dv01 + insights

Quantifying the Accuracy of Rep Line vs. Loan Level Data for Cashflow Generation in Online Lending

ABSTRACT / In this report, we utilize dv01's cashflow engine to study the accuracy and efficiency of cashflow generation methods at different degrees of granularity, from a single loan rep line, to rep lines across multiple attributes, to loan level data. Our findings show that the rep line methodology widely used for securitization deals creates significant inaccuracy for cashflow projections. We conclude with showing how integrating up-to-date securitization collateral data with a scalable loan level cashflow engine addresses these inaccuracies.

October 2017

Wei Wu, Ph.D., CFA, Principal Data Scientist
Yehuda Graber, Quantitative Engineer
Frank Deutschmann, Head of Product

support@dv01.co | 646-854-5258 | 915 Broadway, Suite 802, New York, NY 10010

INTRODUCTION TO CASHFLOW GENERATION / Cashflow projection and analysis is a critical process for investing in fixed-income markets. More accurate cashflow projections are the foundation for making better decisions in valuation, performance analysis, risk analysis, and asset-liability management—especially when analyzing deep credit.

A cashflow engine or cashflow model is the tool used to generate projected future cashflows for a loan/bond portfolio. Inputs to the engine include: the characteristics of the loans or bonds; assumptions for various risk parameters such as default curves, prepayment curves, and recovery rates; and other information such as fees.

For a single fixed rate loan, cashflows are a set of fixed amortized payments for every time period, typically every month until maturity, assuming there is no default and no prepayment. When defaults and prepayments are taken into account, cashflow engines usually use partial loan modeling. The user inputs a set of default and prepayment rates or probabilities and the cashflow engine projects the expected cashflows derived from the amortized payments based on the proportional amount of principal pay-down or charge-off according to the provided prepayment and default rates.

While the above describes cashflow generation for simple online lending loans, projecting cashflows for different tranches of online lending securitization deals involves complicated logic and a system that models the cashflow waterfall of the deals on top of the collateral cashflows. In this report, we concentrate on the accuracy of the projected collateral cashflows of these deals.

dv01 has a cashflow engine optimized for online lending loans and securitization deals, and integrated with direct data feeds from all participants in the deals. We use this engine for the analysis in this report. Readers can also replicate these results using other commercial cashflow engines given appropriate data updates.

REP LINE CASHFLOW ANALYSIS / Traditionally, it has been impractical to generate cashflows on loan level data. You would have to generate cashflows for tens or hundreds of thousands of loans and then aggregate them together, which could take hours with legacy cashflow engines.

The common solution for this is a rep line analysis, which approximates the portfolio with a set of hypothetical loans or assets, called rep lines. Each rep line represents a group or subset of the portfolio with the statistical characteristics of that group. Each individual loan within the group is assumed to be homogeneous and to have same behavior as the rep line.

An effective rep line analysis tries to divide the portfolio into granular enough cohorts

support@dv01.co | 646-854-5258 | 915 Broadway, Suite 802, New York, NY 10010

to capture the diverse characteristics of the loans in the pool, while also minimizing the number of rep lines needed. However, the effectiveness of rep line analysis has not been measured or studied quantitatively, particularly in online lending. We set out to study what makes a good rep line for accurate and efficient cashflow generation.

MEASURING CASHFLOW ACCURACY QUANTITATIVELY / To quantitatively assess the accuracy of a set of projected cashflows, we first define a set of reference cashflows to compare against. We then define a metric to quantitatively measure the inaccuracies against the reference cashflows on the same scale for each projection for an apples-to-apples comparison.

Below is a simple example that illustrates the metrics we are going to use to measure the accuracy of one set of projected cashflows to a set of reference cashflows. In **Figure 1** and **Figure 2**, we use a reference pool composed of 108,663 loans from Lending Club in California. We assume a constant annualized default rate (CDR) of 10% and a constant annualized prepayment rate (CPR) of 20% for all loans. We then generate cashflows on both the loan level and a single loan rep line.

