Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
A flexible framework for intervention analysis applied to credit-card usage during the coronavirus pandemic

Anson T.Y. Ho, Lealand Morin, Harry J. Paarsch, Kim P. Huynh

Ted Rogers School of Management, Ryerson University, Canada
Department of Economics, University of Central Florida, United States of America
Currency Department, Bank of Canada, Canada

ARTICLE INFO

Keywords:
COVID-19
Liquidity
Intervention analysis
Consumer finance
Markov model

ABSTRACT

We develop a variant of intervention analysis designed to measure a change in the law of motion for the distribution of individuals in a cross-section, rather than modeling the moments of the distribution. To calculate a counterfactual forecast, we discretize the distribution and employ a Markov model in which the transition probabilities are modeled as a multinomial logit distribution. Our approach is scalable and is designed to be applied to micro-level data. A wide panel often carries with it several imperfections that complicate the analysis when using traditional time-series methods; our framework accommodates these imperfections. The result is a framework rich enough to detect intervention effects that not only shift the mean, but also those that shift higher moments, while leaving lower moments unchanged. We apply this framework to document the changes in credit usage of consumers during the COVID-19 pandemic. We consider multinomial logit models of the dependence of credit-card balances, with categorical variables representing monthly seasonality, homeownership status, and credit scores. We find that, relative to our forecasts, consumers have greatly reduced their use of credit. This result holds for homeowners and renters as well as consumers with both high and low credit scores.

1. Introduction and motivation

In an interview published in this journal, George E. P. Box described how he and George C. Tiao set out to study air pollution in Los Angeles (Peña, 2001). This analysis led to their seminal paper on intervention analysis, Box and Tiao (1975), in which they studied pollution concentration levels surrounding the introduction of a law restricting the types of gasoline sold. Since then, intervention analysis has been used in such diverse fields as driver safety (Bhattacharyya & Layton, 1979), marketing (Leone, 1987), television viewership (Krishnamurthi et al., 1989), call-center operations (Bianchi et al., 1998), and, more recently, the study of disease outbreaks (Daughton et al., 2017). Intervention analysis has also featured prominently in economics literature—to investigate the effects of government policies on the...
Consumer Price Index (Box & Tiao, 1975), the inflation-targeting strategies of central banks (Angeliz & Arestis, 2008), the effect of local tax policy (Bonham & Gangnes, 1996), the linkages between equity and futures markets (Bhar, 2001), and the impact of natural disasters on capital markets (Worthington & Valadkhani, 2004).

In this paper, we apply intervention analysis to measure changes in credit usage during the COVID-19 pandemic, using Canadian consumer credit data from January 2017 to September 2020. Researchers typically conduct intervention analysis by estimating a time-series model, such as an autoregressive integrated moving average (ARIMA) model, with a long series of aggregated data. In contrast, we take a short but extremely wide panel of individual-level data and use consecutive pairs of monthly observations to identify the law of motion. This approach is the first to take advantage of the variation among individuals within a cross-section, and permits estimation over a much shorter time span or with data that are more widely spaced over time than would otherwise be possible.

Using our method, we discretize the cross-sectional data of individuals in a panel data set to approximate nonparametrically the distribution of measurements. That is, the law of motion in this model governs the time series of the distributions of the individual outcomes. Nonparametric methods have been applied to the mean prediction from a time-series model within an intervention analysis. Stock (1989) applied a semiparametric model for the variable of interest that included dummy variables to measure the effects of the intervention. Park (2012) followed a similar approach, to build a central mean subspace for the time series, which is a dimension-reduction method focused on the conditional mean. To the best of our knowledge, none has been applied to the distribution of the variable of interest.

The unique feature of our approach is that it avoids modeling the moments of the population. The discretized distribution is modeled to follow a Markov process, with transition probabilities represented by a multinomial distribution, permitting these probabilities to depend on covariates. The result is a flexible framework for conducting intervention analysis. Consequently, the effects related to time-varying mean or variance are already admitted. The distribution is fit to the data with little risk of bias from misspecification, creating a reliable benchmark for measuring the effects of the intervention. Furthermore, this framework is resilient to all sorts of otherwise inconvenient features of the data—including point masses as well as natural boundaries or missing observations. Such an approach is critically important when modeling the behavior of a large population of individuals, many of whom have an experience that interacts with these inconvenient features of the variable in question, and would otherwise interfere with the measurement obtained with conventional methods.

In addition to allowing a greater ability to fit the data, the flexibility of our framework allows a researcher to detect responses to interventions that lie outside the canonical set of responses. Typically, the response to an intervention is assumed to take on one of four forms: a gradual change, an abrupt change, an abrupt change that reverses, or an abrupt change that gradually subsides. All four of these forms describe a change in the mean or expected value of the variable and are otherwise only differentiated by the time dimension. Our approach also addresses the nature of the change in the distribution, which can manifest in changes other than the mean. For instance, the response to an intervention could take the form of an increase in the variance or skewness of the distribution, leaving the mean constant, following any of the canonical response patterns in the time dimension. With a framework that is nonparametric, the possibilities extend beyond an analysis of the moments of the distribution. In the language of the time-series literature, the goal of intervention analysis is to estimate a form of structural break in a dynamic process. Our approach is rich enough to detect structural breaks that materialize in the shape of the distribution, rather than in the values of parameters, such as those specifying the mean of the distribution.

This framework is applied to analyze the credit-card balances of Canadians during the COVID-19 pandemic. Our empirical application extends the work in Ho et al. (2021) by entailing a closer inspection of consumer credit usage. In general, our results show large decreases in credit-card balances that are consistent with their findings. Interestingly, we also found that consumers with the tightest credit constraints did not become more indebted during the first wave of COVID-19. Indeed, there was a large increase in the fraction of these consumers with a balance of less than $500. The effect of COVID-19 on their credit usage persisted when the economy reopened after the first wave subsided. On the other hand, we found a substantial decrease in the proportion of creditworthy homeowners with balances between $3000 and $14,000 during the first-wave lockdown periods. This group of consumers also had the strongest recovery in credit usage after the economy reopened, suggesting that the reduction in balances was due to limited spending opportunities. Some of these creditworthy homeowners, however, also appeared in greater proportions in the highest balance categories, although this occurred for a small number of consumers.

Our empirical findings complement literature concerned with the effects of COVID-19 on household finances. Much of this research regarding the crisis has focused on consumer spending. According to Baker et al. (2020), consumers significantly reduced overall spending, and credit-constrained households responded rapidly to the fiscal stimulus payments from the 2020 CARES Act in the United States. In Denmark and Spain, similar effects on consumer spending were found in Andersen et al. (2020) and Carvalho et al. (2020). In Canada, Ho et al. (2021) used a special case of our framework and found that consumers paid down balances of their credit cards and home-equity lines of credit (HELOCs). In other work, Chen, Engert, et al. (2020) showed that consumers also increased cash holdings.

Considering the economic disruption and surging unemployment rate due to COVID-19, our findings are in sharp contrast to existing evidence in the household finance literature that individuals tap into various forms of
credit to smooth their income shocks. For instance, Sullivan (2008) and Agarwal and Qian (2014) found that consumers with few assets smooth their unemployment shocks via unsecured debts. Wealthy homeowners also tend to borrow against their home equity through mortgage refinancing (Chen, Michaux, & Roussanov, 2020; Hurst & Stafford, 2004) or HELOCs (Agarwal et al., 2006).

Our methodology also contributes to the literature on applying nonparametric methods to credit risk management. An early application was tested by Khandani et al. (2010), who constructed nonlinear nonparametric forecasting models of consumer credit risk. They obtained significant improvements in the accuracy of forecasted delinquencies. Kruppaa et al. (2013) also documented improved forecasts from using an implementation of random forests. Yao et al. (2017) used a support vector machine as a classifier in their two-stage loss given default models for credit cards. Jiang et al. (2021) employed large-scale alternative data to construct credit scores to predict consumer delinquency. Although our methodology employs a more rudimentary form of nonparametric analysis, our framework is rich enough to incorporate both the nonlinearity and the dependence present in the data.

The ability to account for dependence affords other possibilities for credit modeling in the time dimension. For policymakers, this framework is particularly useful to analyze the effect of policies that target specific groups of individuals, such as mortgage stress testing and macroprudential policy (see, for example, Siddall, 2016). These policies often concern the tail in a distribution and have important implications on financial stability (Allen et al., 2020) that can be masked over by investigating lower moments. For practitioners, our modeling framework can be used to guide a dynamic model-fitting strategy for credit risk management. The quality of predictions from a credit risk model may degrade over time due to various factors, such as business competition, that shift the client base. Industry practitioners can use our framework to analyze the distribution of their clients’ characteristics. A change in the distribution may indicate that the business landscape has changed over time. For instance, a different population of credit applications may be flowing through the model, or an existing set of clients may be experiencing changes, such that the model specification is different from that under the original model-fitting exercise. Generally, individual-level data are often stored as wide panels of microdata with a limited number of observations. Such data are often highly variable and do not often conform to a parametric distribution. The flexibility of our framework provides a procedure to manage risks in complex and realistic environments.

The remainder of the paper is in four additional sections: In the next section, we first summarize briefly the notions behind intervention analysis. We then develop our econometric model to augment the standard approach to intervention analysis to produce a more flexible framework. In Section 3, we provide simulation evidence for the size and power properties of our test statistics. In Section 4, we apply our technique to a data set that is well suited to this approach—Canadian consumer credit data—and report our empirical results. In the final section of the paper, Section 5, we summarize and conclude.

2. Intervention analysis

In this section, we describe our methodology for conducting intervention analysis. We begin with a description of a standard approach, to contrast with our framework.

2.1. Background

The canonical form of intervention analysis is conducted with a time-series model, most commonly, the autoregressive moving-average (ARMA) model. ARMA models were designed to do two basic things: (1) refine forecasts by using past information, and (2) admit dependence when constructing confidence intervals as well as testing hypotheses. Testing hypotheses concerning events—either anthropogenic or natural—that cause changes in the process can be tricky because distinguishing between the dynamic effects of the change and the dependence in the process is difficult. Researchers interested in investigating the magnitude of the effects of changes have to take a stand: dependence in the errors often either masks the effect itself or makes it difficult to decide whether the effect is statistically significant. To address such difficulties, Box and Tiao developed the tools of intervention analysis.

In general, an ARMA model of the random variable $Y$ in period $t$ can be written as

$$\rho(L)(Y_t - \mu_0) = \phi(L)\varepsilon_t,$$

where $\phi(L)$ and $\rho(L)$ are polynomials in the lag operator $L$, which has the property

$$L^2 Y_t = LLY_t = LY_{t-1} = Y_{t-2},$$

for instance. The elements of the sequence $\{\ldots, \varepsilon_{t-1}, \varepsilon_t, \varepsilon_{t+1}, \ldots\}$ are assumed to be jointly independent, Gaussian innovations that have mean zero and constant variance $\sigma^2_\varepsilon$. An example of an ARMA(1,1), for instance, would involve

$$\phi(L) = (1 + \phi L) \quad \text{and} \quad \rho(L) = (1 - \rho L).$$

In this case, for technical reasons, the following restrictions would be required:

$$|\phi| < 1 \quad \text{and} \quad |\rho| < 1.$$

The above general ARMA model can be rewritten as

$$Y_t - \mu_t = \frac{\phi(L)}{\rho(L)} \varepsilon_t,$$

where we have a subscripted $\mu$, the mean, to highlight the fact that a change, denoted $\delta$ below, will be relative to that object. Suppose, now, that at some time $T$ a change occurs, so that after that date,

$$Y_t - \mu_t = D_t + \frac{\phi(L)}{\rho(L)} \varepsilon_t,$$

where $D_t$ represents the magnitude of the change in period $t \geq T$.

Several possible patterns exist concerning how the intervention may affect the mean of the stochastic process over time: First, a permanent, constant change in the mean may occur; for example, in Panel (1) of Fig. 1,
an increase in every period after $T$ is obtained. Second, a brief constant change in the mean could occur; for example, in Panel (2) of Fig. 1, a temporary decrease in the mean occurs, but has no effect thereafter. Third, a gradual increase (or decrease) might happen; for example, in Panel (3) of Fig. 1, the mean of the process rises to an asymptote. Fourth, a brief initial change, but then a return to the previous mean, may be the outcome; for example, in Panel (4) of Fig. 1, a temporary increase obtains, which then subsides.

