HORIZON
Data Science Applications

# Attrition Study –
# Final Report (redacted)

# Key Questions



**Is there a statistically significant correlation between growth and customer loyalty?**

At the aggregate level, loyalty is **uncorrelated to positively correlated** with growth, depending on how loyalty and growth are measured

- When **loyalty is broadly defined** to include…, then loyalty is **positively** and statistically significantly correlated with growth

# Key Questions



**Is there a statistically significant correlation between growth and customer loyalty?**

- Average attrition risk has actually declined over time on a revenue basis
- “Jolts” (sharp increases in charges over a short period) do occur for individual customers – e.g., a customer could experience, over a period of 6 months, increases of 300% in the number of monthly payments and 50% in total revenue to the company
- The data suggest such jolts may contribute to customer attrition, but there has not been a linear and sequential increase in jolts over time

# Key Questions



**What are the drivers of attrition?**

- **Time in the network/tenure**: Attrition risk is highest…
- **More units** per customer at opt-in, and growth in the number of units, **tend to reduce attrition risk** when the number of additional monthly payments is relatively small; **more monthly payments tends to increase attrition risk**; in cases where the addition of a new unit leads to a large increase in the number of monthly transactions (estimates suggest about 12 or more), then the net effect tends to increase attrition risk
- The occurrence of **very large transactions** (>$XX thousand) may be associated with higher attrition risk

# Agenda

- Understanding the network
- Active customers
- Attrition drivers and interpretations
- Aggregate regression
- Individual customer behaviors and probabilities
- Discussion

# Background and Context

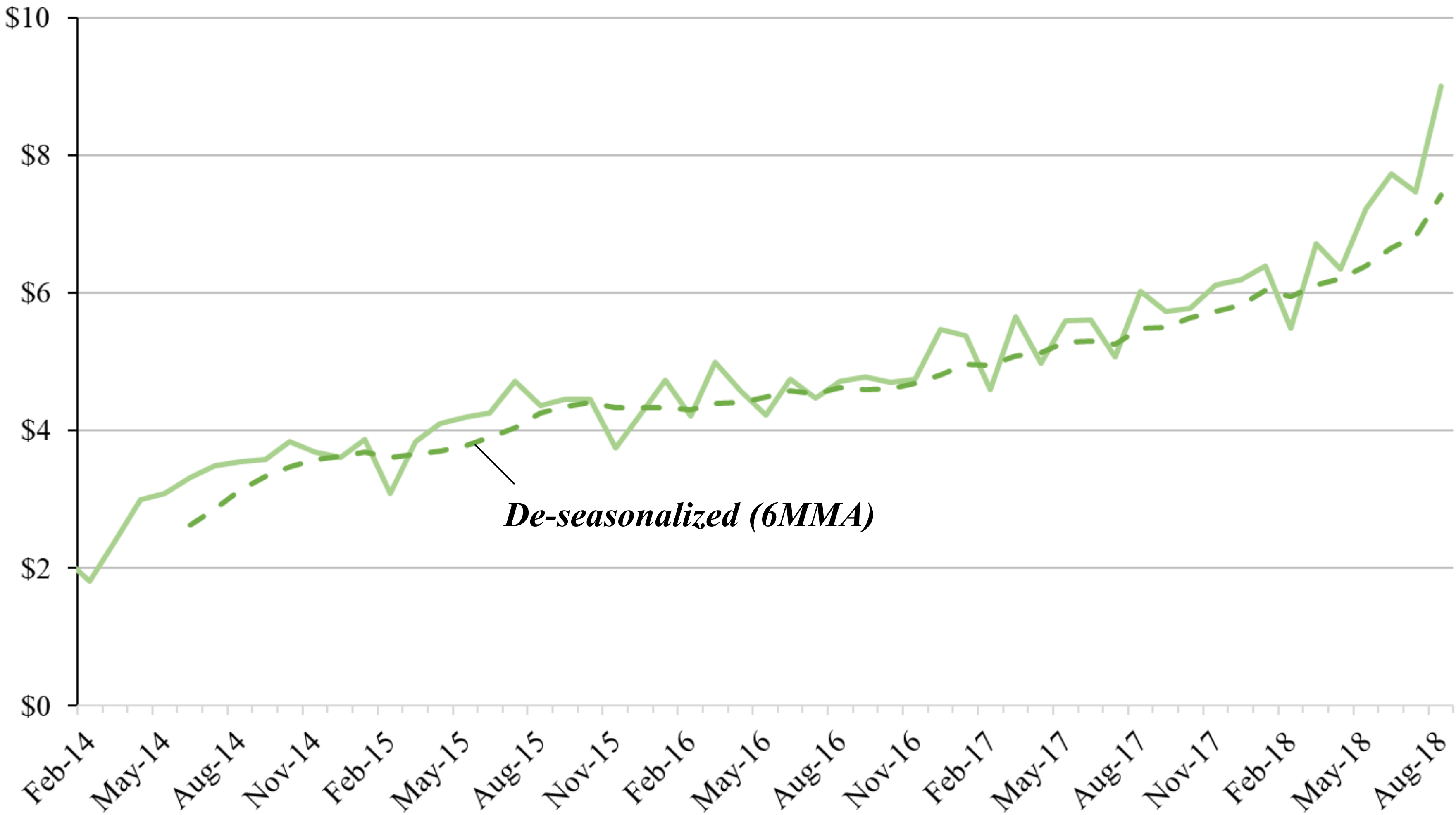
- The team is working to conduct transaction level data analysis
- Specifically, the team would like to look at the factors that drive customer attrition and loyalty
- In particular, the team would like to quantitatively study the impact of growth on customer attrition
- We are also interested in other potential drivers of attrition
- Accordingly, the team would like to commission an outside team to conduct analysis to help inform the company's go-forward strategy