Fig1: Loan level vs. rep line cashflows

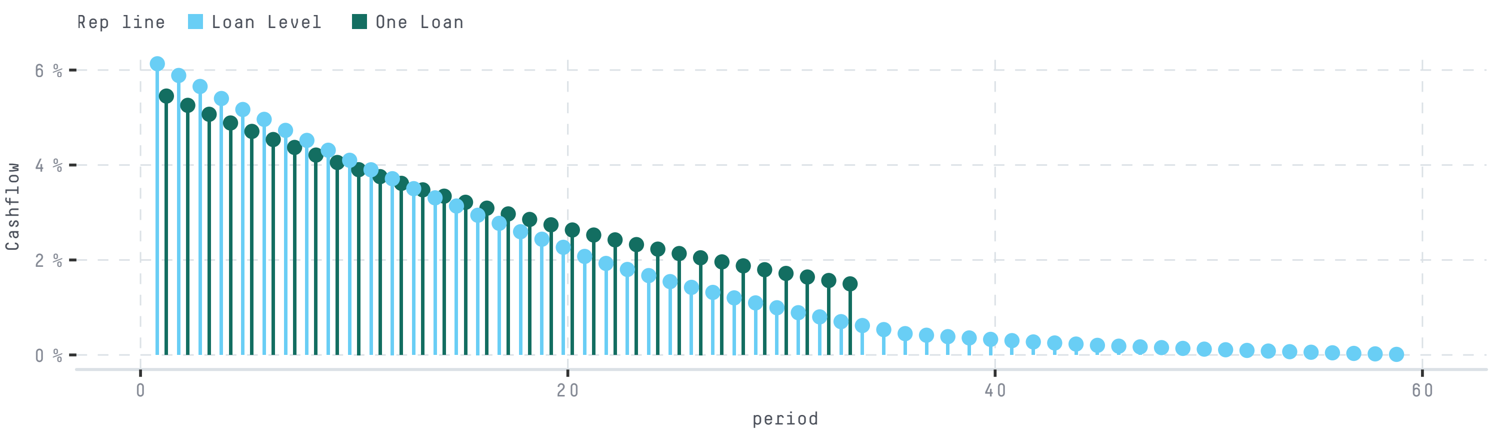
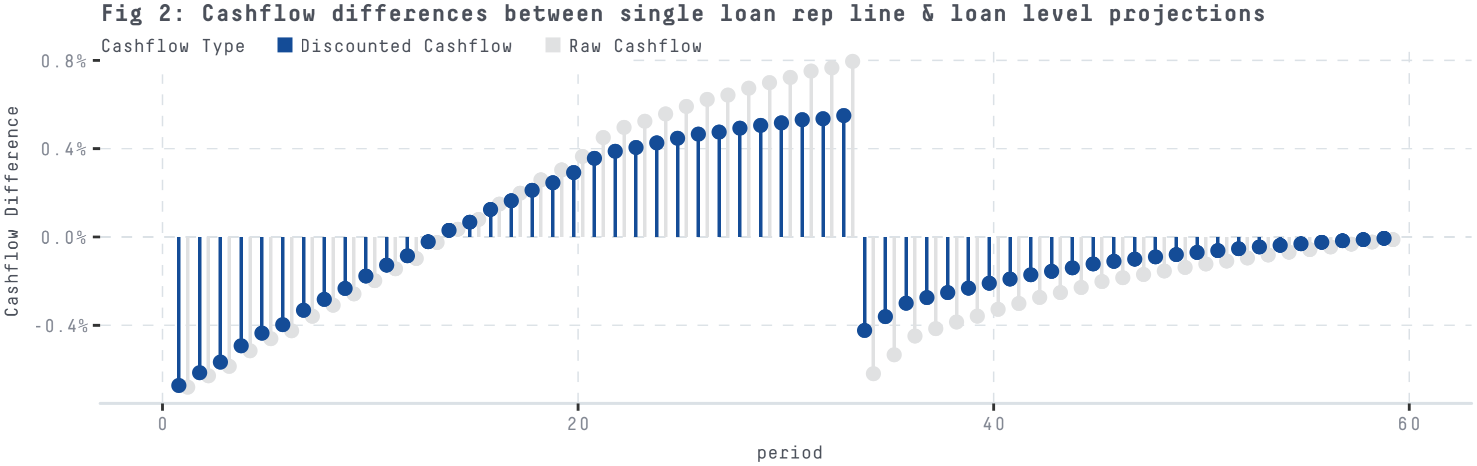