To model the first sort of intervention, simply introduce a dummy variable $I_t$, which equals zero up to $T$ and one thereafter, so

$$ D_t = \delta I_t. $$

To model the second sort of intervention, introduce a similar dummy variable $I_t$, which equals zero up to $T$ and one during the period of effect, and then zero thereafter, so

$$ D_t = \delta I_t. $$

To model the third sort of intervention, introduce a dummy variable $I_t$, which equals zero up to $T$ and one thereafter, but some additional structure is required. Specifically, one needs a model of the gradual effect. One such model involves period-to-period increases that are proportional to the current distance between the old mean $\mu_0$ and the new one $\mu_1$, such as

$$ \mu_T = \mu_{T-1} + \omega(\mu_1 - \mu_{T-1}) $$

$$ 0 < \omega < 1 \text{ and } T = T + 1, T + 2, \ldots $$

where $\mu_T = \mu_0$, and the parameter $\omega$ determines the speed of the gradual change: if $\omega$ is small (say, 0.1), then the change is slow, whereas if $\omega$ is large (say, 0.9), then

the change is relatively fast. This, then, reduces to the following:

$$ D_t = \frac{\delta I_t}{(1 - \omega L)}. $$

where $\mu_1 = \mu_0 + \delta$. To model the fourth sort of intervention, introduce a dummy variable $I_t$, which equals zero up to $T$, one in period $T$, and zero thereafter. Again, some additional structure is required in order to model the gradual return to the old mean. As in the third case, one such model involves a decrease that is proportional to the distance between the new mean $(\mu_0 + \delta)$ and the old one $\mu_0$. This then reduces to the following:

$$ D_t = \frac{\delta I_t}{(1 - \omega L)}. $$

In order to estimate the effect of an intervention, one must know its date—exactly—because one must specify one of the above four models of the intervention beginning at that date. Perhaps belaboring the obvious, if one gets that date wrong, then the entire analysis is at risk: one may totally mismeasure the effect. That noted, once one has determined $T$, one can then identify and estimate the ARMA model in the usual way, using the data from before $T$, and then forecast what would have occurred under the old regime. Based on that forecast, one can then calculate the differences between actual outcomes after the intervention and the forecasted ones. By examining these differences, one can then choose one of the four models of the intervention pattern, and then estimate a particular empirical specification using all of the data. In our work, we do not put structure on the shape of the intervention, choosing instead to estimate the effect nonparametrically.

One of the main limitations of intervention analysis is that the ARMA process is assumed to remain the same
after the intervention as it was before the intervention. Obviously, this is a strong assumption, but it is necessary in order to decompose the dynamic effects of the intervention from the dynamic effects of the dependence. This assumption is reasonable for our analysis, since the true extent of the pandemic was largely unknown before its arrival in Canada, and the policy response was quickly executed, allowing very little time for consumers to act in anticipation.

2.2. A flexible framework for intervention analysis

We now outline a framework for intervention analysis that focuses on the distribution of the cross-section over time. To begin, consider a sample of $N$ outcomes $(Q_1, Q_2, \ldots, Q_N)$ that have been drawn independently from a common cumulative distribution function (CDF) $F^0(q)$. The Glivenko–Cantelli Lemma allows one to estimate the population CDF $F^0(q)$ consistently using the empirical distribution function (EDF):

$$\hat{F}(q) = \frac{1}{N} \sum_{n=1}^{N} I(Q_n \leq q),$$

where $I(A)$ equals one if the event $A$ obtains, and is zero otherwise. That is, the EDF is the proportion of sample outcomes at or below some value $q$. Using transformations of Eq. (1), one can also construct consistent estimates of such population measures as the mean and variance as well as quantiles, such as the median.

This is useful because an intervention need not affect the variable of interest in the same way across the support of the distribution. In other words, nonlinearities may exist. One way in which nonlinearities can manifest themselves is when those who are currently at a boundary (for example, zero credit-card balances) are induced to the interior (for example, positive credit-card balances), whereas those at the interior (for example, those who have already positive credit-card balances) respond differently. Currently, some existing models can deal reasonably effectively with the interior (for example, positive) outcomes, whereas other models can accommodate boundary (for example, zero) outcomes. Integrating the two kinds of models is, however, an open area of development.

Inevitably, external factors beyond our control can affect outcomes, too. Such omitted factors (which may often be unmeasurable) can exhibit strong dependence over time—in short, correlated errors. Correlation in the errors need not imply inconsistency when estimating changes due to events, but correlation can result in estimates of sampling variability that are incorrect, which in turn implies potentially misleading test statistics.

In basic statistics courses, it is common to write the mean squared error (MSE) of an estimator $\hat{\theta}$ of an object $\theta^0$ in terms of the variance ($\mathbb{V}$) of $\hat{\theta}$ as well as the square of its bias ($\text{Bias}^2$). In other words,

$$\text{MSE}(\hat{\theta}) = \mathbb{V}(\hat{\theta}) + \text{Bias}^2(\hat{\theta}).$$

This decomposition in Eq. (2) highlights the tension in some situations where an unbiased estimator has considerable sampling variability, which can be reduced by introducing a biased estimator $\tilde{\theta}$ of $\theta^0$ that has less sampling variability. The research of Stein (1956) is an important example, and helped to spawn the empirical Bayes literature. Elsewhere, building on the research of Andrey Tikhonov in the 1940s and reported initially in Russian, researchers have suggested regularization as a way to reduce variance, but again introducing bias; for example, the LASSO of Tibshirani (1996) is an important method. Recently, Mao and Zheng (2020) noted that economic theory can be used as a form of regularization to reduce sampling variability, but again by introducing bias.

These considerations of the tradeoff between bias and sampling variability are the primary motivation for developing this specific framework for intervention analysis. In order to achieve an accurate benchmark with which to compare performance, we aim to minimize the imposition of theory. This approach allows much more freedom for the data to inform the analysis, producing a representative counterfactual prediction. As long as we have a large data set, we have the luxury of taking on such an agnostic approach to modeling the law of motion while still maintaining a low degree of sampling variability.

An added benefit of our flexible framework is that it allows for a consistent estimate of the intervention response in the presence of bounds, across the entire distribution. To illustrate this point, consider an analysis of a simple autoregressive model with a lower bound of zero. In the context of credit-card balances, this is represented by a discrete time series of the consumer’s net worth. When the net worth is negative, the value is represented by the positive value of the credit-card balance. If the consumer’s net worth is positive, then the consumer holds a zero balance on credit cards and might also hold a positive balance in another account that we do not observe, such as a checking account.

Using this censored series, the estimates of the parameters in the autoregressive model would be both biased and inconsistent. In particular, the estimates of the mean credit balance would be biased upwards, that is, higher credit balances, since the censored series ignores the positive net worth, that is, negative credit balances, at zero. Furthermore, the constraint to non-negative credit-card balances creates the illusion of stronger mean reversion, since the series never moves beyond zero in the negative direction. It is even true that, for example, a random walk with a negative drift would appear to be stationary. In less extreme cases, the speed of mean reversion would be overestimated, as there is, effectively, an infinite degree of mean reversion over the lower bound of zero that is never crossed.

In the above example, the researcher naïvely estimates the model as if the data were not censored. A more sophisticated treatment of censored data can be achieved with a Tobit model, named in honor of the American economics Nobel laureate, James Tobin (1958). This sort of model is designed to analyze random variables in a cross-section that has both continuous and discrete components. In Appendix A.1, we outline an example in which a series conforms to such a specification. Following Amemiya (1984), consider the latent random variable $Q^*_i$, which follows the Gaussian law, having mean $\mu$ and
variance $\sigma^2$. Suppose the observed random variable $Q_n$ is defined by the following:

$$Q_n = \max(0, Q^*_n).$$

In the notation of the example above, $Q^*_n$ represents the negative of the discounted lifetime income of the consumer. For all the observations in which $Q^*_n$ is positive, the consumer holds a credit-card balance. When $Q^*_n$ is negative or zero, however, only a zero balance is recorded in the credit-card account. As a result, a point mass exists at $Q = 0$, which is accounted for in the likelihood function of the Tobit model.

In Appendix A.1, we study a specific example with a first-order autoregressive model—an AR(1) process:

$$(Q^*_n - \mu) = \rho(Q^*_{n-1} - \mu) + \epsilon_n,$$

where $\epsilon_n$ denotes independently and identically distributed Gaussian random variables having mean zero and variance $\sigma^2$. We consider the results of analyzing this data with a Tobit model, while ignoring the dependence. Robin-son (1982) showed that, in this situation, one can estimate $\mu$ consistently, but the asymptotic standard errors are different from those derived from the Hessian matrix for the Tobit model. Furthermore, one cannot consistently estimate the correlation parameter $\rho$. We collected some Monte Carlo evidence that supports these claims. Perhaps most disturbing, however, we found that the error in the estimation of the standard errors grew with the sample size. Resolving these problems would require an estimation method that properly accounts for the dependence. To specify the likelihood function appropriately requires integrating out the missing observation each time a zero is recorded, which can be computationally burdensome, especially in the event of many consecutive zero observations, as is common in our data set.

Aside from dependence in the data, the presence of heteroskedasticity can also complicate inference. Hurd (1979) as well as Arabmazar and Schmidt (1981) demonstrated that heteroskedasticity within the Gaussian family could result in biased and inconsistent Tobit MLEs. In addition, the standard errors based on the Hessian matrix are also incorrect. Arabmazar and Schmidt (1982) as well as Paarsch (1984) demonstrated that when the errors do not follow the Gaussian law, the Tobit MLE, assuming the Gaussian law, is biased and inconsistent; the standard errors based on the Hessian matrix are incorrect, too. In Appendix A.1, we provide some Monte Carlo evidence for these claims in the case when the errors follow the lognormal (Galton) law. These simple models underscore the importance of employing a model with the flexibility to account for the features of the data.

The flexibility of our modeling framework also affords the ability to detect responses that take on a variety of forms. In each of the response patterns depicted in Fig. 1, the response is clearly detectable through a change in the mean of the process. This characterizes the sort of response that can be detected with the traditional forms of intervention analysis. In contrast, in Fig. 2, we expand the set of patterns of intervention responses shown in Fig. 1 to include responses in which the mean remains constant, but changes effect in the higher moments. In these mean-invariant intervention responses, the focus is on the cross-sectional distribution of the response, rather than focusing solely on the time dimension.

As a simple example, consider Panel (1) of Fig. 2, where the intervention manifests in the form of a change in variance with a constant mean. In Panel (2), the intervention response changes the kurtosis of the distribution, leaving all three lower moments unchanged. In our application to credit usage, these distributions were selected to model the following potential response. Suppose the pandemic affected two groups in opposite directions: For instance, suppose consumers with low balances became better off and paid down any debts they had. On the other side of the distribution, suppose consumers who borrowed heavily were made worse off, so that they accumulated more debt. This could leave the mean unchanged and emerge only as a change in variance. Now suppose the population is split into four groups: Some consumers with zero balances begin to borrow, while some consumers with extremely high balances cut back their lavish spending. This chain of events could result in a shift to a more platykurtic distribution, but one that does not change in terms of mean or variance.

Panel (3) of Fig. 2 depicts another type of change: a zero-inflated exponential distribution, where the mean of the exponential is linked to the proportion of zeroes to maintain a constant mean of the mixed distribution. In this example, the variance and skewness change, but the mean remains constant. In our empirical application, this would be consistent with the notion that some consumers pay off their debts, while other consumers are driven deeper into debt.

We also considered the possibility that the mean stays constant, but the mode of the distribution changes, as in Panel (4) of Fig. 2. To achieve this, we used a zero-inflated model with a two-parameter distribution—in this case, the beta distribution. In this example, the proportion of zeroes increased, while the higher mode shifted upward. This would be a type of mean-invariant response that would be easier to detect for a given sample size because it occurs with large proportional changes in probability mass, particularly between the regions containing the modes.

Below, we present a relatively simple, yet flexible, framework that can detect any one of these forms of responses, among many others. Moreover, it is a framework within which the transition from the boundary to the interior is admitted, and within which forms of Markovian dependence can be accommodated.