# Data Overview

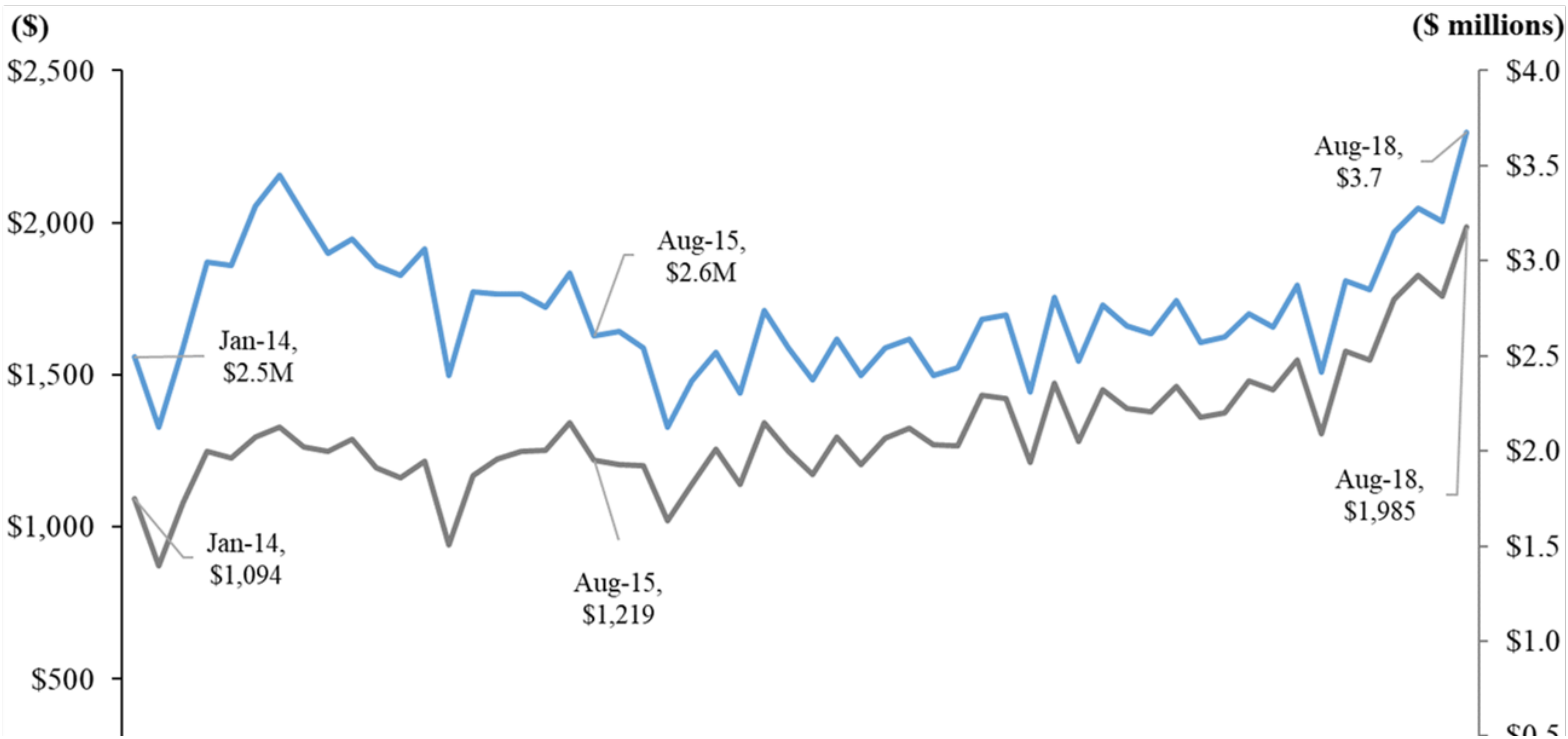
- The company provided the following data:
  - Transaction-level data (millions of observations) including number and size of subscription, revenue, breakdown of fees
  - Data on individual customer characteristics (size, type, products utilized, location),
  - Opt-in/attrition status by customer over time
  - Product details

- The data facilitated customer-level analyses of payments and revenue, by unit, over time, as well as aggregate analyses

