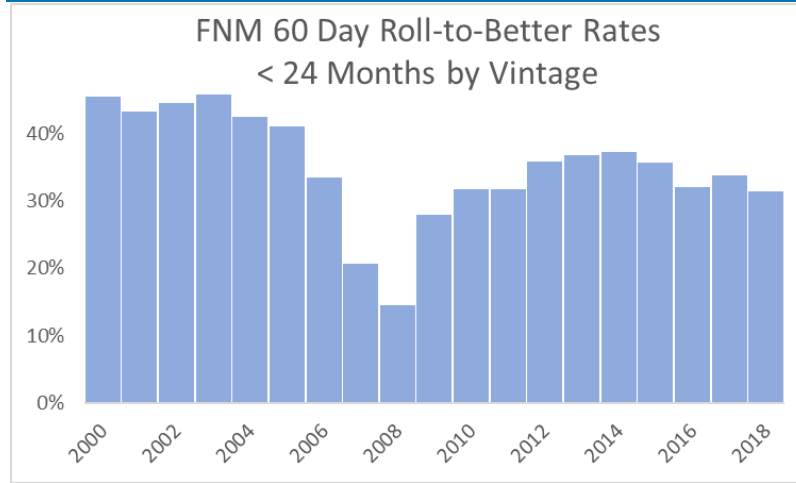
Many households who become delinquent when facing early income interruptions and occasional large expenses become current before 24 months. Thus, it could be premature to conclude ATR violations are present for 60+ delinquent loans younger than 24 months old, and we believe considering borrower ability to repay over longer time periods is more appropriate for evaluating performance. Recent experience with natural disasters and Covid reinforces the idea that many borrowers who experience short-term disruptions are still able to afford their loans over longer horizons.

Figure 3 shows roll-to-better rates for FNMA early 60-day delinquent loans by vintage for each of the last 20 origination years. We focus on recoveries in months 9–24 to achieve sufficiently large samples. Recoveries are impacted by extreme external economics, as reflected by the lower recovery rates for the vintages from the peak of the last financial crisis.

⁴ https://files.consumerfinance.gov/f/documents/cfpb_ability-to-repay-qualified-mortgage_assessment-report.pdf, pp. 83–84

⁵ <https://capitalmarkets.fanniemae.com/tools-applications/data-dynamics>

FIGURE 3: REPERFORMANCE RATES OF 60-DAY DELINQUENT



FNMA loans that become 30 days delinquent before 24 months recover at an even higher rate, so ever-30 would not be a good ATR metric. Loans that are 90 days delinquent recover approximately 15% of the time, so

while the 90-day standard could work, it is somewhat less suitable than 60+. There are half as many early 90+ observations, which reduces statistical confidence for cohorts, especially within 24 months.

C. Drivers: Selecting Performance Indicators

Having defined the desired performance metric (60+ in 24), the next step is to determine an effective but manageable set of drivers. CFPB rightly points out that mortgage credit risk is multi-dimensional, and we show that combining common underwriting drivers improves results. We proceeded by evaluating how well several potential drivers measured at origination presage 24-month serious delinquency across the 20-year sample.

Our approach follows the methodology adopted by the CFPB in the Assessment Report.

For purposes of this assessment, the Bureau assumes that the average “early delinquency rate” and “early foreclosure rate” across a wide pool of Qualified Mortgages (QM) are probative of whether QM loans reasonably assure repayment ability, and that the dependence of these rates on the defining characteristics of QM loans is probative of how those characteristics may influence repayment ability. Likewise, the average “early delinquency rate” and “early foreclosure rate” among a wide pool of non-QM loans are probative of whether such loans reasonably assure repayment ability.⁶

⁶ *Op. cit.*, Assessment report, p. 84.

Figure 4⁷ shows delinquency levels for loans stratified by credit score, DTI, LTV, and loan size. Delinquency rates are shown by the height of the bars, while the lines show the average value of the stratification metric for that bucket. DQ rates are most sensitive to FICO score, moderately sensitive to DTI, and fairly insensitive to LTV even though LTV is strongly related to mortgage credit losses. The

jump in DQ rates for 90 and 100 LTV loans reflects the concentration of FHA and VA loans with lower credit scores. Finally, data suggests that lower balance loans are generally not significantly riskier. While performance cannot be fully reflected by just one or even two credit drivers, we can reasonably focus cohorts on FICO/DTI to provide an indication as to the construction of an ability-to-repay benchmark.