2.3. Baseline model

Even though the EDF is an extremely useful function in statistics, most users of statistics are more familiar with the histogram, which is obviously related to the EDF as well. Specifically, imagine dividing the support of $Q$ into $(K + 1)$ mutually exclusive and exhaustive intervals $[q_k, q_{k+1})$, where $k = 0, 1, \ldots, K$. In short, $Q$ could be zero, whereas the value of $q_{K+1}$ could be infinity.
The histogram is then the fraction of the sample whose values fall within each of those intervals. Unlike the EDF, which can deal with both continuous and discrete values of the outcome variable, when Q is continuous, the histogram is an approximation to the probability density function (PDF) $f_Q(q)$; for discrete random variables, at the appropriate granularity, the histogram is an unbiased and consistent estimator of the probability mass function (PMF). In the case of balances, which are measured to the nearest cent, this estimationtechnique depends on the particular specification. Broadly speaking, we estimate by maximizing the likelihood function, using common tolerance criteria for convergence to an optimum. A quasi-Newton method will often work; as in the case of indicator variables, however, the parameters can sometimes be concentrated out by using sample histograms on subsets of the data. For more complex models with continuous variables, the parameters are estimated by numerically optimizing the likelihood function, but these parameter searches will also be executed on subsets of the data, since the observations previously in a particular state are the only ones relevant to estimate the transition probabilities to the next state. In numerical terms, the Hessian matrix is block diagonal. We describe the calculations involved in performing the numerical optimization in Appendix A.2.

2.4. Dependence

In many applications, persistent, potentially unobserved factors have almost surely been omitted, which can induce dependence among outcomes over time. Many
different forms of dependence can exist. Perhaps the simplest generalization of the independent process assumed above is the first-order Markov process.

To explain the first-order Markov process, consider for observation \( n \) a sequence of random vectors that are generated over the time horizon \( t = 1, 2, \ldots, T \), namely, \( \{Y_{t,1}, Y_{t,2}, \ldots, Y_{t,T}\} \). In general, the distribution of, say, \( Y_{n,T} \) depends on \( \{Y_{n,T-1}, Y_{n,T-2}, \ldots, Y_{n,2}, Y_{n,1}, Y_{n,0}, \ldots\} \), that is, all past realizations of \( Y_n \) not just those observed. In short, the conditional PMF of \( Y_{n,T} \) has the following structure:

\[
  p_{Y_{n,T}}(Y_{n,T-1}; Y_{n,T-2}, \ldots Y_{n,2}, Y_{n,1}, Y_{n,0}, \ldots) = p_{Y_{n,T}}(Y_{n,T-1}; Y_{n,T-2}, \ldots Y_{n,2}, Y_{n,1}, Y_{n,0}, \ldots) 
\]

In order to make any headway in investigating the importance of dependence over time, a device must be introduced to limit the horizon of this dependence. The first-order Markov assumption does just that. Simply put, the first-order Markov assumption states that \( Y_{n,T-1} \) is a sufficient statistic for the entire history of \( Y_{n,t} \) up until period \( (T - 1) \). Put another way, the information in the vector \( \{Y_{n,T-2}, \ldots, Y_{n,2}, Y_{n,1}, Y_{n,0}, \ldots\} \), even if it were observed, can be ignored in an empirical analysis. Therefore, under a first-order Markov process,

\[
  p_{Y_{n,T}}(Y_{n,T-1}; Y_{n,T-2}, \ldots Y_{n,2}, Y_{n,1}, Y_{n,0}, \ldots) = p_{Y_{n,T}}(Y_{n,T-1}; Y_{n,T-2}, \ldots Y_{n,2}, Y_{n,1}, Y_{n,0}, \ldots) 
\]

To illustrate how this process works, consider a simple example: Suppose \( K \) is one; that is, only two states exist—zero and one. Such a specification is often referred to as a mover-stayer model in statistics literature. In the mover-stayer model, the dynamics of movements from state to state are governed by a transition matrix, which is typically written as follows:

\[
  P = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} \tag{6}
\]

where the \( p_{ij} \) elements of \( P \) in Eq. (6) are parameters that satisfy the following restrictions: \( 0 \leq p_{ii} \leq 1 \) and \( p_{00} + p_{11} = 1 \) for \( i = 0, 1 \). Element \( p_{ij} \) of the transition matrix \( P \) is the probability of transitioning to state \( j \) given that the current state is \( i \). Under independence, \( p_{00} = p_{10} \) as well as \( p_{01} = p_{11} \) (by adding up), so the mover-stayer model nests a model in which independence is assumed.

This sort of discrete-state, first-order Markov process can be used to approximate the following linear, first-order autoregressive [AR(1)] model of a continuous, positive random variable:

\[
  \log(Q_t) - \mu = \rho [\log(q_{t-1}) - \mu] + \sigma Z_t \\
  t = \cdots, -1, 0, 1, 2, \ldots. \tag{7}
\]

where \( Z_t \) denotes independent and identically distributed normal random variables having mean zero and variance one. The parameter \( \mu \) in Eq. (7) is a location parameter of sorts, whereas the parameter \( \sigma \) is a scaling parameter, and the parameter \( \rho \) controls the amount of linear dependence. In order for the process to be stationary (that is, for it not to explode as time proceeds), the absolute value of \( \rho \) must be less than one. Under these assumptions, the unconditional distribution of \( Q_t \) then belongs to the lognormal family, having mean \( \mu \) and variance \([\sigma^2/(1 - \rho^2)]\), so its PDF is

\[
  f_Q(q | \mu, \sigma, \rho) = \frac{\sqrt{(1 - \rho^2)}}{q \alpha \sqrt{2\pi}} \exp \left( -\frac{(1 - \rho^2) \log(q) - \mu^2}{2\sigma^2} \right) \tag{8}.
\]

The presence of point masses (for instance, at zero), however, makes it difficult to implement Gaussian autoregressive processes, such as those defined by Eqs. (7) and (8), which is why we chose to discretize the problem.

In keeping with the notation of Eq. (6), for observation \( n \), we specify \( P_n \), in general a \((K + 1) \times (K + 1)\) matrix having the following form:

\[
  P_n = \begin{bmatrix} (1 - \sum_{j=1}^{K} p_{0,j}^n) & h_{0,1}^n & h_{0,2}^n & \ldots & h_{0,K}^n \\ (1 - \sum_{j=1}^{K} p_{1,j}^n) & h_{1,1}^n & h_{1,2}^n & \ldots & h_{1,K}^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (1 - \sum_{j=1}^{K} p_{K,j}^n) & h_{K,1}^n & h_{K,2}^n & \ldots & h_{K,K}^n \end{bmatrix} \tag{9}.
\]

2.5. Introducing feature variables

When a feature variable \( d \) is introduced, the loss function becomes

\[
  \begin{align*}
  f_n(y, \delta_y, d_n) &= y_{n,0} \log \left( 1 - \sum_{j=1}^{K} h_{n,j} \right) \\
  &\quad - \sum_{k=1}^{K} y_{n,k} \left( \gamma_k + \delta_y d_n - \log \left( 1 - \sum_{j=1}^{K} h_{n,j} \right) \right) \\
  &= -\sum_{j=1}^{K} y_{n,j} (\gamma_j + \delta_y d_n) \\
  &\quad + \log \left( 1 + \sum_{j=1}^{K} \exp(\gamma_j + \delta_y d_n) \right).
  \end{align*}
\]

where \( \delta \) denotes parameters collected in the \((K \times 1)\) vector \([\delta_1, \delta_2, \ldots, \delta_K]^T\). In this case, the extra subscript \( n \) followed by a comma is added to the definition of \( h \) to alert the reader to the fact that this function now depends on the value of \( d_n \), that is,

\[
  h_{n,k} = \frac{\exp(\gamma_k + \delta_k d_n)}{1 + \sum_{j=1}^{K} \exp(\gamma_j + \delta_j d_n)}.
\]

The loss function for the sample can be formed by adding up the \( f_n(y, \delta_y, d_n) \) terms. Obviously, including additional feature variables is straightforward, but computationally tedious.

2.6. Dealing with seasonality

One way to deal with seasonality is to assume that the transition probabilities change periodically, such as month to month. This is as simple as introducing a set
of binary feature variables in the above framework. In short, introduce the single-index function \( x_{n,t} \beta_k \), where \( x_t \) is defined to be a \((1 \times 12)\) vector that has zeros everywhere except for a one in the relevant month for period \( t \) and \( \beta_k = [\beta_{k,1}, \beta_{k,2}, \ldots, \beta_{k,12}]^\top \). Whence, under the logit link function, \( \log(h_{n,t}/h_{0,n,t}) = x_{n,t} \beta_k \).

If seasonal indicators are the only covariates, as in Ho et al. (2021), then the maximum likelihood estimator is equivalent to estimating a series of separate transition matrices, as in Eq. (9), except that the sample for each matrix is restricted to the observations corresponding to each season. Further, the rows of these transition matrices are estimated by calculating a series of histograms, conditioning on the category in the previous observation.

### 2.7 Measuring the Effect

How do we decompose the observed distribution of the variable of interest? That is, how do we distinguish between dependence in the data and the effect of the intervention on this variable? Well, we estimate the model up to the date before the intervention. Then, based on those estimates, we calculate the one-, two-, three-, ..., \( \ell \)-month-ahead forecast distributions according to the Markov model:

\[
\mathbf{\hat{p}}_{T+\ell} = \hat{p}_{T+\ell} \mathbf{\hat{p}}_T \quad \text{for} \quad \ell = 1, 2, 3, \ldots, T
\]

We then compare the distributions generated by Eq. (10) to what actually obtained: in our notation, \( \mathbf{\hat{p}}_{T+1}, \mathbf{\hat{p}}_{T+2}, \mathbf{\hat{p}}_{T+3}, \text{and} \mathbf{\hat{p}}_{T+4} \). As a summary measure of the difference, we consider

\[
100 \times \log \left( \frac{\mathbf{\hat{p}}_{T+\ell}}{\mathbf{\hat{p}}_{T+\ell}} \right),
\]

(11)

which is the percentage difference between what actually obtained relative to what is predicted to obtain in cell \( k = 0, 1, \ldots, K \). This object can be plotted on the ordinate versus the various cells on the abscissa to provide the reader with a visual description of how the variables changed relative to what would have been predicted before the intervention. The following statistic,

\[
N \sum_{k=0}^{K} \hat{p}_{T+\ell} \log \left( \frac{\hat{p}_{k,T+\ell}}{\hat{p}_{k,T+\ell}} \right)
\]

facilitates formal statistical testing of the difference in the two distributions.

This statistic was first proposed by Kullback and Leibler (1951) and is commonly referred to as the Kullback–Leibler divergence criterion. It is based on a concept of information introduced by Shannon (1948). Belov and Leibler (2011) demonstrated that, for a pair of continuous distributions, a version of this statistic has a limiting (asymptotic) \( \chi^2 \) distribution with one degree of freedom.

Parkash and Mukesh (2013) investigated the case of discrete distributions, which corresponds to our application, determining that the statistic has a limiting \( \chi^2 \) distribution, but with \( K \) degrees of freedom—that is, the number of categories minus one. Song (2002) demonstrated that the Kullback–Leibler divergence statistic is asymptotically equivalent to the likelihood-ratio statistic for detecting a difference between distributions. In our application of this statistic, however, many parameters must be estimates, so it remains an empirical question as to the sample size that achieves the asymptotic distribution. We document this in the following section.

### 3. Simulation evidence

In this section, we provide two types of simulation evidence concerning the effectiveness of our test statistic. First, we verify the limiting distribution of our statistic under the null hypothesis of a reasonably large sample size. Second, we conduct a power analysis to determine the magnitude of a deviation from the null hypothesis that can be detected with this statistic, as a function of the size of the sample.

#### 3.1. Empirical distribution

We implemented a parametric bootstrap simulation of our forecast statistic. Under the null hypothesis of no change in the transition matrix, first-order asymptotic distribution theory predicts that the Kullback–Leibler divergence statistic follows a \( \chi^2 \) distribution, when comparing the forecasted probabilities to those observed under the data generating process. We verify this claim here.

We begin our investigation by defining a transition matrix \( \mathbf{P} \) with zero-inflated lognormal distributions in the columns. This matrix was chosen to approximate the transition matrix that was actually obtained from the observations in the empirical example in the next section. We consider

For each period, from \( t = 1 \) to \( t = 36 = T \), we generated the state vector of counts for the next period by looping down the categories and taking draws from the multinomial distribution with probabilities from the ergodic distribution defined by the transition matrix. This vector has an integer count of \( N = 50000 \) individuals across the \((K + 1 = 11)\) categories, including the zero category.

For each period, from \( t = 1 \) to \( t = 36 = T \), we generated the state vector of counts for the next period by looping down the categories and taking draws from the multinomial distribution with probabilities defined by the corresponding column of the transition matrix. These are draws of vectors of integers with sums that add up to the bin count from the last period. After the realization of period \( T \), the calculation of \( \mathbf{P} \) was completed, denoted \( \mathbf{P}_B \) in the bootstrap replications.