# Revenue
# ($ millions)



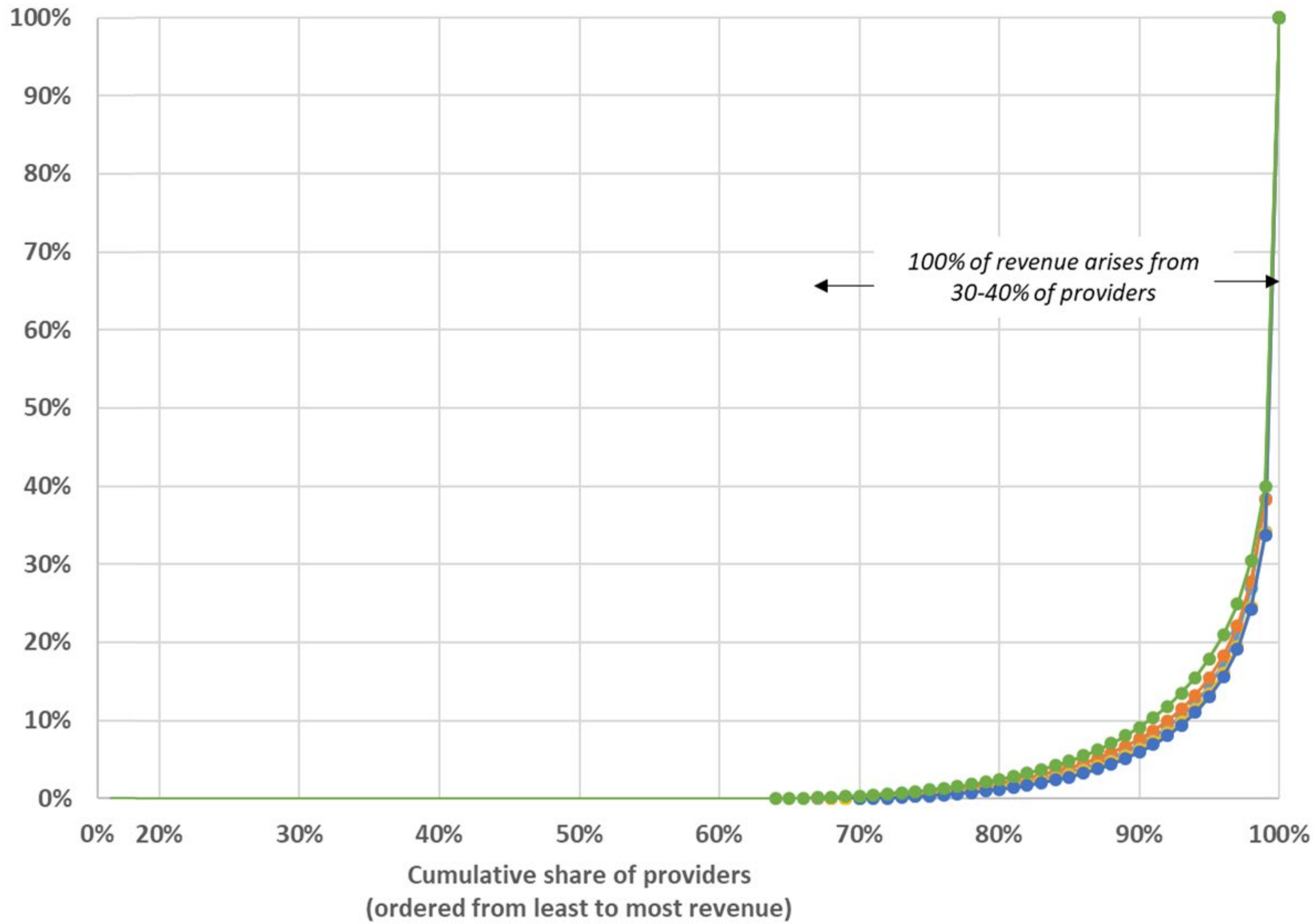
# Aggregate growth



# Revenue generation across customers

- In any given month, only ~10% of customers contribute revenue; the *share* of customers who are revenue-generating has been fairly constant over the company's history
- Constant share of a growing network means that more and more customers are contributing revenue as the company grows over time (i.e., revenue is growing due to increases in revenue per customer and the number of customers)
- The set of revenue-generating customers is not the same from month-to-month; there is substantial volatility in revenue over time for individual customers, with some customers alternating between high and low/zero revenue. Over longer time intervals, however, the identities of the top customers tend to be fairly stable (i.e., those who started out as high revenue generators tend to maintain that status)

# Cumulative distribution of revenue by customer



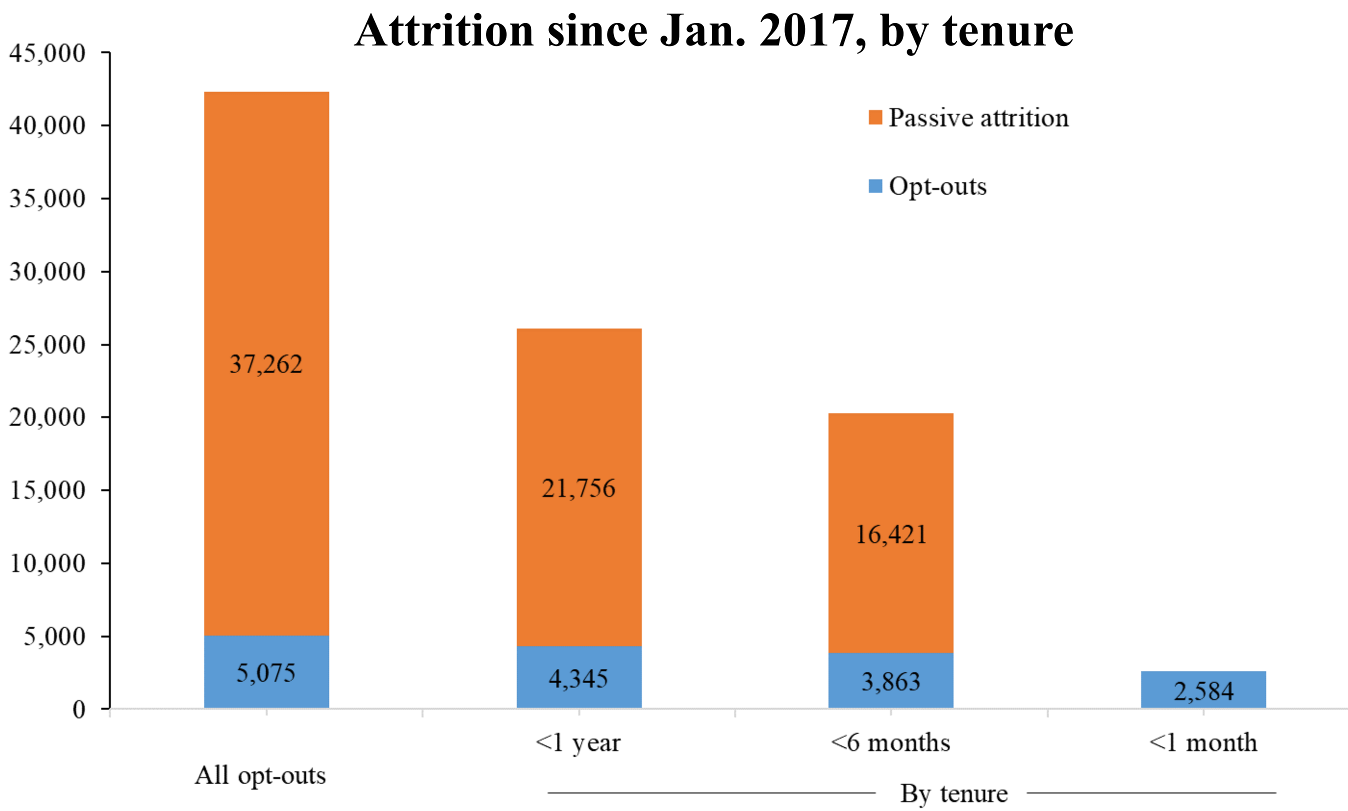
# Defining “active”