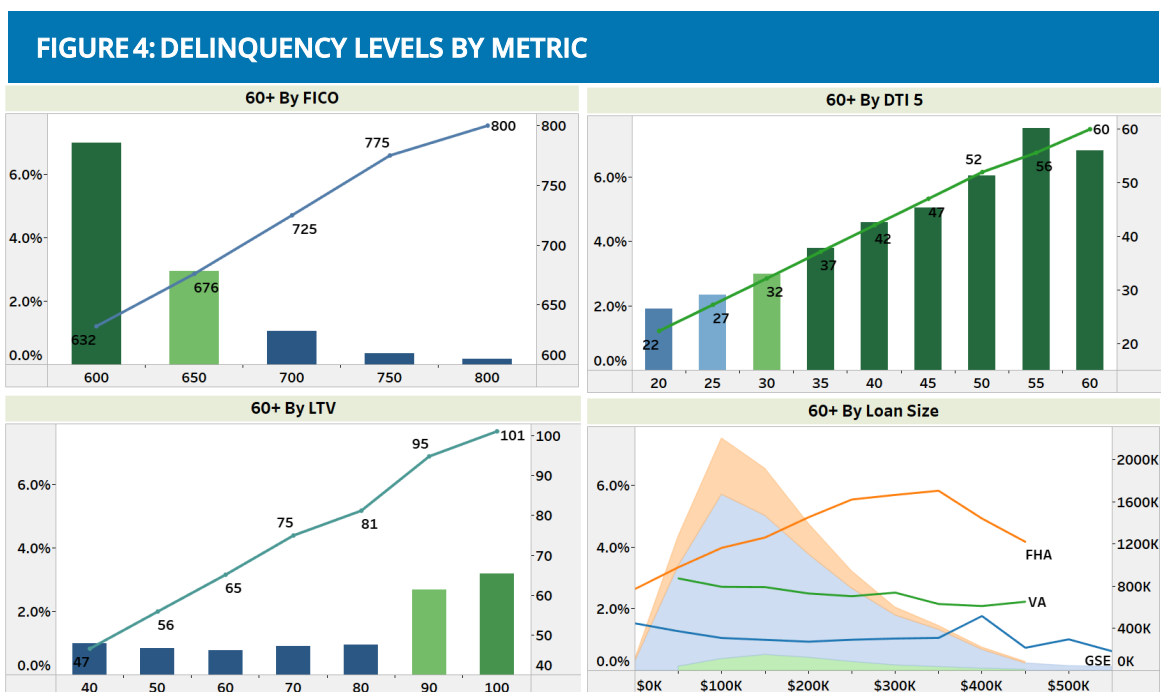


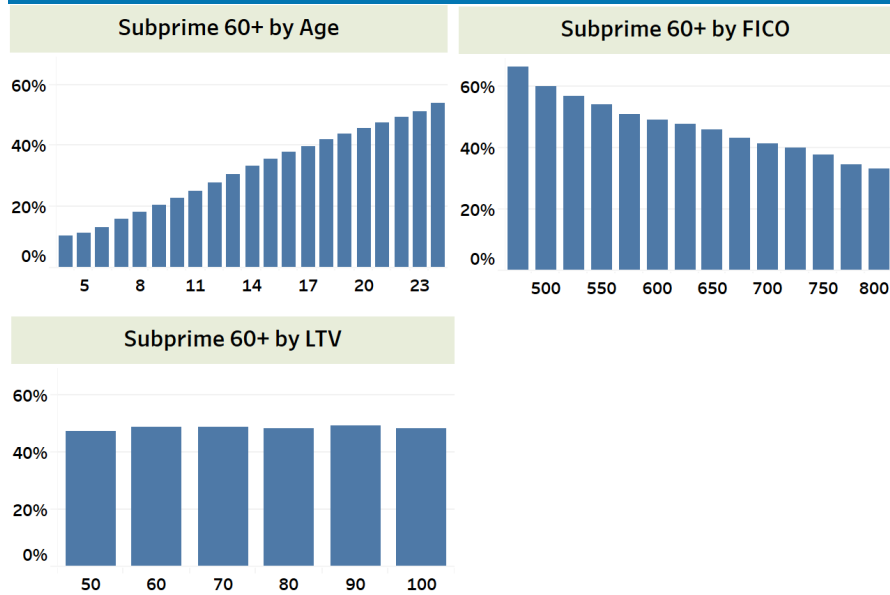
Figure 5 shows 60+ in 24 rates for subprime loans originated from 2004–07,⁸ about 7 million loans. Once again, DQs are sensitive to credit score but less so to LTV. Notably, during the same period, delinquency rates for subprime loans were about five times higher than government lending at comparable FICO and LTV values.

These loans would not meet current ATR requirements due to restrictions on product type, and the poor performance of these loans (illustrated below) confirms a foundational premise of an ATR standard, namely the manufacturing integrity of mortgages. It is not meaningful to focus on specific values of credit drivers unless first ensuring that income, reserves, and house values are reasonably documented and verified.

⁷ Bucketed values for FICO, DTI, and LTV categories are shown at the bottom of the range. The average is generally midway between; so, for FICO: 625, 675, etc., Figure 4 shows the actual averages in greater detail.