For \( L = 12 \) periods after period \( T \), the data generating process continued to produce the next vectors recursively,

---

3 We did not store the values for individuals in a data set size \((N \times T)\). Instead, we added these counts to the cumulative calculation of the maximum-likelihood estimator for the matrix \( \mathbf{P} \), the estimate of \( \mathbf{P} \). This procedure led to the same value that would have obtained were we to have calculated using the large \((N \times T)\) data set, without storing it, and it also preserves the dependence in the data in a way that follows the data generating process under the null hypothesis.
with the true transition matrix $P$, as before. This generated realized counts of individuals in the categories at each period $(T + 1)$, . . . , $(T + L)$. We also computed forecasts by left-multiplying the estimated transition matrix $\hat{P}_T$ with $\hat{p}_T$, the last vector of relative frequencies from the in-sample period. For each realization of the 9999 replications, we calculate the Kullback–Leibler distance between $\hat{p}_{T+m}$ and $\hat{P}_T \times p_T$. As shown in Fig. 3, this distance statistic follows the $\chi^2$ distribution with $K = 10$ degrees of freedom.

### 3.2. Power analysis

We defined local alternatives in a particular fashion to approximate the estimates from the data in our empirical application to credit-card balances. We increased the probability mass in the zero category by a factor of $(1 + c/\sqrt{N})$ by moving probability mass from the other categories. We observe 99% rejection probability under alternatives with $c = 10$, and 55% rejection probability with $c = 5$. This indicates that a sample size of 10,000 individuals can detect deviations of 10% of the probability mass in the zero category with power near one—that is, a change from 15% to 13.5% in the zero category, for instance. A sample size of 1,000, 000 can detect a reallocation of 1% probability mass. A pair of distributions from the null hypothesis ($c = 0$) and the alternative hypothesis ($c = 5$) are depicted in Fig. 4. The dashed vertical line represents the 5% critical value of the $\chi^2$ distribution with $K = 10$ degrees of freedom, with 55% of the probability mass of the distribution of the alternative in the rejection region.

We then considered local alternatives of this form with a set of local alternatives to calculate power curves. That is, we increased the probability mass in the zero category to $(1 + c/\sqrt{N})$ times that probability mass and shifted this proportionately from the other categories, with $c = 0$ (the null hypothesis) and $c = 2.5, 5.0, 7.5,$ and $10.0$ (the alternative hypotheses). The power curve appeared invariant across sample sizes, in terms of the number of individuals each period. In particular, the power curve for a sample size of 20,000 was hardly distinguishable from that from a sample size of 50,000. This confirmed that the test has the power to detect deviations in probability mass inversely proportional to the square root of the sample size.

Although this example illustrates the effectiveness of our method at detecting a change in the distribution, this form of response, with a changing mean, can also be detected with the traditional form of intervention analysis. We also analyzed the mean-invariant intervention responses shown in Fig. 2. The intervention pattern in Panel (3) is closely related to the example above, in that it employs a sequence of zero-inflated exponential distributions. In this example, in contrast, the exponential parameter $\lambda$ is linked to the proportion of zeros $p_{\text{zero}}$ to maintain a fixed mean $\mu$, by setting the exponential parameter to $\lambda = (1 - p_{\text{zero}})/\mu$. We investigated several alternatives in which $p_{\text{zero}} = 0.10(1 + c/\sqrt{N})$ and $\mu = 2.25$, so that the mean of the exponential component of the mixture is set to 0.40 under the null distribution. With a modest sample size of 10,000 in the cross-section, the test achieves 99.3% power under the alternative with $c = 25$, which corresponds to a weight of 0.125 on zero and a parameter on the exponential distribution of 0.38. This sort of change implies a minor change in the variance or skewness of the distribution: the standard deviation rises roughly 2%, from 2.53 to 2.58, and the skewness rises less than 1%, from 2.052 to 2.056. In this example, the changing weight on zero drives the power of the test, although it requires a larger change: 25% more probability mass on zero, rather than 10% in the
zero-inflated lognormal case above, with a changing mean response. Overall, this indicates that our framework can detect such changes even when the mean is unchanged, in which case the traditional form of intervention analysis would be ineffective.

Panel (1) of Fig. 2 illustrates a simple example of a case in which the variance changes in response to an intervention but the mean remains constant. To examine this case, we generated alternatives from normal distributions, each with mean $\mu = 5.0$ and standard deviation $\sigma = 2.5(1 + c/\sqrt{N})$. As in the case in Panel (3) described above, this response to an intervention would not be detected under the traditional framework for intervention analysis. Within our flexible framework, however, the effect can be detected: we found a power of 99.8% for the alternative with $c = 8$, which corresponds to a standard deviation of 2.7, an 8% increase in variance.

For the third case of our power analysis, we considered the model in Panel (2) of Fig. 2, in which the first three moments remain constant and the kurtosis changes. We used a mixture of two normal distributions, with weight $p_{\text{mix}}(c, N) = 0.5(1 + c/\sqrt{N})$ on the first normal distribution, with mean $\mu = 5.0$ and standard deviation $\sigma_1(c, N) = 2.5(1-c/\sqrt{N})$. The remaining probability mass was placed on the second normal distribution, with the same mean, except with standard deviation $\sigma_2(0, N) = 2.5$ under the null distribution. This results in a normal distribution under the null, and an excess kurtosis of zero. For the alternative distribution, we set the variance of the second distribution in a way that imposed constant variance of the mixture distribution over the alternatives, using the formula

$$\sigma_2^2(c, N) = \frac{p_{\text{null}}(0, N)\sigma_1^2(0, N) - p_{\text{null}}(c, N)\sigma_1^2(c, N) + (1 - p_{\text{null}}(0, N))\sigma_2^2(0, N)}{1 - p_{\text{null}}(c, N)}$$

with the arguments $(0, N)$ or $(c, N)$ indicating the parameters of the mixture distribution under the null and alternative, respectively. With a sample size of 10,000, the test had 99.5% power to detect alternatives with $c = 22$, which corresponds to $p_{\text{null}}(c, N) = 0.61$, $\sigma_1^2(c, N) = 1.95$, and $\sigma_2^2(c, N) = 3.17$. The resulting mixture for this alternative had an excess kurtosis of 0.72. Thus, even with the first three moments matching, our framework can detect small changes in the fourth moment with a modest sample size of 10,000 in the cross-section.

Finally, we conducted a simulation to determine the performance of our model to detect a response of the form of Panel (4), in which the mode of the distribution changed and the mean remained the same. This sequence of alternatives also featured a zero-inflated distribution with $p_{\text{zero}} = 0.10(1 + c/\sqrt{N})$ and the remaining probability mass on a beta distribution with parameters $\alpha$ and $\beta$. We fixed $\alpha = 3.0$ under the null and alternatives, and varied $\beta$ to impose a constant mean of $\mu = 3.0/7.0 \approx 0.43$, using the formula $\beta = \alpha(1 - p_{\text{zero}} - \mu)/\mu$. We found a rejection probability of 99.6% with a sample size of 10,000 and parameter $c = 20$, which corresponds to 0.12 probability mass on zero. Although the mean was held the same throughout, the distributions can have a large degree of divergence with this sort of alternative, since the distributions differ greatly in terms of the regions of the support with both high and low probability mass.

In each of these simulations, we took draws from the null and alternative distributions to analyze the performance of our statistic. In our empirical application, however, we also estimated the transition matrices that determine the path to the null distribution. The question remains whether the measurement of the response is still reliable when the forecast distribution is calculated from estimated transition matrices. To study this question, we conducted another set of simulations by

---

**Fig. 4.** Bootstrap distribution under the null and alternative hypotheses.
drawing the transition matrix \( \hat{P}_b \) from the asymptotic distribution of the maximum-likelihood estimators. For the log-odds of probabilities in each of the columns of \( P \), we obtained the estimate and Hessian matrix. In each bootstrap replication, we drew from the multivariate normal distribution with mean at the estimates and covariance matrix the negative of the inverse of the Hessian matrix. These log-odds were then transformed into probabilities and inserted as columns of the bootstrap transition matrix \( \hat{P}_b \). We simulated the model with the zero-inflated lognormal distribution with a changing mean to illustrate a case that more closely matches that found in the credit-card data. For sample sizes in the tens of thousands, the results were not perceptibly different from those above, as would be expected from the asymptotic distribution theory for the maximum-likelihood estimator of the parameters of the multinomial distribution.

4. Empirical application

To apply our framework for intervention analysis, we investigated the credit usage of Canadian consumers during the COVID-19 pandemic. Here, we first describe the data set and then demonstrate that our method is useful and appropriate for analyzing these data.

4.1. Consumer credit data

Through a contract with the Bank of Canada, we acquired access to monthly anonymized data from TransUnion®, one of the two credit bureaus in Canada. The data set contains account-level balances on credit cards from January 2017 to September 2020. We aggregated account-level balances to the individual level, in order to study individuals’ decisions and to avoid complications from individuals’ choice-of-card decisions (Felt et al., 2021). The data set also contains consumer credit scores as well as the encrypted postal code of their primary residential addresses. Following Bhutta and Keys (2016), we defined consumers to be homeowners if they ever had a mortgage or a home-equity line of credit while living at their current postal code. For the analysis in this paper, a random 1% sample of individuals was constructed from the entire data set, based on the power analysis presented above. There are 290,436 credit-card holders in this sample with a total count of 10,528,372 observed monthly balances, in which there are 124,229 homeowners with 4,803,515 monthly observations.

Once a sample data set of individual-level balances was obtained, we assigned the continuous balance variables to discrete balance categories. Instead of using evenly spaced bins, we organized consumer balances into intervals of increasing width to account for the lengthy tail of the distribution. The balance categories were sorted into intervals of width $250$ up to $1500$, intervals of $500$ up to $6000$, $1000$ up to $10,000$, $2000$ up to $20,000$, $5000$ up to $30,000$, one $30,000$–$40,000$ category, and a category of $40,000$ and above in the tail. This grid of $K + 1 = 29$ categories is representative of the variety of credit lines typically offered to different types of consumers in different risk categories.

We defined the histogram bins according to risk categories for this particular application, because this aligns with the most commonly occurring credit lines. We divided consumers into sets of customers that financial institutions treat similarly. In other circumstances, one would start with common techniques for defining histogram bins. These boundaries would be further partitioned with knowledge of the relevant boundaries in the data, such as the bound of zero in our application. Finally, any atoms in the data should be placed in a separate bin, separating the adjacent observations into separate bins. With these factors taken into account, we also recommend that some bins be combined if separating them would result in a sparse transition matrix. For this reason, the bins in our application increase in width for the categories of customers with higher balances.

Our aim is to characterize changes in credit usage during the pandemic. Fig. 5 conveys this information through a series of histograms. The columns furthest back represent the category with the most consumers: those with zero balances. The next most common category comprises consumers with balances between $0$ and $250$. Clearly, an increase in membership in the low-balance categories in the first six months of the pandemic (March to September 2020) occurred. In the spring of 2020, more consumers had a balance of less than $250$ than in any other period in the sample. Over the sample period from January 2017 to June 2020, the average monthly balance was $4064$, which dropped to nearly $3205$ by June 2020. This represents a decline of 20% following a sustained pre-pandemic year-over-year growth of 2.7%. In addition to the COVID-19 pandemic effect, a clear seasonal pattern also exists in the low-balance categories: there are higher average balances in the fourth quarter of every year (the holiday season), and the average balances decrease in the first quarter every year. Our statistical framework aims to quantify these changes during the pandemic, while accounting for the counterfactual path over time.

At the individual level, four main features of the data set are accommodated nicely by our empirical framework. First, the distribution in a given month has a nonstandard shape, which is not characterized adequately by a small number of moments. Most notably, the highest bars at zero in Fig. 5 show that a substantial fraction (perhaps as much as 15%) of individuals in the sample have zero balances, precisely on the boundary. Second, the series exhibit strong seasonality: the distribution is expected to shift leftward during the spring months, which is precisely when we aim to measure the effect of the pandemic. Third, considerable dependence exists over time; for example, if an account had a zero balance at the end of last month, then it is very likely (in general, over 50% of the time) to have a zero balance at the end of this month.\footnote{Credit scores are well known to reflect the creditworthiness of consumers. Credit bureaus construct their own credit scores, so there may be different credit ratings across them.}

\footnote{We estimated an autoregressive model for the stacked panel of balances, conditioning on past balances for each individual, when observed. We found that a higher-order autoregressive model fit the data better than a first-order model.}

\footnote{We estimated an autoregressive model for the stacked panel of balances, conditioning on past balances for each individual, when observed. We found that a higher-order autoregressive model fit the data better than a first-order model.}
The framework that we have proposed has the flexibility to accommodate all of these characteristics of credit-card balances.