- Focusing on active customers (the ones that did start generating revenue), the best predictor of attrition (passive or affirmative) is also one or more consecutive months of zero revenue

- However, there is inter-month volatility in revenue even among active customers; zero revenue in one and even more consecutive months is not unusual (*see next slide)*, so **there is a need to identify additional markers for/deeper drivers of unit and dollar attrition of the business**

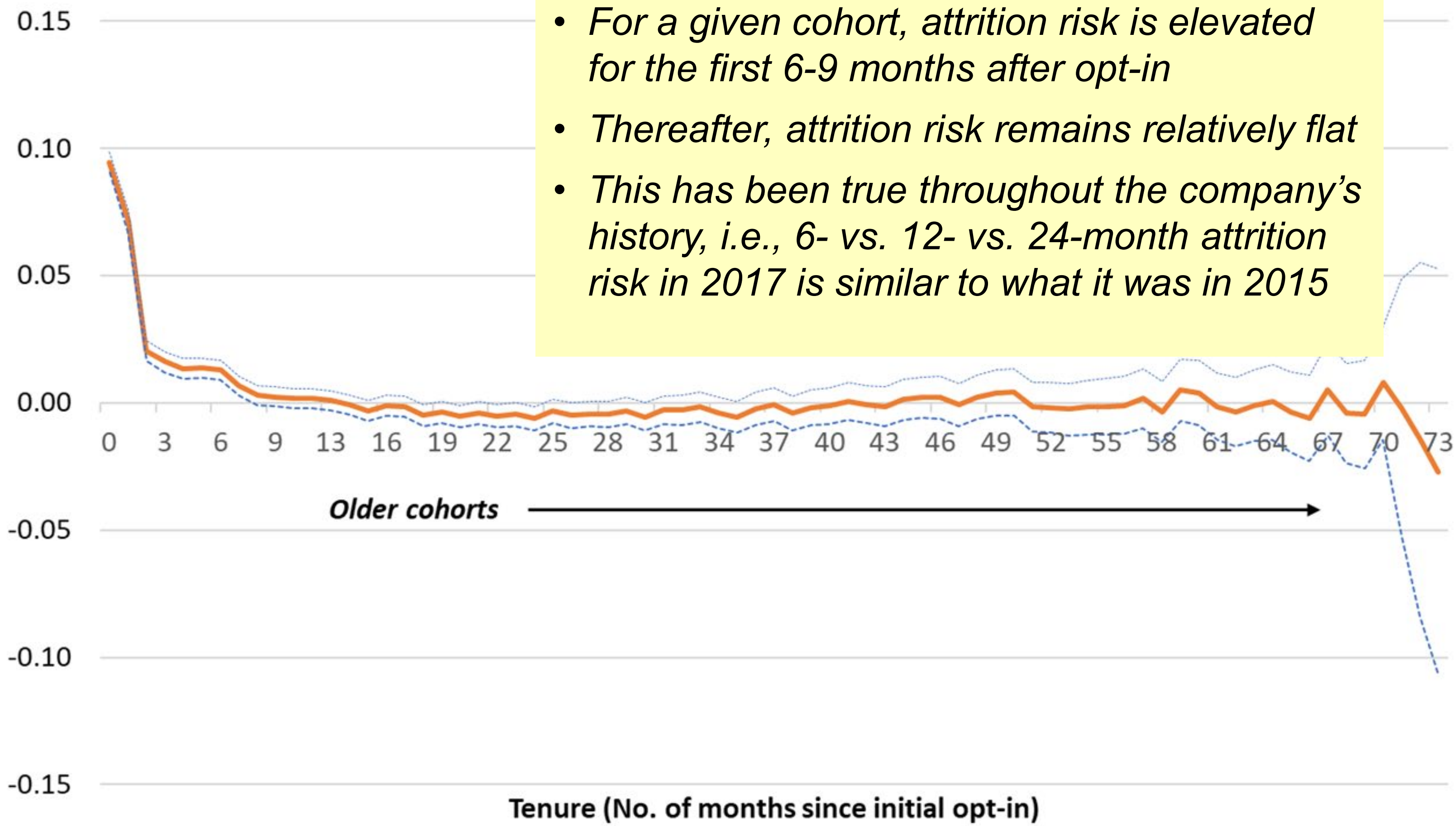
# Predicting attrition

A large share of customers attrite within the first few months after opting in, either passively (by ceasing/never starting to generate revenue) or affirmatively (by opting out)




Note: Based on first-time attriters only.