⁸ Intex

FIGURE 5: SUBPRIME DELINQUENCY RATES

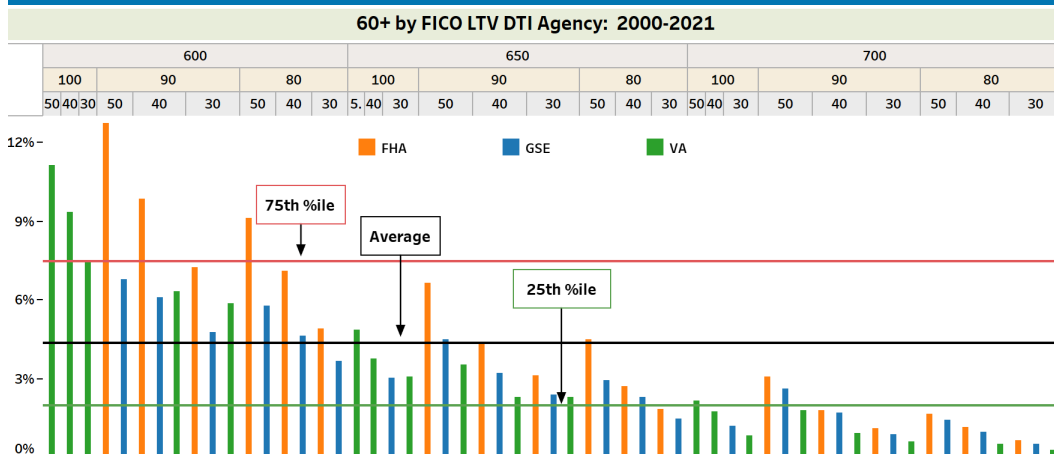


D. Measuring Performance of Cohorts

We selected the performance metric and the pair of drivers that are most predictive of performance. The next step is to construct cohorts across a wide range of FICO/DTI combinations and measure 60+ at 24 performance. Figure 6 shows DQ rates for over 50 cohorts by FICO/LTV/DTI and Agency. We further identify the unweighted average, 25th

and 75th percentile 60+ DQ rates. There is some variation across Agencies, but the data runs across different time frames (we also compare Agency DQ rates across the same time frame later in the paper). DQ rates generally vary by these drivers, with FICO dominating.

FIGURE 6: DELINQUENCY RATES BY COHORT



Next, we consider Agency level cohorts in more detail, and thus by before and after 2014, when FHA and VA loans became available in our dataset. Results are shown to the latest common date, mid-2020, and through Q1 2021 for FHA/VA data. We then examine performance across various driver combinations to illuminate the sets that best correlate to delinquency, and then finally the set that could reasonably serve as the ATR benchmark.

Figure 7 breaks down performance for the major channels by three time periods. We also identify the **FHA Average** cohort: this is a narrow cohort of FHA loans that generates about the same delinquency rate as the entire FHA portfolio. The non-linear relationship between metrics and drivers means that the **FHA Average** cohort will have “riskier” characteristics than the average loan in the FHA portfolio. This candidate for the ATR benchmark has an average credit score of 675, DTI of 45, and LTV of 95. The proof of concept is to construct this benchmark using loans for each Agency and then determine if performance is consistent across them. A useful benchmark will have consistent performance across the sub-markets. The following are the details of the data sets.

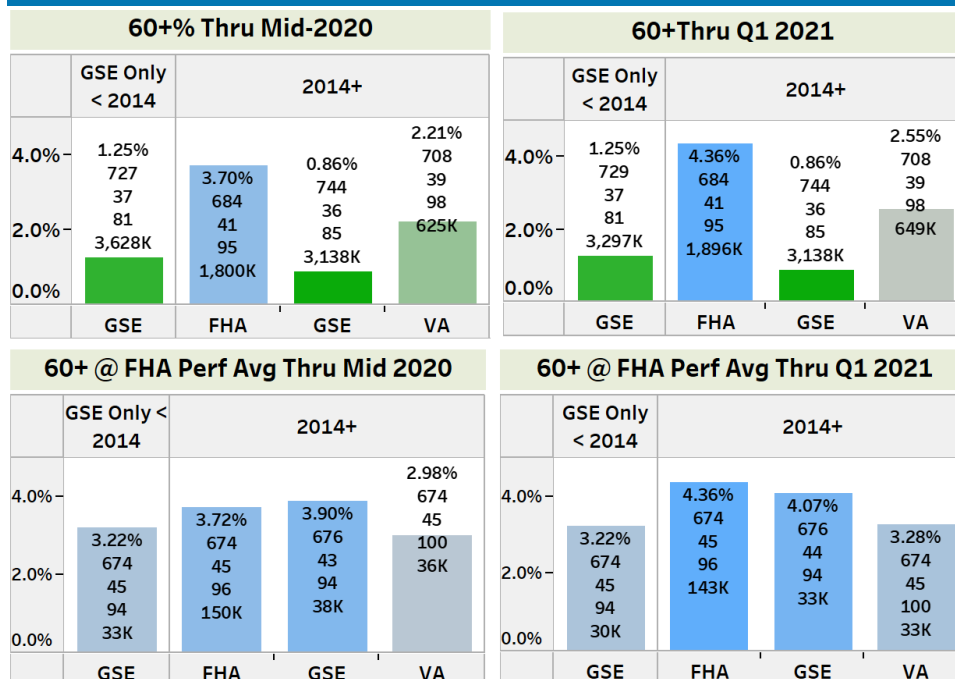
- Up to 2014 for GSEs and before FHA and VA data are available
- 2014–mid-2020 when all three datasets are available⁹
- Adding FHA and VA data for 3Q 2020–1Q 2021