Fig 2: Cashflow differences between single loan rep line & loan level projections



support@dv01.co | 646-854-5258 | 915 Broadway, Suite 802, New York, NY 10010

Figure 1 shows loan level cashflows (Blue) and single loan rep line cashflows (Green). We can see that there are significant deviations between these two sets of cashflows. Present value is used to quantify the differences in the magnitude and timing of future cashflows. We define first type of cashflow matching error, present value error, as:

$$PV = \sum_{m=1}^{M} \frac{cf_m}{\left(1 + {}^{i}/_{12}\right)^m}$$

$$error_{PV} = \frac{\left|PV_{projection} - PV_{reference}\right|}{PV_{reference}} = \frac{1}{PV_{reference}} \cdot \left|\sum_{m=1}^{M} \frac{\left(cf_m^{projection} - cf_m^{reference}\right)}{\left(1 + {}^{i}/_{12}\right)^m}\right|$$

The present value error is the sum of the differences in cashflows across all periods as a percentage of present value of the reference pool.

Figure 2 shows the raw and discounted cashflow differences between the single loan rep line cashflow projection and the loan-level reference cashflows we show in **Figure 1**. Note that the deviations can be positive or negative for different periods and that, based on the definition of present value error, the positive deviation can be cancelled out by a negative deviation at a different period. Therefore the present value error is not affected by the timing of the cashflows, just the total present value of cashflows. However, investors care about both the magnitude and the timing of the cashflow. We address this by defining a second type of cashflow matching error, shape error:

$$error_{shape} = \frac{1}{PV_{reference}} \cdot \sum_{m=1}^{M} \frac{\left|cf_m^{projection} - cf_m^{reference}\right|}{\left(1 + {}^{i}/_{12}\right)^m}$$

Shape error incorporates timing mismatches in cashflows, ensuring that positive and negative mismatches do not offset each other.

CASHFLOW ACCURACY UNDER UNIFORM DEFAULT AND PREPAYMENT ASSUMPTIONS / With the definitions of errors in place, we evaluate the accuracies of various rep line cashflows. We will assume all loans have the same default curve and prepayment curve (flat 10% CDR and 20% CPR) and run projections on the same loan pool used above, Lending Club Standard Platform loans issued in California. **Table 1** shows the results for cashflow projections using rep lines at different levels of granularity. The first line shows the cashflows generated at loan level that will serve as the reference cashflows rep line errors will be benchmarked against. It takes 1.5 seconds to generate cashflows for all 108,663 loans in the loan pool. (Note that dv01's cashflow engine was optimized for accurate loan-level predictions without sacrificing speed.)

Table 1: Cashflow accuracies for different rep lines for Lending Club in California

Loan Pool	# Loans	Rep Line Method	NPV Error	Shape Error	Time(sec)
Lending Club in CA	108,663	Loan Level	0	0	1.54
Lending Club in CA	1	Single Loan	0.00862	0.16831	0.02
Lending Club in CA	2	Group by matching terms	0.00327	0.11280	0.05
Lending Club in CA	94	Group by matching term+age	0.00087	0.00087	2.05
Lending Club in CA	828	Group by matching term+age+coupon	0.00002	0.00002	20.68

The rep lines in **Table 1** are based on several factors that affect cashflows. In addition to matching average coupon, the loans used in this projection:

1. match only the average term and average age with one loan
2. match separately for loans with different original terms
3. match original terms and loan ages

First, we can see that the present value error is small for all rep lines. This is because the cashflows for rep lines are generated based on the weighted average coupon of the loans in the rep lines. The same weighted average rate is used to discount these cashflows. Therefore, the present values for different rep lines are all close to par, even though the timing of the cashflows can be significantly different. So the more appropriate measurement of cashflow accuracy in this case is the shape errors.