# Relative attrition risk by customer tenure
## (Baseline: Risk of customers with 12 months of tenure)




- *For a given cohort, attrition risk is elevated for the first 6-9 months after opt-in*
- *Thereafter, attrition risk remains relatively flat*
- *This has been true throughout the company's history, i.e., 6- vs. 12- vs. 24-month attrition risk in 2017 is similar to what it was in 2015*

# Overview of the aggregate model

- We developed models of aggregate unit and revenue attrition over time
- The models were designed to fit the historical data well, but also to facilitate forecasting going forward; we didn’t use a time trend because its usefulness for forecasting purposes is unclear
- The different variables tested were informed by results from modeling of the drivers of attrition at the individual customer level
- The aggregate model also allows us to directly test the question of whether aggregate growth and loyalty are correlated

# Library of potential attrition drivers/ explanatory variables



| **Potential driver** | **Vars.: Customer-level model** (dep. var: yes/no transit in given month) | **Vars: Aggregate model** (dep. vars: total unit and $ attrition in given month) |
|---|---|---|
| Cohort | Cohort dummies (month-years) | In given month, the numbers of active customers who are 1 month old, 6 months old, 12 months old, 18 months old, 24 months old, ... |
| Location | State dummies | In given month, the numbers of active customers in each state/region |
| Month | Month dummies | Month dummies |
| Year | Year dummies | Year dummies |
| Type | Type dummies | Nos. of A's, B's, C's, and other types |
| Size | Size dummies | Nos. of small, medium, and large |
| Product (based on largest source of revenue) | Product dummies (ABC, XYZ, other) | Nos. of customers in ABC, XYZ, and other products |

# Equation: Aggregate revenue attrition

| Total monthly revenue attrition (logs) = | | Role in model |
|---|---|---|
| | $\beta_1$ Month dummies | Accounts for seasonality |
| + | $\beta_2$ # of active customers (log) | Accounts for level of attrition |
| + | $\beta_3$ 12-month trailing revenue (log) | Accounts for level of attrition (specifically revenue attrition) |
| + | $\beta_4$ # of ABC customers (log) | Accounts for increasing uptake of ABC vs. XYZ over time |
| + | $\beta_5$ Peak fee paid in a single month as of 6 months ago (log of mean across customers) | Benchmarks peak fee (revenue) paid to company prior to 6 months ago |
| + | $\beta_6$ Peak fee paid in a single month from opt-in up to 6 months ago (log of mean across customers) | Captures effect of "jolt", defined higher peak revenue in past 6 months |

# Aggregate regression and forecast

- Does growth drive attrition? No, not aggregate growth
    - When we added the most intuitive measures of the aggregate size of the network to the regression (e.g., total purchases, total purchases per customer), these variables did not have a positive and statistically significant correlation with aggregate attrition (either on a unit or revenue basis)

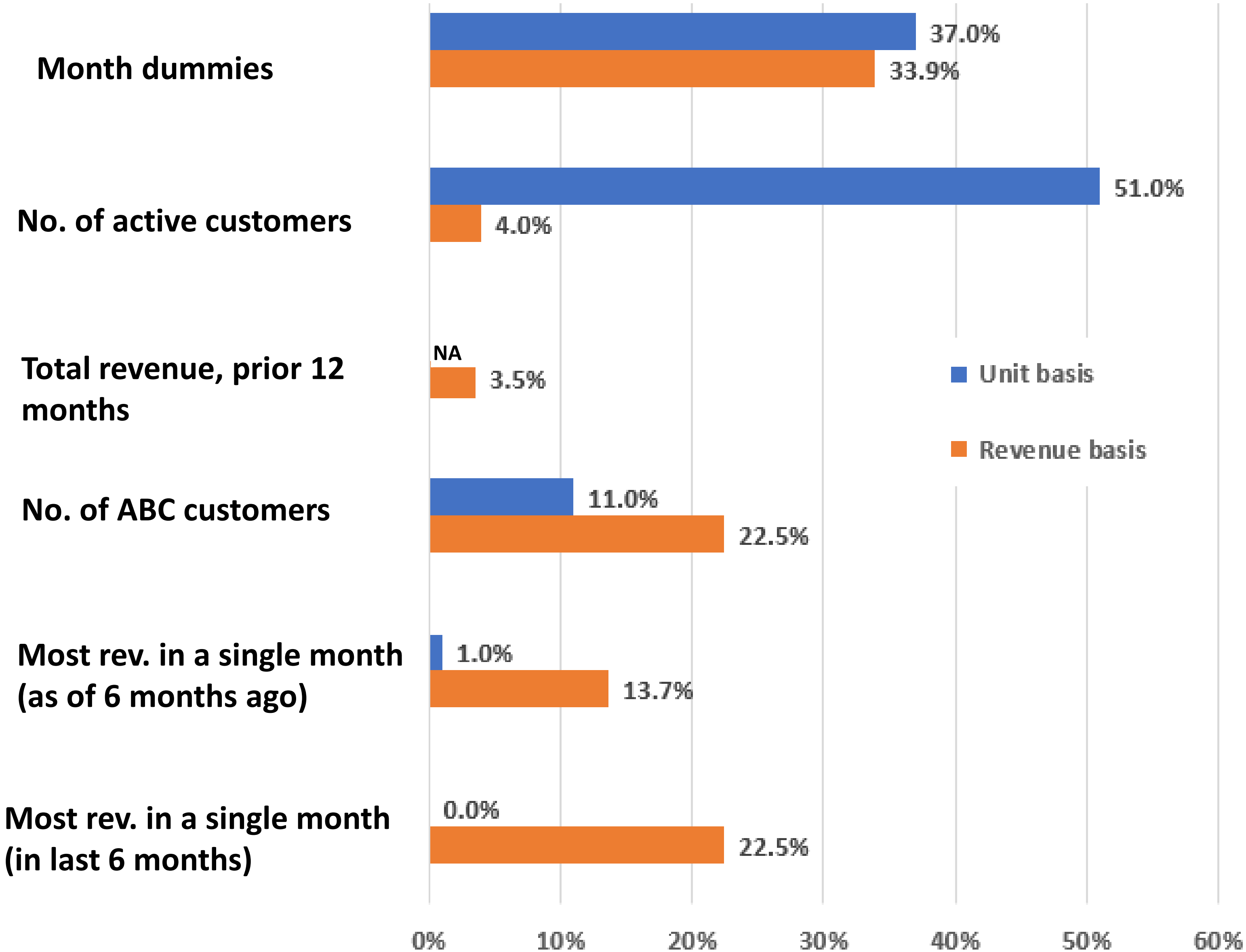
- This is unsurprising when one compares the shape of aggregate network growth over time vs. where our model fit of attrition misses
    - The model does a decent job of capturing the choppy monthly fluctuations in revenue. Where it misses most is in capturing more gradual – but non-monotonic – changes in attrition. By contrast, size of customer base has generally increased monotonically over this period