4.2. Modeling procedure

With the data set of credit-card balances, we show by example how our framework is used to detect changes in the distribution in response to the intervention. A researcher should follow this modeling procedure:

1. Specify the explanatory variables to predict transition probabilities.
2. Estimate the transition probabilities by calculating the parameter values that maximize the value of the likelihood function.
3. Use the estimated parameter values to calculate the predicted transitions of the mass of consumers in each category, beginning with the initial proportions.
4. Calculate the actual proportions of consumers in each category throughout the sample.
5. Calculate the KLD statistic for the comparison of the predicted and actual proportion vectors for each time period within the sample.
6. The next step depends on whether the KLD statistic detects anomalies in the proportion of consumers in each category.

   (a) If the KLD statistic detects differences in proportions, then modify the list of explanatory variables. Normally, this will involve adding new explanatory variables that have variation related to the anomalies found in the tests with the KLD statistic.

   (b) If the KLD statistic does not detect anomalies, then the model is adequate to detect changes out of sample.

Although the model that results from Step 6(b) is adequate to detect a response to the intervention, the researcher may wish to employ a richer model, with additional explanatory variables. This would provide the added benefit of more precise predictions of the transition probabilities, thus allowing the researcher to detect changes of a smaller magnitude or to detect similar changes with a smaller sample size in the cross-section.

Furthermore, the addition of explanatory variables would be appropriate for testing hypotheses involving differences in responses by subsets of the population or in a way that is related to other explanatory variables. In fact, a measurement of the differences in response between subsamples without including related variables would confound the measurement of the law of motion with any differences in the distribution across those subsamples. This would introduce inconsistency into the estimation of the response from misMeasuring the counterfactual post-intervention law of motion. This suggests that the researcher should consider a model that goes beyond simply being adequate to detect changes without false positives within the pre-intervention sample. The inclusion of additional variables should be guided by the question at hand. In our empirical application, the pandemic effect must be measured against a benchmark that includes the differences in the distribution of credit-card balances across credit-score categories.

4.3. Models

We estimated four models that seek to characterize the law of motion of credit-card balances before the pandemic. These were then used to construct a counterfactual forecast with which to compare consumer behavior during the pandemic. The approach we utilized is similar to that for “excess mortality” in the demography literature; see Statistics Canada (2020).

Our simplest model, labeled “Histograms”, serves as a benchmark for the others. We compared the histograms month-by-month from those during 2020 to those estimated over the sample period from January 2017 to January 2020. This model takes into account the seasonality over the months throughout the calendar year. It is also simple to estimate: it requires the estimation of only $12K = 336$ free parameters in 12 histograms, which are estimated very precisely given our sample size. The histograms, however, do not take into account the dependence noted in Fig. 5.

The next model, labeled “Fixed”, accounts for dependence between balances in consecutive months. It assumes a first-order Markov process for the transitions between months. The parameters in this model are the transition probabilities that govern the movement of consumers between balance categories each month, a total of $K(K+1) = 28 \times 29 = 812$ parameters. A single transition matrix $\hat{P}_t = \hat{P}$ is estimated from pairs of observations over the entire sample. Using the transition matrix, we...
constructed counterfactual forecasts for the distribution of credit-card balances. We initialized the forecasts with the proportion of consumers in each balance category that is observed in January 2020, reflecting activity recorded during the month of December 2019.\(^6\) We then calculated a series of forecasts by left-multiplying the vector with the transition matrices for each month. The forecast for the next month is calculated in the same way, using the forecast from the period before.

We considered a third model, labeled “Monthly”, by augmenting the “Fixed” model with a simple set of monthly indicators.\(^7\) A separate transition matrix \(\hat{P}_t\) is estimated each month, using monthly indicators as covariates. It combines the characteristics of the “Histograms” and the model with the “Fixed” transition matrix. Since these monthly indicators are mutually exclusive and exhaustive, each transition matrix is estimated using only the observations that correspond to the particular pairs of months. This alleviated the computational burden of estimating many more parameters. The model with monthly transition matrices has 12 times as many as that in the “Fixed” model—some 97,444 parameters.

In the fourth model, labeled “Covariates”, we used two variables to serve as covariates to estimate transition probabilities in addition to the seasonal indicators. We constructed a categorical variable on creditworthiness by dividing credit scores into three categories. Consumers with a credit score below 700 are placed in the “Low” credit-score category; those with credit scores between 700 and 839 are considered “Medium”; and those with credit scores of 840 or above are allocated to the “High” category.\(^8\) The other variable included is a home-ownership indicator. Including an interaction term to separate homeowners within each credit-score category, there are 10 times as many parameters, adding up to a total of 97,440 parameters.

### 4.4. Empirical results

We evaluated these models by performing a series of tests to detect differences between the observed distributions and the forecasted ones. In this set of tests, we compared the benchmark with an \(\ell\)-step-ahead forecast. In Table 1, we collect the results of this series of comparisons for the four models. The column labeled “Histograms” shows the comparison to sample histograms; “Fixed” denotes the comparison to the forecast with a fixed transition matrix; the “Monthly” column refers to the forecasts using separate transition matrices each month; and the “Covariates” column refers to the same statistic using separate transition matrices, each estimated with credit-score and home-ownership categories. The statistic is the Kullback–Leibler divergence statistic comparing the \(\ell\)-step-ahead forecasted distribution with the observed sample distribution. The \(p\)-value columns show the probability of observing a statistic more extreme from the \(\chi^2\) distribution with \(K = 28\) degrees of freedom. For reference, these divergence statistics should be compared to the critical values of 41.34, 48.28, and 56.89, corresponding to the 5\%, 1\%, and 10 basis-point levels of significance.

All models detected a statistically significant shift in the distributions that occurred in April 2020. This was followed by a larger deviation in May 2020 that subsided over the subsequent months, although a large difference persisted thereafter. The models differed in their characterization of the period before the pandemic. For the model with monthly transitions, the statistics up to March did not wander far from those expected under the null hypothesis of no change, especially considering that they were calculated with tens of millions of observations. The fit was even closer for the forecasts from the model with credit-score and home-ownership covariates.

As a measure of goodness of fit, we also compared the deviations from predictions from one-step-ahead forecasts. We document in Appendix A.3 that the model with a fixed transition matrix is not suited to the data, as it erroneously detects changes in the months before the pandemic, even when conditioning on the distribution in the previous month. This problem was largely absent from the one-step-ahead forecasts with the model with monthly transition matrices: the distance statistics were all within a reasonable distance of the critical values at conventional levels of significance. This suggests that the model with monthly transition matrices is as good as it needs to be to detect a difference in distributions, and

| \(\ell\)-step-ahead forecasts from alternative models. |
|-----------------------------------------------|
| **Histogram** | **p-value** | **Fixed** | **p-value** | **Monthly** | **p-value** | **Covariate** | **p-value** |
| January 2020 | 357.1 | 0.0000 | 508.4 | 0.0000 | 60.3 | 0.0004 | 56.2 | 0.0012 |
| February 2020 | 329.8 | 0.0000 | 1,142.7 | 0.0000 | 68.4 | 0.0000 | 55.9 | 0.0013 |
| March 2020 | 261.0 | 0.0000 | 2,078.4 | 0.0000 | 234.6 | 0.0000 | 134.5 | 0.0000 |
| April 2020 | 3,981.9 | 0.0000 | 8,124.9 | 0.0000 | 5,326.5 | 0.0000 | 4,168.5 | 0.0000 |
| May 2020 | 6,718.1 | 0.0000 | 10,174.7 | 0.0000 | 8,693.4 | 0.0000 | 6,577.4 | 0.0000 |
| June 2020 | 4,416.8 | 0.0000 | 6,529.4 | 0.0000 | 6,453.2 | 0.0000 | 4,405.6 | 0.0000 |
| July 2020 | 2,679.1 | 0.0000 | 5,138.6 | 0.0000 | 4,494.3 | 0.0000 | 2,741.5 | 0.0000 |
| August 2020 | 2,358.3 | 0.0000 | 4,115.0 | 0.0000 | 3,874.1 | 0.0000 | 2,566.6 | 0.0000 |
| September 2020 | 2,219.5 | 0.0000 | 3,809.5 | 0.0000 | 3,919.0 | 0.0000 | 2,705.0 | 0.0000 |

\(^6\) Although this may be early, it avoided initializing our forecasts while the pandemic was active in Canada. Also, it allowed for a few months of out-of-sample tests of the modeling framework in the months before March 2020.

\(^7\) The “Monthly” model is estimated in Ho et al. (2021).

\(^8\) Credit bureaus usually sort consumers into categories such as subprime, prime, or above-prime. Consumers in lower-quality categories, e.g. subprime, are typically individuals with adverse credit histories. Since credit delinquency rates are low in Canada, the sample sizes for lower-quality categories are small enough that those estimates would be much more variable than the others. The “High”, “Medium”, and “Low” definitions of categories allocate consumers into similar-sized groups that better reflect the variability in consumer creditworthiness.
that no further complexity is warranted to answer the question of whether a change has occurred.

The decision over which model is appropriate, if any, can be settled by analyzing the performance of the model within the pre-intervention period. We conducted this comparison for two of the models above and present the results in Appendix A.4. The model with a fixed transition matrix appears to detect changes that follow a pattern through the seasons. The tendency for this model to detect false positives will degrade the accuracy of the measurement of the intervention effect. The richer models with monthly seasonality do not appear to raise false positives through the pre-pandemic period. The fact that these models produce forecasts that do not detect deviations in the pre-pandemic period shows that the models provide a reliable benchmark for analyzing the effects of the pandemic. The model with covariates, however, is the appropriate choice, since we aim to detect effects specific to subsets of consumers with different home-ownership status and risk profiles.

This is evident in that the histograms and the fixed, non-seasonal model erroneously detect differences in credit-card balances in the early months. On closer inspection of the fixed model, we found that this behavior is also observed each year in the sample, since the fixed transition model does not account for the annual pattern of declining balances after the holiday season up to March and April, when many consumers receive tax returns. The model with histograms estimated over the sample period does not account for the growth of credit usage year-over-year.

We then analyzed the path of credit-card balances throughout the pandemic, using home-ownership status and credit scores as explanatory variables, with the predictions from the “Covariates” model. This provided deeper insight to the findings of Ho et al. (2021) that consumers are reducing their overall level of borrowing from credit cards and home-equity lines of credit (HELOCs). It remains an empirical question whether the pattern differs for consumers without access to home equity and/or other forms of credit.

The changes in the distribution of credit usage for homeowners with medium credit scores are depicted in Fig. 6.9 The results in May and August 2020 form bookends on the transition throughout the first wave of the pandemic, since most Canadian provinces implemented some sort of economic lockdown in April and gradually reopened in July.10 In May, the proportion of these consumers with balances from $0 to $500 (the lowest three categories) increased by nearly 30%. There was a larger fraction of consumers with balances less than $2000, while the largest decreases were those with balances ranging from $8000 to $10,000. By August, the effect was weakened but still significant. Generally, smaller changes are observed in balance categories smaller than $14,000, while the greatest decreases occurred for balance categories greater than $16,000. The findings are similar for non-homeowners, except that the changes take place in lower-balance categories. Overall, the changes in credit usage largely mimic those of the full population, similar to the results in Ho et al. (2021). Our results provide evidence that the decrease in credit-card balances can be attributed to reduced spending during the lockdown.11 Although the gradual reopening of the economy may have attenuated the effect, the big spending that commonly occurs in the summer months was still missing during the pandemic.

We further investigated the changes for consumers with the tightest credit constraints—that is, those with low credit scores and no home equity to use as collateral for a loan. Fig. 7 depicts the proportional changes in balance categories for these consumers in May and August 2020. The scale is magnified because the proportional changes are much larger: nearly 40% more consumers had positive balances below $250. The change in the proportion of consumers with zero balances is even larger. The statistic for August 2020 compares a forecast of 1187.8 consumers with the 2367 consumers who are actually observed to have balances of zero, although the statistic shows a difference of log(2367/1187.8) × 100 = 68.95%. Overall, the May and August distributions for these consumers are more similar than that of the middle-credit-worthiness homeowners in Fig. 6. This suggests a stronger and sustained reduction in credit usage by consumers with the tightest credit constraints. This is an interesting finding because, first, consumers with tight credit constraints did not become more indebted, despite the unemployment rate surging from 7.9% in March to 13.7% in May 2020. Second, these consumers did not change their credit usage when unemployment subsided to 10.2% in August as the economy reopened.