From **Table 1**, we can see that the shape error with a single loan rep line is significant at 0.168, which took 0.02 seconds to generate. As we increase the granularity of the rep lines, the errors decrease, 0.11 for rep lines matching two types of original terms, and 0.0009 for rep lines matching both terms and ages, with a total of 94 rep lines. As the number of rep lines increases, so do the execution times.

> Cashflows will closely mirror loan level cashflows if we use rep lines that match loan terms and loan ages.

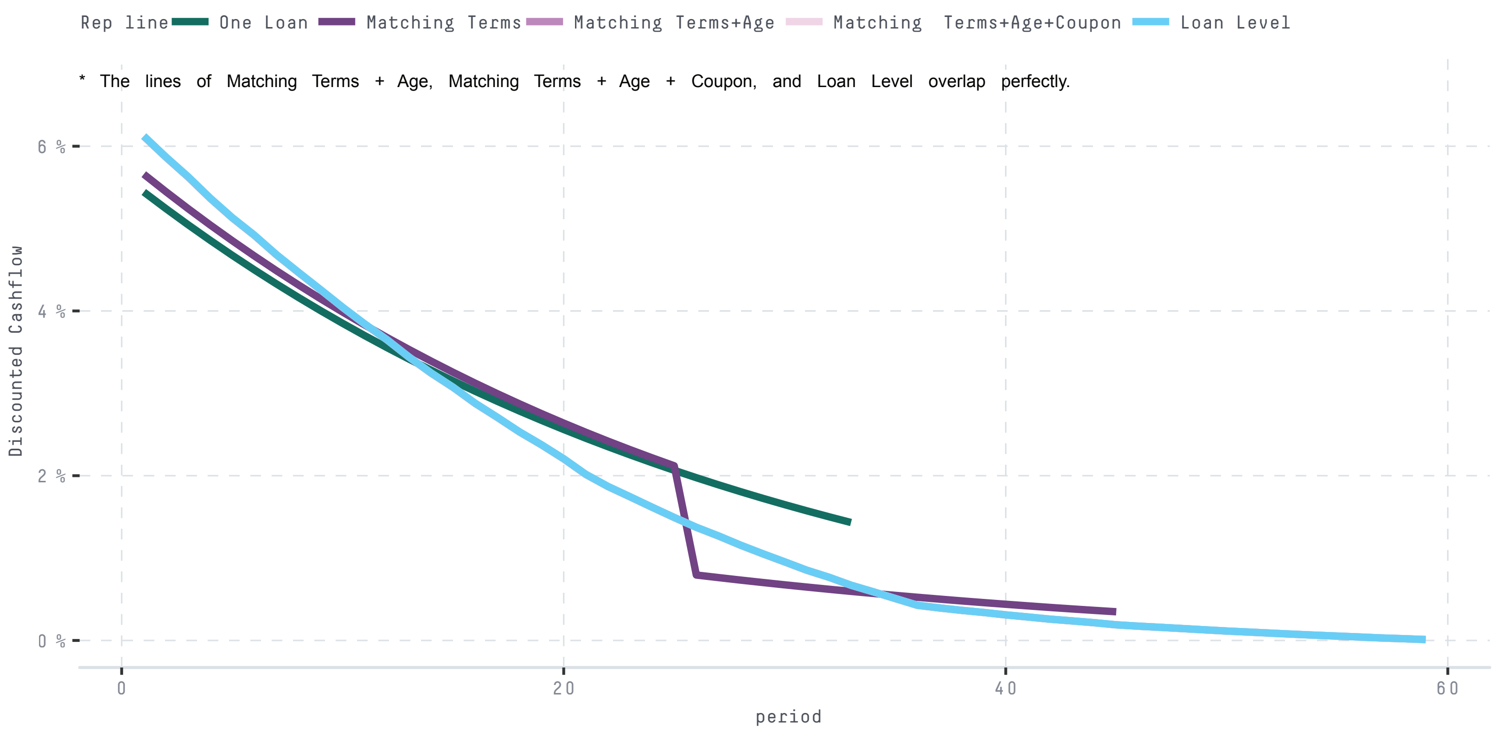
Notice that rep lines grouped by terms and ages have a very small error. As we stated previously, the main variables that determine cashflows are: basic loan characteristics (loan rate, loan term, loan age or remaining term), and default and prepayment rates. When the default and prepayment rates are the same, the cashflows will closely mirror loan level cashflows if we use rep lines that match loan terms and loan ages.