# Coefficient estimates

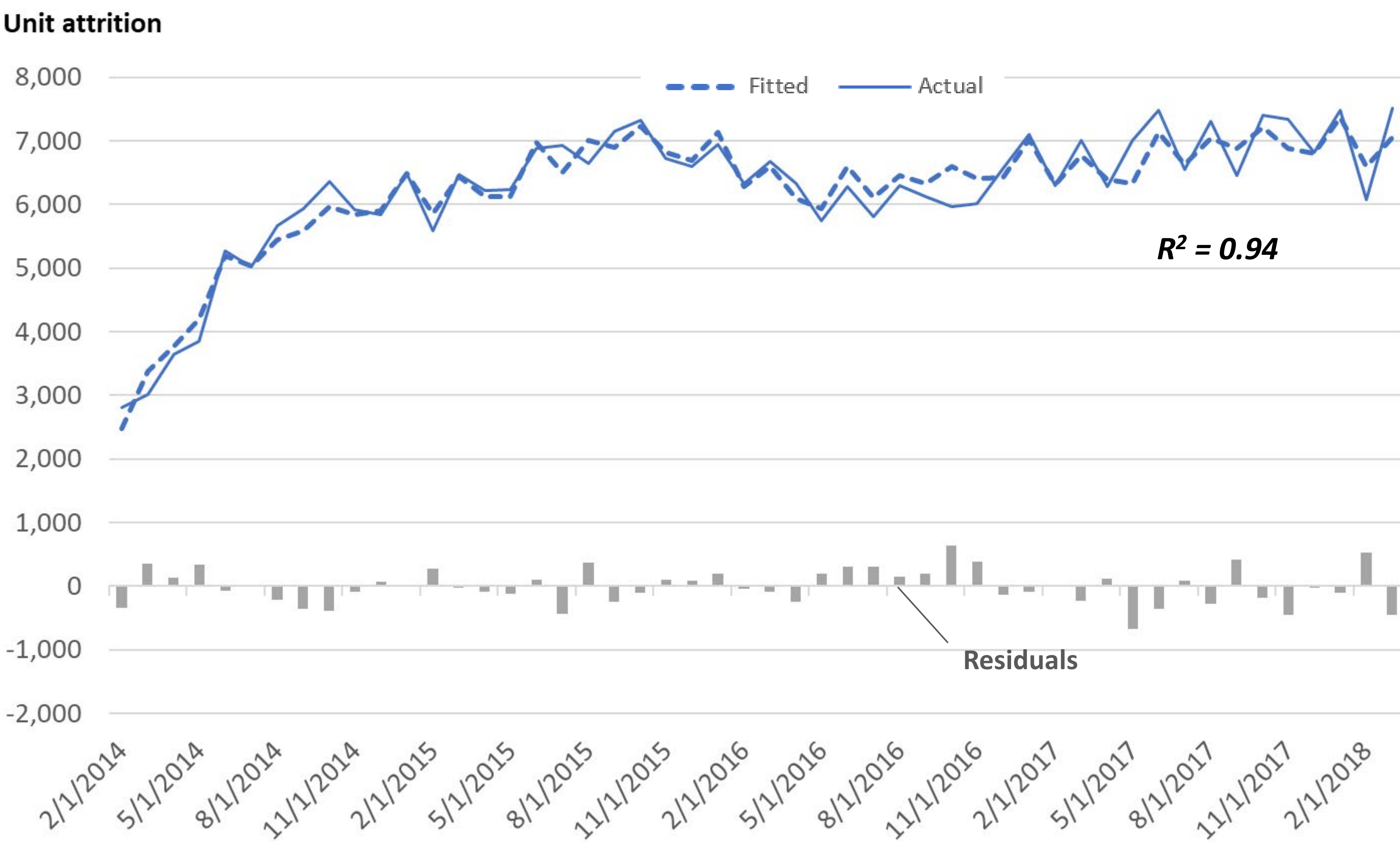
| Dependent variable (log): | Unit attrition | | Revenue attrition | | Interpretation |
|---|---|---|---|---|---|
| **No. of active customers (log)** | 1.200 | *** | 2.905 | | 1% increase in # of active customers → 1.2% increase in unit att'n (insig. effect for rev) |
| **Total revenue, prior 12 months (log)** | | | -1.338 | | No stat. sig. association |
| **No. of ABC customers (log)** | -0.126 | ** | -1.370 | *** | 10% increase in # of customers using ABC (holding total # of customers fixed) → 1.3% decrease in unit att'n, 13.7% decrease in rev att'n |
| **"Jolt" effect** | | | | | |
| Most rev. in a single month to date as of 6 months ago (avg. across customers) | -0.002 | | 0.059 | ** | 10% higher peak monthly fees in more distant past → 0.6% higher rev att'n |
| Most rev. in a single month during the past 6 months (avg. across customers) | -0.001 | | 0.103 | *** | 10% increase in peak monthly revenue (fees) (holding constant fees paid in the more distant past) → 1% higher rev att'n |
| **Month dummies?** | Yes | | Yes | | |
| **No. of obs. (months)** | 50 | | 50 | | |
| **$R^2$** | 0.94 | | 0.85 | | Models explain 94% and 85% of variation in aggregate unit and revenue attrition over time, respectively |