The top two panels in Figure 7 show unadjusted DQ rates and associated data for FICO, DTI, LTV, and Count. GSE DQ rates are consistently low, FHA and VA loans have higher DQ rates through mid-2020, and they rise 15–20% further when including the latest nine months. The significant differences in key credit drivers such as credit score and DTI generally account for the differences in DQ rates.

The lower panels show results for the **FHA Average** cohort for each Agency. Delinquency differences between Agencies narrow dramatically, in the range of 3% to 4%, with 2014 GSE and FHA DQs differing by only 0.18% through mid-2020, before the impact of the pandemic. Even with the pandemic included, there is consistent performance between the Agencies. These results begin to define a cohort that can serve as benchmark. We note that this GSE cohort accounts for only about 1% of the GSE sample.

⁹ GSE loan level performance datasets are updated with varying timeliness.

FIGURE 7: DELINQUENCY RATES BY TIME PERIOD AND AGENCY



Given these results, why can't the benchmark be 4–5% 60+ rate? The CFPB recognized that a specific cutoff of delinquencies would not work for all economic environments.

... this analysis does not define or otherwise identify any acceptable limits of delinquencies and defaults for QM and non-QM loans. Delinquencies are measured but are not assessed against any assumed benchmark. Defining or otherwise identifying benchmarks for acceptable levels of delinquencies for new loans...is difficult in part because the level of delinquencies at a given time (and thus for vintages of loans made around that time) will depend not only on the characteristics and underwriting of the loans themselves but also on the subsequent health of the economy as a whole.¹⁰

Yet even with this caveat, we can draw from the CFPB reports an indication of what they view as acceptable levels of delinquency for loans that meet the ability-to-repay requirements, recognizing that these levels can vary significantly in different environments.

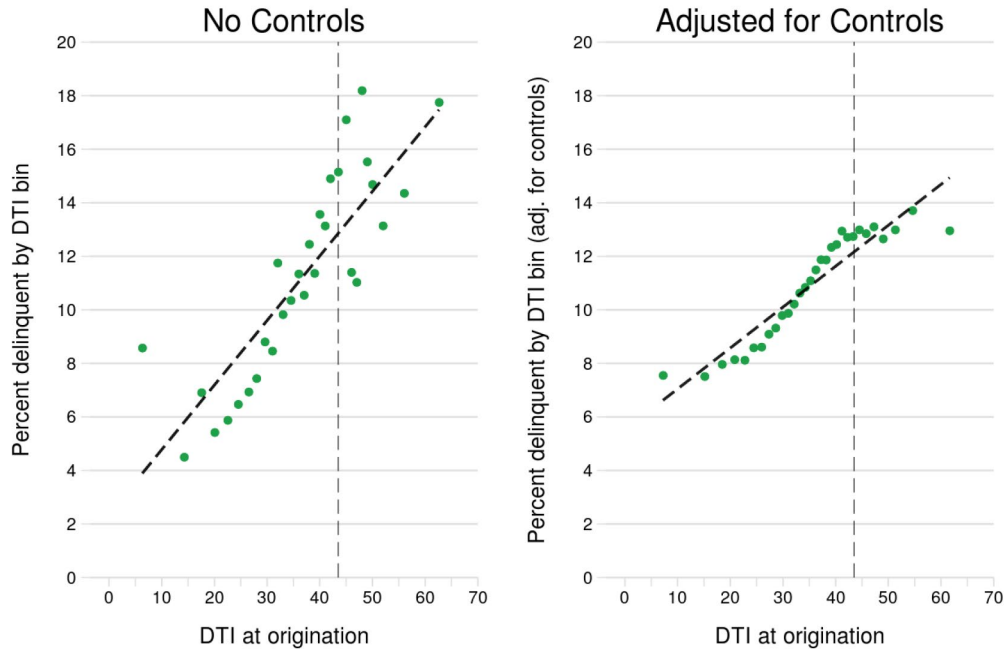
For example, Figure 8 from the assessment report¹¹ showed that loans that met the 43% DTI requirement for otherwise qualifying loans had early delinquencies of 12 - 14% during the period from 2006 to 2008. And as shown in Figure 9, this same category of loans had early delinquencies below 1% after the financial crisis.¹²

¹⁰ Op. cit., Assessment report, p. 84.

¹¹ Ibid., p. 102.