Including coupon in addition to term and age as a grouping variable further reduces shape error close to 0, as seen in the last line of **Table 1**. However, the number of rep lines increases significantly, from 94 to 828, causing the run time to be significantly longer than running loan level projections. The run time overhead is caused by the communication inefficiencies of dynamically generating 828 replines outside of the cashflow engine and passing the data into the engine. In other words, the efficiency

support@dv01.co | 646-854-5258 | 915 Broadway, Suite 802, New York, NY 10010

advantage of rep lines is lost when the rep lines become complicated and the magnitude of improvement in projections is minor, as can be seen in **Figure 3**.

Fig 3: Cashflows using various rep lines



Observant readers may also notice that when term and age are included in the rep line grouping, the present value error matches the shape error. This means that the cashflow differences at every time period have the same sign. See footnotes for more detailed information on this observation.*

The necessity of rep lines is more clearly illustrated in larger loan pools. In **Table 2**, we replicate the process described above but with a much larger loan pool, consisting of more than 1 million loans. You can see that the shape error, known as cashflow error from this point forward, is similar to **Table 1**, even though this pool is much larger. Here, as in Table 1, rep lines that include term and age produce acceptable error and shorter run time than loan level cashflows.

" The efficiency advantage of rep lines is lost when the rep lines become complicated and the magnitude of improvement in projections is minor.

** We can analytically prove that, for an amortized loan pool, cashflows generated using rep lines that include term and age as a grouping are always smaller than the sum of cashflows from individual loans at every time period due to the convex property of the amortized payments over coupon rate. Contact intelligence@dv01.co for further details.*

Table 2: Cashflow accuracies for different rep lines for full Lending Club platform

Loan Pool	# Loans	Rep Line Method	NPV Error	Shape Error	Time(sec)
All Lending Club	1,108,160	Loan Level	0	0	17.26
All Lending Club	1	Single Loan	0.00691	0.16363	0.02
All Lending Club	4	Group by matching terms	0.00370	0.10883	0.10
All Lending Club	136	Group by matching term+age	0.00145	0.00145	3.10
All Lending Club	1,010	Group by matching term+age+coupon	0.00002	0.00002	24.51

> The accuracy of the cashflows increases with the granularity of the rep lines.

The conclusions in **Table 1** and **Table 2** still hold when we run the above rep lines using CPR and CDR curves with various shapes and magnitudes. The accuracy of the cashflows increases with the granularity of the rep lines, with the inclusion of term and age being essential to accuracy.

However, the above assumptions do not hold in the real world. Errors in cashflow projection can come from two main sources:

1. Significant discrepancies of the input curve assumptions from actual default/prepay statistics.
2. Mismatches in loan term/age distributions between rep lines and the exact loan pool.

We will now look closely at the second reason in the context of online lending securitization deals. Discrepancies caused by the first reason will be examined in a follow-up report.

SIGNIFICANT CASHFLOW ERRORS FROM A COMMONLY USED REP LINE GENERATION METHOD: STATIC REP LINES AND ROLLED BALANCES / As concluded in the above section, grouping by term and age, or remaining term, when constructuring rep lines is essential for cashflow accuracy. As a consequence, rep lines that do not accurately reflect the distribution of the remaining terms of underlying collaterals can cause significant errors. This section will highlight how traditional practices for generating collateral cashflows for ABS securitizations exhibit the above tendencies, causing inaccurate cashflow results.