At the other extreme, homeowners with high credit scores show a different pattern of changes, as seen in Fig. 8. The proportion of these consumers in the very-low-balance categories does increase, just as that of the consumers in the other groups, but to a lesser degree. In May, there was a large decrease in these homeowners with balances from $3000 to $14,000. The shift in these intermediate balance categories is much more pronounced than in the rest of the population, with changes of up to 50%. A subset of those creditworthy homeowners is increasing their use of credit during the pandemic. We recorded 30% more consumers in this group with balances from $25,000 to $30,000 and even more consumers with

---

9 The sample size is large enough that the visible changes in these figures are statistically significant, since the standard error bounds, were we to include them, would be approximately the width of a pixel. Our sample-size decision was inspired by the power analysis in Section 3.2 to ensure that this was the case. If there were constraints on the size of the sample, for example, due to data availability, then it would be appropriate to show the variability of the estimated changes with standard error bounds.

10 We present other figures for the intervening months in Appendix A.5.

11 Unfortunately, we do not have sufficient information to separate spending versus revolving balances in credit-card accounts. However, existing research has shown that credit-card balances in Canada are largely attributed to consumer spending. Bilyk and Peterson (2015) showed that the growth in credit-card balances from 2000 to 2015 reflected increased spending rather than increased short-term borrowing. Henry et al. (2018) used survey data in 2017 to show that about 30% of credit-card holders have positive revolving balances.
balances above $30,000. Although we observed large percentage changes, very few consumers indeed experienced these changes: most homeowners with high credit scores do not hold high credit-card balances.\footnote{In August 2020, only 76 consumers had balances from $25,000 to $30,000, compared to the forecast of 57.4; 69 consumers had balances from $30,000 to $40,000, compared to the forecast of 37.3; and 36 consumers had balances greater than $40,000, compared to the forecast of 20.1.} Also note that in August 2020, the effect of COVID-19 on the credit usage of these consumers was more moderate when compared to other groups. These results suggest that affluent homeowners reduced their use of credit during the peak periods in the first wave of COVID-19, possibly due to limited spending opportunities in an economic lockdown. This group of creditworthy consumers had the strongest responses in credit usage during the economic recovery.

5. Summary and conclusions

We developed an approach to intervention analysis that avoids modeling the moments of the population. Our approach features a discretized distribution that is modeled to follow a Markov process, with transition probabilities that may depend on covariates. This tool proved
very useful in modeling transitions with seasonality and conditioning on covariates. The result is a flexible framework for conducting intervention analysis, with little risk of bias from misspecification. The flexibility of this approach creates a reliable benchmark for measuring the effects of an intervention and is resilient to the effects of what would otherwise be inconvenient features of the data. This approach is useful when modeling the behavior of a large population of individuals when conventional methods of intervention analysis may fail.

Our procedure can be extended to many other applications. For instance, this modeling approach could be used to detect responses to policy changes. Leverage ratios in mortgage lending, such as the loan-to-value ratio, are known to have unique clustered distributions due to lending regulations (Bilyk et al., 2017). Changes in mortgage stress tests, such as lowering the debt–service ratio, may induce nontrivial distributional effects at the tail of the distribution that may not be noticed in lower moments (Bilyk & teNyenhuis, 2018). Changes in the distribution of loan balances, such as an increasing portion of loans with a very high balance or a deterioration in loan quality, may have important tail-risk implications on a financial institute’s loss-given-default modeling (for example, Yao et al., 2015), consequently affecting its capital buffer requirement.

In addition, this modeling approach can be used to detect changes in a population to guide a dynamic modeling strategy: perhaps a change will trigger a model rebuild or, at least, a further evaluation of the performance of an existing credit risk model. This framework could also be applied to a number of risk models designed as inputs for particular investment decisions. For example, it could be used in credit approval decisions to determine whether there is a change in the composition of customers. For credit-line assignment, an institution would want to know whether consumers need higher or lower lines of credit. In the marketing of financial products, firms can form their competitive strategy by understanding trends in consumer characteristics. In any of these modeling situations, there may be an unknown change in the competitive landscape, resulting in a different population of customers that are evaluated by the model.

One should, however, be careful to apply this method only to situations in which the pre-intervention period is unaffected by the intervention before it appears. With incorrect timing of this intervention response, this approach would not identify the counterfactual path, because the estimate would be confounded with changes in anticipation of the intervention, thus leading to a biased estimate of the response. For our application to the pandemic, the initial changes took place over a timeline that was short relative to the time between observations.

We used this framework to provide a plausible forecast of the distribution of credit-card balances, under the counterfactual state in which the COVID-19 pandemic did not take place. We found a significant downward shift in consumer credit usage in Canada, slashing billions of dollars off credit-card balances across the country, to an unprecedented level. Moreover, this result applies to both homeowners and non-homeowners alike, while the change in distribution represents a greater reduction of debt for homeowners. Further, we found an overall reduction in credit usage for consumers with credit scores at all levels—a finding in stark contrast to the experience over the years leading up to the pandemic. Allen et al. (2021) provided a structural econometric approach to quantify the amount left on the table for households who do not request a deferral on credit-card balances. Future work can attempt to link the intervention analysis to structural models.
Declarative competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix

A.1. Tobit models of time-series data

Random variables in a cross-section that have both continuous and discrete components are often specified as Tobit models, a term that honors the research of the American economics Nobel laureate, James Tobin (1958). Following Takeshi Amemiya (1984), consider the latent random variable \( Q^* \), which follows the Gaussian law, having mean \( \mu \) and variance \( \sigma^2 \). Suppose the observed random variable \( Q \) is defined by the following:

\[
Q_n = \max(0, Q^*_n).
\]

In this case, a point mass exists at \( Q = 0 \), having probability

\[
\Pr(Q \leq 0) = \Pr(\mu + \sigma Z \leq 0) = \Pr\left(Z \leq -\frac{\mu}{\sigma}\right) = \int_{-\infty}^{\frac{-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} \exp\left(-z^2/2\right) \, dz = \int_{-\infty}^{\frac{-\mu}{\sigma}} \phi(z) \, dz = \Phi\left(-\frac{\mu}{\sigma}\right),
\]

where \( Z \) is a standard normal variate, having associated probability density function \( \phi(\cdot) \) and cumulative distribution function \( \Phi(\cdot) \), while \( Q \) follows the Gaussian law above the threshold \((-\mu/\sigma)\). The discrete-continuous probability mass/density function of \( Q \) induced by this structure is the following:

\[
f_Q(y|\mu, \sigma) = \left\{ \begin{array}{ll}
\Phi\left(-\frac{\mu}{\sigma}\right) & \text{if } Q = 0, \\
\frac{1}{\sigma} \phi\left(\frac{q - \mu}{\sigma}\right) & \text{if } Q > 0,
\end{array} \right.
\]

where \( \mathbb{I}(A) \) denotes the indicator function, which equals one when the event \( A \) obtains, and zero otherwise.

A representative relative frequency distribution of some 200 observations from the censored Gaussian law is depicted in Fig. A.1. From a sequence of \( N \) observations \( \{q_n\}_{n=1}^N \), collected in the vector \( q \), one can estimate the parameter vector \( \theta = [\mu, \sigma]^T \) using the method of maximum likelihood by optimizing the following logarithm of the likelihood function:

\[
\log \mathcal{L}(\theta|q) = \sum_{n=1}^N \left\{ \mathbb{I}(q_n = 0) \left( \log \left[ \phi\left(-\frac{\mu}{\sigma}\right)\right] \right) + \mathbb{I}(q_n > 0) \left( \log \phi\left(\frac{q_n - \mu}{\sigma}\right) - \log(\sigma) \right) \right\}.
\]

For the latent random variable \( Q^*_n \), consider a special case of an ARIMA\((p, d, q)\), the AR(1), which can be written as follows:

\[
(Q^*_n - \mu) = \rho(q^*_{n-1} - \mu) + \varepsilon_t,
\]

where \( \varepsilon_t \) denotes independently and identically distributed Gaussian random variables having mean zero and variance \( \sigma^2 \). A representative graph of these time-series data is depicted in Fig. A.2. From \( \{q^*_t\}_{t=0}^T \), a sequence of \((T+1)\) observations concerning the latent random variable, one can consistently estimate the parameter vector \( \theta = [\mu, \rho, \sigma]^T \) using the method of maximum likelihood by optimizing the following function:

\[
\mathcal{L}(\theta|q^*) = \prod_{t=1}^T \frac{1}{\sigma} \phi\left(\frac{q^*_t - \mu(1-\rho) - \rho q^*_{t-1}}{\sigma}\right).
\]

As above, now suppose that the observed random variable \( Q_t \) is determined by the following rule:

\[
Q_t = \max(0, Q^*_t).
\]

This series is depicted as the solid line in Fig. A.3, while the negative values are shown by the “+” symbol.
If the errors follow the Gaussian law, then by maximizing the
function, one cannot consistently estimate the correlation parameter ρ.

For the Gaussian law, in Table A.1, we collect some
Monte Carlo evidence demonstrating these claims when
µ₀ = 0.5, ρ₀ = 0.9, and the variance of ε is one, so σ₀ = 2.29416,
for samples of size T = 100, 200, 500, 1000,
where the number of replications is 9999. Notice how the
bias in ˆµ, the Tobit maximum-likelihood estimator of µ₀,
grows smaller as the sample size T increases. In Fig. A.4,
we depict the extent of the bias, |β(µ₀)/T|, which is
approximately O(1/T). Perhaps most disturbing,
however, is how different the estimated variance of the MLE ˆµ
is, when compared to the first-order asymptotic approximation
calculated from elements of the inverse of the negative
of the estimated Hessian matrix and reported as V(ˆµ) in the table: the extent of the error rises as the
sample size gets larger. The ratio of the variance of ˆµ₁₀₀,
0.92045, to its asymptotic approximation, 0.06938, is 13.3,
while the ratio of the variance of ˆµ₂₀₀, 0.69396, to its
asymptotic approximation, 0.03421, is 20.3. The ratio of
the variance of ˆµ₅₀₀, 0.45701, to its asymptotic approxi-
mation, 0.01372, is 33.3, and the ratio of the variance of
ˆµ₁₀₀₀, 0.32192, to its asymptotic approximation, 0.00685,
is 47.0. Not only are the standard errors underestimated,
but the extent appears to increase with the sample size,
making inference untenable.

Table A.1

| Parameter | Mean | Variance | Minimum | Maximum |
|-----------|------|----------|---------|---------|
| µ₀ = 0.5  |      |          |         |         |
| T = 100   | 0.45192 | 0.92045 | -3.76182 | 3.81443 |
| T = 200   | 0.46592 | 0.69396 | -3.39578 | 2.91524 |
| T = 500   | 0.48435 | 0.45701 | -1.61577 | 2.20638 |
| T = 1000  | 0.49304 | 0.32192 | -0.92277 | 1.56088 |

σ₀² = 2.29416

| T = 100   | 2.01571 | 0.50511 | 0.50365 | 4.77354 |
| T = 200   | 2.14657 | 0.39701 | 1.15388 | 5.07263 |
| T = 500   | 2.23906 | 0.26690 | 1.47906 | 3.66571 |
| T = 1000  | 2.26828 | 0.19464 | 1.67941 | 3.13936 |

V(µ)²

| T = 100   | 0.06938 | 0.06399 | 0.01117 | 1.34104 |
| T = 200   | 0.03421 | 0.01999 | 0.00732 | 0.52426 |
| T = 500   | 0.01372 | 0.00435 | 0.00470 | 0.05856 |
| T = 1000  | 0.00685 | 0.00149 | 0.00339 | 0.01645 |

To specify the likelihood function appropriately re-
quires integrating out the unobserved εₜ, for every zero-
valued q. To start, consider the case where the first T
observations (counting the observation for t = 0) are
positive, but the last one is zero. The likelihood function in
this case is

L(θ|q) = \prod_{t=1}^{T-1} \frac{1}{\sigma} \phi \left( \frac{q_t - \mu(1 - \rho - \rho q_{t-1})}{\sigma} \right) \times \int_{-\infty}^{-\mu(1 - \rho - \rho q_{T-1})} \phi(z_{T-1}) \, dz_{T-1}.

Now, consider the case where the first (T−1) observations
(again counting the observation for t = 0) are positive,
but the last two are zero. The likelihood function in this
case is

L(θ|q) = \prod_{t=1}^{T-2} \frac{1}{\sigma} \phi \left( \frac{q_t - \mu(1 - \rho - \rho q_{t-1})}{\sigma} \right) \times \int_{-\infty}^{-\mu(1 - \rho - \rho q_{T-2})} \int_{-\infty}^{-\mu(1 - \rho - \rho q_{T-1})} \phi(z_{T-1}) \phi(z_{T}) \, dz_{T-1} \, dz_{T}.