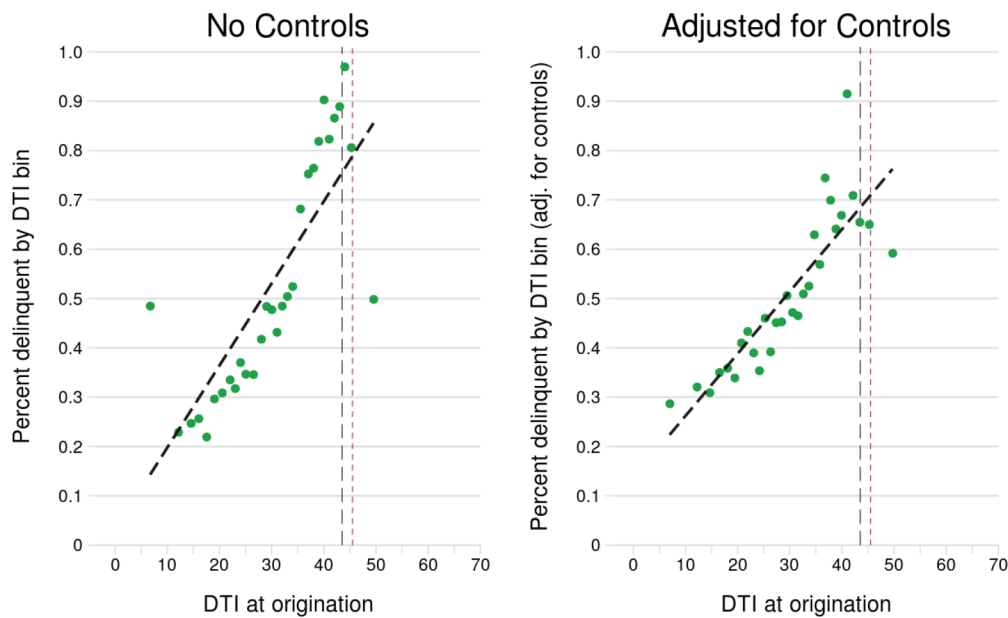
¹² Ibid., p. 103.

FIGURE 8: EARLY DELINQUENCIES BY DTI, 2006-2008



Data Source: McDash Loan Current and Loan Delinquency History Files, DTI topcoded at 70

FIGURE 9: EARLY DELINQUENCIES BY DTI, 2012-2015



Data Source: McDash Loan Current and Loan Month Files
 Notes: DTI topcoded at 70, vertical lines reflect General QM DTI threshold (gray dashed) and GSE DTI limit without compensating factors (purple short-dashed)

Likewise in the proposed rulemaking,¹³ the CFPB showed in Figure 10 that loans with a spread of less than 2% over the PMMS had an early delinquency rate of less than 14%, but

that delinquency rate dropped to around 3.5% for loans originated in 2018 (by the end of 2019).

FIGURE 10: EARLY DELINQUENCIES BY RATE SPREAD

TABLE 1—2002–2008 ORIGINATIONS, EARLY DELINQUENCY RATE BY RATE SPREAD

Rate spread (interest rate + PMI approximation—PMMS ¹⁹¹) in percentage points	Early delinquency rate (percent)
< 0	2
0–0.24	2
0.25–0.49	4
0.50–0.74	5
0.75–0.99	6
1.00–1.24	8
1.25–1.49	10
1.50–1.74	12
1.75–1.99	13
2.00–2.24	14
2.25 and above	14

TABLE 2—2018 ORIGINATIONS, EARLY DELINQUENCY RATE BY RATE SPREAD

Rate spread over APOR in percentage points	Early delinquency rate (as of Dec. 2019) (percent)
< 0	0.2
0–0.49	0.2
0.50–0.99	0.6
1.00–1.49	1.7
1.50–1.99	2.7
2.00 and above	4.2

The CFPB methodology showed that the level of delinquency can be affected by the economic environment. In our analysis we generalize this method to allow for evaluating

any cohort of loans against a cohort of loans that is consistent with the definitions of ability to repay already used by the CFPB.

¹³ <https://www.federalregister.gov/documents/2020/07/10/2020-13739/qualified-mortgage-definition-under-the-truth-in-lending-act-regulation-z-general-qm-loan-definition>, Tables 1 and 2, printed page 41732.

ESTABLISHING A BENCHMARK

Rather than a specific cutoff, we recommend that the standard for evaluating Ability-to-Repay be the performance of a well-defined cohort of loans. A benchmark should be reliable and relevant. This means it should display consistent performance across time in relation to other cohorts and be reasonably placed within the outcome distribution. For example, cohorts at the extremes of the performance distribution would exclude either nearly all or none of the universe of loans and thus would not be relevant standards.

A benchmark cohort is like a passive stock index against which performance can be measured. The performance of the passive index changes every year as economic conditions change and funds of different strategies seek to mimic or outperform passive index returns. Similarly, a benchmark mortgage cohort will be clear, stable, and time-tested, even as its performance changes every

year. This approach permits innovation in developing new combinations of underwriting standards, including new drivers. Such new cohorts can “demonstrate” Ability-to-Repay status with a few years of delinquency performance that is comparable or better than the benchmark. Such an objective performance metric, when augmented by qualitative evidence of sound underwriting, would presumably be evidence that a borrower’s ability to repay had been adequately examined prior to the making of a loan.