Note: Asterisks denote statistical significance (***, p<0.01; **, p<0.05)

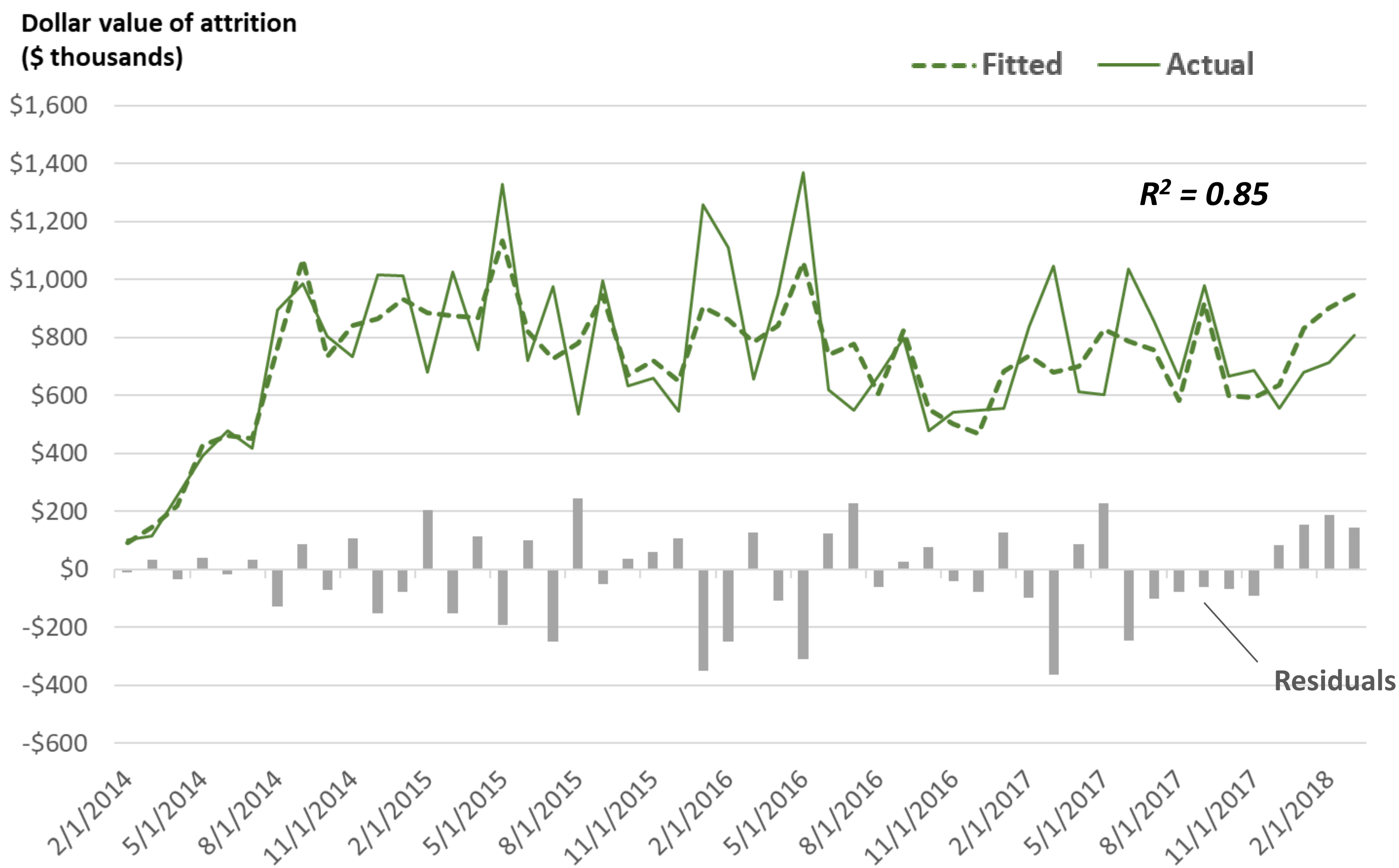
# Contribution to fit



# Fit: Unit attrition model



# Fit: Revenue attrition model



# Relationship between customer-level and aggregate models of attrition

- Attrition is a highly idiosyncratic process, so even while we can use the company's rich data gain meaningful insights into drivers, most of what drives the attrition status of any individual customer in a given month remains unexplained (low $R^2$)

- However, to the extent the individual model is "on average" correct, we can use it to inform a model of *aggregate* unit attrition and associated revenue attrition (high $R^2$ in fit with historical data)

# Baseline model

Our "Baseline" model of attrition risk at the customer level included:

- Year and month dummies to allow for a time trend and account for seasonality
- Cohort dummies to allow for differential attrition risk among customer cohorts
- Product and segment dummies