One of the most commonly used methodologies for generating cashflows for securitization deals relies on the rep lines stated in the deal's OM or offering memorandum. OM rep lines are constructed based on the statistical loan pool before deal closing, and thus often do not accurately reflect the final loan pool in the deal. This is because there is a lag between the preparation of the OM and the final cut-off date that determines the final composition of the collateral portfolio. In addition, the

support@dv01.co | 646-854-5258 | 915 Broadway, Suite 802, New York, NY 10010

exact variables used to generate OM rep lines are not given, making it difficult to verify how closely these rep lines capture the statistical profile of the final collateral pool.

Another source of inaccuracy is the way current balances of the rep lines are determined after closing date. Traditionally, investors and cashflow vendors only receive the aggregate balances of the collateral loan pool as reported in the monthly trustee reports. The conventional method used to get the balances of each rep line is the following: First, calculate the scheduled amortized loan balances for each rep line. Then, calculate the balance difference between the total scheduled balance and the trustee reported total balance. The balances for all the rep lines is determined by proportionally distributing this total difference according to the scheduled balance amounts in each rep line. This procedure is carried from one period to another. This method is called “rolled balance” method.

Our research shows that the prepay and default rates can vary greatly for different rep lines with different characteristics.

This process assumes the same default and prepay rates for each rep line for each time period. While nonoptimal, this is the only reasonable method for updating rep line balances monthly given the traditional lack access to updated loan level data.

However, our research shows that the prepay and default rates can vary greatly for different rep lines with different characteristics. Therefore, discrepancies in the loan balances for each rep line used as the starting point for cashflow generation can occur. Thus, resulting cashflows diverge from accurate cashflows generated using timely loan level data. As time passes, discrepancies will grow and amplify.

Table 3: Cashflow accuracies for different rep lines for a securitization deal

Rep Line Method	Loan Pool Balance	**# Loans** 09/30/16	12/31/16	3/31/17	6/30/17	**Cashflow Error** 09/30/16	12/31/16	3/31/17	6/30/17
Loan Level	Current	16,268	15,360	14,275	13,075	0.00000	0.00000	0.00000	0.00000
Term+Age+Grade, Dynamic	Current	268	267	264	261	0.00094	0.00085	0.00077	0.00070
Term+Age+Grade, Static at Closing	Rolled	267	267	267	267	0.00062	0.00206	0.00505	0.00837
OM Replines, Static at Closing	Rolled	226	226	226	226	0.02927	0.02977	0.02948	0.02829

In **Table 3**, we present cashflows generated using different rep line methods for a securitization deal from 2016. (This was replicated with another deal which produced similar results, not included here for brevity.) Row 1 uses up-to-date loan level data at run time to project cashflows, ensuring no discrepancies between the rep lines and the underlying collateral. (This data is accessible through dv01.)

> Using the "rolled balance" method, we see a cashflow error of 2.9% two months after the closing of the deal.

Rows 2 and 3 are the results using rep lines constructed by grouping on loan term, remaining terms, and loan grade. Row 2 uses these groupings and actual loan balances for each rep line compiled from loan level data from the source above, and dynamically generates rep lines for each date as time passes. In Row 3, rep lines use the same grouping, but the rep lines are only generated once at deal closing and the "rolled balance" method described above is used to set balances of the rep lines. Row 4 uses OM rep lines, which are static at closing. For this deal, the OM rep lines provided enough information to determine that they were constructed approximately based on the following variables: original loan term, remaining loan term, and loan grade.

Using these different methods, we generate cashflows at various points in time throughout the lifetime of the deal. In this study, we used simplified default and prepay assumptions, using a constant 20% CPR and 10% CDR. The deal's closing date was 7/29/2016. We started the cashflow generation run immediately after the first payment day on 9/30/16. We then generated cashflows in three month intervals. The first line in the table (Row 1) serves as the reference since it is based on loan level data and actual loan balances on those dates. All other rep line methods are measured against the loan level cashflows. The results are shown in order of accuracy.