Figure 11 shows a time series consideration of early serious delinquencies by FICO and LTV. DQs do not vary much by LTV, as shown earlier, but they do vary quite widely over time. DQs vary consistently by FICO as well as across time. The use of a cohort rather than a specific level of delinquency allows the benchmark to remain valid over time.

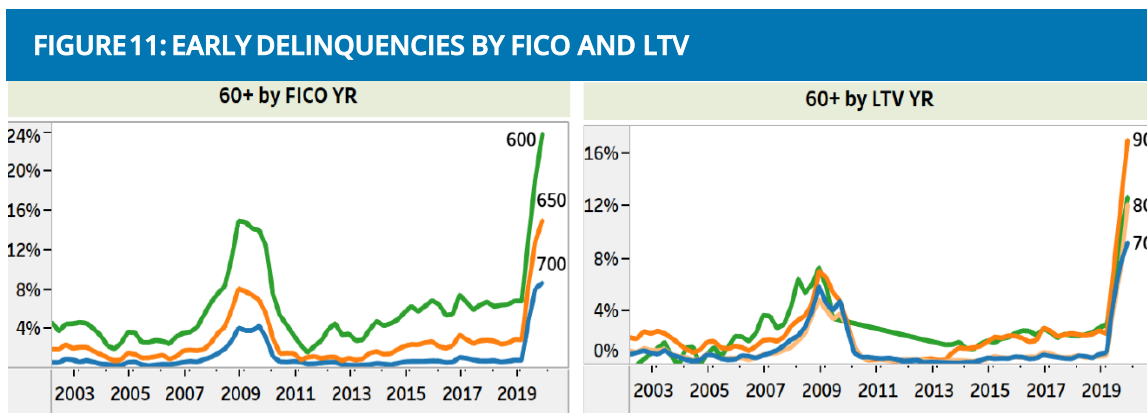


Figure 12 shows a time series view by unadjusted Agency cohorts in the left panel and then for loans from each of the three sources with characteristics similar to the **FHA Average** cohort on the right. This limits the sample for each agency to loans with an average credit score of 675, DTI of 45, and LTV of 95. The left panel shows unadjusted Agency

performance differences, which are persistent and large. The second panel again shows that when selecting loans from each Agency that are within the target cohort, differences mostly disappear. This is a crucial result because it indicates that the selection of drivers (FICO, LTV, DTI) for fully documented agency and government loans is sufficient to construct a benchmark.

FIGURE 12: DELINQUENCY RATES BY AGENCY VS FHA

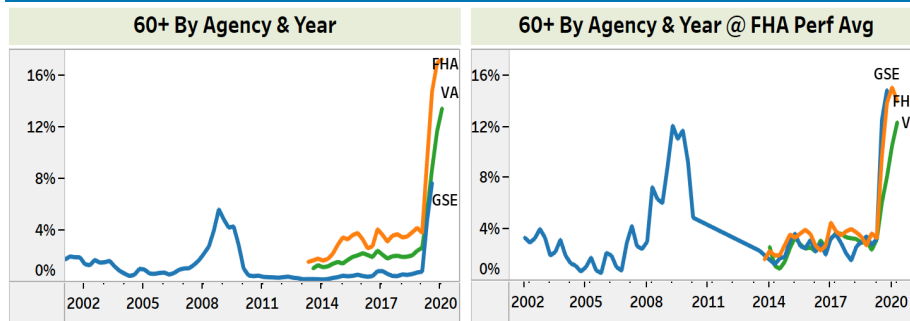
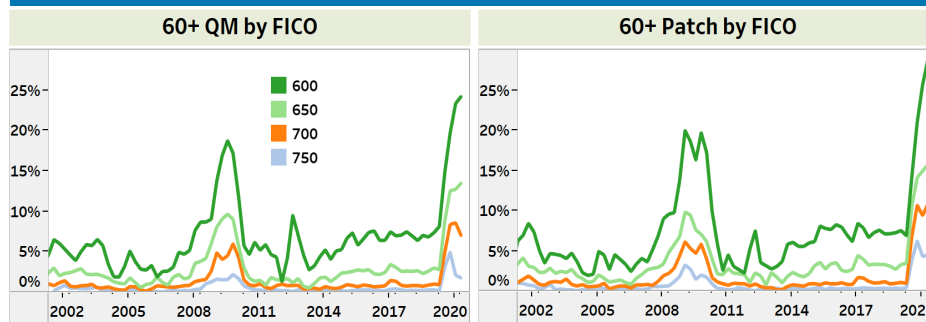


Figure 13 shows the time series performance of GSE loans near the QM DTI cutoff of 43% DTI. The “QM Standard” cohort includes DTI 40–43, loans that presumably would have met the requirements of the 43% DTI standard, and “Patch” loans with DTI from 44–49%. Predictably, delinquency rates within each group vary quite widely by credit score and over time, but not materially across these DTI

groups. A benchmark cohort would provide a reasonable method to reflect the time dimension of the performance data. These charts also show the shortcomings of the DTI cutoff. Loans with high credit scores could have higher DTIs without severe risk of violating an ATR benchmark, while lower credit score loans may have had unacceptably high delinquencies, even with a DTI below 43%.