All specifications reported in this section were estimated by OLS (linear probability models); logistic models yielded qualitatively similar results

# Baseline model results: Time trend

*Baseline model*

| Year (ref: 2014) | | |
|---|---|---|
| 2015 | -0.024 | *** |
| | 0.001 | |
| 2016 | -0.042 | *** |
| | 0.001 | |
| 2017 | -0.054 | *** |
| | 0.001 | |
| 2018 | -0.072 | *** |
| | 0.001 | |

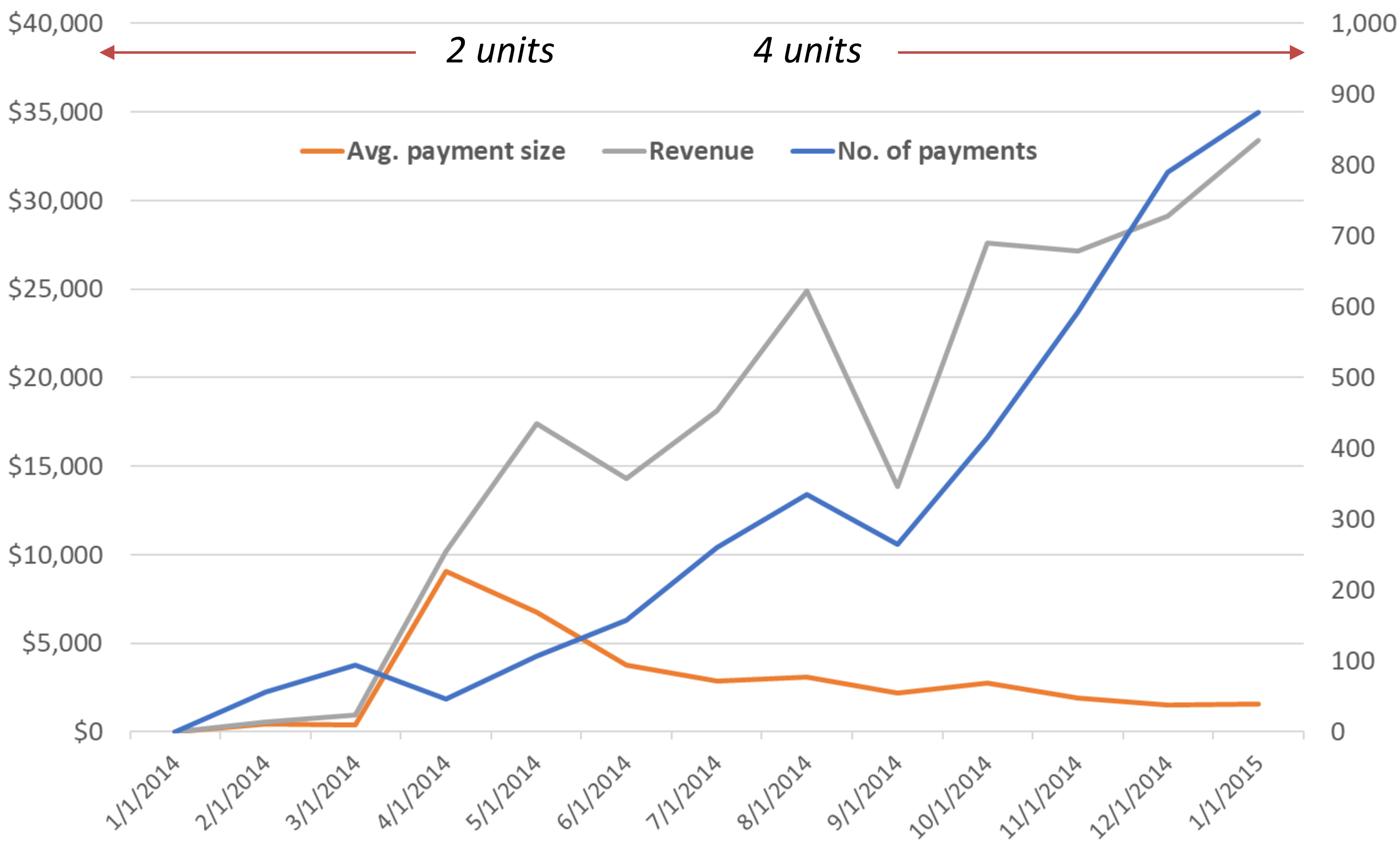
Notes:

1. Asterisks denote statistical significance (***, $p<0.01$)
2. Std. errors reported directly below coefficient estimates
3. Model includes dummies for month, cohort, and state
4. No. of observations: 719,926; $R^2$: 0.048

**Interpretation**

- Year dummies (time trend) allows for average attrition risk to differ from year to year
- **The increasingly negative coefficients shown here point to a decrease in average attrition risk over time (holding other factors in the model constant)**
- 2018 coefficient of -0.072 → avg. attrition risk is 7.2% lower in 2018 vs. 2014 and (0.072-0.054)=1.8% lower in 2018 vs. 2017

# Jolt effect?
# Facility 1734A, opted out in 1/2015




Implied price, %

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2.41% | 2.41% | 2.39% | 2.41% | 2.41% | 2.41% | 2.41% | 2.42% | 2.41% | 2.42% | 2.42% | 2.42% |

# Knowledge of attrition drivers can be leveraged

A better understanding of unit and dollar attrition drivers is useful in two ways:

1. Aids attrition mitigation at individual customer level (e.g., facilitates earlier intervention to prevent attrition, informs strategies to target higher NPV customers)

2. Allows for business-level analyses based on forecasts and simulations of aggregate attrition