When one or two zeros occur in a row, this is relatively
straightforward to do using methods of quadrature, but
in our data the number of zeros in a row could be in
the tens or twenties. Integrals of dimension ten or twenty
are arduous to calculate accurately because in dimensions
of four or more, simulation methods must be used to
approximate the integrals at O(1/\sqrt{S}) accuracy, where S
is the number of simulation draws.
Table A.2
Some Monte Carlo evidence with lognormal errors.

| Parameter | Mean   | Variance | Minimum | Maximum |
|-----------|--------|----------|---------|---------|
| $\mu^0 = 1.0$ |        |          |         |         |
| $T = 100$ | 9.10341| 2.02235  | 3.36693 | 29.30399|
| $T = 200$ | 9.53405| 1.47786  | 5.11464 | 19.70764|
| $T = 500$ | 9.81798| 0.94632  | 6.71972 | 14.55416|
| $T = 1000$| 9.91227| 0.66267  | 7.68366 | 12.91488|
| $\sigma^0 = 4.98723$ |        |          |         |         |
| $T = 100$ | 4.54009| 1.74165  | 1.34228 | 42.71136|
| $T = 200$ | 4.70636| 1.38829  | 2.09365 | 31.85160|
| $T = 500$ | 4.83995| 0.97589  | 2.69117 | 20.95802|
| $T = 1000$| 4.89231| 0.74422  | 3.17284 | 15.35153|

That noted, Michael D. Hurd (1979) as well as Abbas Arabmazar and Peter Schmidt (1981) demonstrated that heteroskedasticity within the Gaussian family could result in biased and inconsistent Tobit MLEs; the standard errors based on the Hessian matrix are also incorrect. Arabmazar and Schmidt (1982) as well as Harry J. Paarsch (1984) demonstrated that when the errors do not follow the Gaussian law, the Tobit MLE, assuming the Gaussian law, is biased and inconsistent; the standard errors based on the Hessian matrix are incorrect, too.

To illustrate this claim here briefly, for completeness, we assumed that the errors follow the lognormal (Galton) law, where $\mu$ has mean zero and variance 4.7257. To get the same amount of censoring (around 35%), we needed to adjust $\mu^0$. In Table A.2, we collect some Monte Carlo evidence demonstrating these claims when $\mu^0 = 1.0$, $\rho^0 = 0.9$, and $\sigma^0 = 4.9872$, for samples of size $T = 100, 200, 500, 1000$, where the number of replications was again 9999. Notice how biased $\hat{\mu}$, the Tobit maximum-likelihood estimator of $\mu^0$, is, and this bias does not disappear as the sample sizes get larger. In short, in the presence of censoring, getting the shape of the distribution correct is extremely important, which is why we chose to model the shape semiparametrically in the research reported below.

A.2. Training the model

For a particular row $k$ of the transition matrix, collect the parameters from $y_k$ and $\delta_k$ in the $(M \times 1)$ vector $\theta_k = [y_k \delta_k]^{T}$. Also, represent the collected labels and features by the matrix $W_k$. Training the model involves finding

$$\hat{\theta}_k = \arg\min_{\theta_k} f(\theta_k | W_k).$$  \hspace{1cm} (14)

Because the objective function in Eq. (14) is continuous, convex, and differentiable in the argument vector $\theta_k$, the unique minimizer $\hat{\theta}_k$ is characterized by the following vector of first-order conditions:

$$\nabla_{\theta_k} f(\theta_k | W_k) = 0_M.$$  \hspace{1cm} (15)

We demonstrated above that the individual elements of the gradient vector $\nabla_{\theta_k} f(\theta_k | W_k)$ have convenient, closed-form expressions. Nevertheless, these estimates can only be calculated numerically.

A.2.1. Numerical methods

One approach is to employ Newton’s method. For notational parsimony, we suppress the $k$ subscript in what follows. The gradient function, which we denote by $g(\theta)$ for notational parsimony, is expanded in a Taylor series about some initial guess $\hat{\theta}_0$ to yield

$$0_M = g(\theta) + \nabla g(\theta_0) (\theta - \hat{\theta}_0) + R_2.$$

Ignoring the remainder vector $R_2$, and imposing the zero equality, allows one to solve for

$$\hat{\theta}_1 = \hat{\theta}_0 - [\nabla g(\theta_0)]^{-1} g(\theta_0).$$

Replacing $\hat{\theta}_0$ with $\hat{\theta}_1$, one can iterate to convergence according to some criterion.

To implement Newton’s method requires inverting the matrix of second partial derivatives $\nabla^2 g(\theta_0)$, which is often referred to as the Hessian matrix. In the multinomial logit case, the Hessian matrix has a convenient closed-form expression, although one can also approximate it using the so-called outer product of the gradient (OPG) approximation. Unfortunately, inverting either the Hessian matrix or the OPG approximation is an $O(M^3)$ calculation, which can be computationally arduous when M is large.

A commonly used way to solve Eq. (15) when $M$ is large is a method referred to as gradient descent, and sometimes as steepest descent. Under this method, a direction

$$d(\theta) = \frac{g(\theta)}{||g(\theta)||_2}$$

is chosen. Updating the weights then involves calculating the following sequence iteratively:

$$\hat{\theta}_{r+1} = \hat{\theta}_r + \alpha d(\hat{\theta}_r) \quad r = 0, 1, 2, \ldots$$

where $\alpha$ is a positive parameter (which may be indexed by $r$) that controls the rate of descent. When $\alpha$ is indexed by $r$, the sequence of $\alpha_r$ typically decreases as $r$ increases. Convergence is attained when the norm of the first partial derivative vector is less than some prescribed accuracy criterion. For example, if the Euclidean norm is used, then when

$$||g(\hat{\theta}_r)||_2 < \varepsilon$$

where the convergence criterion $\varepsilon$ could be $10^{-6}$. A second way to evaluate convergence is to calculate the norm of the difference between two successive iterants, and to stop when it is less than some prescribed accuracy criterion. For example, if the Euclidean norm is again used, then this reduces to

$$||\hat{\theta}_r - \hat{\theta}_{r-1}||_2 < \varepsilon$$

where $\varepsilon$ does not need to be the same as $\varepsilon$. Alternatively, calculate

$$\frac{||\hat{\theta}_r - \hat{\theta}_{r-1}||_2}{||\hat{\theta}_{r-1}||_2} \quad \text{or} \quad \frac{||\hat{\theta}_r - \hat{\theta}_{r-1}||_2}{||\hat{\theta}_r||_2}.$$
and stop iterating when the relative changes are small. A third way to evaluate convergence is to calculate the relative improvement in the loss function evaluated at adjacent iterants. For example,

$$\log \left[ \frac{f(\hat{\theta}_{r-1}|y, W)}{f(\hat{\theta}_r|y, W)} \right] < \Delta$$

where $\Delta$ might be $10^{-d}$. In words, when changes in the loss function are less than a basis point, convergence is attained.

The method of gradient descent can often be slow to converge, whereas Newton's method can be impractical when $M$ is large. The conjugate-gradient method offers a suitable alternative in convex problems like this one. As before, introduce the direction vector $d(\theta)$ and the gradient vector $g(\theta)$, but for notational parsimony denote these by $d_r$ and $g_r$, respectively, when evaluated at $\theta_r$. Thus, for an initial estimated $\hat{\theta}_0$, one obtains

$$d_0 = - \frac{g_0}{\|g_0\|_2}.$$ 

For $r = 0, 1, 2, \ldots$ and for a positive definite matrix $Q$, while $g_r \neq 0_M$, calculate the following sequence:

$$\begin{align*}
\alpha_r &= \frac{g_r^T g_r}{d_r^T Q d_r} \\
\hat{\theta}_{r+1} &= \hat{\theta}_r + \alpha_r d_r \\
g_{r+1} &= g_r - \alpha_r d_r \\
\omega_r &= \frac{g_{r+1}^T g_{r+1}}{g_r^T g_r} \\
d_{r+1} &= g_{r+1} + \omega_r d_r.
\end{align*}$$

In practice, however, $g_r$ will never attain the zero vector $0_M$, so one typically continues to iterate only until one (or more) of the above convergence criteria have been attained.

A.2.2. Quantifying the sampling variability

In large samples, for $k = 0, 1, \ldots, K$,

$$\hat{\theta}_k \stackrel{d}{\rightarrow} N \left[ \hat{\theta}_k, V(\hat{\theta}_k) \right],$$

where $V(\hat{\theta}_k)$ can be approximated by

$$\hat{S}_k = \left[ \nabla_{\theta} f(\hat{\theta}_k|W) \right]^{-1}.$$ 

Collect the parameters for the $k = 0, 1, \ldots, K$ rows in the vector $\theta = [\theta_0^T, \theta_1^T, \ldots, \theta_K^T]^T$ and denote the trained estimate by $\hat{\theta}$. The estimates of each row will be asymptotically independent of one another, so the aggregate $\hat{S}$ can be approximated by

$$\hat{S} = \begin{bmatrix}
\hat{S}_0 & 0_{M,M} & \cdots & 0_{M,M} \\
0_{M,M} & \hat{S}_1 & \cdots & 0_{M,M} \\
\vdots & \vdots & \ddots & \vdots \\
0_{M,M} & 0_{M,M} & \cdots & \hat{S}_K
\end{bmatrix}$$

where $0_{M,M}$ is an $(M \times M)$ matrix of zeros.

How does one calculate the sampling variability of some object, for instance, under this framework, $\hat{p}_{T+1} = \mathbf{P}(\hat{\theta}|\hat{p}_T)$? First, using the Cholesky decomposition, calculate $\hat{F}$, a lower-triangular matrix that satisfies the following:

$$\hat{F} \hat{F}^T = \hat{S}.$$ 

Because $\hat{S}$ is block diagonal, one can form $\hat{F}$ by stacking the individual Cholesky factors $\hat{F}_k$ that solve

$$\hat{F}_k \hat{F}_k^T = \hat{S}_k \quad k = 1, 2, \ldots, K.$$ 

Specifically,

$$\hat{F} = \begin{bmatrix}
\hat{F}_0 & 0_{M,M} & \cdots & 0_{M,M} \\
0_{M,M} & \hat{F}_1 & \cdots & 0_{M,M} \\
\vdots & \vdots & \ddots & \vdots \\
0_{M,M} & 0_{M,M} & \cdots & \hat{F}_K
\end{bmatrix}.$$ 

Next, consider, for example, $B = 9999$ bootstrap replications. For each simulation $b = 1, 2, \ldots, B$,

1. draw $k = (k+1) (M \times 1)$ vectors of standard Gaussian variates denoted by $[Z_{b,k}]_{k=0}^K$ for realization $b$;
2. for row $k = 0, 1, \ldots, K$, form

$$\hat{\theta}_{b,k} = \hat{\theta}_k + \hat{F}_k Z_{b,k};$$
3. evaluate $\hat{p}_{b,T+1} = \mathbf{P}(\hat{\theta}_b|\hat{p}_T)$ and save;
4. continue (1) to (3) until $B$ is attained.

A.3. Goodness-of-fit tests

In Table 1 in Section 4.4, we present statistics to measure the difference between forecasted distributions and the observed proportions in each balance category. This approach is ideal because it begins the forecast from the month of January, in which the consumers in Canada were largely unaffected by the COVID-19 pandemic.

In Table A.3, we collect a similar comparison, except that the forecasts are calculated as one-step-ahead forecasts from the previous month. The model with a fixed transition matrix is not suited to the data, even when conditioning on the distribution in the previous month. For the model with monthly transition matrices, the $p$-values are at conventional levels of significance with Kullback–Leibler divergence statistics measured in double digits. During the months of April and May, the statistic shows a marked deviation from that expected from the $\chi^2$ distribution. Even when conditioning on the proportions observed during the previous months, there is strong evidence that the distributions have changed. This also suggests that the model with monthly transition matrices is as good as it needs to be to detect a difference in distributions, and that no further complexity is warranted to answer the question of whether a change has occurred.

It is also recommended that the researcher calculate the distance statistic for time periods within the sample to determine the goodness of fit of the model specifications. A significant statistic will indicate that the model should be revised. In Table A.4, we collect these statistics for all months of the sample period for which the estimates are available. The in-sample fit of the monthly model (labeled “Monthly”) is superior to that with the fixed transition matrix (labeled “Fixed”). For the monthly model,
the in-sample fit is acceptable, which is expected for a well-specified model. It is also clear that the model with a fixed transition matrix is inadequate. This is especially true for the first few months of the year. Since the performance of this model suffers during the time period in which we are attempting to measure the effects of the pandemic, we reject this model in favor of the model with monthly transition matrices.