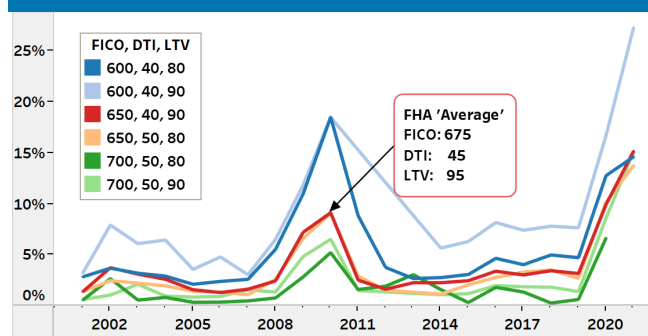
FIGURE 13: GSE LOAN DELINQUENCY RATES



We believe a good candidate for the Ability-to-Repay benchmark is the cohort of loans with underwriting characteristics that generate performance similar to the FHA portfolio. For this benchmark cohort, early serious DQ rates reach 10% twice in 20 years but are quite low in the intervening years and are consistent with the delinquency rates of loans that met the

Qualified Mortgage tests using the DTI and the spread over APOR tests during stressful economic periods. Figure 14 shows the performance of the “FHA Average” cohort in comparison to other cohorts over the past 20 years. Such a cohort provides a clear benchmark for evaluation of whether any cohort of loans meets the performance requirements expected of ability to repay loans.

FIGURE 14: DELINQUENCY RATES FOR FHA AVERAGE VS OTHER COHORTS



Comparison to Alternatives

A fixed delinquency rate is an attractive metric because it's so simple, but delinquency rates, like interest rates and stock market returns, vary every year. So how can ATR be determined at origination by a fixed delinquency standard if it depends on borrower performance two years later? The preceding four figures reinforce two themes. Cohort performance is consistent over time across economics, that is, they don't cross, and performance varies widely over time as economics change even after controlling for key drivers.

Fixed delinquency standards would be too strict in weaker economies and could approve borrowers who won't be able to withstand financial stress even in normal times. Alternatively, varying the standard as economic conditions change is complicated and creates uncertainty for the mortgage market. These results reinforce the conclusion that fixed standards won't work, whereas benchmarks provide the reliability and clarity required for lending standards.

An approach the CFPB recently adopted is the APOR spread over average mortgage rates. While, as shown above, APOR does correlate with delinquencies, this approach may also have shortcomings as an ATR

standard in the long term. Market spreads are a result, not a measure, of individual household ability to repay. They contain prepayment and liquidity risk, which are unrelated to ability to repay, and further, are closely tied to loan to value ratios (LTV), which may reflect loss severity more than ability to repay. While this approach has many benefits, it could be enhanced with a data-driven approach similar to what we have examined here.

Liquidity risk is especially asymmetric, as reflected by the last two financial stresses in the US economy. Agency mortgage spreads widened somewhat, but private mortgage market spreads widened significantly. In a future market disruption, it is possible that private mortgages could generally fail the APOR standard while the Agency market continues to operate.

Finally, spreads may be reliable in the center of markets, but less so where confidence is needed, at the edges of markets, for innovative underwriting drivers, and for marginal, riskier, non-traditional, and underserved borrowers. While an APOR based standard is a significant improvement over the Appendix Q approach, it may not offer the same degree of consistency and flexibility through all market environments as the benchmark approach that we are recommending that the industry consider.