Starting with the least accurate, in the last line of the table (Row 4), we have the OM reported rep lines. As the balances of OM rep lines are only given once and the statistical variables used to generate these rep lines are not disclosed, these rep lines are static, with balances updated through the "rolled" balance method. Using this method, we see a cashflow error of 2.9% on 9/30/16, only two months after the closing of the deal. This error is significant as it is the error as a percentage of present value. The cashflow errors do not improve for subsequent dates and stay consistent at 2.9%.

The reason for this significant error is because of the mechanics of how securitization works. There is a significant time delay between the formation of the rep line statistics and the closing of the deal. The composition of the final loan pool can be different from the statistical loan pool the OM used to generate the rep lines. In Table 4, we show the overall statistics on the loan term. The overall loan balances are closely matched. The percentage allocations along the loan terms are also close between the OM rep lines and "real" loan level statistics. However, the loan count is different, which means loans went in and out of the loan pool between the "statistical cut-off date" and the closing date.

If we look at the distribution of the loan balance over the remaining terms as shown in **Figure 4**, we see significant differences between the loan balances of OM stated rep lines along the remaining term. (We used dv01's Securitizations functionality to

support@dv01.co | 646-854-5258 | 915 Broadway, Suite 802, New York, NY 10010

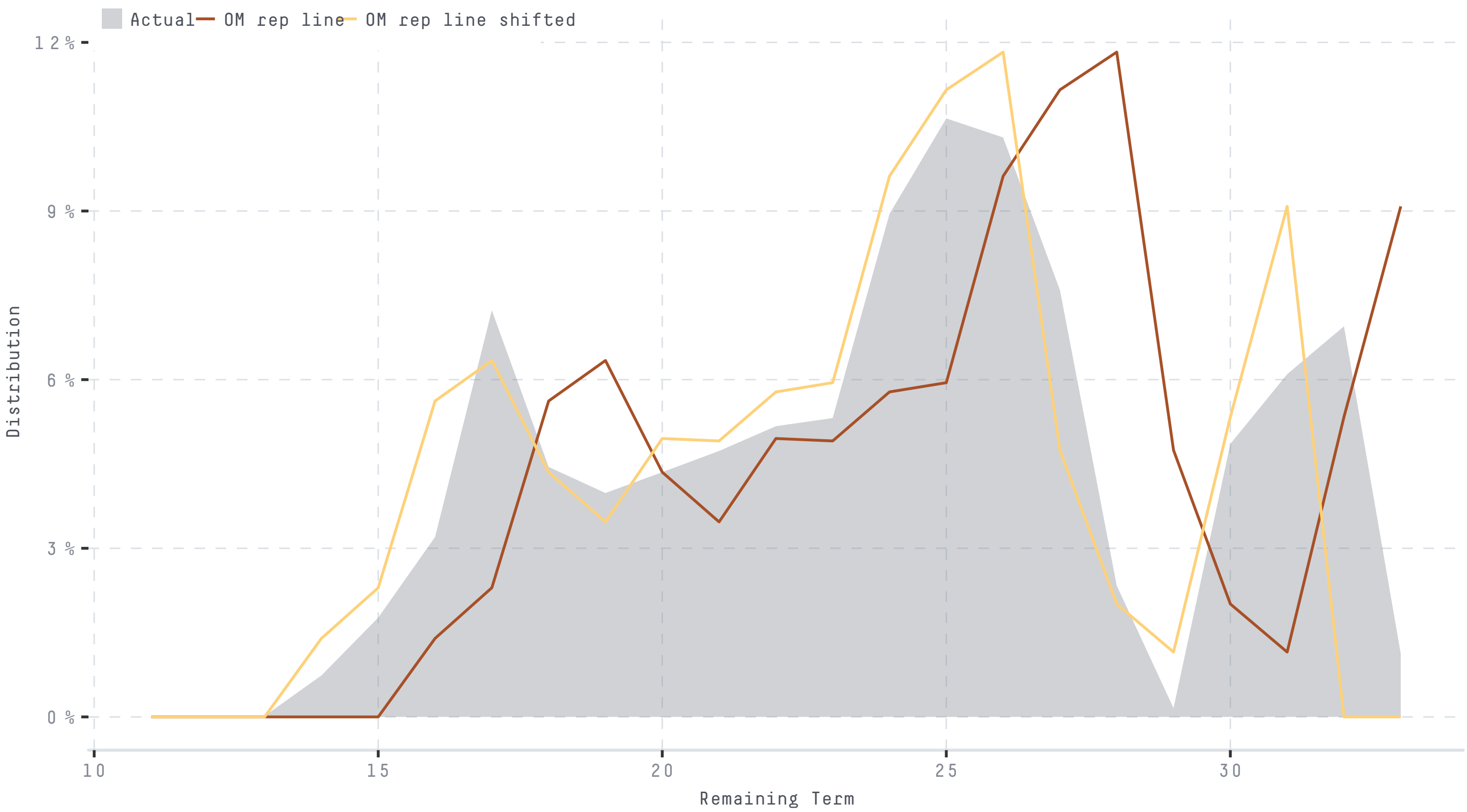
Table 4: Distributions of loan terms in OM rep lines vs. Actual