A.4. Comparison to model with fixed transition matrix

If we were to ignore the results in Table A.4 above, we might inspect the deviations from forecasts for the fixed-transition model, which are depicted in Fig. A.5 and Fig. A.6. It appears that the effect measured with the fixed model is similar to that measured with the seasonal model. When, however, one compares the results from this model in March, as in Panel A.5(a) of Fig. A.5, a similar effect appears. This feature is not present when one considers the distributions in March using the model with separate transition matrices, as shown in Fig. A.7, or using the model with covariates, as in Fig. A.8. The most extreme deviations from the forecasted proportions in each category are five to six percent. Without seasonal effects, the change in distributions is confounded with the usual pattern of repayments for holiday spending. It also measures the reduction of balances when many consumers receive their tax refund. This is further evidence that the restricted model specification is inadequate, but both models that include monthly variation are candidates for analyzing the response to the pandemic. The model with covariates, however, has the added advantage of measuring the different effects for subsamples with, for instance, different credit scores, without confounding the pandemic response with the differences between subsamples with different credit scores.

A.5. Intervention pattern over time

In the text in Section 4.4, we present figures showing the deviations from the forecasted distributions of balances during the months of May and August. Although these two months are representative of the changes at notable checkpoints in the pandemic, we document the entire sequence of the deviations in what follows. The comparisons are made using the full model with covariates, not just the restricted model. The effect of the pandemic appears strongest in May, as shown in Panel A.10(a) of Fig. A.10. When, however, one compares the results from this model in March, as in Panel A.5(a) of Fig. A.5, a similar effect appears. This feature is not present when one considers the distributions in March using the model with separate transition matrices, as shown in Fig. A.7, or using the model with covariates, as in Fig. A.8. The most extreme deviations from the forecasted proportions in each category are five to six percent. Without seasonal effects, the change in distributions is confounded with the usual pattern of repayments for holiday spending. It also measures the reduction of balances when many consumers receive their tax refund. This is further evidence that the restricted model specification is inadequate, but both models that include monthly variation are candidates for analyzing the response to the pandemic. The model with covariates, however, has the added advantage of measuring the different effects for subsamples with, for instance, different credit scores, without confounding the pandemic response with the differences between subsamples with different credit scores.

A.6. Specification error from using a lower-order Markov process

In our analysis, we employed a first-order Markov model where the transition probabilities are assumed to follow the multinomial logit model. This choice was based on the analytical tractability and reduced complexity of the optimization of the likelihood function. Another source of computational complexity is the order of the Markov process. Suppose we divide the support of the source of computational complexity is the order of the Markov process. Suppose we divide the support of the sources of the transition matrix $K$ into $K + 1$ bins. For a first-order Markov process, we would need to estimate a vector of parameters that define probabilities in a transition matrix $P$. Strictly speaking, each column of the matrix $P$ is implied by a unit restriction, so each column requires the estimation of $K$ free parameters per row. Still, the estimation of the
parameter vector is of order $K^2$, without employing any explanatory variables. If we were to increase the order of the Markov process to order $p$, this would raise the complexity of the estimation to the order $(K^p)^2$. With a large number of observations, an increase in the Markov order results in a substantial increase in computational complexity.

In preliminary analysis, we determined that consumer balances were governed by a higher-order autoregressive process. The predictions from models of different orders did not, however, differ substantially from those of a first-order autoregressive model. Faced with a scarcity of computing resources, we must make a tradeoff sample size and process order. This decision depends on the magnitude of the loss of model accuracy from restricting attention to a lower-order Markov process.

To this end, we conducted some simulations to compare the prediction accuracy of the first- and second-order Markov chain model when the true process is of second order. We generated a panel of data to approximate the nature of our observed data. For each consumer, we generated realizations of an AR(2) process with autoregressive parameters $(\rho_1, \rho_2) = (0.6, 0.2)$, which approximate the values estimated from the data. Starting values were...
drawn from the unconditional distribution implied by these parameters, with a burn-in of 50 periods. We then transformed these series by taking the exponential to get a panel of observations \( y_{nt} \). This resulted in a series of observations that approximately match the features of the data. To reduce computational requirements we divided the cross-section into 11 bins, resulting in a histogram, shown in Fig. A.13, calculated from one realization of the panel. This demonstrates that the artificial data has a distribution that is comparable to that of the observed data.

Under this DGP, we generated 100 individual series, each with 500 observations, with 1000 replications of the entire panel. For each panel, we assigned observations to the same \((K + 1) = 11\) bins as in the histogram in Fig. A.13, for every individual and every time period. The eleventh bin captures the observations in the tail of the distribution. We then estimated a pair of transition matrices on a training sample of the first 250 observations. The first-order model is characterized by the \([ (K + 1) \times (K + 1) ]\) transition matrix \( \hat{P} \), with element \( \hat{p}_{ij} \) denoting the probability of an individual moving to state \( i \) from state

---

**Fig. A.7.** Deviations from forecasted credit-card balances (monthly).

**Fig. A.8.** Deviations from forecasted credit-card balances (covariates).
We also estimated a second-order model by expanding the state space to dimension $(K + 1)^2$, instead of $(K + 1)$, by categorizing observations according to membership in the pair of histogram bins over the last two periods. This requires the estimation of a $[(K + 1)^2 \times (K + 1)]$ transition matrix $\hat{Q}$, which takes the form of a stack of $[(K + 1) \times (K + 1)]$ transition matrices similar to $\hat{P}$ but conditioning on the state two periods back.

We compared the prediction accuracy of these two estimates by recording the difference between the predicted proportion in each of the 11 bins with the actual proportions observed. We calculated statistics for this comparison over the remaining 250 time periods in the validation sample. In principle, we could have determined the expected proportions in each probability bin, conditioning on each log $q_{int}$, a continuous variable. In the interest of reducing the computational burden, however, we chose to compare to the observed histograms in each period of the validation sample. The predictive accuracy of the estimates of proportions in the histogram is well known and serves as a fair benchmark with which to compare the two estimates.
The results of this comparison are depicted in Figs. A.14 and A.15. Fig. A.14 illustrates the mean prediction error across balance categories with the true model, with the second-order Markov model shown in black dotted lines, and the underspecified first-order Markov model shown in solid grey lines. For both models, the predictions lie right on the horizontal axis at zero, with barely perceptible variation in the balance categories with higher proportions and corresponding higher variance. This indicates little bias from underspecifying the model of the transition probabilities. To compare the variation in prediction error, Fig. A.14 also illustrates confidence intervals of the
prediction error, calculated as $\pm 1.96 \times \hat{\sigma}^2_k$, where $\hat{\sigma}^2_k$ is the sample standard deviation of the prediction error for the proportions in balance bin $k$. Again, very little difference exists between these models, except in the boundary categories, with higher proportions and higher variances. Fig. A.15 indicates the distribution of the prediction error by showing the quintiles of the prediction error. The distributions are very similar between the two models, with differences in variance supporting the results in Fig. A.14, with little difference in skewness. We perceive this as sufficient evidence that the order of the Markov process can be safely underspecified with minimal loss of predictive accuracy. With scarce computing resources, additional computational capacity can be allocated to increasing the sample size.
References

Agarwal, S., Ambrose, B. W., & Liu, C. (2006). Credit lines and credit utilization. *Journal of Money, Credit and Banking*, 38(1), 1–22. http://dx.doi.org/10.1353/mcb.2006.0010, URL https://ideas.repec.org/a/mcb/jmcb/v38y2006ip1p1-22.html.

Agarwal, S., & Qian, W. (2014). Consumption and debt response to unanticipated income shocks: Evidence from a natural experiment in Singapore. *American Economic Review*, 104(12), 4205–4230. http://dx.doi.org/10.1257/aer.104.12.4205.

Allen, J., Clark, R., Li, S., & Vincent, N. (2021). Debt-relief programs and money left on the table: Evidence from Canada’s response to COVID-19. *Canadian Journal of Economics*, (forthcoming).

Allen, J., Grieder, T., Peterson, B., & Roberts, T. (2020). The impact of macroprudential housing finance tools in Canada. *Journal of Financial Intermediation*, 42, Article 100761. http://dx.doi.org/10.1016/j.jfi.2017.08.004.

Amemiya, T. (1984). *Advanced econometrics*. Cambridge, MA: Harvard University Press.

Andersen, A. L., Hansen, E. T., Johannesen, N., & Sheridan, A. (2020). Consumer responses to the COVID-19 crisis: Evidence from bank account transaction data. *Covid Economics*, 7, 88–114.

Angeriz, A., & Arestis, P. (2008). Assessing inflation targeting through intervention analysis. *Oxford Economic Papers*, 60(2), 293–317.

Arabmazar, A., & Schmidt, P. (1981). Further evidence on the robustness of the Tobit estimator to heteroskedasticity. *Journal of Econometrics*, 17, 253–258.

Arabmazar, A., & Schmidt, P. (1982). An investigation of the robustness of the Tobit estimator to non-normality. *Econometrica*, 50, 1055–1063.

Baker, S. R., Farrokhnia, R. A., Meyer, S., Pagel, M., & Yannelis, C. (2020). How does household spending respond to an epidemic? Consumption during the 2020 COVID-19 pandemic. In Working paper series. no. 26949, National Bureau of Economic Research.

Below, D. L., & Armstrong, R. D. (2011). Distributions of the Kullback–Leibler divergence with applications. *British Journal of Mathematical and Statistical Psychology*, 64, 291–309.

Bhar, R. (2001). Return and volatility dynamics in the spot and futures markets in Australia: An intervention analysis in a bivariate EGARCH-X framework. *Journal of Futures Markets*, 21(9), 833–850.

Bhattacharyya, M. N., & Layton, A. P. (1979). Effectiveness of seat belt legislation on the Queensland road toll—An Australian case study in intervention analysis. *Journal of the American Statistical Association*, 74(367), 596–603.

Bhutta, N., & Keys, B. J. (2016). Interest rates and equity extraction during the housing boom. *American Economic Review*, 106, 1742–1774.

Bianchi, L., Jarrett, J., & Choudary Hanumara, R. (1998). Improving forecasting for telemarketing centers by ARIMA modeling with intervention. *International Journal of Forecasting*, 14(4), 497–504.

Bilyk, O., & Peterson, B. (2015). Credit cards: Disentangling the dual use of borrowing and spending: Staff analytical note 2015–3, Bank of Canada, URL https://www.bankofcanada.ca/wp-content/uploads/2015/12/san2015-3.pdf.

Bilyk, O., & teNyenhuis, M. (2018). The impact of recent policy changes on the Canadian mortgage market: Staff analytical Note 2018–35, Bank of Canada, URL https://www.bankofcanada.ca/2018/11/staff-analytical-note-2018-35/.

Bilyk, O., Ueberfeldt, A., & Xu, Y. (2017). *Analysis of household vulnerabilities using loan-level mortgage data: Technical report, Bank of Canada Financial System Review*, URL https://www.bankofcanada.ca/wp-content/uploads/2017/11/fsr-november2017-bilyk.pdf.

Bonham, C. S., & Gangnes, B. (1996). Intervention analysis with cointegrated time series: The case of the hawai hotel room tax. *Applied Economics*, 28(10), 1281–1293.

Box, G. E. P., & Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*, 70, 70–79.

Carvalho, V. M., Garcia, J. R., Hansen, S., Ortiz, A., Rodrigo, T., Mora, S. R., & Ruiz, P. (2020). Tracking the COVID-19 crisis with high-resolution transaction data: Technical report DPI-4642, Centre for Economic Policy Research.

Chen, H., Engert, W., Huynh, K. P., Nicholls, G., Nicholson, M., & Zhu, J. (2020). *Cash and COVID-19: The impact of the pandemic on demand for and use of cash*, Staff discussion paper 2020-6, Bank of Canada, Ottawa, Canada.

Chen, H., Michaux, M., & Roussanov, N. (2020). Houses as ATMs: Mortgage refinancing and macroeconomic uncertainty. The *Journal of Finance*, 75(1), 323–375. http://dx.doi.org/10.1111/jofi.12842, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.12842.

Daughton, A., Generous, N., Priedhorsky, R., & Deshpande, A. (2017). An approach to and web-based tool for infectious disease outbreak intervention analysis. *Scientific Reports*, 7.

Felt, M.-H., Hayashi, F., Stavins, J., & Welte, A. (2021). *Distributional effects of payment card pricing and merchant cost pass-through in Canada and the United States: Staff working paper 21-8*, Bank of Canada, URL https://www.bankofcanada.ca/wp-content/uploads/2021/02/swp2021-8.pdf.