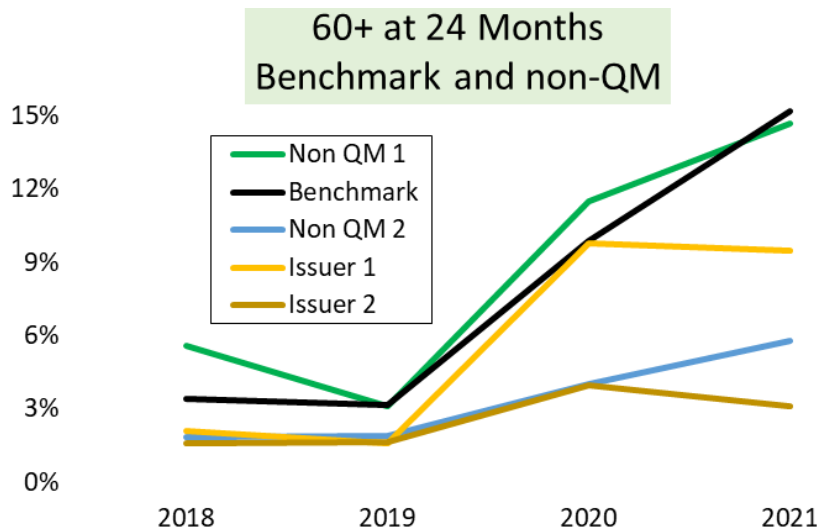
APPLYING THE PROPOSED BENCHMARK TO THE NON-QM MARKET

A crucial test of an ATR benchmark is how it works outside the umbrella of government lending for non-qualifying loans. Thus, we now compare the proposed benchmark to a sample of non-QM mortgage cohorts originated and securitized from 2016–2019, for consideration around 24 months old during 2018–2021. Since these mortgages are generally underwritten to expanded credit lending criteria and reduced documentation of income, we include ARMs and do not exclude layered risks. In contrast, DTI was not included, since many loans are reduced documentation. Loans with terms under 20 years and over 40 years were excluded for comparability to the initial proposed benchmark which only included 30-year mortgages.

In Figure 15 we show two examples of how this proposed ATR benchmark could provide an

objective, data-driven validation of the ATR performance for the non-QM market: by cohort and by issuer. We show two cohorts: Non QM1 has 650 FICO and 90 LTV and performs about the same as the proposed benchmark; Non QM2 outperforms the proposed benchmark and has 750 FICO and 80 LTV. We also selected originations for two large non-QM securitizers—which outperformed the proposed benchmark slightly and significantly, respectively. The results illustrate that it's possible to apply this cohort approach to non-QM loans which may use alternative underwriting. The results further suggest that this benchmark would not disrupt the non-QM market and would provide much needed clarity to consumers and originators on the ATR standard.

FIGURE 15: NON-QM COHORT PERFORMANCE AND THE BENCHMARK



CONCLUSION

In this Proof-of-Concept for an ATR standard we have used an extensive (not exhaustive) data set of mortgages and shown that mortgage cohorts can be constructed using FICO and DTI. These cohorts show consistent performance over time in relation to each other and demonstrate that combining drivers provides better correlation to performance than using one driver alone. A fixed performance standard cannot be used because delinquency rates vary over time even when controlling for age and other risk drivers.

Generally, a cohort of loans with well-defined documentation, verification, and underwriting can be a benchmark to evaluate whether other cohorts of loans meet the performance requirements of Ability-to-Repay. For instance, a cohort that performs similarly to the average of the FHA portfolio could be a relevant benchmark. The benchmark is set so that it is consistent with current lending practices, but the subprime cohort shown earlier clearly fails.

Crucially, we show how the proposed benchmark can be applied outside of government lending to the current non-QM market. Most non-QM lending of the past several years performs as well or better than the proposed benchmark, suggesting this ATR standard would not disrupt the current non-QM market and could provide needed clarity around ATR loans.

A cohort that reliably performs like the FHA portfolio average is consistent with a wide range of loans currently being originated. Its delinquency performance is also similar to loans that meet either the prior 43% DTI standard or the more recent 2% spread over APOR standard adopted by the CFPB. Adopting such a cohort as the benchmark will facilitate the evaluation of new approaches to underwriting and documentation. Performance-based benchmarks also ensure fair access to a wider range of market participants (originators, investors, and households) and prevent consumer exploitation.

This publication is believed to be reliable, but its accuracy, completeness, timeliness, and suitability for any purpose are not guaranteed. All opinions are subject to change without notice. Nothing in this publication constitutes (1) investment, legal, accounting, tax, or other professional advice or (2) any recommendation or solicitation to purchase, hold, sell, or otherwise deal in any investment. This publication has been prepared for general informational purposes, without consideration of the circumstances or objectives of any particular investor. Any reliance on the contents of this publication is at the reader's sole risk. All investment is subject to numerous risks, known and unknown. Past performance is no guarantee of future results. For investment advice, seek a qualified investment professional. Note: An affiliate of Andrew Davidson & Co., Inc. engages in trading activities in securities that may be the same or similar to those discussed in this publication.

Andrew Davidson & Co., Inc. (AD&Co) was founded in 1992 by Andrew Davidson, an international leader in the development of financial research and analytics, mortgage-backed securities product development, valuation and hedging, housing policy and GSE reform, and credit-risk transfer transactions. Since its inception, the company has provided institutional fixed-income investors and risk managers with high quality models, applications, consulting services, research, and thought leadership, aimed at yielding advanced, quantitative solutions to asset management issues. AD&Co's clients include some of the world's largest and most successful financial institutions and investment managers. For more information, visit www.ad-co.com.



New York
65 Bleecker Street, Fifth Floor
New York, NY 10012-2420

Raleigh
150 Fayetteville Street, Suite 1030
Raleigh, NC 27601-2957

212.274.9075
support@ad-co.com
www.ad-co.com