	OM Rep Lines			**Loan Level Stats**		
Term	Balance	Percent	# Loans	Balance	Percent	# Loans
36	91,960,979	44.76%	8,476	91,197,929	44.39%	8,886
60	113,479,299	55.24%	7,404	114242,345	55.61%	7,643
Total	205,440,278	100.00%	15,880	205,440,274	100.00%	16,529

see the balance at each value of the remaining term based on loan level data.) Upon further investigation, we see that the biggest cause of difference is the inaccurate remaining terms that were reported for the rep lines. One way to "fix" this is to shift the rep line distribution along remaining term by two months, which approximates the two month period between the statistical cut-off and the closing date for the deal. We can see in **Figure 4** that after shifting the remaining terms, the distribution of the OM rep lines is much closer to the actual loan pool. Without this correction, there will be significant differences in loan balance distributions and thus significant cashflow error.

Examining the results from Row 2 and 3 in Table 3, we see that both use rep lines

Figure 4: Distributions by remaining term of actual loan pool, OM rep line, and shifted OM rep line at deal closing



> The rolled balance method generates significant errors, accumulating as the deal seasons.

grouped by the same attributes but employ different methods for updating balances. For the rolled balance method in Row 3, the error is fairly small at 0.00062 at the beginning of the deal. But as time goes by, the cashflow errors get bigger and bigger: 0.00206, 0.00505, 0.00837. On the other hand, errors for cashflows generated using rep lines with up-to-date stats from loan level balances, summarized in Row 2, stay around the same level: 0.0009 to 0.0007. Thus, we can conclude that the rolled balance method generates significant errors, accumulating as the deal seasons.

The errors in projected cashflows due to the rolled balance approach are especially significant on the residual tranche of the deals. The errors we measure here are the deviation in collateral cashflows measured against the percentage of the total value of the collateral pool. Using the rolled balance method, the cashflow error of 0.8% at 11 months after the deal closes has a significant impact on the value of the residual tranche. Fluctuations in collateral cashflows at the tail end all go directly into the residual, thus relatively small changes can impact the residual tranche significantly.

CONCLUSION / Common approaches to cashflow generation for securitization deals have many shortcomings and can potentially produce significant inaccuracies. These problems are caused by the lags in the traditional securitization reporting process, as well as the static nature of legacy cashflow tools. Specifically, the commonly used “rolled balance” method of generating cashflows from static rep lines can result in significant errors, especially as time passes.

Modern cashflow engines, such as dv01’s solution, are closely integrated with originators, servicers, trustees, and data providers, allowing them to access up-to-date data for securitization deals at the loan level. This enables users to generate cashflows in loan level detail with up-to-date data with no noticeable delay, as well as have the flexibility to create accurate dynamic rep lines based on updated collateral data, which results in higher accuracy

In a follow-up report, we will examine the effect that discrepancies between input curve assumptions and true default/prepay statistics have on cashflow generation for online lending securitizations.

support@dv01.co | 646-854-5258 | 915 Broadway, Suite 802, New York, NY 10010

support@dv01.co | 646-854-5258 | 915 Broadway, Suite 802, New York, NY 10010